

# Rewriting the rules of machine-generated art

August 19 2020, by Kim Martineau

---



A new GAN-editing tool developed at MIT allows users to copy features from one set of photos and paste them into another, creating an infinite array of pictures that riff on the new theme — in this case, horses with hats on their heads. Credit: David Bau

Horses don't normally wear hats, and deep generative models, or GANs, don't normally follow rules laid out by human programmers. But a new

tool developed at MIT lets anyone go into a GAN and tell the model, like a coder, to put hats on the heads of the horses it draws.

In a new study appearing at the European Conference on Computer Vision this month, researchers show that the deep layers of [neural networks](#) can be edited, like so many lines of code, to generate surprising images no one has seen before.

"GANs are incredible artists, but they're confined to imitating the data they see," says the study's lead author, David Bau, a Ph.D. student at MIT. "If we can rewrite the rules of a GAN directly, the only limit is human imagination."

Generative adversarial networks, or GANs, pit two neural networks against each other to create hyper-realistic images and sounds. One neural network, the generator, learns to mimic the faces it sees in photos, or the words it hears spoken. A second network, the discriminator, compares the generator's outputs to the original. The generator then iteratively builds on the discriminator's feedback until its fabricated images and sounds are convincing enough to pass for real.

GANs have captivated artificial intelligence researchers for their ability to create representations that are stunningly lifelike and, at times, deeply bizarre, from a receding cat that melts into a pile of fur to a wedding dress standing in a church door as if abandoned by the bride. Like most deep learning models, GANs depend on massive datasets to learn from. The more examples they see, the better they get at mimicking them.

But the new study suggests that big datasets are not essential. If you understand how a [model](#) is wired, says Bau, you can edit the numerical weights in its layers to get the behavior you desire, even if no literal example exists. No dataset? No problem. Just create your own.

"We're like prisoners to our training data," he says. "GANs only learn patterns that are already in our data. But here I can manipulate a condition in the model to create horses with hats. It's like editing a genetic sequence to create something entirely new, like inserting the DNA of a firefly into a plant to make it glow in the dark."

Bau was a software engineer at Google, and had led the development of Google Hangouts and Google Image Search, when he decided to go back to school. The field of deep learning was exploding and he wanted to pursue foundational questions in computer science. Hoping to learn how to build transparent systems that would empower users, he joined the lab of MIT Professor Antonio Torralba. There, he began probing deep nets and their millions of mathematical operations to understand how they represent the world.

Bau showed that you could slice into a GAN, like layer cake, to isolate the artificial neurons that had learned to draw a particular feature, like a tree, and switch them off to make the tree disappear. With this insight, Bau helped create GANPaint, a tool that lets users add and remove features like doors and clouds from a picture. In the process, he discovered that GANs have a stubborn streak: they wouldn't let you draw doors in the sky.

"It had some rule that seemed to say, 'doors don't go there,'" he says. "That's fascinating, we thought. It's like an 'if' statement in a program. To me, it was a clear signal that the network had some kind of inner logic."

Over several sleepless nights, Bau ran experiments and picked through the layers of his models for the equivalent of a conditional statement. Finally, it dawned on him. "The neural network has different memory banks that function as a set of general rules, relating one set of learned patterns to another," he says. "I realized that if you could identify one

line of memory, you could write a new memory into it."

In a short version of his ECCV talk, Bau demonstrates how to edit the model and rewrite memories using an intuitive interface he designed. He copies a tree from one image and pastes it into another, placing it, improbably, on a building tower. The model then churns out enough pictures of tree-sprouting towers to fill a family photo album. With a few more clicks, Bau transfers hats from human riders to their horses, and wipes away a reflection of light from a kitchen countertop.

The researchers hypothesize that each layer of a deep net acts as an associative memory, formed after repeated exposure to similar examples. Fed enough pictures of doors and clouds, for example, the model learns that doors are entryways to buildings, and clouds float in the sky. The model effectively memorizes a set of rules for understanding the world.

The effect is especially striking when GANs manipulate light. When GANPaint added windows to a room, for example, the model automatically added nearby reflections. It's as if the model had an intuitive grasp of physics and how light should behave on object surfaces. "Even this relationship suggests that associations learned from data can be stored as lines of memory, and not only located but reversed," says Torralba, the study's senior author.

GAN editing has its limitations. It's not easy to identify all of the neurons corresponding to objects and animals the model renders, the researchers say. Some rules also appear edit-proof; some changes the researchers tried to make failed to execute.

Still, the tool has immediate applications in computer graphics, where GANs are widely studied, and in training expert AI systems to recognize rare features and events through data augmentation. The tool also brings

researchers closer to understanding how GANs learn visual concepts with minimal human guidance. If the models learn by imitating what they see, forming associations in the process, they may be a springboard for new kinds of machine learning applications.

Provided by Massachusetts Institute of Technology

Citation: Rewriting the rules of machine-generated art (2020, August 19) retrieved 4 April 2024 from <https://techxplore.com/news/2020-08-rewriting-machine-generated-art.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.