

New tool highlights what generative models leave out when reconstructing a scene

November 11 2019, by Kim Martineau



A new tool reveals what AI models leave out in recreating a scene. Here, a GAN, or generative adversarial network, has dropped the pair of newlyweds from its reconstruction (right) of the photo it was asked to draw (left). Credit: Massachusetts Institute of Technology

Anyone who has spent time on social media has probably noticed that



GANs, or generative adversarial networks, have become remarkably good at drawing faces. They can predict what you'll look like when you're old and what you'd look like as a celebrity. But ask a GAN to draw scenes from the larger world and things get weird.

<u>A new demo</u> by the MIT-IBM Watson AI Lab reveals what a model trained on scenes of churches and monuments decides to leave out when it draws its own version of, say, the Pantheon in Paris, or the Piazza di Spagna in Rome. The larger study, Seeing What a GAN Cannot Generate, was presented at the International Conference on Computer Vision last week.

"Researchers typically focus on characterizing and improving what a machine-learning system can do—what it pays attention to, and how particular inputs lead to particular outputs," says David Bau, a graduate student at MIT's Department of Electrical Engineering and Computer Science and Computer Science and Artificial Science Laboratory (CSAIL). "With this work, we hope researchers will pay as much attention to characterizing the data that these systems ignore."

In a GAN, a pair of neural networks work together to create hyperrealistic images patterned after examples they've been given. Bau became interested in GANs as a way of peering inside black-box neural nets to understand the reasoning behind their decisions. An earlier tool developed with his advisor, MIT Professor Antonio Torralba, and IBM researcher Hendrik Strobelt, made it possible to identify the clusters of artificial neurons responsible for organizing the image into real-world categories like doors, trees, and clouds. A related tool, GANPaint, lets amateur artists add and remove those features from photos of their own.

One day, while helping an artist use GANPaint, Bau hit on a problem. "As usual, we were chasing the numbers, trying to optimize numerical reconstruction loss to reconstruct the photo," he says. "But my advisor



has always encouraged us to look beyond the numbers and scrutinize the actual images. When we looked, the phenomenon jumped right out: People were getting dropped out selectively."

Just as GANs and other neural nets find patterns in heaps of data, they ignore patterns, too. Bau and his colleagues trained different types of GANs on indoor and outdoor scenes. But no matter where the pictures were taken, the GANs consistently omitted important details like people, cars, signs, fountains, and pieces of furniture, even when those objects appeared prominently in the image. In one GAN reconstruction, a pair of newlyweds kissing on the steps of a church are ghosted out, leaving an eerie wedding-dress texture on the cathedral door.

"When GANs encounter objects they can't generate, they seem to imagine what the scene would look like without them," says Strobelt. "Sometimes people become bushes or disappear entirely into the building behind them."

The researchers suspect that machine laziness could be to blame; although a GAN is trained to create convincing images, it may learn it's easier to focus on buildings and landscapes and skip harder-to-represent people and cars. Researchers have long known that GANs have a tendency to overlook some statistically meaningful details. But this may be the first study to show that state-of-the-art GANs can systematically omit entire classes of objects within an image.

An AI that drops some objects from its representations may achieve its numerical goals while missing the details most important to us humans, says Bau. As engineers turn to GANs to generate synthetic images to train automated systems like self-driving cars, there's a danger that people, signs, and other critical information could be dropped without humans realizing. It shows why model performance shouldn't be measured by accuracy alone, says Bau. "We need to understand what the



networks are and aren't doing to make sure they are making the choices we want them to make."

More information: Seeing What a GAN Cannot Generate: <u>ganseeing.csail.mit.edu/</u>

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: New tool highlights what generative models leave out when reconstructing a scene (2019, November 11) retrieved 26 April 2024 from <u>https://techxplore.com/news/2019-11-tool-highlights-reconstructing-scene.html</u>

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.