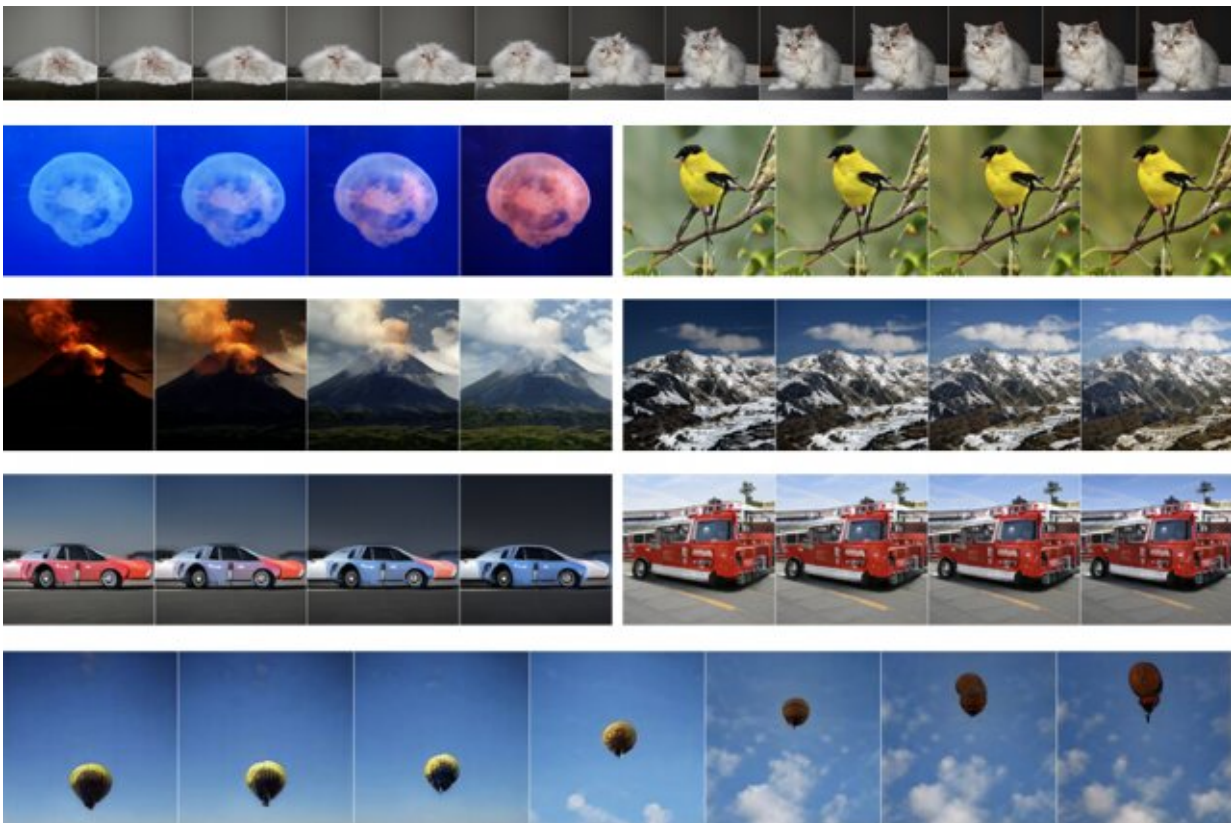


Training AI to generate varied poses and colors of objects and animals in photos

May 7 2020, by Kim Martineau



MIT researchers have developed a kind of creativity test for generative models to see how far they can go in visualizing objects in photos from various angles and in different colors. The tool shows how much the model's imagination depends on the variety of images it has seen. Credit: Ali Jahanian/MIT

Most firetrucks come in red, but it's not hard to picture one in blue.

Computers aren't nearly as creative.

Their understanding of the world is colored, often literally, by the data they've trained on. If all they've ever seen are pictures of red fire trucks, they have trouble drawing anything else.

To give computer vision models a fuller, more imaginative view of the world, researchers have tried feeding them more varied images. Some have tried [shooting objects](#) from odd angles, and in unusual positions, to better convey their real-world complexity. Others have asked the models to generate pictures of their own, using a form of artificial intelligence called GANs, or generative adversarial networks. In both cases, the aim is to fill in the gaps of image datasets to better reflect the three-dimensional world and make face- and object-recognition models less biased.

In [a new study](#) at the International Conference on Learning Representations, MIT researchers propose a kind of creativity test to see how far GANs can go in riffing on a given image. They "steer" the [model](#) into the subject of the photo and ask it to draw objects and animals close up, in bright light, rotated in space, or in different colors.

The model's creations vary in subtle, sometimes surprising ways. And those variations, it turns out, closely track how creative human photographers were in framing the scenes in front of their lens. Those biases are baked into the underlying dataset, and the steering method proposed in the study is meant to make those limitations visible.

"Latent space is where the DNA of an image lies," says study co-author Ali Jahanian, a research scientist at MIT. "We show that you can steer into this abstract space and control what properties you want the GAN to express—up to a point. We find that a GAN's creativity is limited by the diversity of images it learns from." Jahanian is joined on the study by co-

author Lucy Chai, a Ph.D. student at MIT, and senior author Phillip Isola, the Bonnie and Marty (1964) Tenenbaum CD Assistant Professor of Electrical Engineering and Computer Science.

The researchers applied their method to GANs that had already been trained on ImageNet's 14 million photos. They then measured how far the models could go in transforming different classes of animals, objects, and scenes. The level of artistic risk-taking, they found, varied widely by the type of subject the GAN was trying to manipulate.

For example, a rising hot air balloon generated more striking poses than, say, a rotated pizza. The same was true for zooming out on a Persian cat rather than a robin, with the cat melting into a pile of fur the farther it recedes from the viewer while the bird stays virtually unchanged. The model happily turned a car blue, and a jellyfish red, they found, but it refused to draw a goldfinch or firetruck in anything but their standard-issue colors.

The GANs also seemed astonishingly attuned to some landscapes. When the researchers bumped up the brightness on a set of mountain photos, the model whimsically added fiery eruptions to the volcano, but not a geologically older, dormant relative in the Alps. It's as if the GANs picked up on the lighting changes as day slips into night, but seemed to understand that only volcanoes grow brighter at night.

The study is a reminder of just how deeply the outputs of deep learning models hinge on their data inputs, researchers say. GANs have caught the attention of intelligence researchers for their ability to extrapolate from data, and visualize the world in new and inventive ways.

They can take a headshot and transform it into a Renaissance-style portrait or favorite celebrity. But though GANs are capable of learning surprising details on their own, like how to divide a landscape into clouds

and trees, or generate images that stick in people's minds, they are still mostly slaves to data. Their creations reflect the biases of thousands of photographers, both in what they've chosen to shoot and how they framed their subject.

"What I like about this work is it's poking at representations the GAN has learned, and pushing it to reveal why it made those decisions," says Jaako Lehtinen, a professor at Finland's Aalto University and a research scientist at NVIDIA who was not involved in the study. "GANs are incredible, and can learn all kinds of things about the physical world, but they still can't represent images in physically meaningful ways, as humans can."

More information: On the "Steerability" of Generative Adversarial Networks: openreview.net/pdf?id=HylsTT4FvB

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: Training AI to generate varied poses and colors of objects and animals in photos (2020, May 7) retrieved 10 April 2024 from <https://techxplore.com/news/2020-05-ai-varied-poses-animals-photos.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.
