The Synapse: Intercollegiate science magazine

Volume 20 | Issue 1 Article 14

2019

How to Fake a Face: AI Systems Create Realistic Faces From Scratch

Reuben Duebester

Follow this and additional works at: https://digitalcommons.denison.edu/synapse

 Part of the Life Sciences Commons, and the Physical Sciences and Mathematics Commons

Recommended Citation

Duebester, Reuben (2019) "How to Fake a Face: AI Systems Create Realistic Faces From Scratch," *The Synapse: Intercollegiate science magazine*: Vol. 20: Iss. 1, Article 14.
Available at: https://digitalcommons.denison.edu/synapse/vol20/iss1/14

This Article is brought to you for free and open access by Denison Digital Commons. It has been accepted for inclusion in The Synapse: Intercollegiate science magazine by an authorized editor of Denison Digital Commons.

How to Fake a Face

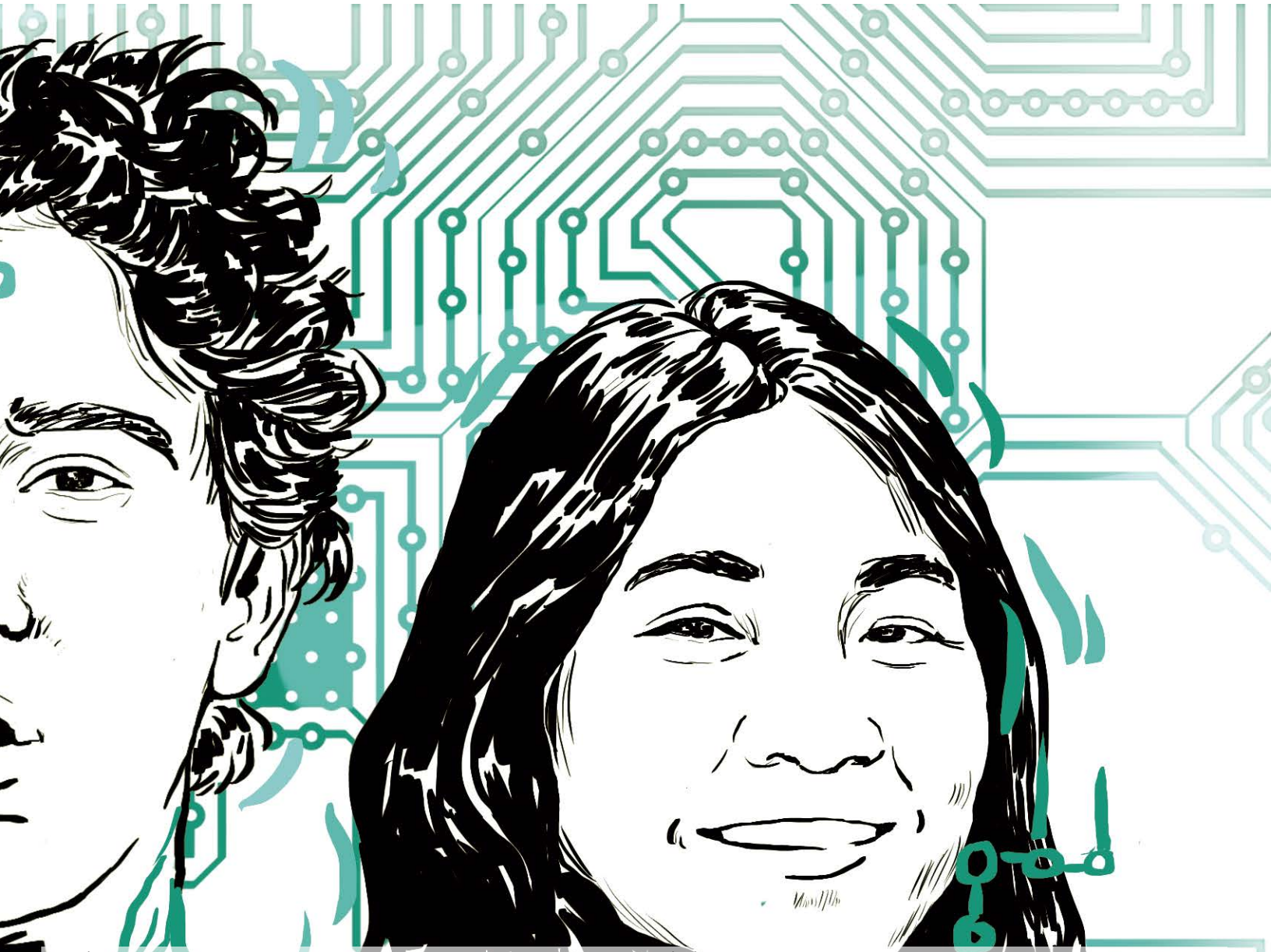
AI Systems Create Realistic Faces From Scratch

Written by Reuben Duebester
Illustrated by Cecilia Larson

How do you know what you see, when you see it? This seemingly simple question has been a significant problem in computer science for almost 50 years. Image recognition, also referred to as computer vision, is the task of extracting high-level information from an image or video. Humans are very talented at performing this task, capable of recognizing objects in images even when they have been distorted or partially obscured. Developing an algorithm that matches this competence has proven surprisingly difficult. Attempts at creating image processing algorithms began in the late 1960s alongside the emergence of artificial intelligence as a field, but many of the current algorithms and methods were not developed until the late 1980s and early 1990s. Even the best algorithms in use today don't match human perception.

One particular area of success in image processing technology has been facial recognition. If you have ever used FaceID on a smartphone, you have taken advantage of this success. According to National Public Radio, the facial recognition algorithms developed at Facebook and Google have recognition rates of 98 percent and 94 percent respectively, while the current algorithm used by the Federal Bureau of Investigation only scores an 85 percent success rate. Artificial intelligence systems are now very good at recognizing faces, and this progress has led researchers to ask a new question: If computers can recognize a face, can they generate their own pictures of faces that are indistinguishable from the real thing? In the last few years, Generative Adversarial Networks have emerged as a promising tool for creating highly convincing faces entirely from scratch. In April 2018, a remarkably powerful new algorithm developed at a large American technology company, NVIDIA, led to a flurry of activity in the field. These algorithms are already impacting society in subtle ways, and raise serious questions in digital ethics that we will soon need to address.

The best way to understand Generative Adversarial Networks is to break the phrase down into its parts. First of all, what is a network? Networks, commonly referred to as neural networks, are modeled on how neurons work in nature. In the human brain, thousands of neurons are wired together. When a neuron receives a stimulus in the form of an electric charge, it distributes that stimulus to neighboring neurons. These complex sequences of neuron activation form the basis for cognition and sensory processing. At a basic level, an artificial neural network works in a similar way. First, simulated "neurons" are arranged into layers. Each neuron in a layer receives a stimulus value from neurons in the layer above, it then combines these inputs, and processes them using some variable weight, before feeding the result to the neurons in the layer below until it reaches the final layer. The final layer of the network produces a numeric output that a human can interpret. These networks can be trained to do a specific task—such as return a "yes" value if an image contains a cat—by changing the weights of each neuron. A common strategy is to feed the network a dataset with known values —such as pictures prelabeled "cat" or "not cat"—and then modifying the network until



it makes correct guesses.

Next, we need to understand what it means for a network to be adversarial. Adversarial learning is a different strategy for teaching a neural network to perform a task. Two neural networks are created and pitted against each other in competition. One neural network—the discriminator—is trained in advance to perform a specific task, such as recognizing photos containing cats. A second network—the generator—is then given the following challenge: produce output that will fool the discriminator! A fierce competition ensues. In our example, the discriminator will constantly improve at distinguishing cats from non-cats, but the generator will also constantly improve at producing cat-like images. This naturally brings us to the final part of the definition. The system is referred to as generative because the end result is a neural network that has been trained to create original images that are—hopefully—indistinguishable from the ones originally used to train the discriminator.

Generative Adversarial Networks still have a lot of room for improvement. They can produce novel static images quickly and reliably but are incapable of producing completely original and believable video footage. GANs have also been used successfully to transfer facial data from one video source onto another. This technology broke into mainstream news in early 2018, when users on Reddit created a community dedicated to sharing fake pornographic content featuring celebrities. Although this content was quickly banned, the tools remain freely available for anyone to experiment with them.

Fake video clips produced by these tools are currently limited to just a few seconds and contain glitches and artifacts that make them easy to spot. Additionally, producing these false videos is a very difficult computational task, requiring hours of processing time on dedicated cloud-computing systems. Unfortunately, the technology doesn't need to be perfect to be dangerous. Security experts and lawmakers are becoming increasingly concerned that video faking technology could be used in online disinformation campaigns as soon as the 2020 elections. The best GANs may never replicate the subtle nuances of human behavior, like blink frequency and microexpressions, but that probably won't matter if the content is politically charged and disseminated over social media.

Along with its potential as a powerful weapon for disinformation and smear campaigns, GANs could also create an environment where the authenticity of any legitimate photo or video could be called into question. Over time, this technology will only get more convincing and more user-friendly. Soon anyone with a basic technological understanding might be able to fabricate convincing video footage on their personal computers. It will take increasing vigilance to sort real footage from fake, especially video content shared on social networks with the intent to shape political opinion. Which brings us back to the same question with which we began: How do you know what you see, when you see it? • • •