

Beyond deep fakes: Transforming video content into another video's style, automatically

11 September 2018



Researchers at Carnegie Mellon University have devised a way to automatically transform the content of one video into the style of another, making it possible to transfer the facial expressions of one person to the video of another person, or even a cartoon character. Credit: Carnegie Mellon University

Researchers at Carnegie Mellon University have devised a way to automatically transform the content of one video into the style of another, making it possible to transfer the facial expressions of comedian John Oliver to those of a cartoon character, or to make a daffodil bloom in much the same way a hibiscus would.

Because the data-driven method does not require human intervention, it can rapidly transform large amounts of [video](#), making it a boon to movie production. It can also be used to convert black-and-white films to color and to create content for virtual reality experiences.

"I think there are a lot of stories to be told," said Aayush Bansal, a Ph.D. student in CMU's Robotics Institute. Film production was his primary motivation in helping devise the method, he explained, enabling movies to be produced more

quickly and cheaply. "It's a tool for the artist that gives them an initial model that they can then improve," he added.

The technology also has the potential to be used for so-called "deep fakes," videos in which a person's image is inserted without permission, making it appear that the person has done or said things that are out of character, Bansal acknowledged.

"It was an eye opener to all of us in the field that such fakes would be created and have such an impact," he said. "Finding ways to detect them will be important moving forward."

Bansal will present the method today at ECCV 2018, the European Conference on Computer Vision, in Munich. His co-authors include Deva Ramanan, CMU associate professor of robotics.

Transferring content from one video to the style of another relies on artificial intelligence. In particular, a class of algorithms called generative adversarial networks (GANs) have made it easier for computers to understand how to apply the style of one image to another, particularly when they have not been carefully matched.

In a GAN, two models are created: a discriminator that learns to detect what is consistent with the style of one image or video, and a generator that learns how to create images or videos that match a certain style. When the two work competitively—the generator trying to trick the discriminator and the discriminator scoring the effectiveness of the generator—the system eventually learns how content can be transformed into a certain style.

A variant, called cycle-GAN, completes the loop, much like translating English speech into Spanish

and then the Spanish back into English and then evaluating whether the twice-translated speech still makes sense. Using cycle-GAN to analyze the spatial characteristics of images has proven effective in transforming one image into the [style](#) of another.

That spatial method still leaves something to be desired for video, with unwanted artifacts and imperfections cropping up in the full cycle of translations. To mitigate the problem, the researchers developed a technique, called Recycle-GAN, that incorporates not only spatial, but temporal information. This additional information, accounting for changes over time, further constrains the process and produces better results.

The researchers showed that Recycle-GAN can be used to transform video of Oliver into what appears to be fellow comedian Stephen Colbert and back into Oliver. Or video of John Oliver's face can be transformed a cartoon character. Recycle-GAN allows not only facial expressions to be copied, but also the movements and cadence of the performance.

The effects aren't limited to faces, or even bodies. The researchers demonstrated that video of a blooming flower can be used to manipulate the image of other types of flowers. Or clouds that are crossing the sky rapidly on a windy day can be slowed to give the appearance of calmer weather.

Such effects might be useful in developing self-driving cars that can navigate at night or in bad weather, Bansal said. Obtaining video of night scenes or stormy weather in which objects can be identified and labeled can be difficult, he explained. Recycle-GAN, on the other hand, can transform easily obtained and labeled daytime scenes into nighttime or stormy scenes, providing images that can be used to train cars to operate in those conditions.

More information:

www.cs.cmu.edu/~aayushb/Recycle-GAN/

Provided by Carnegie Mellon University

APA citation: Beyond deep fakes: Transforming video content into another video's style, automatically (2018, September 11) retrieved 3 October 2022 from <https://techxplore.com/news/2018-09-deep-fakes-video-content-style.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.