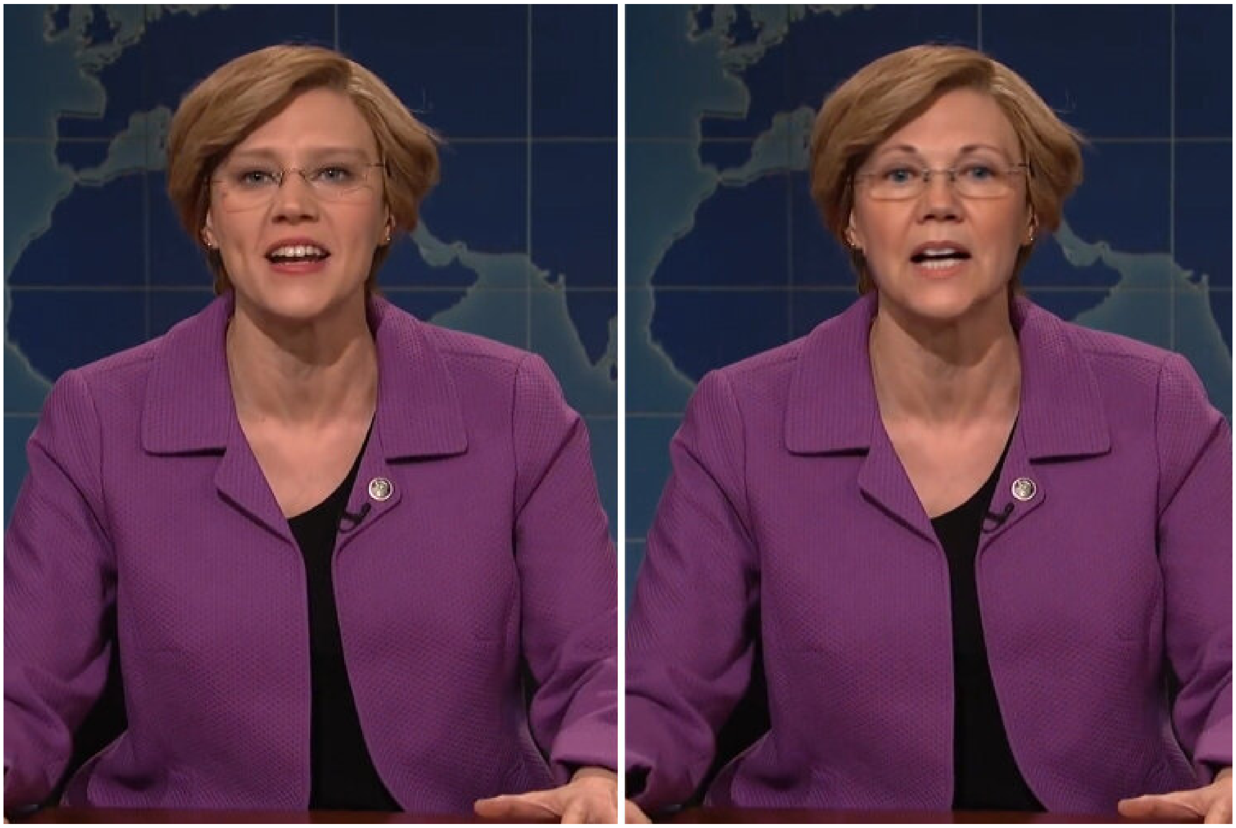
TechXplore

Researchers use facial quirks to unmask 'deepfakes'

June 19 2019, by Kara Manke



On the left, Saturday Night Live star Kate McKinnon impersonates Elizabeth Warren during a skit, and on the right, face swap deepfake technology has been used to superimpose Warren's face onto that of McKinnon. Credit: UC Berkeley photo by Stephen McNally

After watching hours of video footage of former President Barack

Obama delivering his weekly address, Shruti Agarwal began to notice a few quirks about the way Obama speaks.

"Every time he says 'Hi, everybody,' he moves his head up to the left or the right, and then he purses his lips," said Agarwal, a computer science graduate student at UC Berkeley.

Agarwal and her thesis advisor Hany Farid, an incoming professor in the Department of Electrical Engineering and Computer Science and in the School of Information at UC Berkeley, are racing to develop digital forensics tools that can unmask "deepfakes," hyper-realistic AI-generated videos of people doing or saying things they never did or said.

Seeing these patterns in the real Obama's speech gave Agarwal an idea.

"I realized that there is one thing common among all these deepfakes, and that is that they tend to change the way a person talks," Agarwal said.

Agarwal's insight led her and Farid to create the latest weapon in the war against deepfakes: a new forensic approach that can use the subtle characteristics of how a person speaks, such as Obama's distinct head nods and lip purses, to recognize whether a new video of that individual is real or a fake.

Their technique, which Agarwal presented this week at the Computer Vision and Pattern Recognition conference in Long Beach, CA, could be used to help journalists, policy makers, and the public stay one step ahead of bogus videos of political or economic leaders that could be used to swing an election, destabilize a financial market, or even incite civil unrest and violence.

"Imagine a world now, where not just the news that you read may or may not be real—that's the world we've been living in for the last two years, since the 2016 elections—but where the images and the videos that you see may or may not be real," said Farid, who begins his tenure at UC Berkeley on July 1. "It is not just about these latest advances in creating fake images and video. It is the injection of these techniques into an ecosystem that is already promoting fake news, sensational news and conspiracy theories."

The new technique works because all three of the most common deepfake techniques—known as "lip-sync," "face swap," and "puppet-master,"—involve combining audio and video from one source with an image from another source, creating a disconnect that may be uncovered by a keen viewer—or a sophisticated computer model.

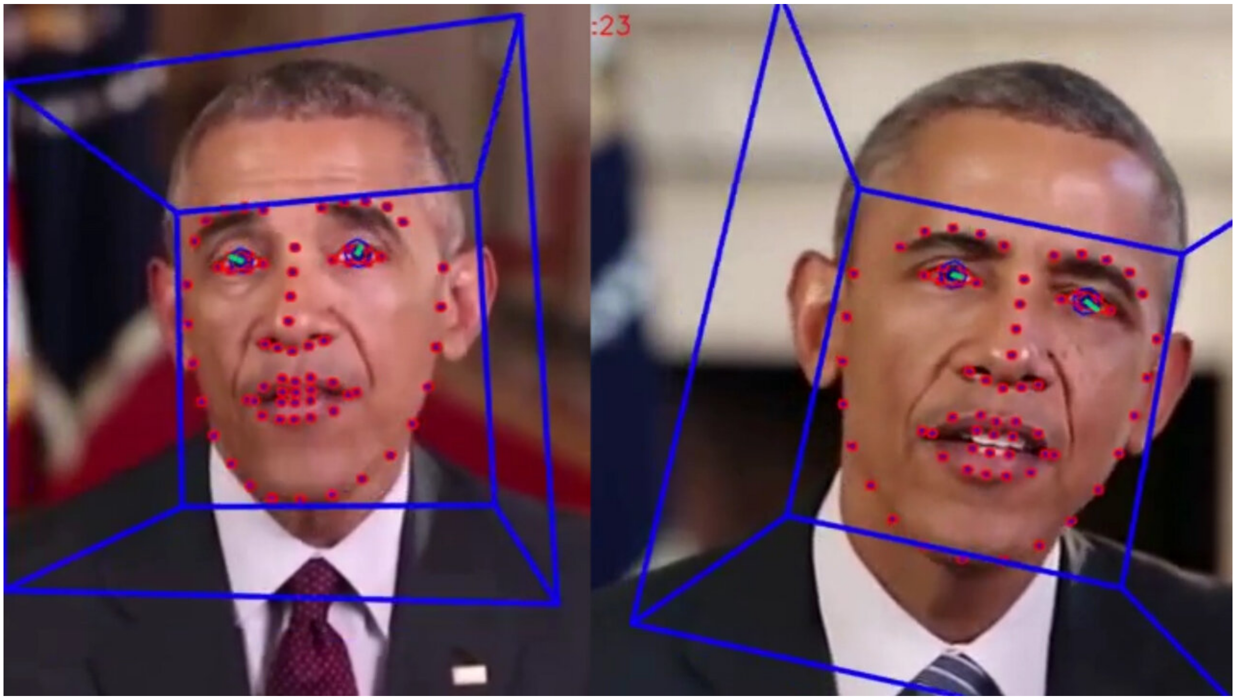
Using the "face swap" technique, for example, one could create a deepfake of Donald Trump by superimposing Trump's face onto a video of Alec Baldwin doing an impersonation of Trump, so that it is almost as if Baldwin is wearing a skin-tight Trump mask. But Baldwin's facial expressions will still show through the mask, Agarwal said.

"The new image that is created will have the expressions and facial behavior of Alec Baldwin, but the face of Trump," Agarwal said.

Likewise, in a "lip-sync" deepfake, AI algorithms take an existing video of a person talking, and alter the lip movements in the video to match that of a new audio, where the audio may be an older speech taken out of context, an impersonator speaking, or synthesized speech. Last year, actor and director Jordan Peele used this technique to create a viral video of Obama saying inflammatory things about president Trump.

But in these videos, only the lip movements are changed, so the expressions on the rest of the face may no longer match the words being spoken.

To test the idea, Agarwal and Farid gathered video footage of five major political figures—Hillary Clinton, Barack Obama, Bernie Sanders, Donald Trump and Elizabeth Warren—and ran them through the open-source facial behavior analysis toolkit OpenFace2, which picked out facial tics like raised brows, nose wrinkles, jaw drops and pressed lips.



OpenFace tracking software analyzes a real video of President Obama on the left, and a “lip-sync” deepfake on the right. Credit: UC Berkeley photo by Stephen McNally

They then used the outputs to create what the team calls "soft biometric" models, which correlate facial expressions and head movements for each political leader. They found each leader had a distinct way of speaking and, when they used these models to analyze real videos and deepfakes created by their collaborators at the University of Southern California, they found the models could accurately tell the real from the fake between 92 and 96 percent of the time, depending on the leader and length of the video.

"The basic idea is we can build these soft biometric models of various world leaders, such as 2020 presidential candidates, and then as the videos start to break, for example, we can analyze them and try to determine if we think they are real or not," Farid said.

Unlike some digital forensics techniques, which identify fakes by spotting image artifacts left behind during the fabrication process, the new method can still recognize fakes that have been altered through simple digital processing like resizing or compressing.

But it's not foolproof. The technique works well when applied to political figures giving speeches and formal addresses because they tend to stick to well-rehearsed behaviors in these settings. But it might not work as well for videos of these people in other settings: for example, Obama may not give his same characteristic head nod when greeting his buddies.

Deepfake creators could also become savvy to these speech patterns and learn to incorporate them into their videos of world leaders, the researchers said.

Agarwal says she hopes the new approach will help buy a little time in the ever-evolving race to spot deepfakes.

"We are just trying to gain a little upper-hand in this cat and mouse game of detecting and creating new deepfakes," Agarwal said.

Provided by University of California - Berkeley

Citation: Researchers use facial quirks to unmask 'deepfakes' (2019, June 19) retrieved 23 April 2024 from https://techxplore.com/news/2019-06-facial-quirks-unmask-deepfakes.html