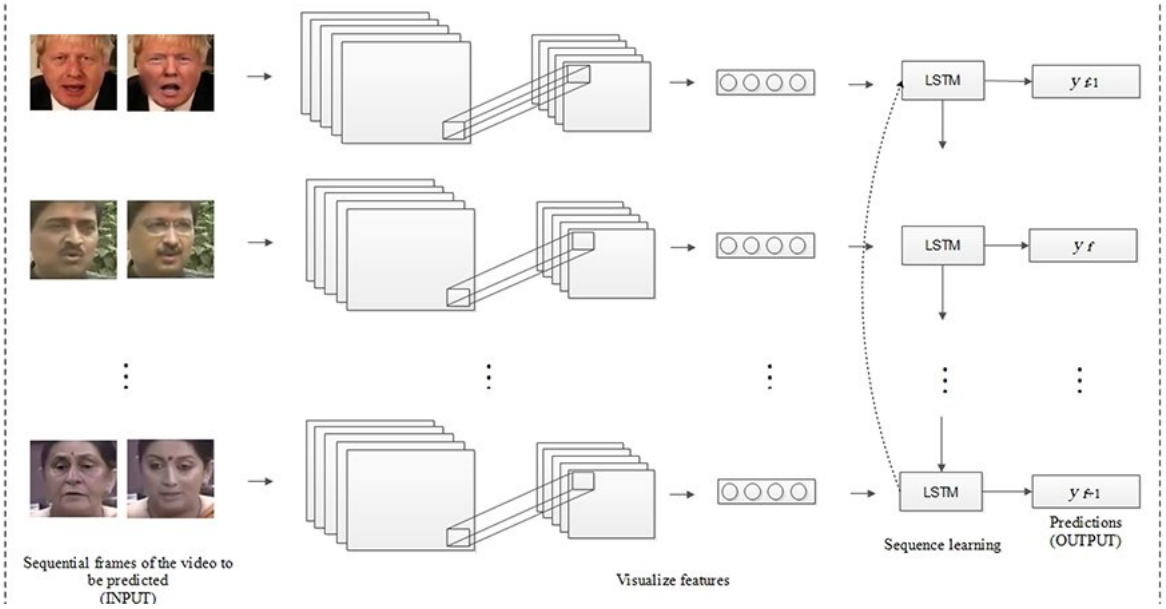
AI algorithm detects deepfake videos with high accuracy

July 28 2020



AI-based technology proposed by Kaur et al. can detect deepfake videos within seconds. Credit: SPIE

Artificial intelligence (AI) contributes significantly to good in the world. From reducing pollution to making roads safer with self-driving cars to enabling better healthcare through medical big-data analysis, AI still has plenty of untapped potential. Unfortunately, just like any technology in the world, AI can be used by those with less noble intentions.

Deepfake face swapping

Such is the case with a certain AI-based technique called "deepfake" (combination of "deep learning" and "fake"), which uses deep neural networks to easily create fake videos in which the face of one person is superimposed on that of another. These tools are easy to use, even for people with no background in programming or video editing. This technique can be used to create compromising videos of virtually anyone, including celebrities, politicians, and corporate public figures. In this era of unprecedented connectedness and instant communication-when news can become viral in a matter of hours-such videos can cause great harm to those in the videos as well as to the social and cultural psyche of the associated communities.

AI to the rescue

"Deepfakes" are out there. But the harm can be mitigated if these videos can be automatically detected. And what better way to do that than to use AI itself?

While AI-based deepfake video detection methods do exist, researchers from Thapar Institute of Engineering and Technology and Indraprastha Institute of Information Technology in India have developed a new algorithm with increased accuracy and precision. Their work could be a milestone in the quest to counter one of the many kinds of infodemics that we face today.

Broadly, their approach consists of a classifier that decides whether an input video is real or (deep) fake. To make accurate decisions possible, like all AI-based applications, their algorithm had to be trained first. To this end, the researchers began by creating a dataset of 200 videos of similar-looking pairs of politicians; 100 of these were real and the other

100 were generated using deepfake. A portion of the frames of these videos were labeled and fed to the algorithm as training data; the rest were used as a validation dataset to test if the program could correctly catch face-swapped videos.

Deepfake detection method

The algorithm itself can be divided into two levels. At the first level, the video frames undergo some light image processing, such as rescaling, zooming, and horizontal flipping, as preparation for subsequent stages. The second level comprises two main components: a convolutional neural network (CNN) and a long short-term memory (LSTM) stage.

The CNN is a special type of neural network that can automatically extract features from sequential video frames. What features are extracted and how they are defined is known only to the CNN. The LSTM is a type of recurrent neural network that is especially useful for processing time-series data (in this case, sequential video frames). Upon comparing the original and deepfake videos, the LSTM network easily detects inconsistencies in the frames of the latter. The algorithm can identify them using only about two seconds of video material.

The performance of this new method was tested and compared with that of other existing state-of-the-art AI-based techniques for detecting deepfake videos. For a total of 181,608 real and deepfake frames retrieved from the dataset of videos collated in this study, the proposed method achieved higher accuracy (98.21%) and precision (99.62%), with a much lower total training time.

This study highlights the boon and the bane of AI. While AI can be used to damage reputations and spread misinformation, it can also be used to prevent infodemics, the harms of which are sharply felt amid crises like the COVID-19 pandemic.

More information: Sawinder Kaur et al. Deepfakes: temporal sequential analysis to detect face-swapped video clips using convolutional long short-term memory, *Journal of Electronic Imaging* (2020). DOI: 10.1117/1.JEI.29.3.033013

Provided by SPIE