TechXplore

Research introduces new approach for detecting deepfakes

February 27 2024, by David Bradley

Credit: AI-generated image

Research published in the *International Journal of Ad Hoc and Ubiquitous Computing*introduces a new approach to tackling the challenges posed by deepfake technology, which generates manipulated media content that closely resembles authentic footage.

The novel method combines the miniXception and long short-term memory (LSTM) models to analyze suspicious content more effectively and identify deepfake images with greater than 99% accuracy.

While fake and fraudulent videos and images have been with us for many years, the term "deepfake" more commonly refers to manipulated videos or images that have been created using artificial intelligence and deep learning techniques. These technologies allow users to superimpose or replace, the original contents of an image or video with other content.

Commonly a person's face and voice might be faked in a video. Such deepfakes might be used for entertainment purposes as is the case with many apps that allow everyday users to create "amusing" content featuring their friends and family or indeed celebrities.

However, the more insidious use of deepfakes has gained popular attention because of the potential to deceive viewers, often leading to concerns about misinformation, privacy infringement, and the manipulation of public and political discourse.

Such videos represent a significant threat to democracy where voters and consumers alike might be exposed to seemingly legitimate political content that is faked propaganda with malicious intent. Identifying deepfake content is more important than ever at a time of heightened political tensions and fragility. There is an urgent need for powerful detection methods and awareness about their existence and potential consequences.

Until now, deepfake detection has been hindered by low accuracy rates and difficulties in generalizing across different datasets. Yong Liu, Xu Zhao, and Ruosi Cheng of the PLA Strategic Support Force Information Engineering University in Henan, Tianning Sun of the Zhejiang Lab, Zonghui Wang of Zhejiang University, China, and Baolan Shi of the

University of Colorado Boulder in Boulder, Colorado, U.S., have proposed a model that improves on the accuracy of earlier approaches.

The team conducted cross-dataset training and testing, employing transfer learning methods to improve the model's ability to generalize across various datasets. They used focal loss during training to balance samples and enhance generalization still further.

Their tests demonstrate the promise of this approach, showing a detection accuracy of 99.05% on the FaceSwap dataset. This is better than previous methods, such as CNN-GRU, and requires fewer parameters to achieve this level of success.

More information: Yong Liu et al, Detection of deepfake technology in images and videos, *International Journal of Ad Hoc and Ubiquitous Computing* (2024). DOI: 10.1504/IJAHUC.2024.136851

Provided by Inderscience

Citation: Research introduces new approach for detecting deepfakes (2024, February 27) retrieved 12 May 2024 from https://techxplore.com/news/2024-02-approach-deepfakes.html