Cat-and-Mouse: Adversarial Teaming for Improving Generation and Detection Capabilities of Deepfakes

Prospective research effort
Project Leader: Dr. Yajie Zhao

FIGURE 1. VGL Virtual human quality demonstration. Left: CG character captured, reconstructed, rendered using ICT - VGL light stage and process technologies; Right: Real Photo.

Background

The rapid evolution of artificial intelligence (AI) technologies, particularly in the realm of computer vision and Generative AI, has given rise to both innovative opportunities and potential threats. Deepfake, deep capture and GenAI technology, in particular, pose a significant risk to national security by enabling the creation of highly convincing, yet entirely fabricated, multimedia content. Ensuring national security against misinformation and disinformation is critical, particularly with the rise of generative AI (GenAI). Traditional detection models, trained on data from older techniques, face challenges in identifying deepfakes produced by advanced AI systems. This underscores the need for ongoing research and the development of adaptive defense mechanisms to counter the evolving threats posed by malicious GenAI applications.

Objectives

The objective of this research is to establish a co-development environment for a CG+AI pipeline for rapid, high-quality deepfake database generation. This dynamic framework will enable continuous updates and exploration of generation quality limits, providing high-quality training and testing data for detection models. By integrating generation and detection advancements, this approach ensures deepfake detection remains resilient against evolving techniques. To test the quality of our generator and database, we will evaluate it using off-the-shelf and state-of-the-art deepfake detectors available in the market.

Deepfake Generation

Our goal is to generate photorealistic deepfake videos using advanced computer graphics and AI to test deepfake detectors. We will create three types: (1) Motion-driven videos, where a target mimics a source's expressions using 3D face reconstruction and pixel-level tracking; (2) Text-driven videos, leveraging text-to-audio and lip-sync techniques; and (3) Multimodal videos, combining synthesized voices with video to challenge detectors that lack robust audiovisual datasets.

FIGURE 2. CG to Real demonstration. Left: CG character created by our AI model with semantic attributes: female, Slavic, 35 years old. Right: After an AI pass to add details and photo realism to the CG character.

To achieve our goal, we propose creating a novel deepfake generation pipeline that combines CG algorithms with AI models. Pure AI generative models typically lack precise control over video details (e.g., lip sync, motion, expressions). While traditional CG pipelines used in movie studios can achieve perfect control, realism, and quality, they are extremely expensive even for creating small clips. To bridge this gap, we propose using 3D virtual humans/CG characters as controls and AI models as the final rendering medium to create an efficient, low-cost deepfake video generator. We breakdown this task into three subtasks:

- High-Quality 3D Avatar Generative Model
- Text Driving sequence generation
- CG to Real

For more information, please visit https://vgl.ict.usc.edu/

Project Leader: Dr. Yajie Zhao

The project or effort depicted was or is sponsored by the U.S. Government and that the content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.