Self-Learning and Self-Evolving Detection of Altered, Deceptive Images and Videos

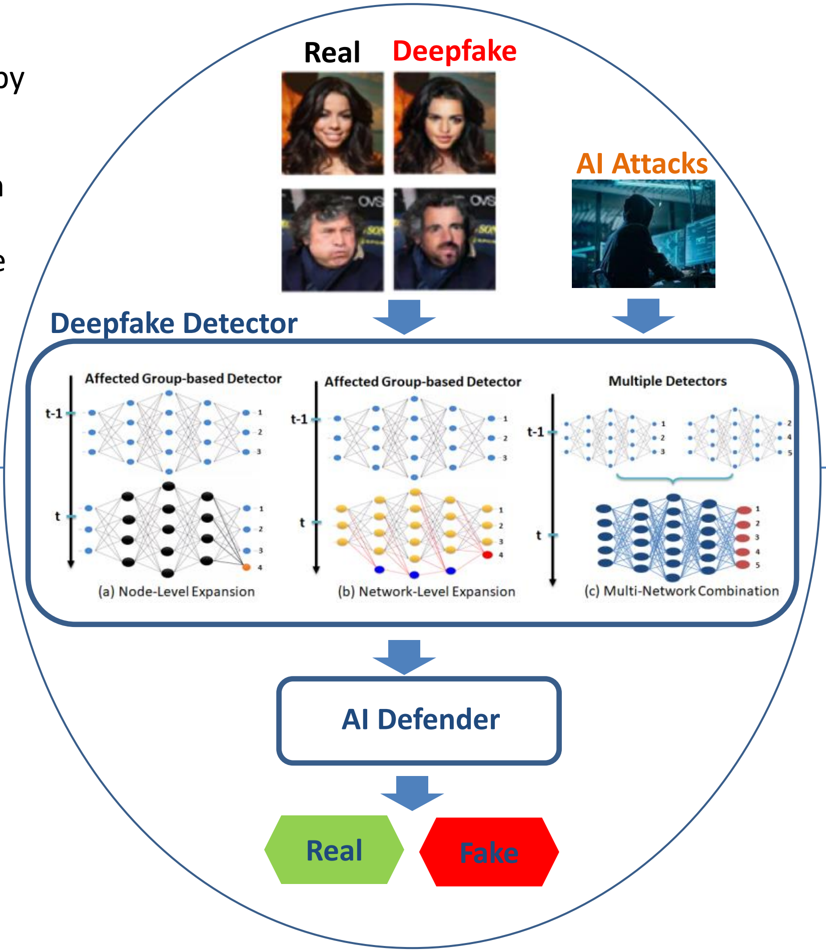
Challenges

- Fake images generated by deep neural networks are hard to be distinguished from real images by human eyes
- How can a deepfake detector achieve high detection rate with limited training samples?
- How can a deepfake detector be robust against AI attacks?
- How can a deepfake detector recognize new types of fake images that it has not been trained on?

Solution:

- Design efficient self-evolving mechanisms to gradually grow deepfake detectors
- Design robust self-defense mechanism against AI attacks on deepfake detectors
- Design lifelong self-learning mechanisms to predict new deepfakes

CNS-2027398: Dan Lin (University of Missouri)
CNS-2027114: Liyue Fan (UNCC)



Real

Deepfake

AI Attacks

Deepfake Detector

AI Defender

Real

Fake

Affected Group-based Detector
t-1
t
(a) Node-Level Expansion
Affected Group-based Detector
t-1
t
(b) Network-Level Expansion
Multiple Detectors
t-1
t
(c) Multi-Network Combination

Scientific Impact:

- Advance deep learning techniques to detect various types of fake images
- Advance AI defense techniques to identify and mitigate threats to AI algorithms
- Advance the capability of AI algorithms towards self-improvement

Broader Impact and Broader Participation:

- Prevent potential crisis caused by wide spread of fake information online
- Create a trustworthy and healthy social networking environment
- Provide research training to students especially those from under-represented groups
- Summer security camp for K-12 teachers