VIMAL

VISUAL INTELLIGENCE AND MULTIMEDIA ANALYTICS LABORATORY

FAKE VISUAL EVIDENCE

CREATION, DETECTION AND ADMISSIBILITY

Wael AbdAlmageed, Ph.D.

Research Director, **Information Sciences Institute**

Research Associate Professor, **Electrical and Computer**

Viterbi School of Engineering

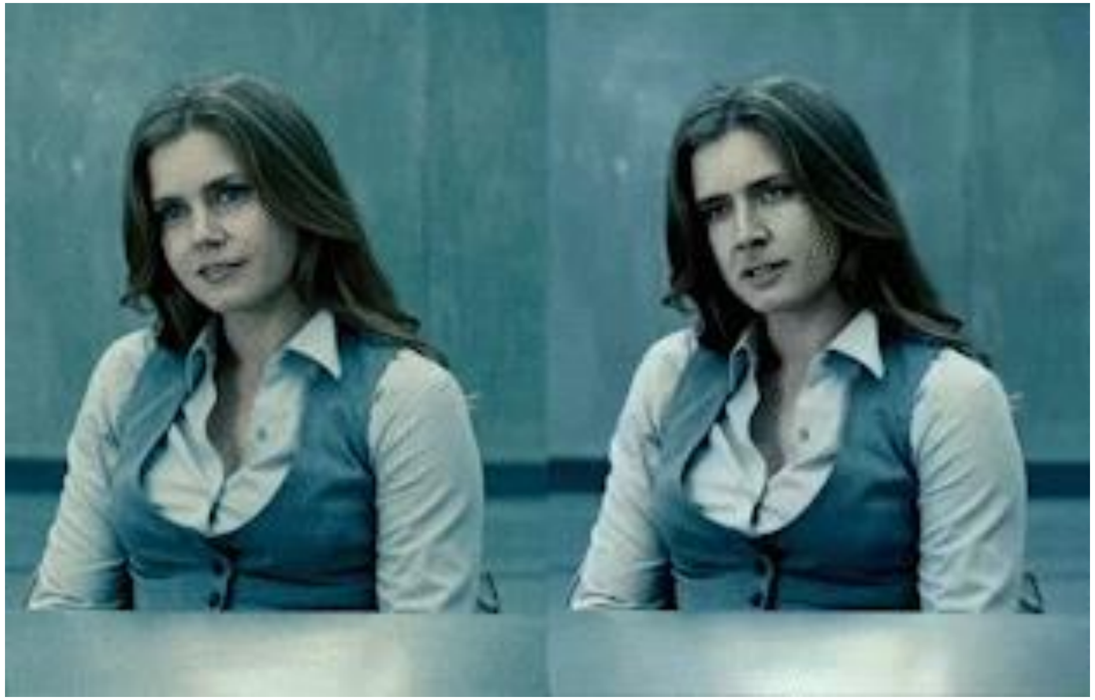
University of Southern California

USC Viterbi *Information Sciences Institute*

WHAT ARE DEEPFAKES?

DEEPFAKES

Fake video of a target person (i.e. victim) created using ***deep*** neural network



HOW LONG DID DEEPFAKES EXIST?

- Forever
- Paul Walker in Fast & Furious 7 replacement cost $50M

SO WHAT IS THE BIG DEAL?

ARTIFICIAL INTELLIGENCE (AI)

DEMOCRATIZATION OF AI

- Accessible
- Sharable

SOCIAL NETWORKS

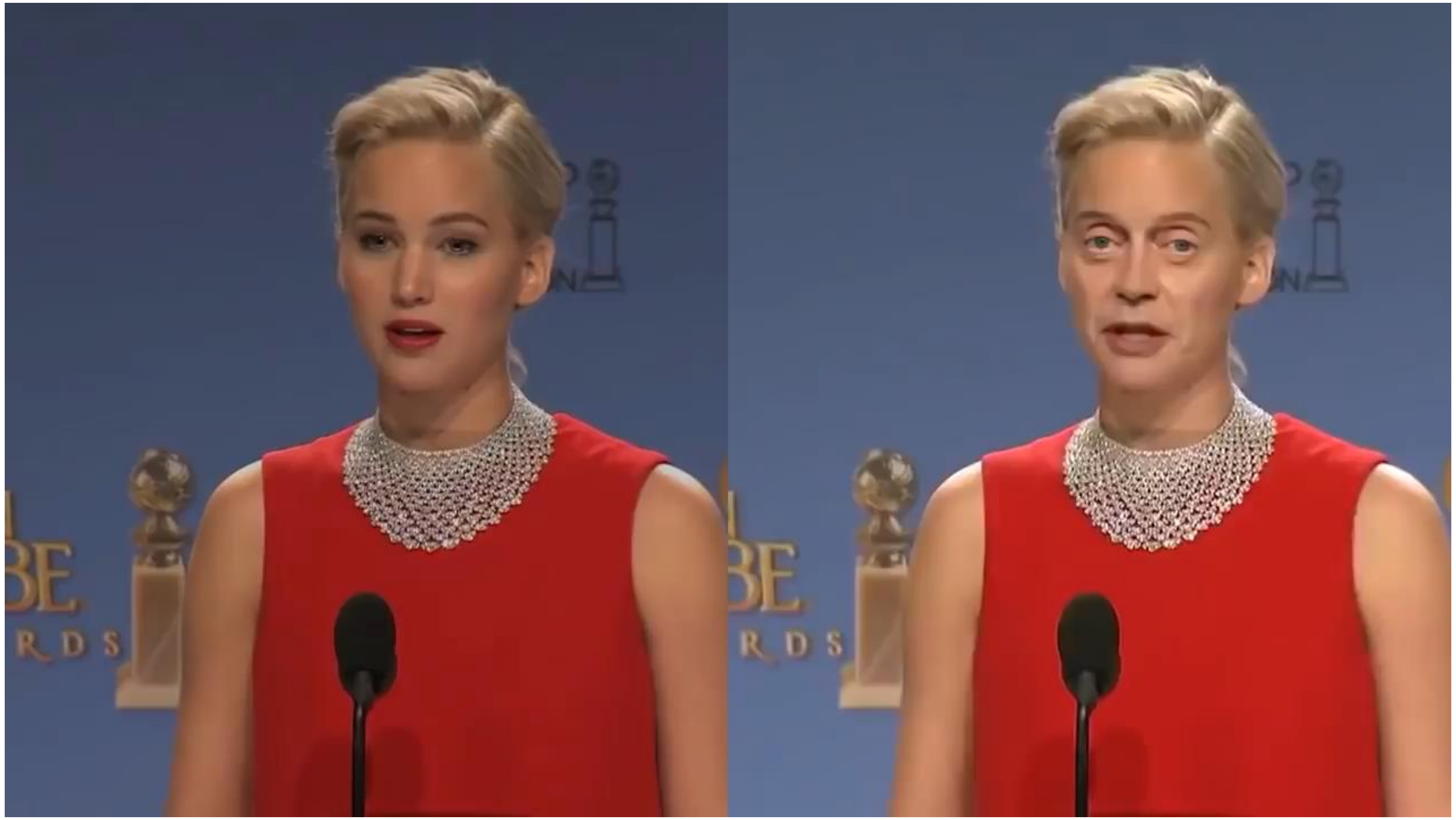
- Echo chambers
- Confirmation bias

DEEPFAKE TYPES

VIMAL

VISUAL INTELLIGENCE AND MULTIMEDIA ANALYTICS LABORATORY

FACE REPLACEMENT



USC Viterbi
Information Sciences Institute

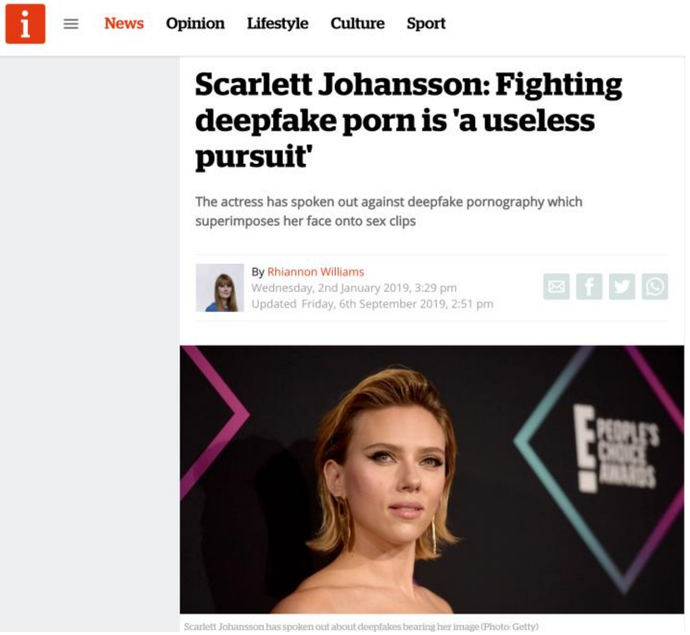
FACE REENACTMENT

DANGERS OF DEEPFAKES

LEGAL IMPLICATIONS

Jeremy Corbyn backing Boris Johnson

National security



Scarlett Johansson: Fighting deepfake porn is 'a useless pursuit'

The actress has spoken out against deepfake pornography which superimposes her face onto sex clips

By Rhiannon Williams
Wednesday, 2nd January 2019, 3:29 pm
Updated Friday, 6th September 2019, 2:51 pm

Scarlett Johansson has spoken out about deepfakes bearing her image (Photo: Getty)

Revenge porn

- Public relations
- Child exploitation
- Stock market
- Court room evidence
- Insurance fraud

STATISTICS

- **14,678** deepfake videos online
- **96%** of which were pornographic in nature
- **100%** featured women
- **134,364,438** views
- **Doubles** every **six** months

* The State Of The Deepfakes, Deeptrace, September 2019

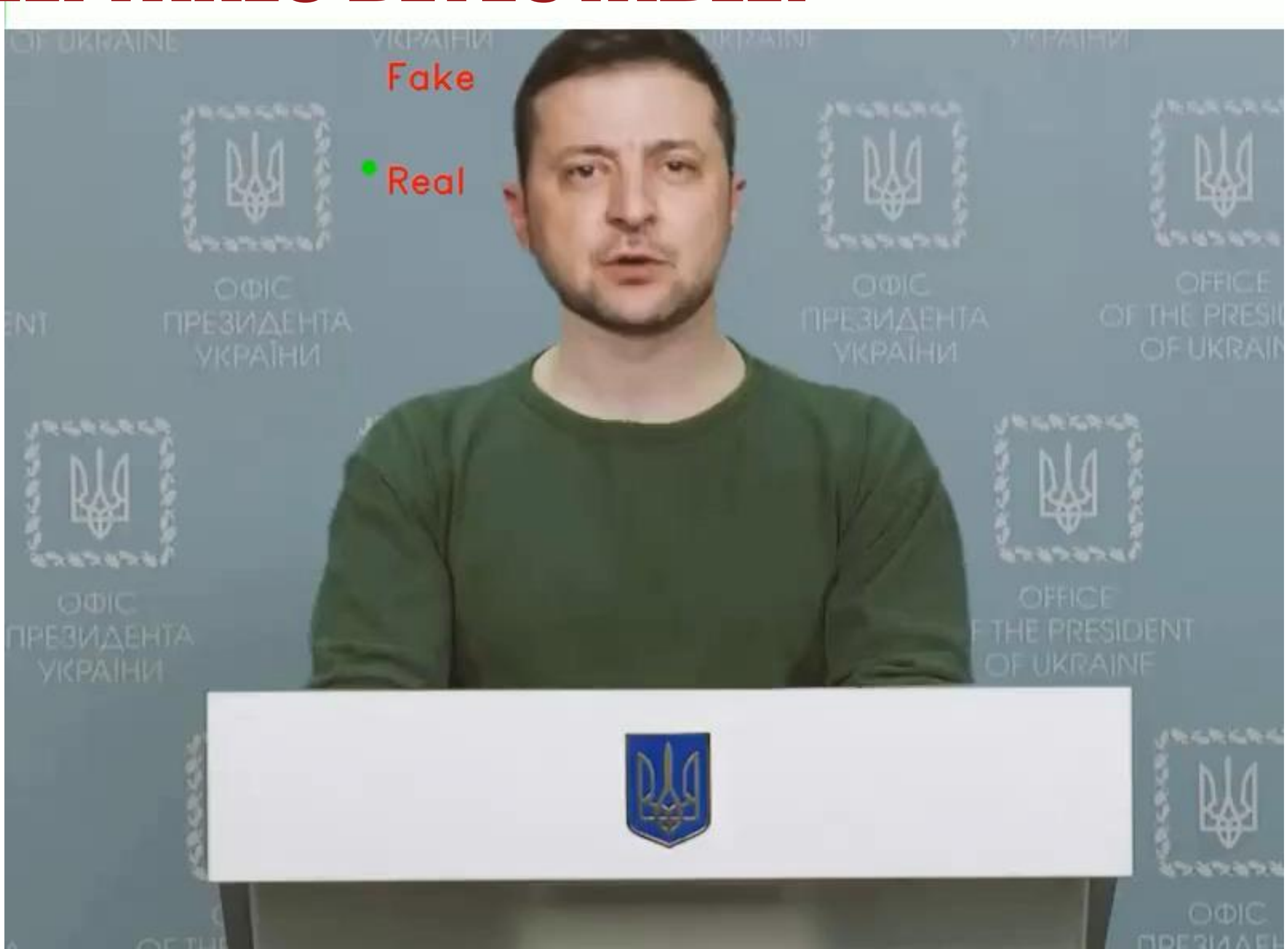
IDENTIFYING DEEPFAKES

ARE DEEPFAKES DETECTABLE?

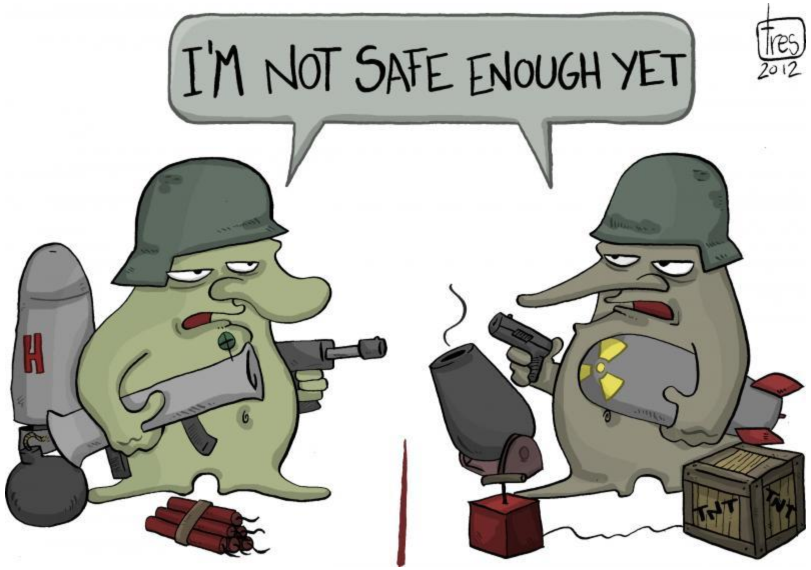
CSU
The times when we could completely count on others, they are over to a certain extent.

ARE DEEPFAKES DETECTABLE?

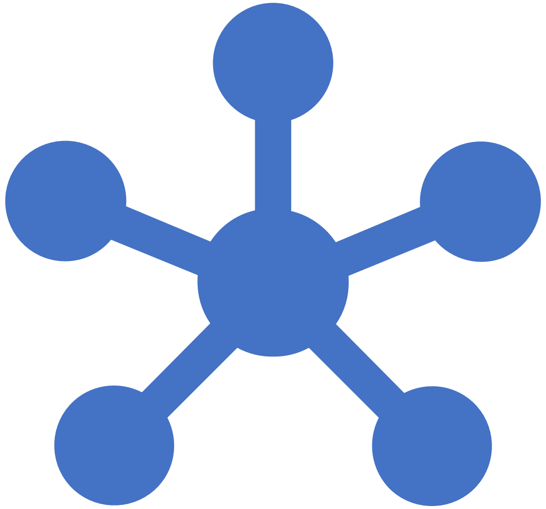
ARE DEEPFAKES DETECTABLE?



CHALLENGES



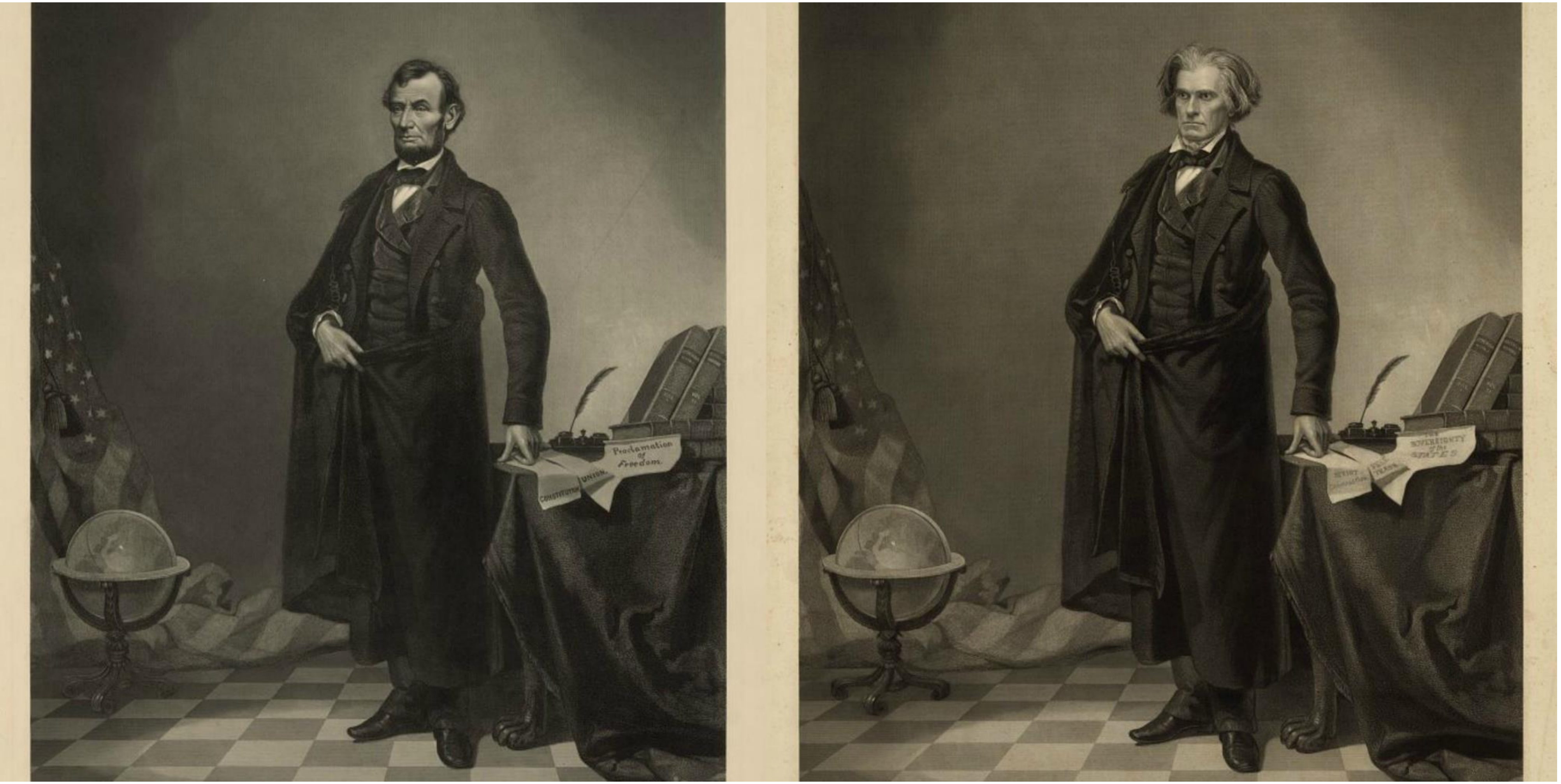
Detecting new deepfake generation methods



Social networks

IMAGE MANIPULATION

WHAT IS IMAGE MANIPULATION?



OBJECTIVE

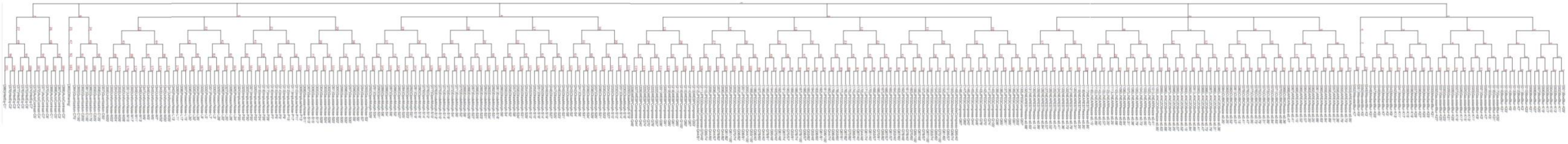
- Works across wide range of manipulations
  - At different granularities
- Capable of localizing manipulations
  - Can be used as forensics evidence
- Performs well on different evaluation datasets
- Does not need fine tuning
- Robust against anti-forensic techniques

BASIC MANIPULATION FAMILIES

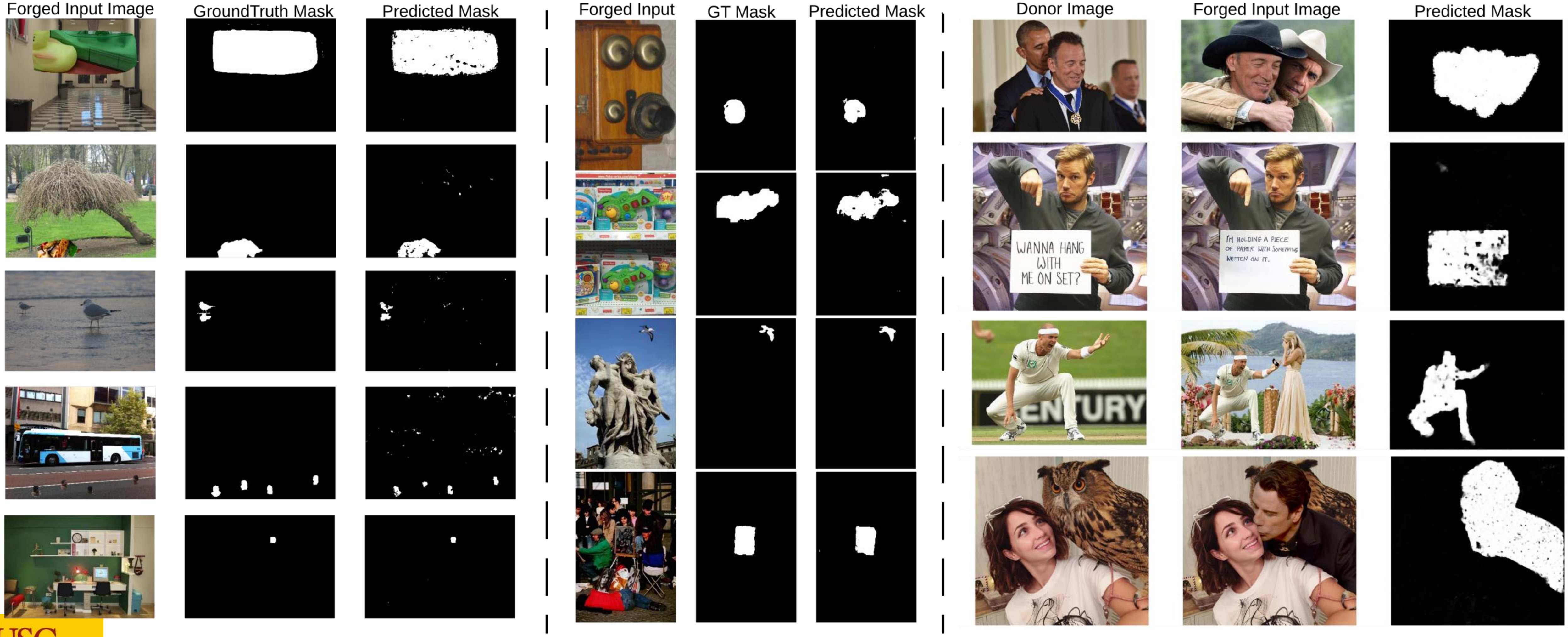
- Compression (JPEG, WebP, …)
- Blurring (Gaussian, media, wavelet, …)
- Morphology (Dilation, erosion, …)
- Contrast manipulation (Histogram equalization, Auto contrast, …)
- Additive noise (Poisson, Gaussian, …)
- Resampling (Lancoz, linear, …)
- Quantization

MANIPULATION HIERARCHY

- Fine-grained manipulations
- Six-level hierarchy
- 385 manipulations
- Dresden Image Database for creating training dataset
- Kaggle Camera Model Identification for testing



QUALITATIVE RESULTS



VIMAL
VISUAL INTELLIGENCE AND MULTIMEDIA ANALYTICS LABORATORY

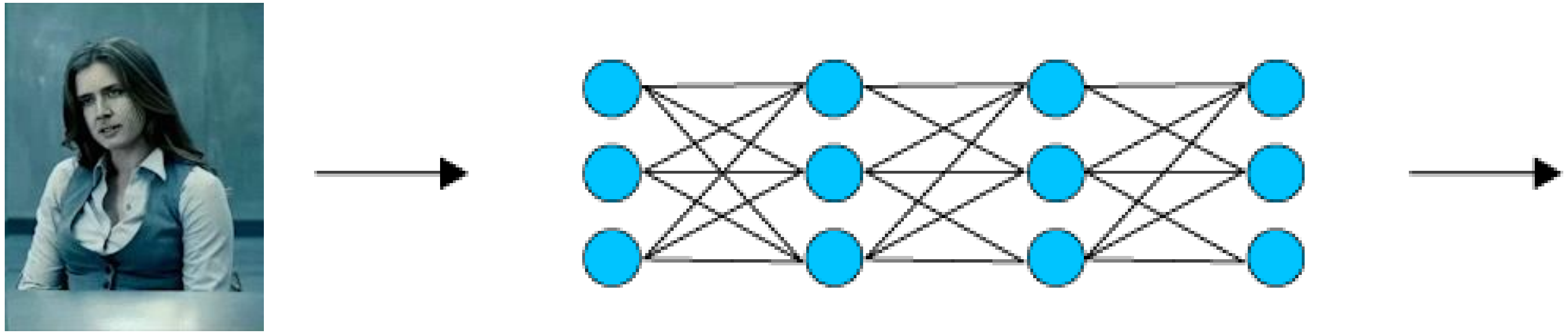
ADMISSIBILITY OF MANIPULATION DETECTION EVIDENCE

USC Viterbi
Information Sciences Institute

ADMISSIBILITY OF SCIENTIFIC EVIDENCE

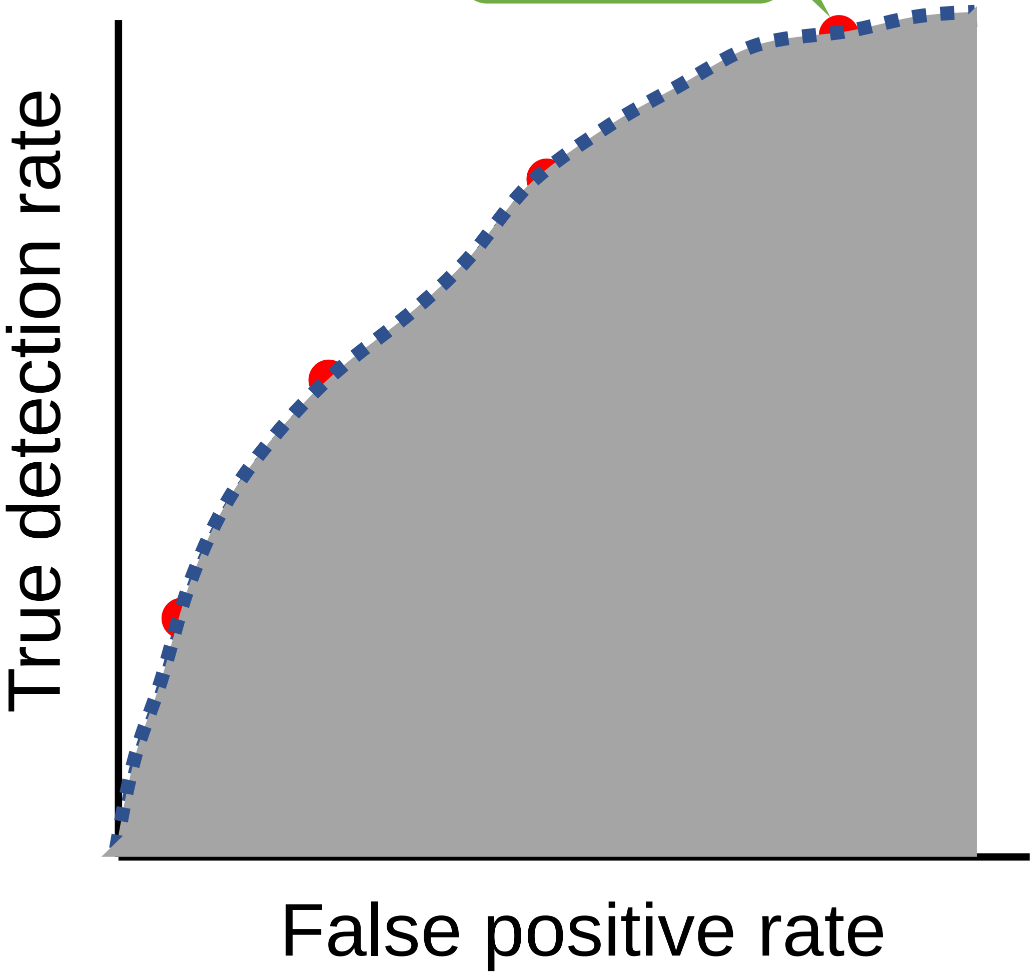
- ***Frye*** standard (general acceptance test)
  - Expert opinion based on a scientific technique is admissible only when the technique is generally accepted as reliable in the relevant scientific community
- ***Daubert*** standard
  - *A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:*
    - *(a) The expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;*
    - *(b) The testimony is based on sufficient facts or data;*
    - *(c) The testimony is the product of reliable principles and methods; and*
    - *(d) The expert has reliably applied the principles and methods to the facts of the case.*

HOW DO DETECTION ALGORITHMS WORK?



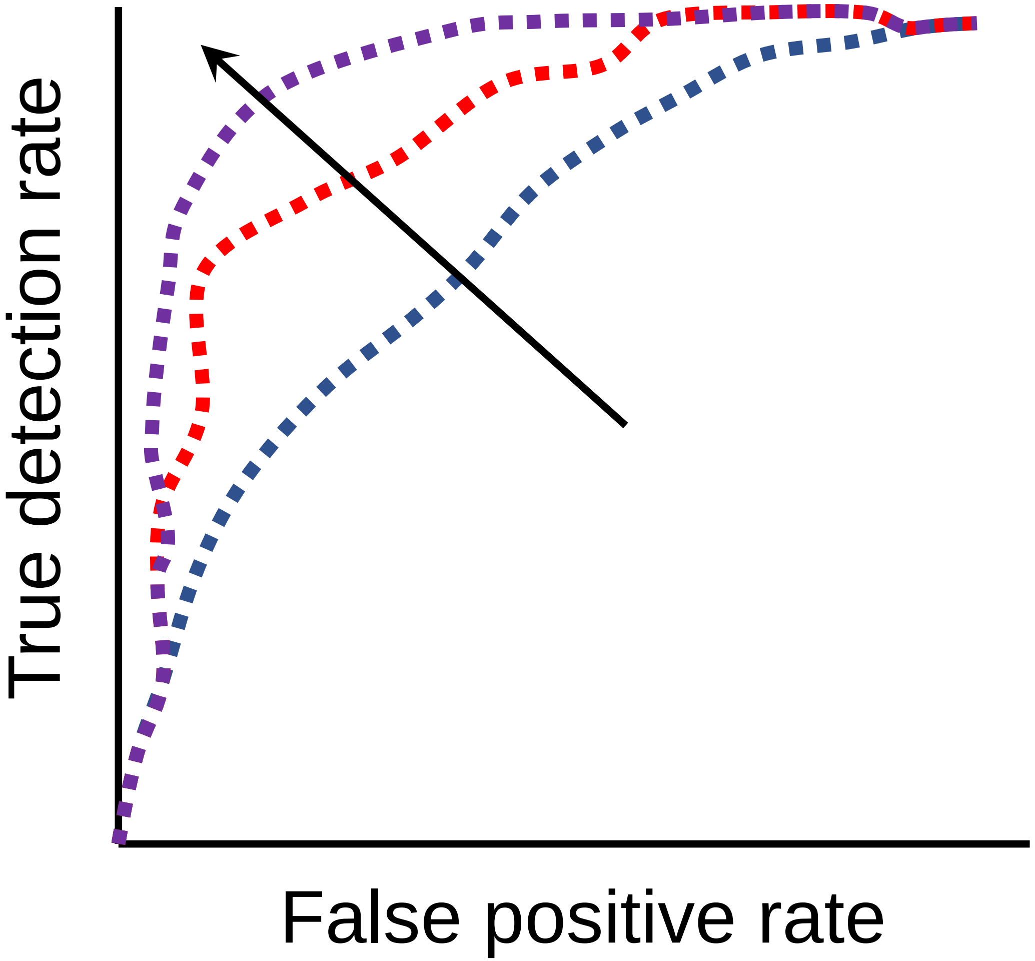
real or fake?
0 – 1

HOW DO DETECTION ALGORITHMS WORK?



Performance at another decision threshold

HOW DO DETECTION ALGORITHMS WORK?



FINAL THOUGHTS

shutterstock.com • 125542328

isn't
SEEING IS
BELIEVING

THANK YOU