AI Gone Rogue: Exterminating Deep Fakes Before They Cause Menace

Vijay Thaware
Symantec
@021vj

Niranjan Agnihotri
Symantec
@agnihotrins
Agenda:

● Deep Fakes: Should we trust what we see?
● Ingredients of a Deep Fake
● Cutting poison with poison
● Looking at Deep Fakes through biological microscope
● Limitations
● How does the future look?
● Black Hat Sound Bytes
Deep Fakes: Should we trust what we see?
Deep Fakes: The many faced evil

- Cyber propaganda
- Fake news=Deep fake news
- Trust issues
- Disinformation campaigns
- Emotional distress
- Can become ubiquitous
- Morality vs Legality
What it takes to prepare a Deep Fake?
Ingredients of a Deep Fake

★ Autoencoders

Image Source: https://www.alanzucconi.com/2018/03/14/understanding-the-technology-behind-deepfakes
How are deep fakes created?
Cutting poison with poison

★ Our Solution

★ Demo!
Face Net Model [1]

Triplet Loss

(128,1) encoding

Anchor

Positive

Negative

LEARNING

Anchor

Positive

Negative
Results

- Mr. Nick: 7.5355
- Mr. Putin: 0.0
- Mr. Robert: 0.0
- Mr. Trump: 0.0
- Unknown: 92.4645

- Mr. Nick: 89.2754
- Mr. Putin: 0.0
- Mr. Robert: 0.0
- Mr. Trump: 0.0
- Unknown: 10.7246

- Mr. Nick: 0.0
- Mr. Putin: 0.0
- Mr. Robert: 0.0
- Mr. Trump: 92.3602
- Unknown: 7.66398
Scope for Improvement

- Face liveness detection
- Contextual intel mechanisms
- Texture investigation
- User interaction

Cutting poison with poison

★ Merits
- Scalable
- Small, simple, fast
- Implementation on social media

★ Limitations
- Performs best only when deep fakes are produced by face swapping/tampering
- Won’t work on videos created without face tampering
Looking at Deep Fakes through biological microscope

★ Eye blinking :-

- Mean blinking rate is 17 blinks/minute (rest)
- While during conversing it gets to 26 blinks/minute
- While reading it gets to 4.5 blinks/minute
Looking at Deep Fakes through biological microscope

★ Observation
  ○ Generally the actors in the deep fake videos are not seen blinking. If not this, their blinking patterns are weird.

★ Key approach:
  ○ Extract the face region from the frames of the video
  ○ Use LRCN (Long Term Recurrent Convolutional Network) to detect eye blinks

Siwei Lyu et al. “In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking”
Limitations

- Difficult but not impossible to create deep fakes with a normal eye blink behavior with adequate data
- Videos of a very small length (< min)
- No front facing angle
- Scaling up
- No face tampering (deep fakes v2.0)
How does the future look?

Videos WITHOUT face tampering
Videos WITHOUT face tampering
Black Hat Sound Bytes

- Video watermarking
- Think before you forward THAT video
- Credibility of the source
- Robust Laws
Thank You!

Vijay Thaware
Symantec
@021vj

Niranjan Agnihotri
Symantec
@agnihotrins