in ictu oculi: Exposing DeepFake Videos by Detecting Eye Blinking

Yuezun Li, Ming-Ching Chang and Siwei Lyu
Computer Science Department
University at Albany, State University of New York
What is DeepFake?
AI-Assisted Fake Porn Is Here and We’re All F ***ed

Someone used an algorithm to paste the face of ‘Wonder Woman’ star Gal Gadot onto a porn video, and the implications are terrifying.

Deepfake technology could create huge potential for social unrest and even trigger wars

The only way to counteract the threat of deepfakes is to rely on the evidence of our own direct experience or authoritative proven sources, writes Rashmee Roshan Lall
How DeepFake works?

Unsupervised training:

Original

Face detection

Landmarks extraction

Face alignment

DeepFake converting

Encoder

“code”

Decoder

x1

x2

y1

y2

Fake

Smooth boundary

Affine warp

+
Why not use traditional forensic methods?

- Traditional forensic methods
  - Signal based: JPEG, CFA, PRNU
  - Physics based: lighting, shadow, reflection
  - Semantic based: where, when, who
- These methods may not be the best solution for detecting AI generated fake videos.
Other works

- [1] exploited the color disparity between GAN generated images and real images.


How to spot a DeepFake?

Real videos

*in ictu oculi*

DeepFake synthesized

Faces do not blink!

DeepFake videos
Why no blinking?

reason: training data may not include closed eyes
Detecting blinking with ML algorithm

- Train a Long term Recurrent CNN (LRCN) [3] model to detect open/closed eye
- Apply this model to estimate blinking rate in a video to determine its authenticity

The pipeline of our method

\[ f_t = \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_f) \]
\[ i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \]
\[ g_t = \tanh(W_{ch}h_{t-1} + W_{cx}x_t + b_c) \]
\[ C_t = f_t \odot C_{t-1} + i_t \odot g_t \]
\[ o_t = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \]
\[ h_t = o_t \odot \tanh(C_t) \]

\[ \sigma(x) = \frac{1}{1+e^{-x}} \]
\[ \tanh(x) = \frac{e^x-e^{-x}}{e^x+e^{-x}} \]

Training LRCN

1. Data preparation
   - 50 videos downloaded from Internet and annotate eye state

2. Training CNN
   - Input size: 224x224
   - Batch size: 16
   - Learning rate: 0.01
   - Decay: 0.9
   - Optimizer: SGD
   - Epoch: 100

3. Training LRCN jointly
   - Input size: 224x224xN
   - Batch size: 4
   - Learning rate: 0.01
   - Decay: 0.9
   - Optimizer: ADAM
   - Epoch: 100
The performance of LRCN

![ROC curve](image)

<table>
<thead>
<tr>
<th>Video</th>
<th>Average video length</th>
<th>FPS</th>
<th>Rate of blinks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin</td>
<td>10 seconds</td>
<td>30</td>
<td>34.1 / min</td>
</tr>
<tr>
<td>Fake</td>
<td>10 seconds</td>
<td>30</td>
<td>3.4 / min</td>
</tr>
</tbody>
</table>

The blinking rate of normal human is set to 10/min [6]


Spot a DeepFake *in ictu oculi*
Spot a DeepFake *in ictu oculi*
Is this the end of DeepFake?

Forgery technology catches up quickly

- e.g., blinking can be fixed with using video frames as training data

- Despite this, our discovery can still increase the difficulties of DeepFake video generation. Now we are developing more effective method to expose the fake videos
Summary

• Technologies for creating DeepFake videos advance rapidly and can cause serious impacts to society

• Digital media forensics are catching up to control the negative effects of DeepFake videos

• The rivalry between forgeries and forensics will continue for coming years
Thank you for your attention!

All non-public domain images/videos used in this talk belong to the creators.