

# Introduction to Deepfakes

Georgia Fargetta

*Erasmus +  
Master Degree in Computer Science  
University of Rouen, France*



International Forensics Summer School

**ETHICAL AND LEGAL CHALLENGES IN AI-DRIVEN FORENSIC SCIENCE**



**JULY 14-20, 2024**

[Watch a preview!](#)

## School Directors



**PROF. SEBASTIANO BATTIATO, PH.D.**  
*University of Catania*



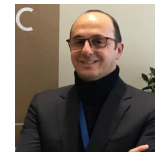
**PROF. DONATELLA CURTOTTI, PH.D.**  
*University of Foggia*



**PROF. GIOVANNI ZICCARDI, PH.D.**  
*University of Milan*

## Speakers

**others coming soon..**



**ALESSANDRO TRIVILINI**  
*Scuola universitaria professionale della svizzera italiana (SUPSI)*



**MARTIN DRAHANŠKÝ**  
*Faculty of Information Technology, Brno University of Technology*



**PROF. DR. DIDIER MEUWLY**  
*University of Twente*

## School location

The school will take place at Sampieri, Sicily  
<https://www.hotelbaiasamuele.it/en/>



## Social Network



IFOSS



@ifoss\_official



@ifoss\_official



IFOSS



[www.ifoss.it](http://www.ifoss.it)



[info@ifoss.it](mailto:info@ifoss.it)

# Deepfakes

The objective of this presentation is to provide an understanding of:

- How deepfakes are created and detected
- The current trends and advancements in this domain
- The shortcomings of the current defense solutions
- The areas that require further research and attention

# Deepfakes

*DeepFakes refers to all those multimedia contents (images, videos, audio) synthetically altered or created by exploiting machine learning generative models.*

We can define a deepfake as:

*“believable media generated by a deep neural network”*

# Deepfakes

Click on the person who is real.



<https://www.whichfaceisreal.com/>

# Deepfakes – which face is real ?



# Deepfakes – which face is real ?

**FAKE**



**REAL**



# Deepfakes – which face is real ?





# Deepfakes – which face is real ?

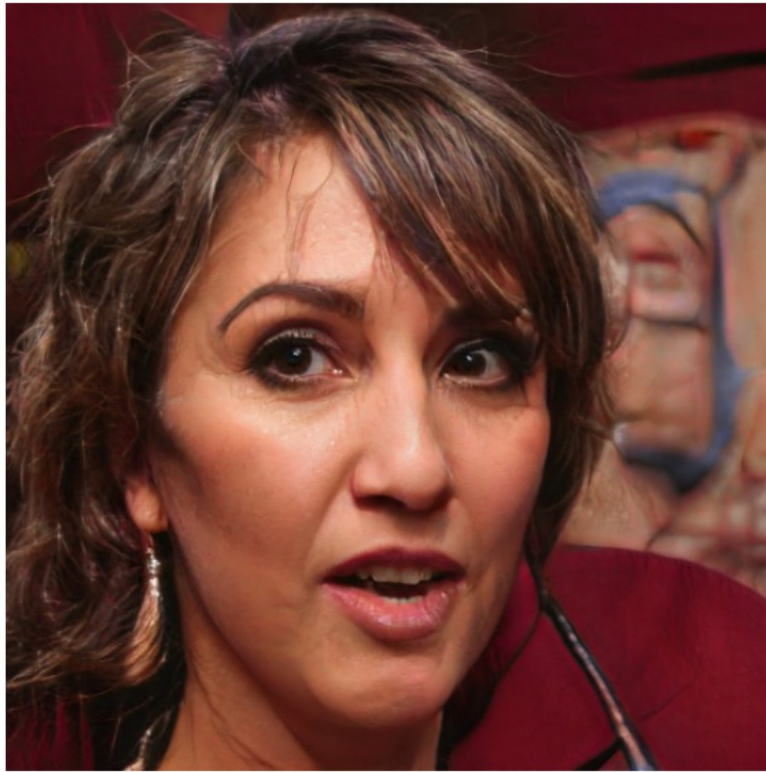
**REAL**



**FAKE**



# Deepfakes – which face is real ?



# Deepfakes – which face is real ?

**FAKE**



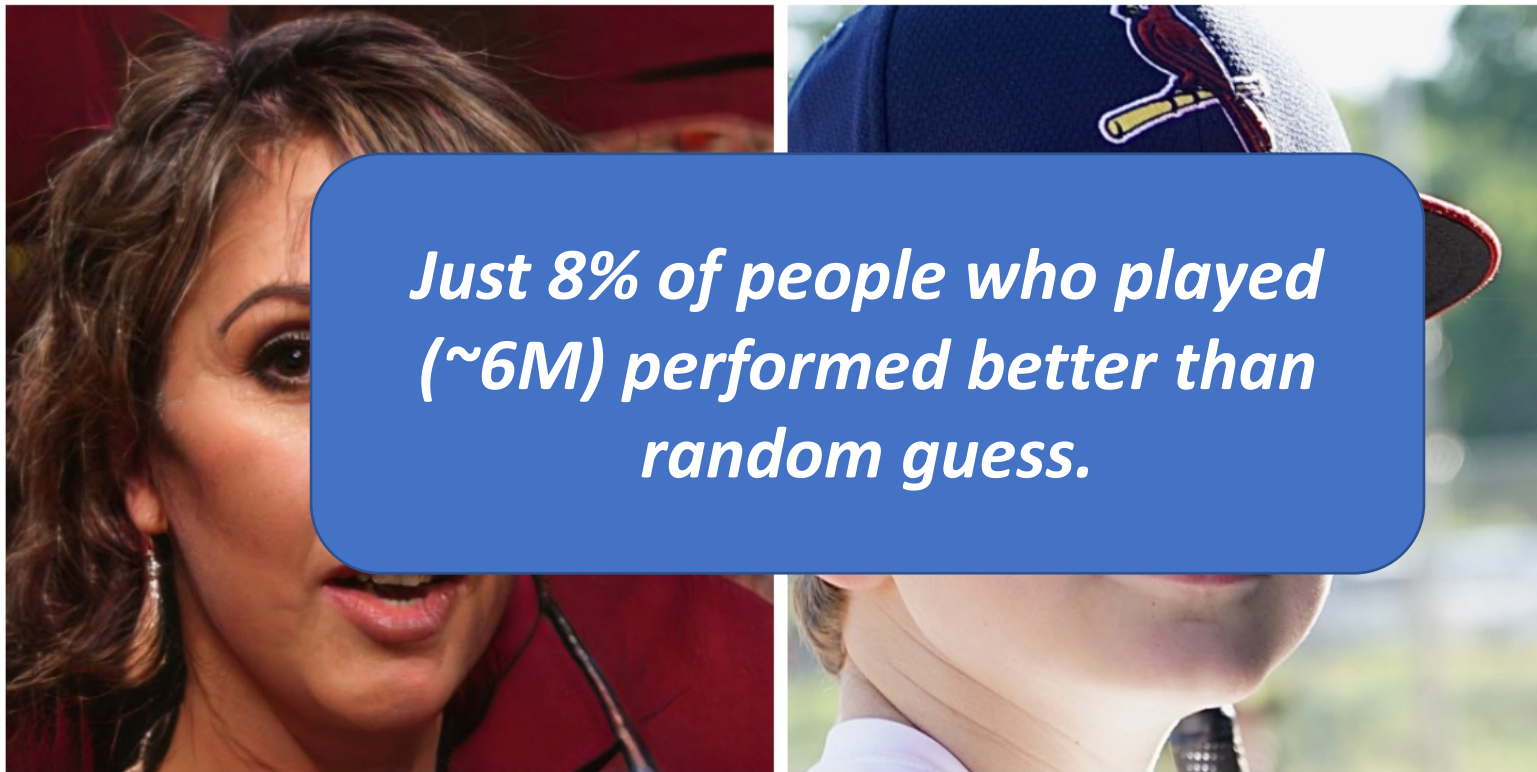
**REAL**



# Deepfakes – which face is real ?

**FAKE**

**REAL**



# Deepfakes

*DeepFakes refers to all those multimedia contents (images, videos, audio) synthetically altered or created by exploiting machine learning generative models.*

We can define a deepfake as *“believable media generated by a deep neural network”*

*NB: We will focus on DFs pertaining to the human face and body.*



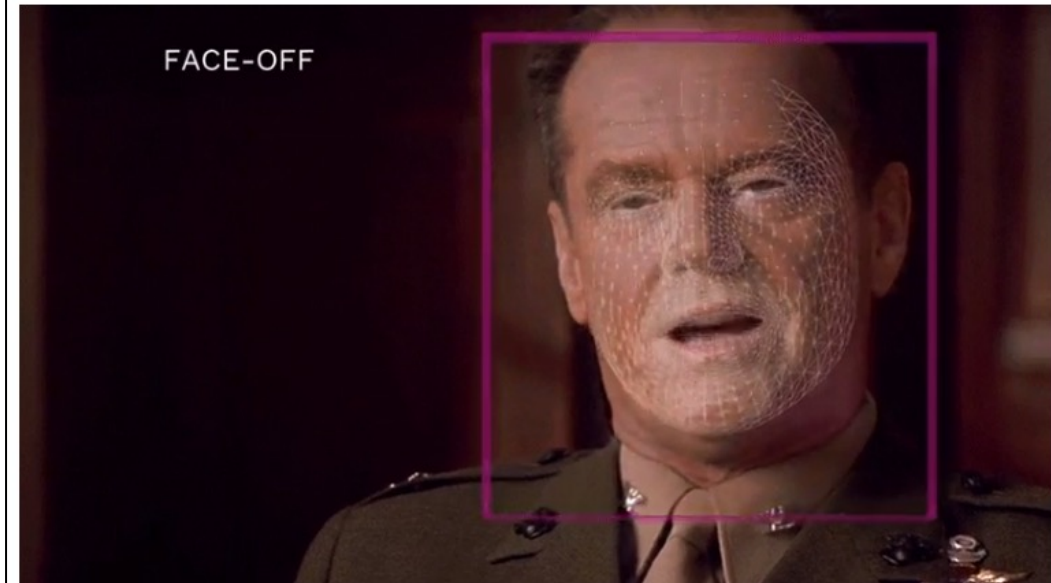
# Deepfakes - (positive) applications

Realistic video dubbing of foreign films.

Source: <https://gizmodo.com/deepfake-lips-are-coming-to-dubbed-films-1846840191>

## Deepfake Lips Are Coming to Dubbed Films

By Andrew Liszewski | 5/06/21 4:15PM | Comments (141) | Alerts



Gif: [Vimeo - Flawless](#)

- f For those who find reading a foreign film's subtitles too distracting, movies are often dubbed into different languages. But that can be equally distracting when the movements of an actor's mouth are completely out of sync with what they're saying. So a company called Flawless has created an [AI-powered solution](#) that will replace an actor's facial performance to match the words in a film dubbed for foreign audiences.
- t
- l
- e
- u

# Deepfakes - (positive) applications

Reanimation of historical figures.

Fake Lincoln



Abraham Lincoln was also brought back to life, with added colour and speech

Source: <https://www.bbc.com/news/technology-56210053>  
<https://derivative.ca/community-post/deepfake-salvador-dal%C3%AD-interacts-museum-visitors-takes-selfies>

## DEEPPFAKE SALVADOR DALÍ INTERACTS WITH MUSEUM VISITORS, TAKES SELFIES!

Salvador Dalí once wrote, "If someday I may die, though it is unlikely, I hope the people in the cafes will say, 'Dalí has died, but not entirely.'" Now 30 years after his alleged death, in a precedent-setting exhibition where a museum has used artificial intelligence-based techniques and TouchDesigner to bring an artist back to life, it turns out that Dalí's prescience was on point!

With the opening of the *Dalí Lives* exhibition at the Dalí Museum in St. Petersburg, Florida, Dalí has quite authentically been resuscitated with a 'deepfake' where the man himself appears to banter pleasantly with museum visitors, taking selfies with people and sending them text messages!



# Deepfakes - (positive) applications

Reanimation of historical figures.



Realistic video dubbing of foreign films.





# Deepfakes – (positive) applications

Virtually trying on clothes while shopping.



Generating realistic images of people wearing clothes by selecting specific outfit, pose, gaze, gender, hair colour etc. for advertising.

# Deepfakes – (positive) applications

Creating deepfakes for entertainment (e.g., memes). Such as music videos with the face of Nicolas Cage or Tom Cruise.



# Deepfakes – (positive) applications

Creating deepfakes for entertainment (e.g., memes).



Reface: Face swap videos and memes with your photo

NEOCORTEXT, INC. Entertainment

★★★★★ 1,471,711

Mature 17+

Contains Ads · Offers in-app purchases

This app is available for your device

*Wombo.AI* – make your selfies sing



*Deep Nostalgia*



# Deepfakes – (malicious) applications

Swap faces of celebrities into pornographic videos and post them online.

In 2017 a Reddit user published used an algorithm to paste the face of 'Wonder Woman' star Gal Gadot onto a porn video and published it online.

---

**MOTHERBOARD**  
TECH BY VICE

## **This Horrifying App Undresses a Photo of Any Woman With a Single Click**

The \$50 DeepNude app dispenses with the idea that deepfakes were about anything besides claiming ownership over women's bodies.



Sources: <https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn>  
[https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman?utm\\_source=vicetwitterus](https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman?utm_source=vicetwitterus)

# Deepfakes – (malicious) applications

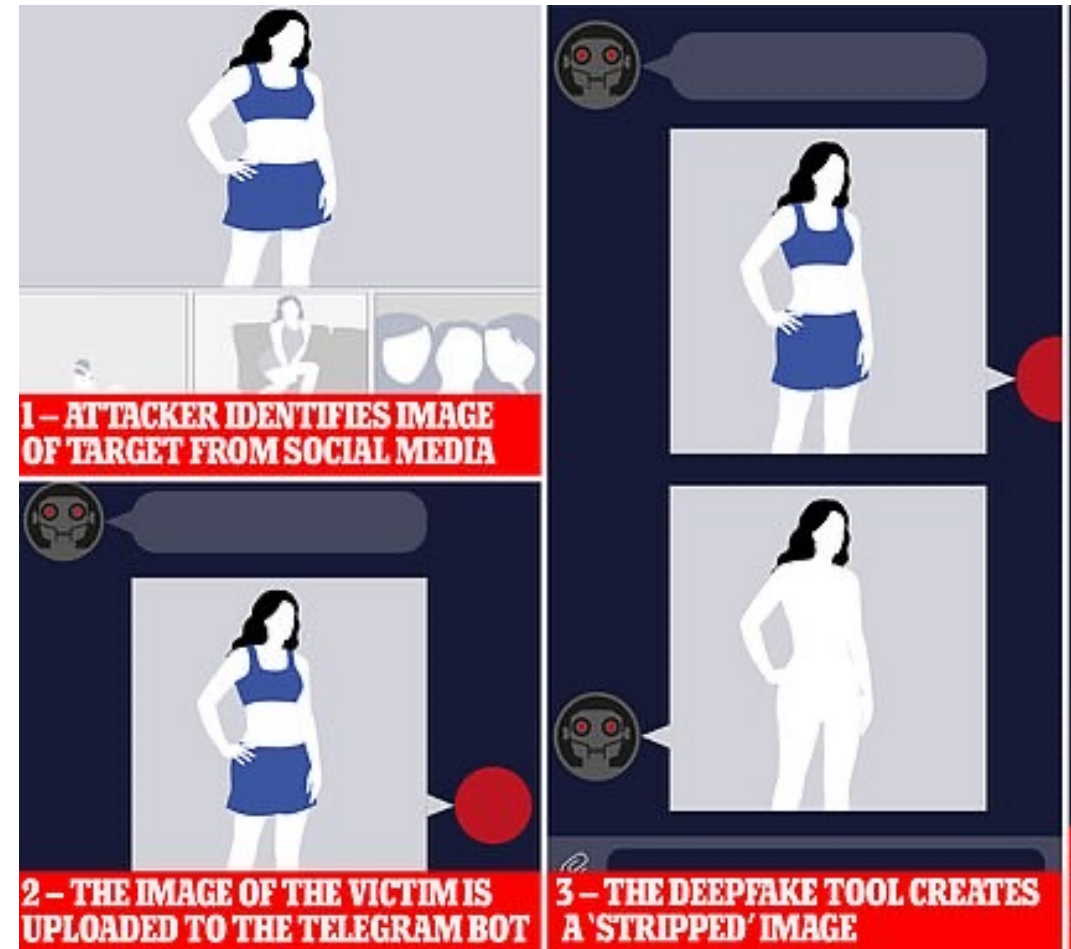
Automated image abuse.

## Disturbing deepfake tool on popular messaging app Telegram is forging NUDE images of underage girls from clothed photos on their social media

- The unnamed 'bot' works using artificial intelligence and machine learning
- It operates on the Telegram messaging app with over 100,000 images shared
- Report authors say anyone is at risk of having an image taken and a nude faked
- The majority of the women and girls 'faked' by the bot were private individuals

By RYAN MORRISON FOR MAILONLINE 

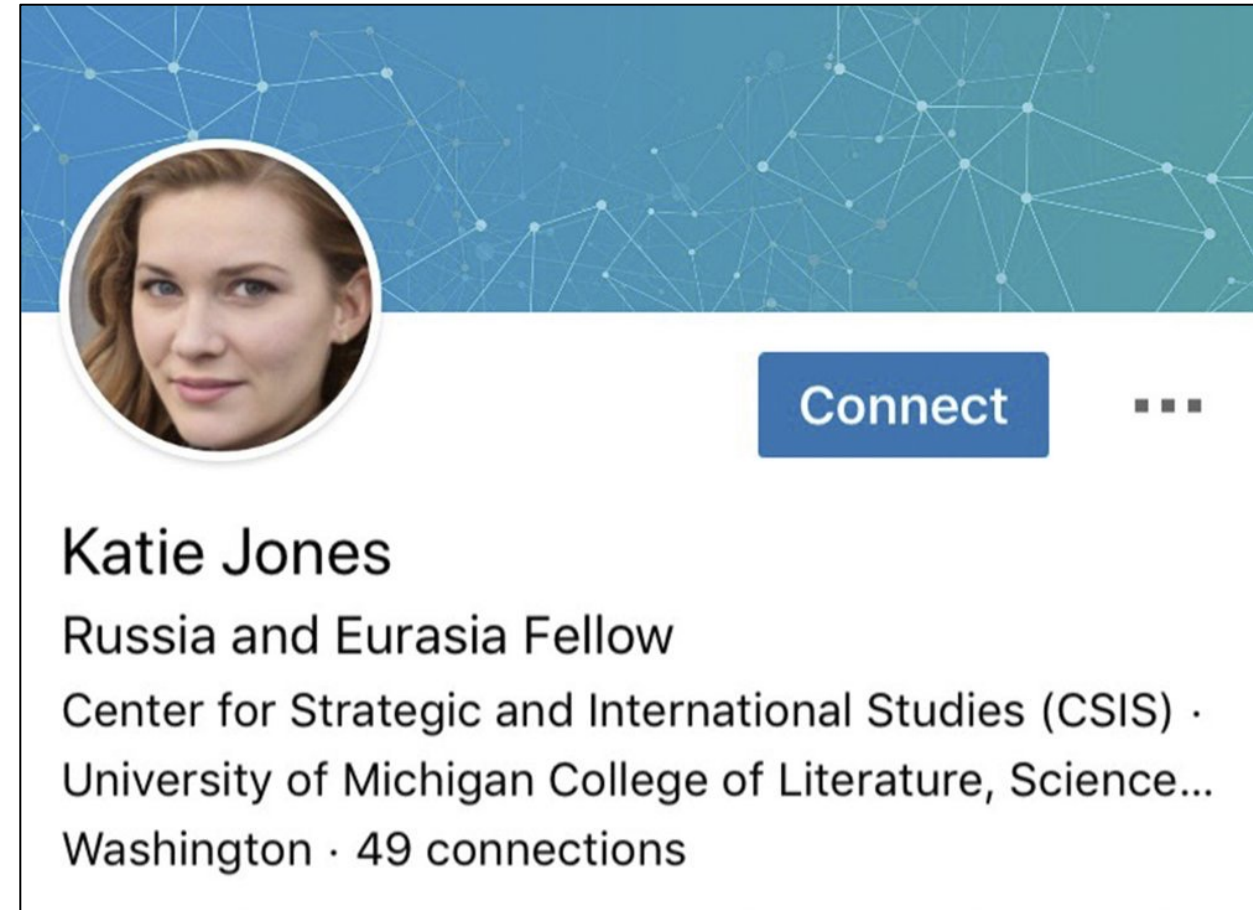
PUBLISHED: 11:55 BST, 21 October 2020 | UPDATED: 13:23 BST, 21 October 2020



Source: <https://www.dailymail.co.uk/sciencetech/article-8863233/Disturbing-deepfake-tool-popular-messaging-app-Telegram-forging-NUDE-images-underage-girls.html>

# Deepfakes – (malicious) applications

- Job experience in innovation
- Network of pundits and experts
- Connections with prestigious USA institutions and politicians amongst which:
  - a deputy assistant secretary of state
  - a senior aide
  - a senator
  - a famous economist considered for a seat on Federal Reserve



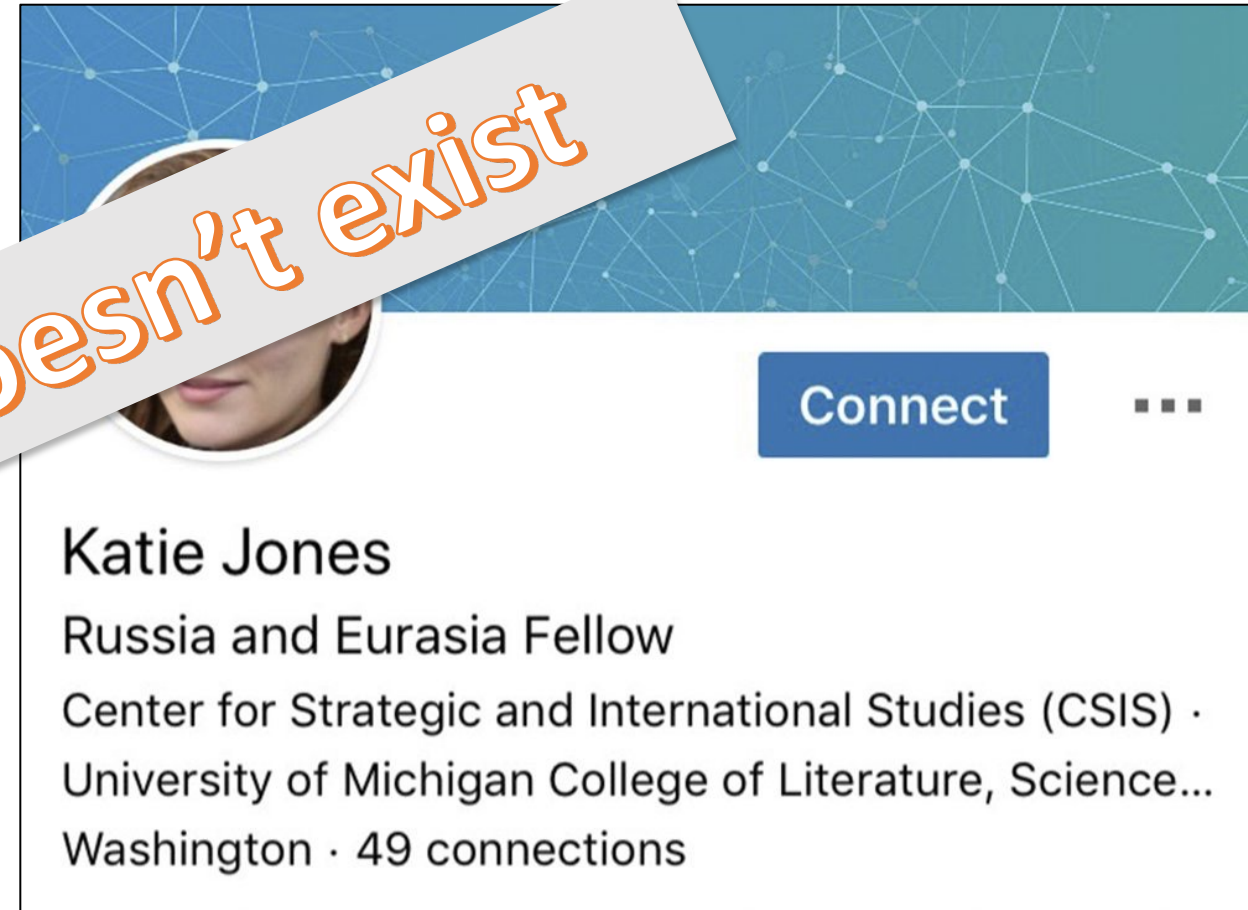
# Deepfakes – (malicious) applications

- Job experience in innovation
- Network of pundits and experts
- Connections with prestigious USA institutions

and politicians amongst which:

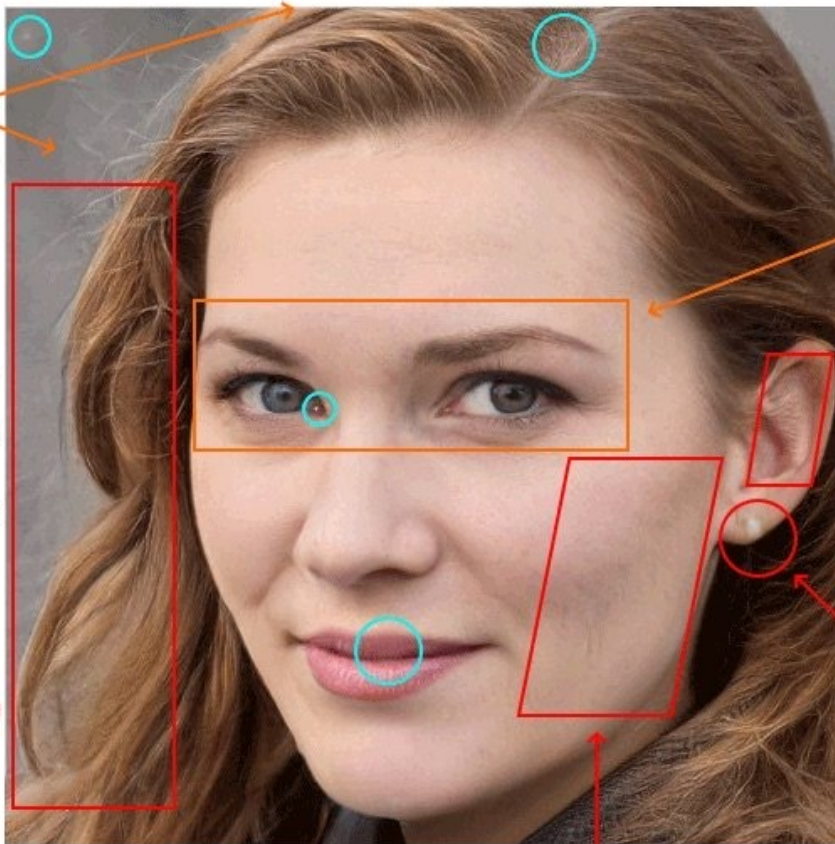
- a deputy assistant secretary
- a senior aide
- a senior advisor
- a senior economist considered for a seat on Federal Reserve

**Katie Jones doesn't exist**



# Deepfakes – (malicious) applications

Indistinct background, close crop



Other potential artifacts

Halo effect or painterly quality around hair

Smudges/striations/drip marks on cheek

## Experts: Spy used AI-generated face to connect with targets

By RAPHAEL SATTER June 13, 2019

Cockeyed, heterochromatic eyes

Strange scale-like effect on upper earlobe

Earring is blurry or melted

Source: <https://apnews.com/article/ap-top-news-artificial-intelligence-social-platforms-think-tanks-politics-bc2f19097a4c4fffaa00de6770b8a60d>



# Deepfakes – (malicious) applications

Financial fraud.

## Listen carefully: The growing threat of audio deepfake scams

AI software capable of cloning voices is proving a useful weapon for fraudsters.

By **Greg Noone** 04 Feb 2021 (Last Updated 15 Mar 2021)

Sep 3, 2019, 04:42pm EDT | 48.818 views

## A Voice Deepfake Was Used To Scam A CEO Out Of \$243,000



**Jesse Damiani** Contributor

Consumer Tech

*I cover the human side of VR/AR, Blockchain, AI, Startups, & Media.*

According to a new report in *The Wall Street Journal*, the CEO of an unnamed UK-based energy firm believed he was on the phone with his boss, the chief executive of firm's the German parent company, when he followed the orders to immediately transfer €220,000 (approx. \$243,000) to the bank account of a Hungarian supplier.

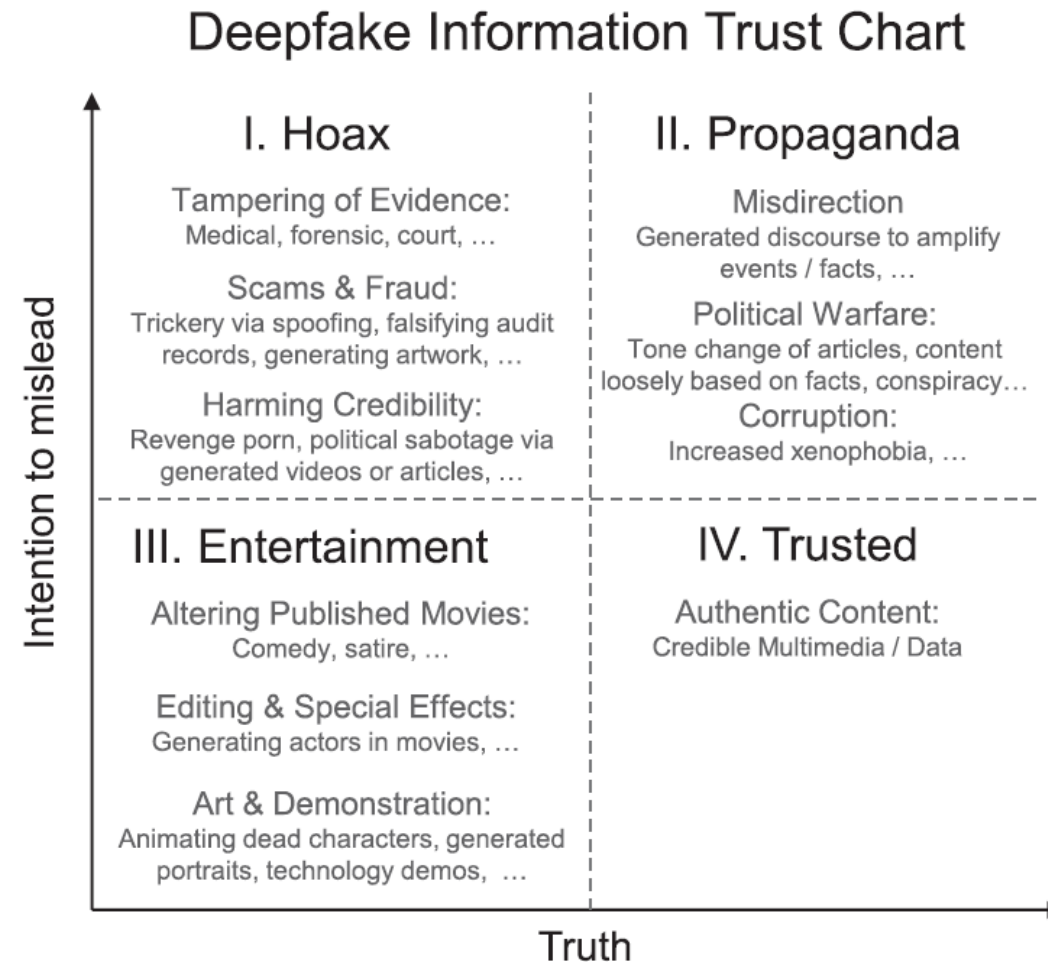
<https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/>

<https://techmonitor.ai/techonology/cybersecurity/growing-threat-audio-deepfake-scams>

# Deepfakes – (malicious) applications

- Fake porn of famous people
- Automated image abuse
- Foreign Espionage
- Financial fraud
- Disinformation
- Defamation
- ...

# Deepfakes - intention vs truth



# Deepfakes – easy to access

Several consumer apps are already available for free:

- RefaceApp
- Wombo AI
- etc..

But also professional-level tools:

## **DeepFaceLab**

*“the leading software for creating deepfakes”*

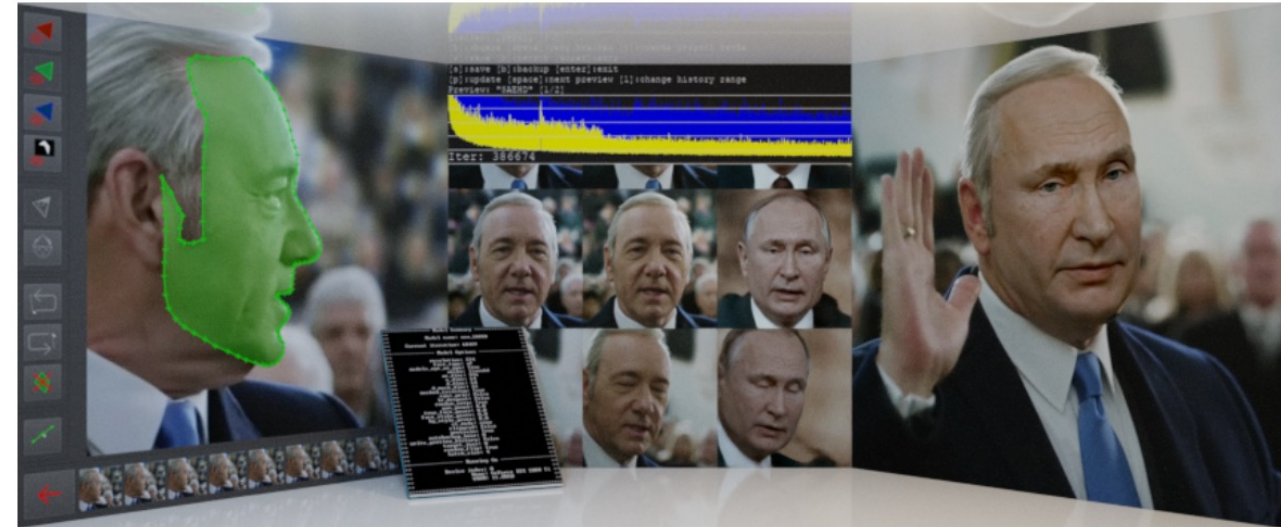
<https://github.com/iperov/DeepFaceLab>

# Deepfakes – easy to master

## DeepFaceLab

*“the leading software for creating deepfakes”*

<https://github.com/iperov/DeepFaceLab>



## DeepFaceLive



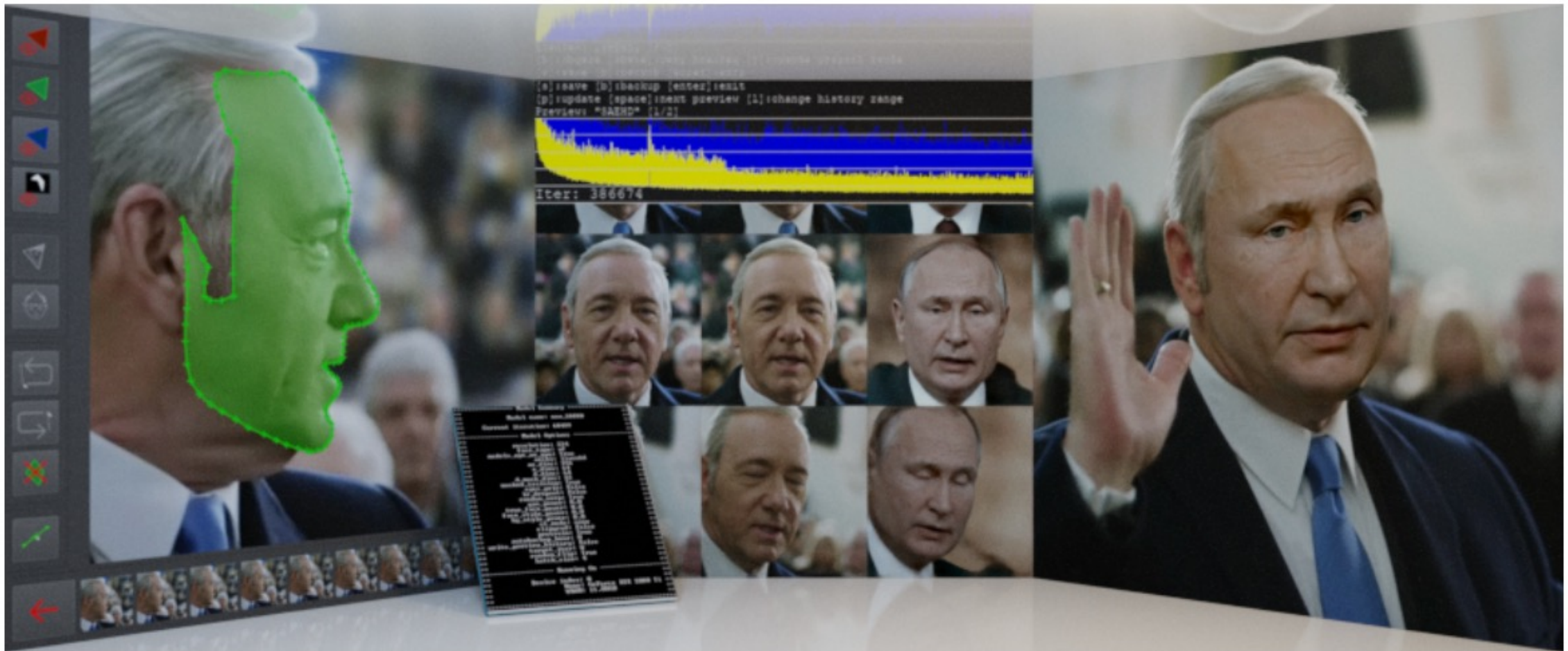
## DeepFaceLive

*Realtime face swap for PC streaming or video calls*

<https://github.com/iperov/DeepFaceLive>

# Deepfakes – easy to access

## *DeepFaceLab*



# Deepfakes – easy to access

## ***DeepFaceLab***

Available tools:

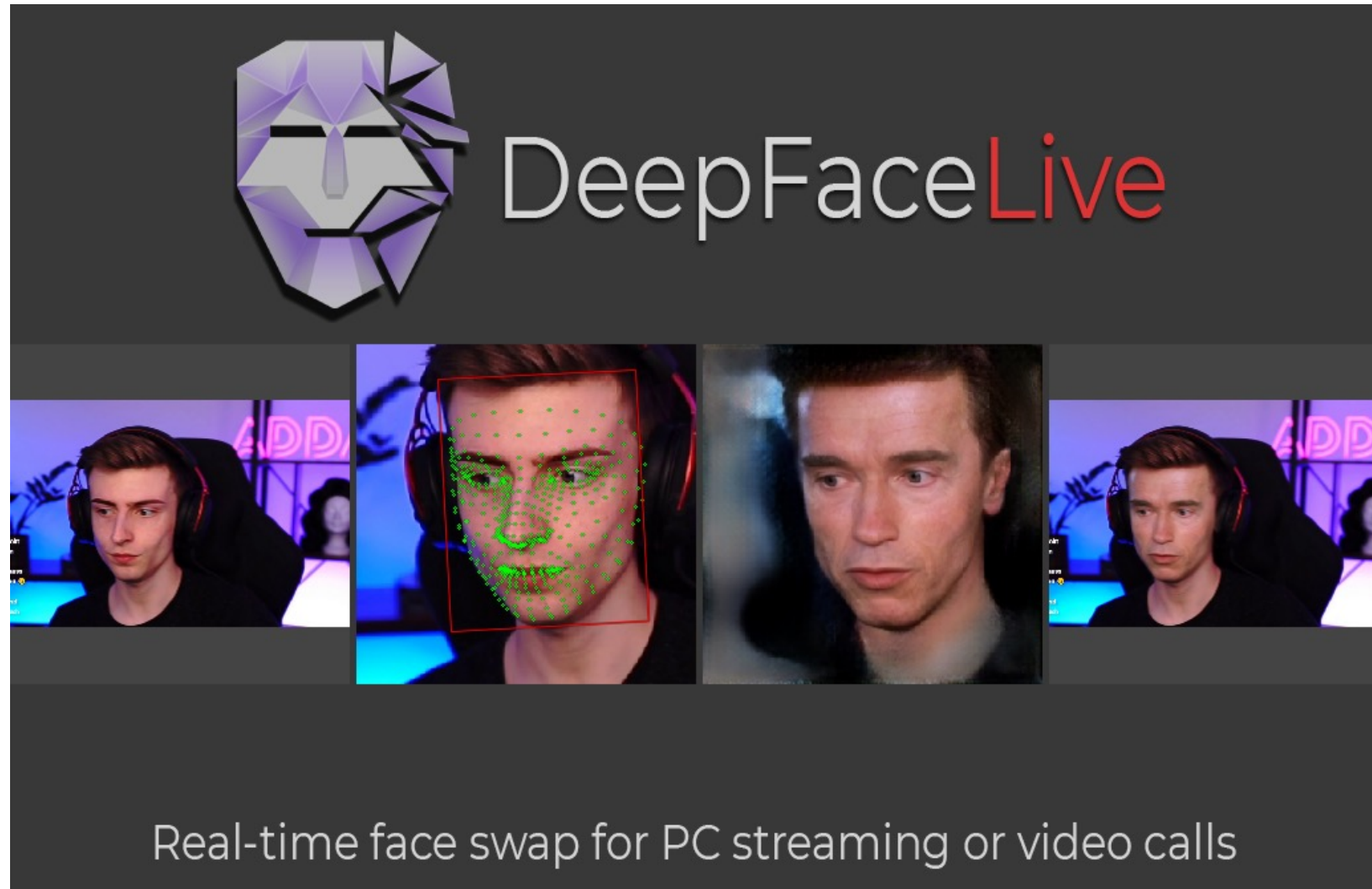
- Replace the face
- De-age the face
- Replace the whole head
- Manipulate politicians lips

Additional available material:

- Ready to work facesets
- Pretrained models
- Forums and support groups



# Deepfakes – easy to access



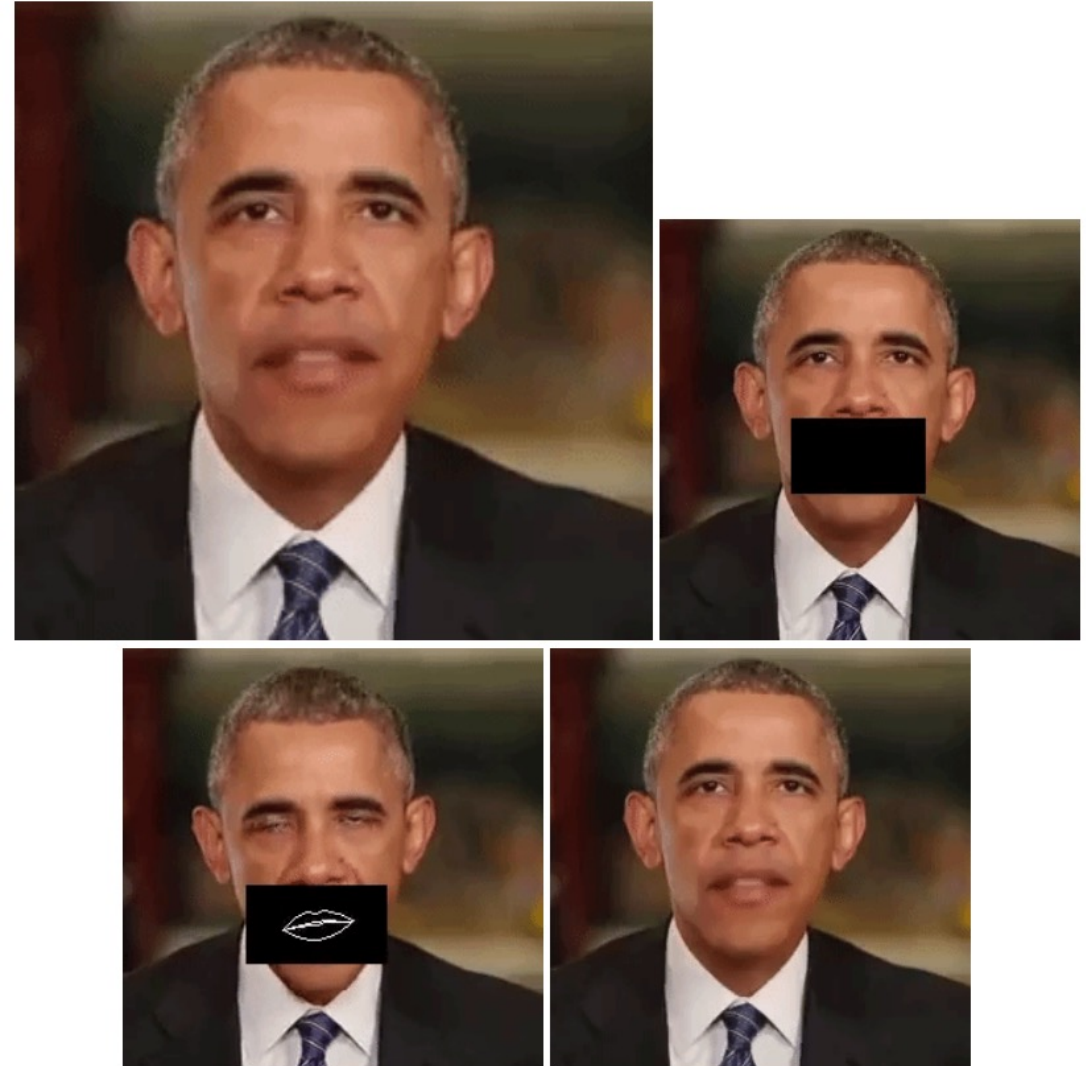
<https://github.com/iperov/DeepFaceLive>



# Deepfakes – easy to access

## ***ObamaNet***

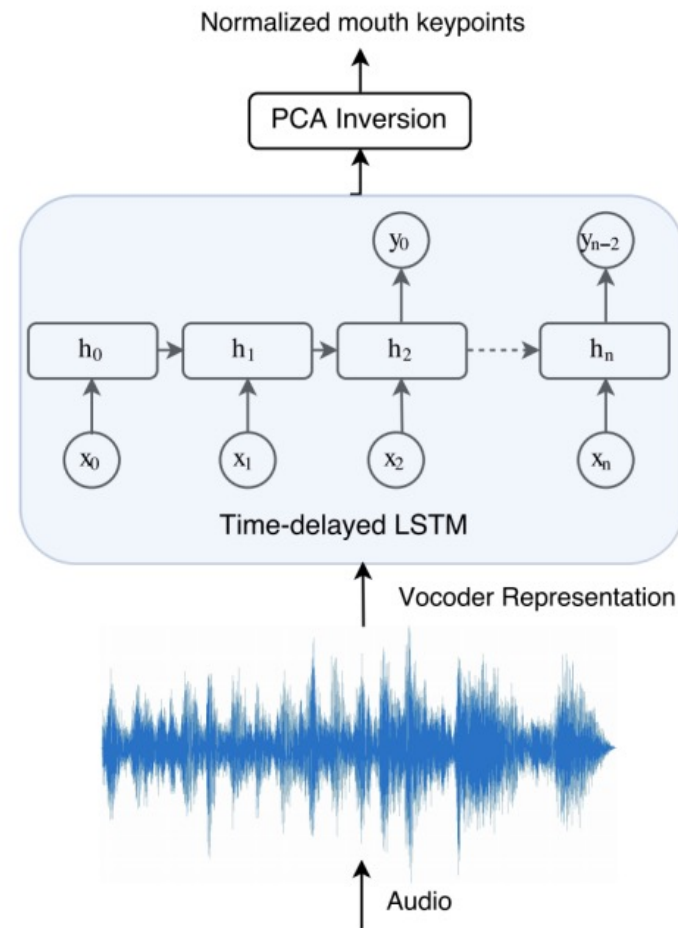
takes any text as input and generates both the corresponding speech and synchronized photo-realistic lip-sync videos.



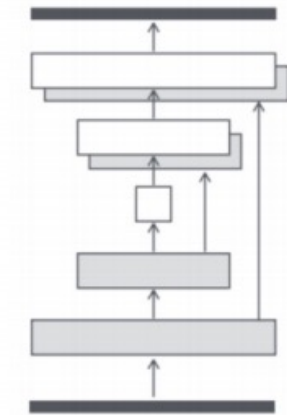
# Deepfakes – easy to access

**ObamaNet:** reenacts an individual's mouth and voice using text as input.

Text to audio + mouth keypoints movements + mouth inpainting



Synthesized Image



U-Net



Target video image with drawn keypoints

# Why deepfakes are a problem

- Online content is used for criminal, reputation or influence operations (trust abuse)
- It has not been possible to easily fake people in videos as it has today, therefore people trust it more than images that can be faked with photoshop
- Is possible to easily spread convincing deepfakes just before an election
- The asymmetry of cost to impact could be high

# Why deepfakes are a problem

- Easy to access
- Easy to master
- Inexpensive
- Scalable

Often the quality is very high.




# Deepfakes – Attack categories

We will focus on deepfakes regarding the human face and body. We can identify four categories.


- **Reenactment:** the source image (e.g., attacker's face) is used to drive the expression, mouth, gaze, pose or body of the target image.
- **Replacement:** the target is replaced with the source, preserving the identity of the source.
- **Editing:** the attributes of the target are added, altered or removed.
- **Synthesis:** the deepfake content is created with no target as basis.

# Deepfakes – Attack categories

Source  $x_s$




Target  $x_t$




  

Face Editing




Hair   Article   Age   Beauty   Ethnicity

Facial Reenactment



Face Replacement




  

●: Always  
○: Sometimes

Transfers:	Gaze	Mouth	Expression	Pose	Complete	Transfer	Swap
Gaze	●		○		○		○
Mouth		●	○		●		●
Expression		○	●		●		●
Pose				●	●		
Identity						●	●

Face Synthesis

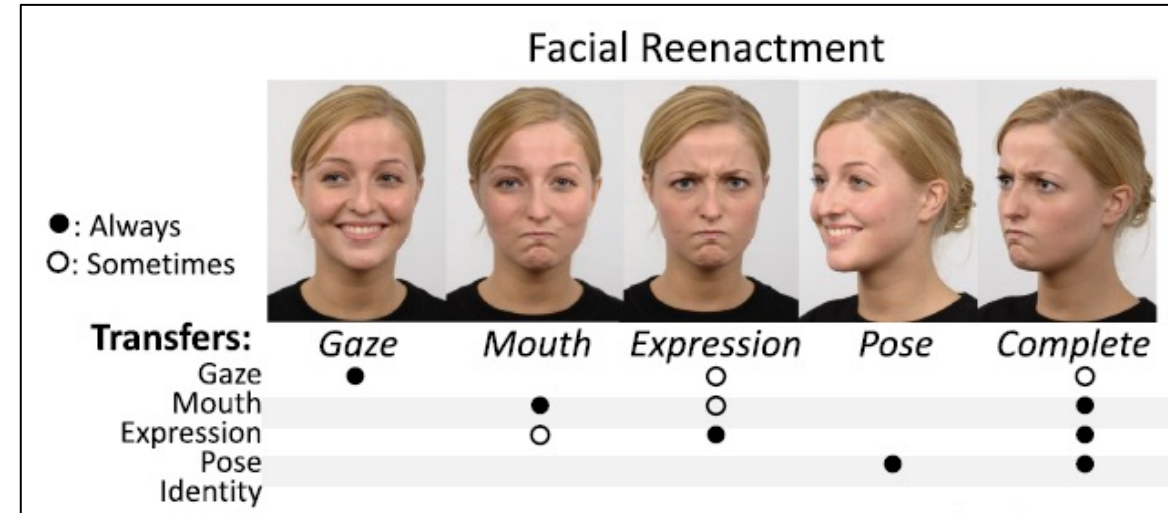


Reenactment and replacement are the greatest concern because they give an attacker control over one's identity.

# Deepfakes – reenactment details

The reenactment can involve specific aspects:

Expression (video games and movie industry)  
 mouth (realistic voice dubbing - lips sync),  
 Gaze (improve photographs), Pose, Body.



**Attacks:** attackers can impersonate an identity, controlling what he/she says or does. He can also generate embarrassing content, tamper surveillance footage or impersonate someone in a conversation.

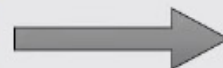
# Deepfakes – reenactment details

Zakharov, Egor, et al. "Few-shot adversarial learning of realistic neural talking head models." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.

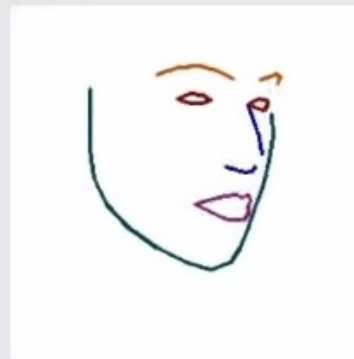
Training frames:



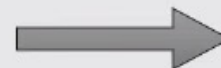
Driving sequence



Landmarks  
detector



Face landmarks



Learned talking head



# Deepfakes – reenactment details



# Deepfakes – reenactment details

Forrest Gump (1994) - President Kennedy Vs Forrest Gump



# Deepfakes – reenactment details

*Forrest Gump (1994) - President Kennedy Vs Forrest Gump*



This is something not new...

*Matteo Renzi (2020)*



# Deepfakes – reenactment details

*Forrest Gump (1994) - President Kennedy Vs Forrest Gump*



This is something not new...

- Very well founded Hollywood studios
- Extreme expensive technology
- Skilled specialist engineers
- Professional equipment

# Deepfakes – replacement details

Replacement: the target is replaced with the source, preserving the identity of the source. In particular:

- **Transfer:** the content of the target is replaced with the source, e.g.. face transfer to visualize a person in different outfits (fashion industry).
- **Swap:** the content transferred to the target is driven by the source. The most popular is “face swap” used to generate memes or satirical content or blurring of a face for anonymization.

**Attacks:** creating fake porn contents.



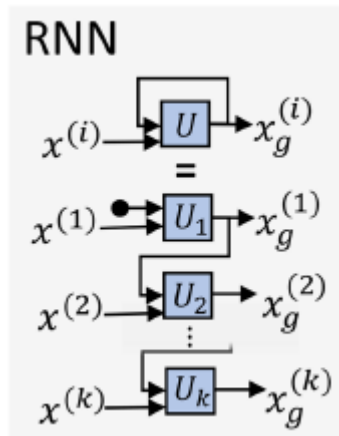
Break time!

# Deepfakes – technical background

Deepfakes are usually created using some basic networks combined with CNNs.

# Deepfakes – technical background

Deepfakes are usually created using some basic networks combined with CNNs.

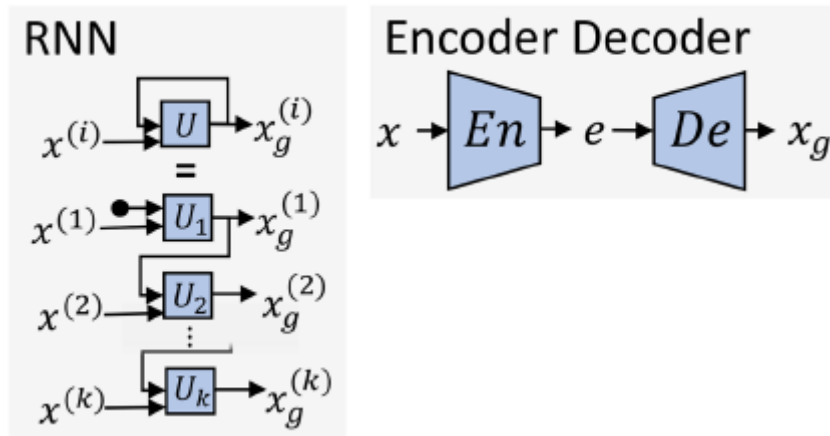


**Recurrent Neural Networks (RNN):** able to handle sequential data (e.g., audio/video). The network remembers its internal state and uses it to process subsequent inputs. LSTM are the most used type of RNN.

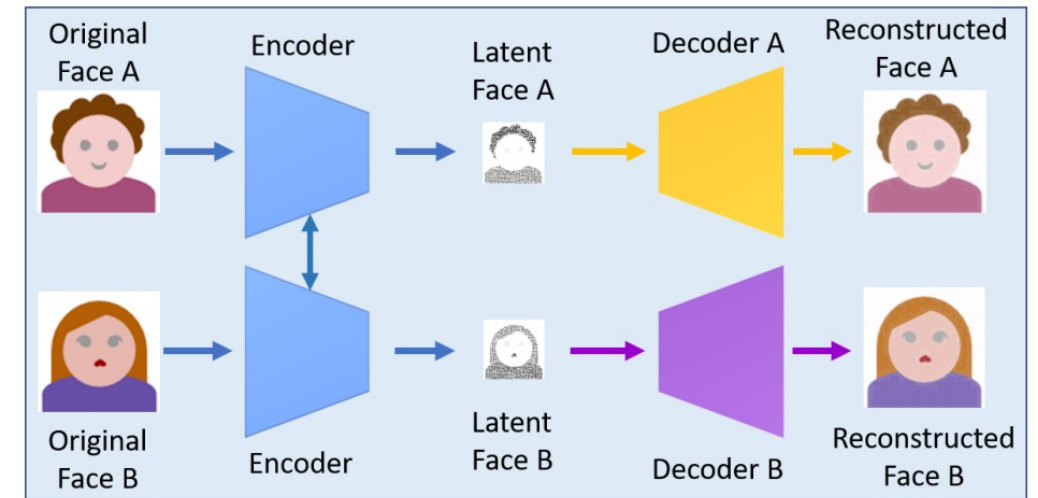
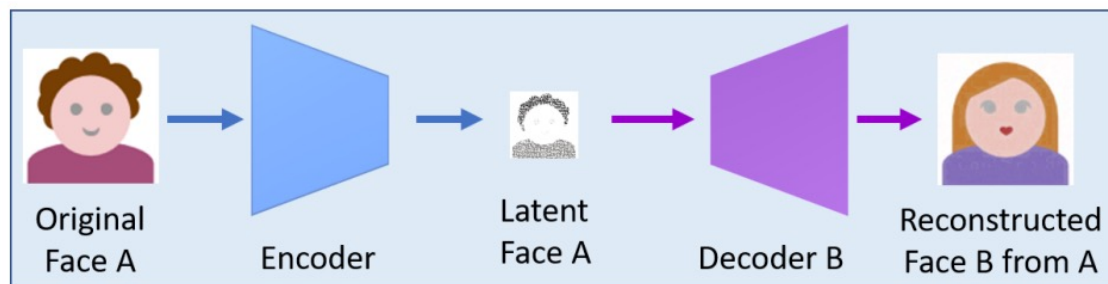


# Deepfakes – technical background

Deepfakes are usually created using some basic networks combined with CNNs.



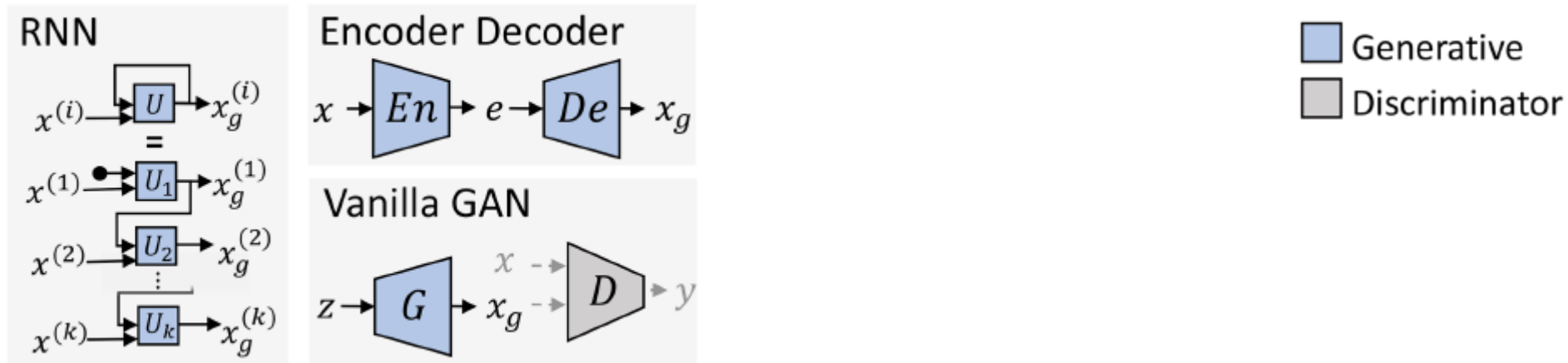
- **Recurrent Neural Networks (RNN), e.g., LSTM**



**Encoder-Decoder (ED):** consists of (at least) an encoder  $En$  and a decoder  $De$ . The  $En$  summarizes the input to a vector (or encoding/embedding), whereas  $De$  transforms such encoding to influence the output. When the network is trained to produce the original input the ED is called Autoencoder. Deepfake tech exploits multiple ED manipulating the input to produce specific decoded outputs.

# Deepfakes – technical background

Deepfakes are usually created using some basic networks combined with CNNs.



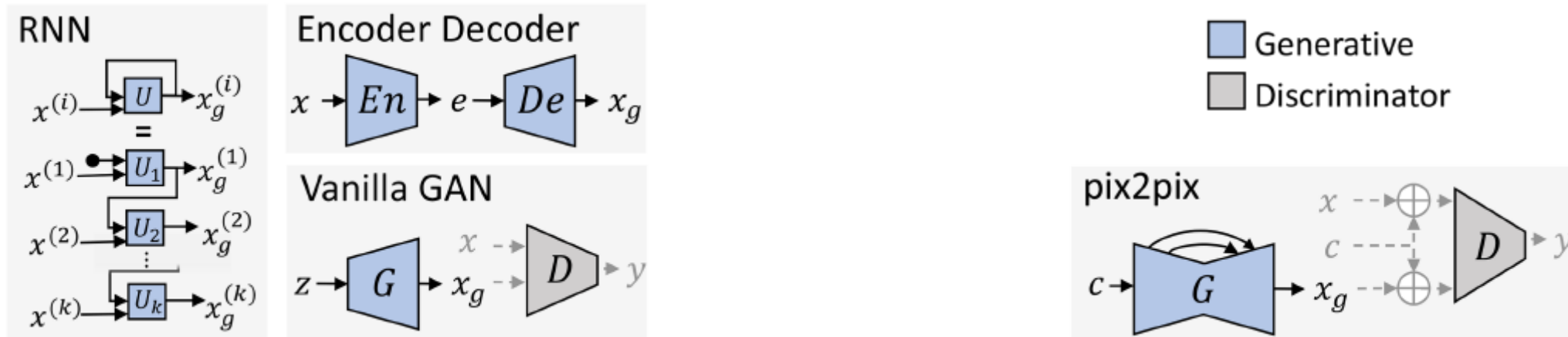
- **Recurrent Neural Networks (RNN), e.g., LSTM**
- **Encoder-Decoder (ED)**

**Generative Adversarial Network (GAN):** consists of two neural networks. The generator creates fake examples with the aim of fooling the discriminator, which learns to discriminate between real and fake examples.

Numerous GANs have been proposed over the years. In particular, there are two popular image translation nets that represent the fundamental principle of GANs used in Deepfake technology: pix2pix and CycleGAN.

# Deepfakes – technical background

Deepfakes are usually created using some basic networks combined with CNNs.

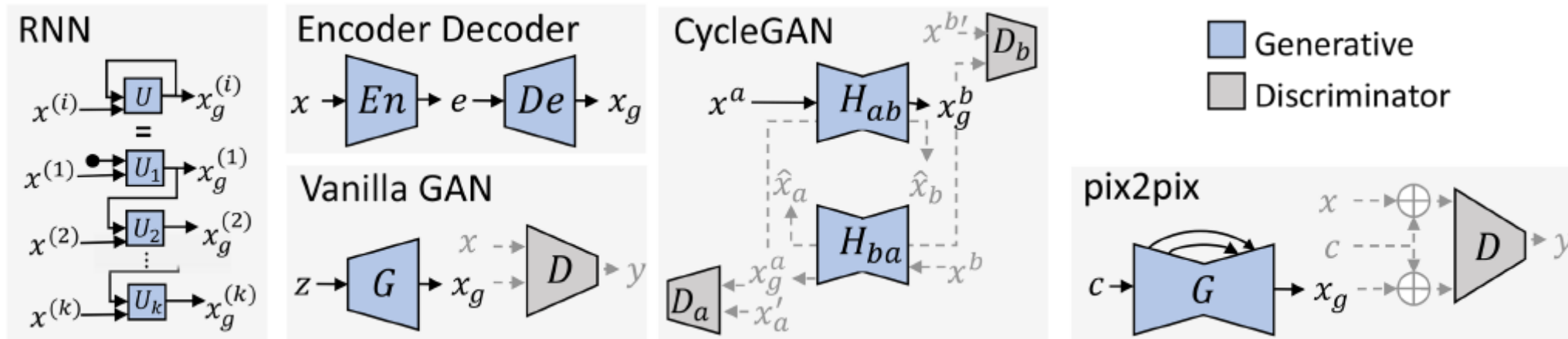


- **Recurrent Neural Networks (RNN), e.g., LSTM**
- **Encoder-Decoder (ED)**
- **Generative Adversarial Network (GAN)**

**Pix2pix:** a GAN which enables paired transformation from one image domain to another (e.g., horse to zebra). The generator is an ED-CNN based on U-Net, which has skip connections that enable the generator to bypass the compression layers when needed. Later, pix2pixHD has been released.

# Deepfakes – technical background

Deepfakes are usually created using some basic networks combined with CNNs.

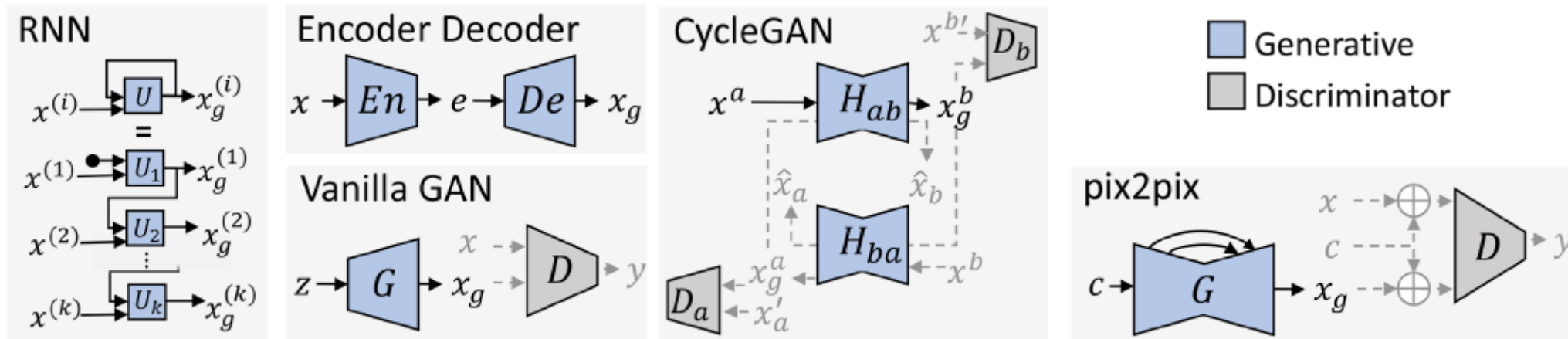


- **Recurrent Neural Networks (RNN), e.g., LSTM**
- **Encoder-Decoder (ED)**
- **Generative Adversarial Network (GAN)**
- **Image-to-Image Translation (pix2pix)**

**CycleGAN:** an improvement of pix2pix allowing unpaired image translation. The network forms a cycle consisting of two GANs used to convert images from one domain to another and then back again to ensure consistency (thanks to the cycle consistency loss).

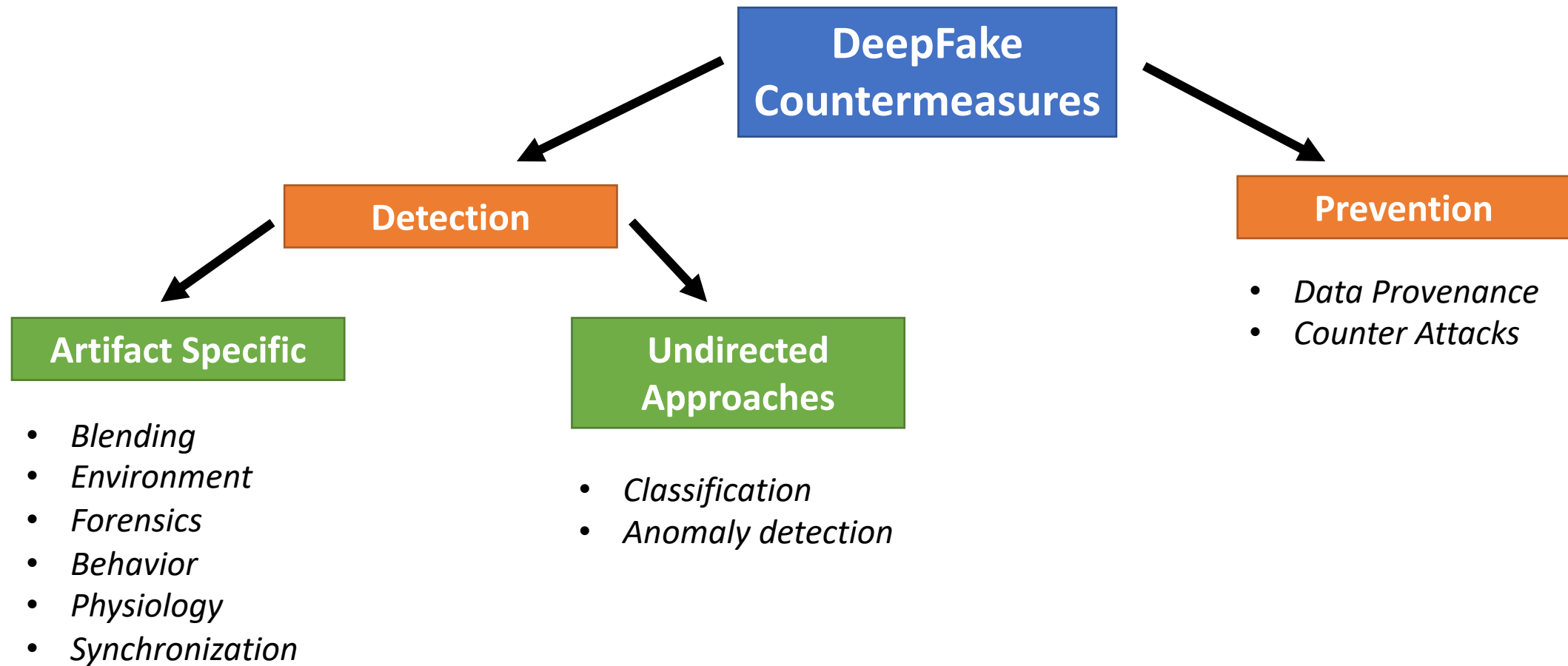
# Deepfakes – technical background

Deepfakes are usually created using some basic networks combined with CNNs.



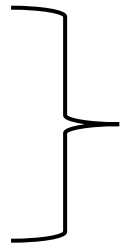
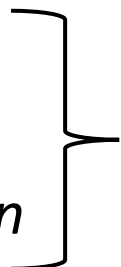
- **Recurrent Neural Networks (RNN), e.g., LSTM**
- **Encoder-Decoder (ED)**
- **Generative Adversarial Network (GAN)**
- **Image-to-Image Translation (pix2pix)**
- **CycleGAN**

# Deepfakes - countermeasures



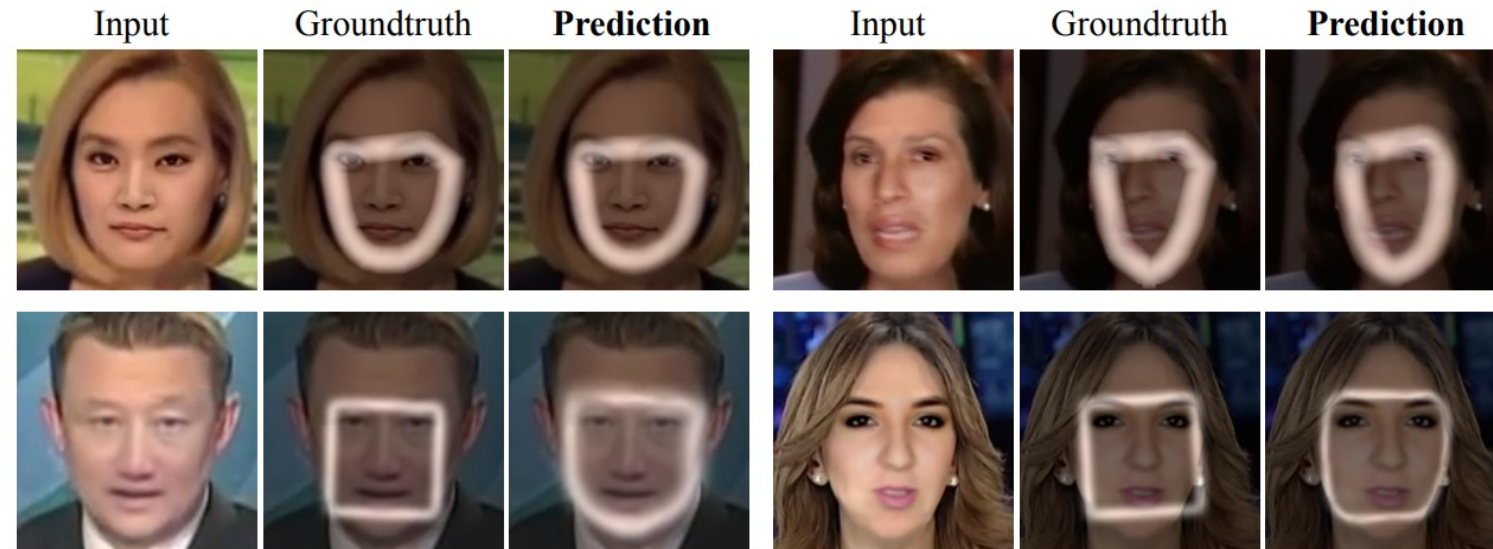
# Deepfakes - countermeasures

Deepfakes often generate artifacts that may be subtle to humans but can be easily detected using ML and forensics analysis. Some works identify deepfakes by searching for specific artifacts.

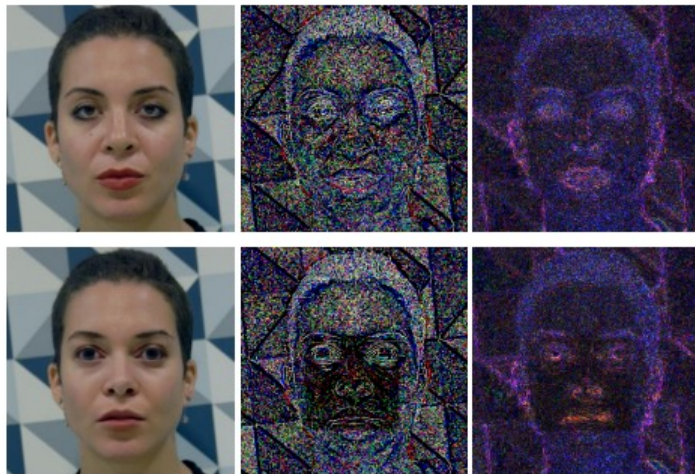
- *Blending*
  - *Environment*
  - *Forensics*
- 
- Spatial artifacts*
- *Behavior*
  - *Physiology*
  - *Synchronization*
- 
- Temporal artifacts*

# Deepfakes - countermeasures

## *Blending artifacts*



## *ELA*

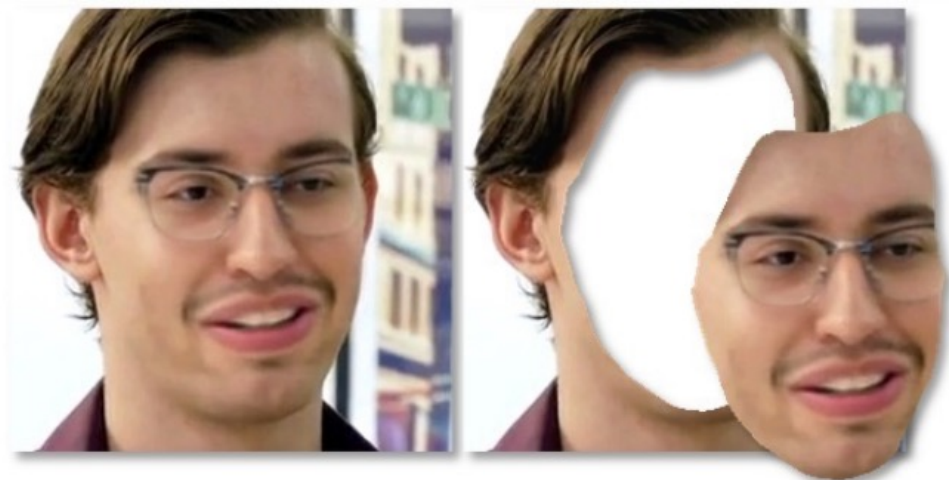
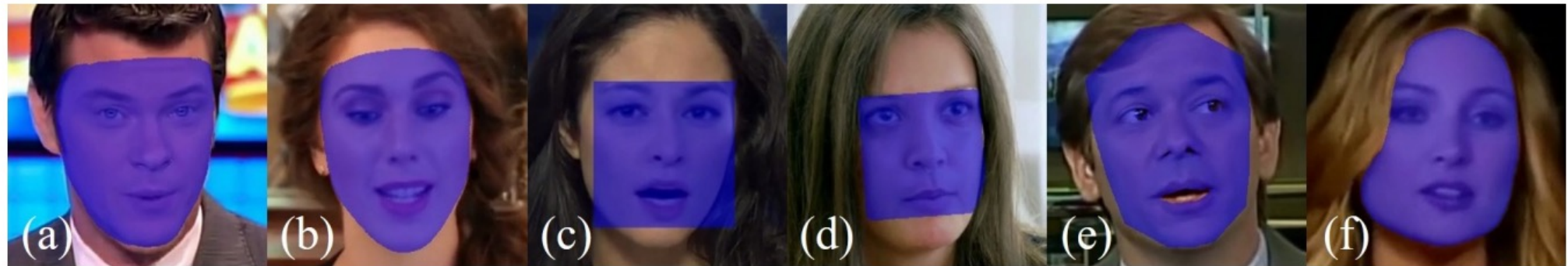


Some artifacts appear where the generated content is blended back into to the original frame.



# Deepfakes - countermeasures

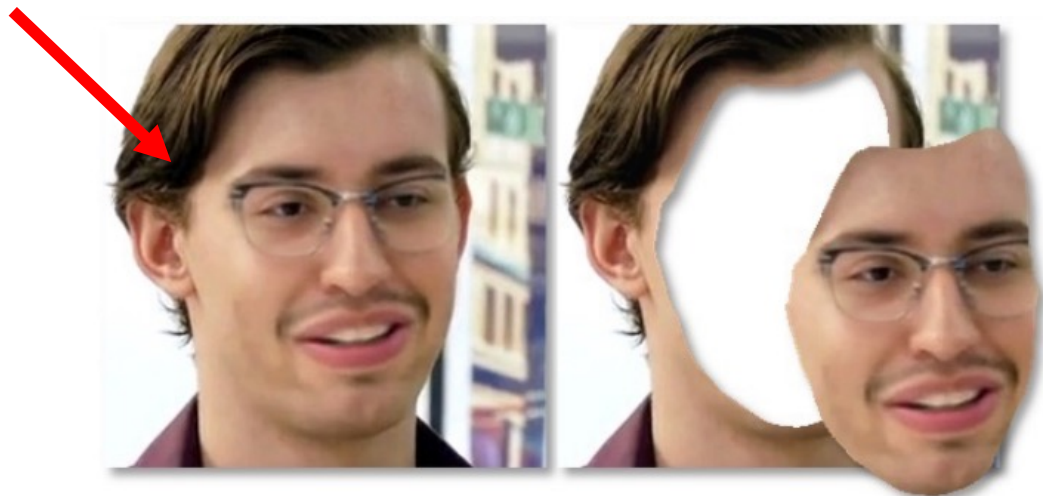
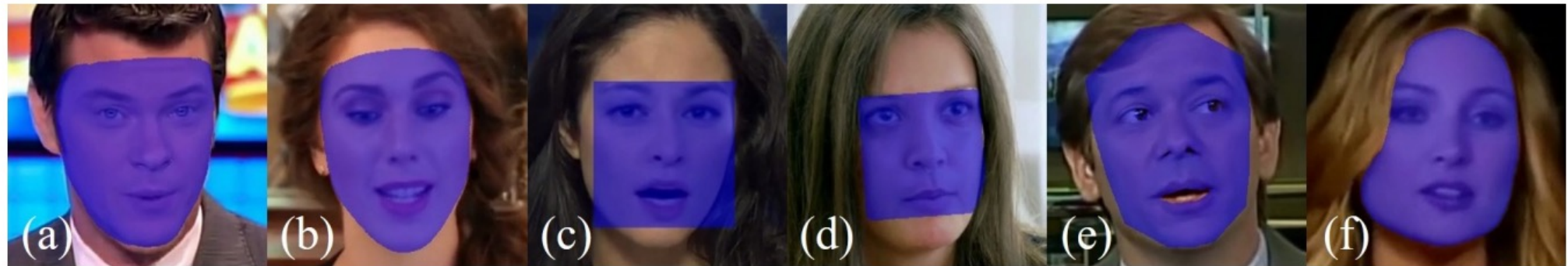
## *Environment artifacts*



The content of a fake face can be anomalous in context to the rest of the image (foreground vs. background).

# Deepfakes - countermeasures

## *Environment artifacts*

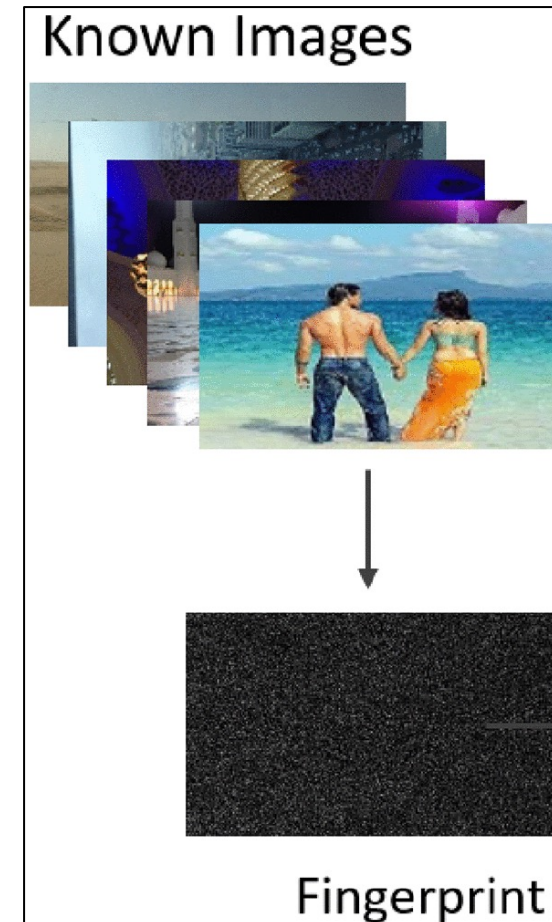
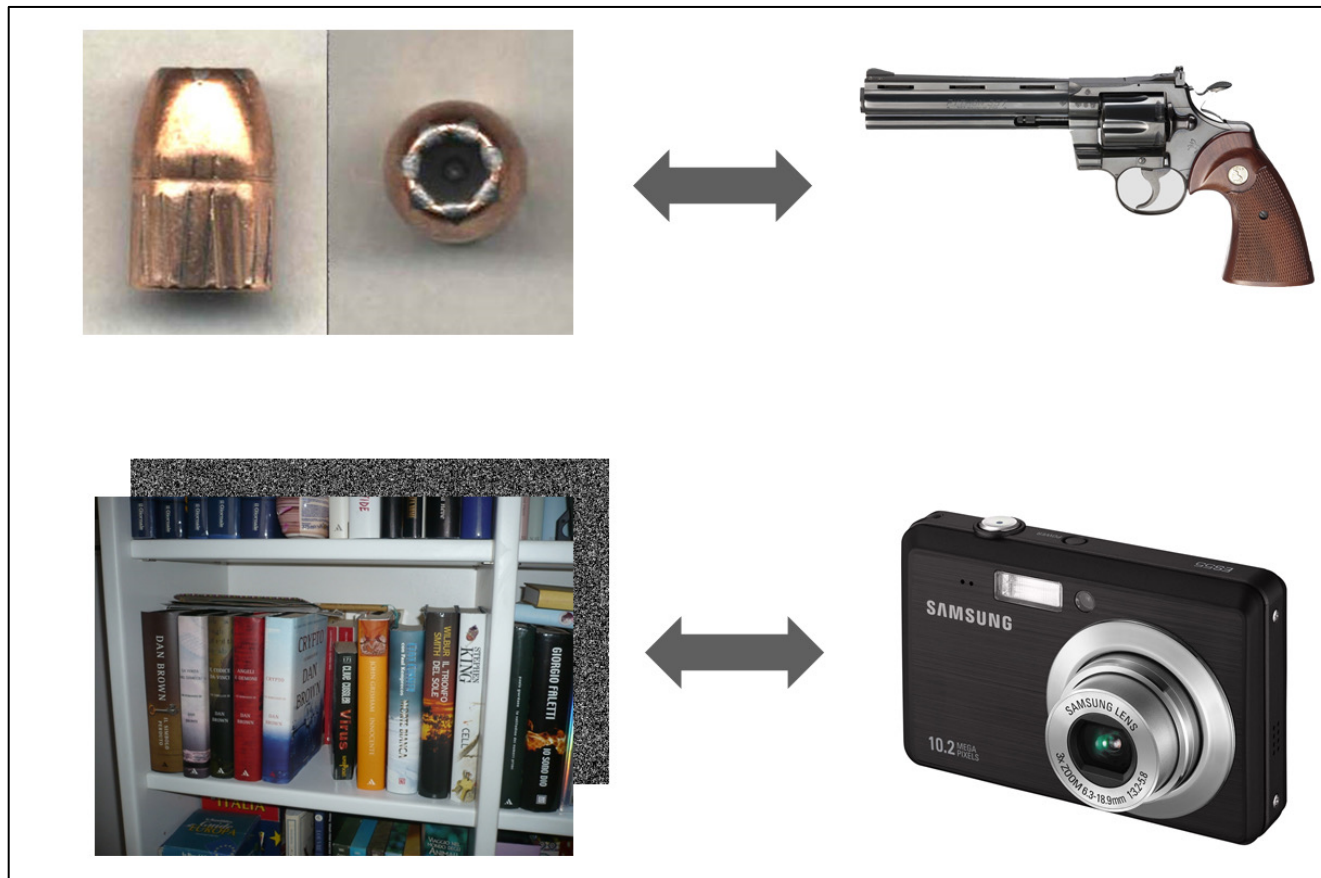


The content of a fake face can be anomalous in context to the rest of the image (foreground vs. background).

# Deepfakes - countermeasures

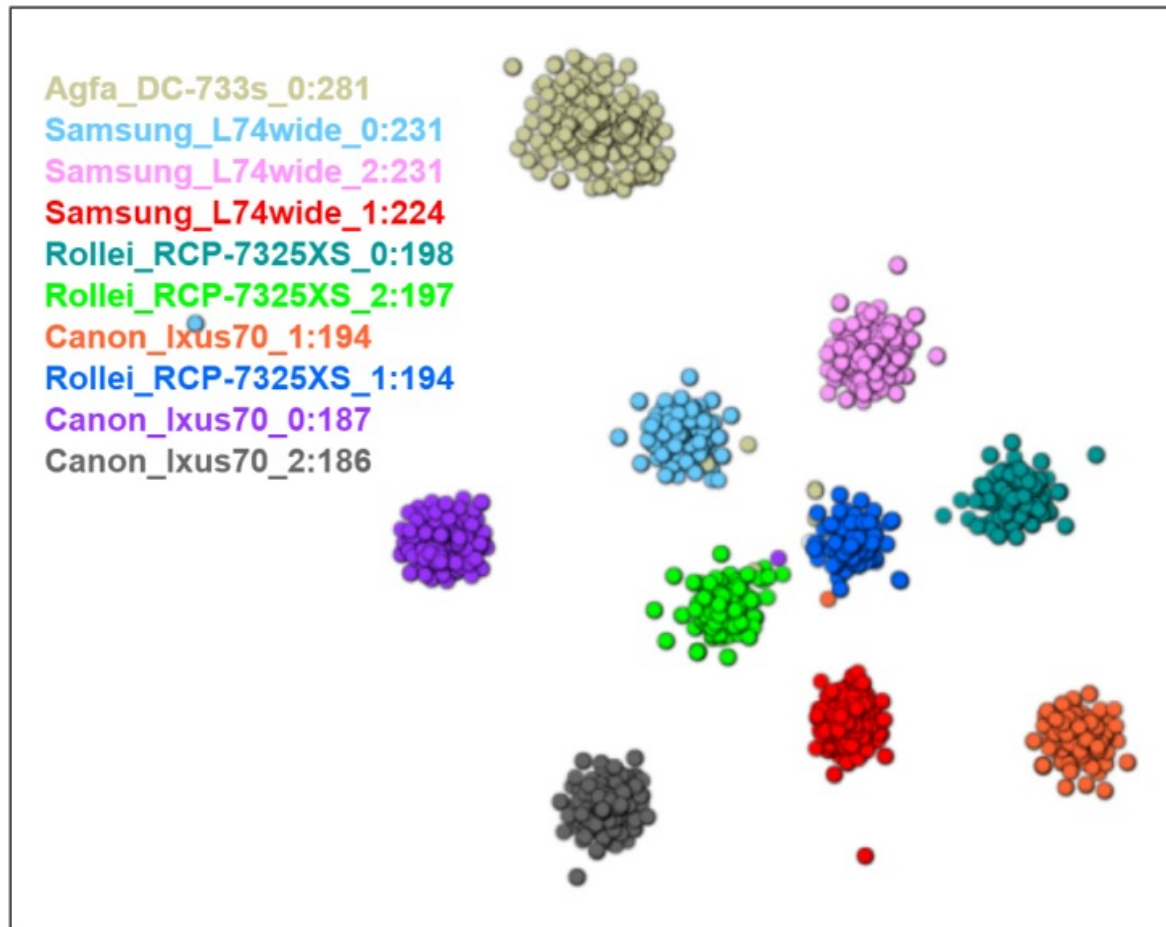
## Forensics

Traditional forensics methods can exploit the so-called PRNU (Photo Response Non-Uniformity) which can be considered as a sort of **camera fingerprint**.



# Deepfakes - countermeasures

## Forensics



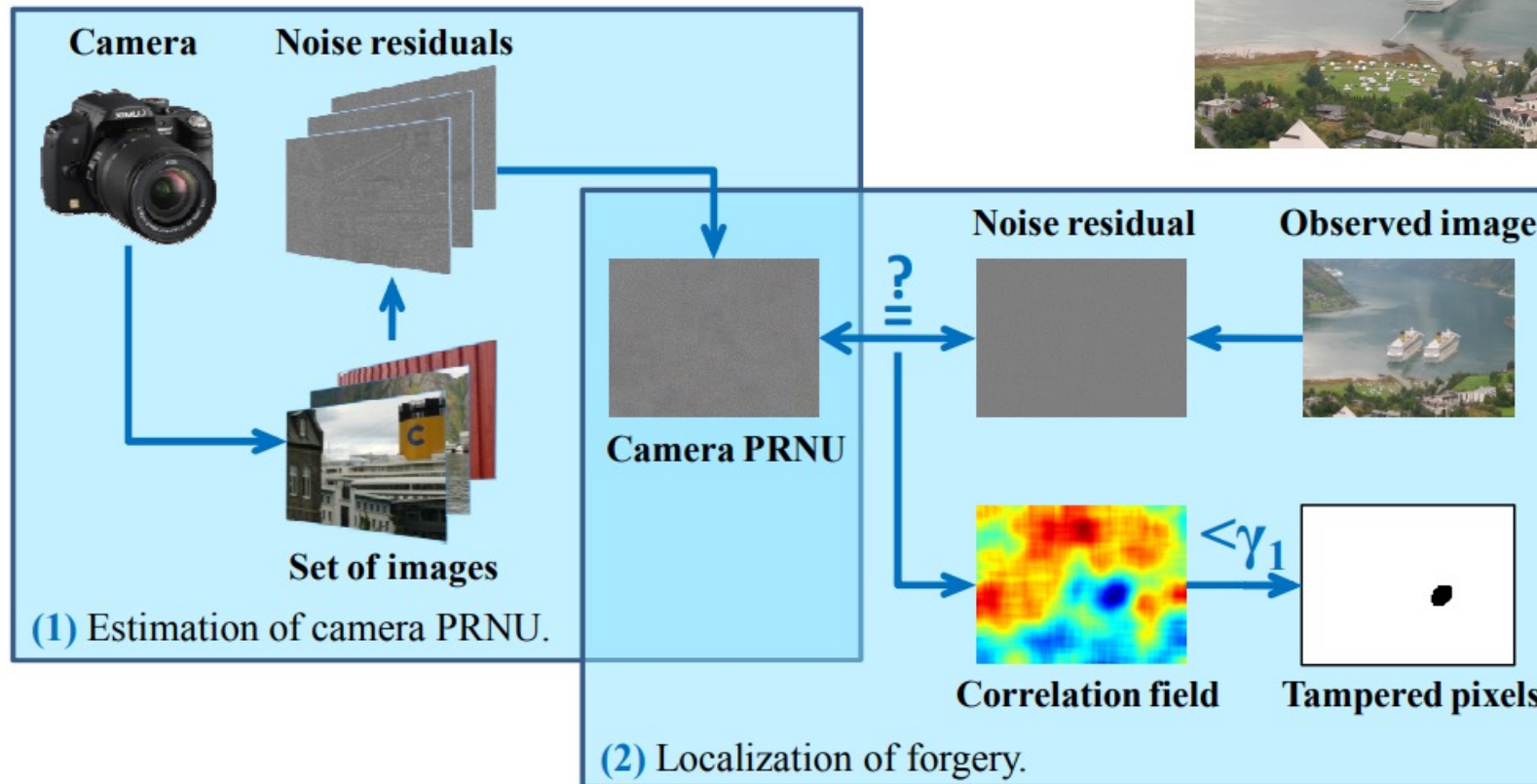
Traditional forensics methods can exploit the so-called PRNU (Photo Response Non-Uniformity) which can be considered as a sort of **camera fingerprint**.

$$y_i = (1 + k_i)x_i + \theta_i = x_i k_i + x_i + \theta_i$$

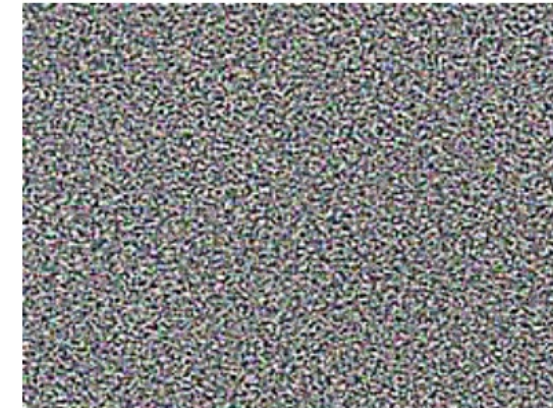
PRNU                      Additive noise  
 ↓                              ↓  
 ↑                              ↑  
*i*-th pixel value                      Ideal (noise-free) image

# Deepfakes - countermeasures

## Forensics



Camera PRNU



Tampered parts of the image can be detected by analyzing the inconsistencies of the PRNU.

# Deepfakes - countermeasures

## Forensics

The work by *Guarnera et al.*, aims to extract a Deepfake fingerprint from images, by exploiting a model trained to detect and extract a **fingerprint** that represents the **Convolutional Traces (CT) left by GANs** during image generation.

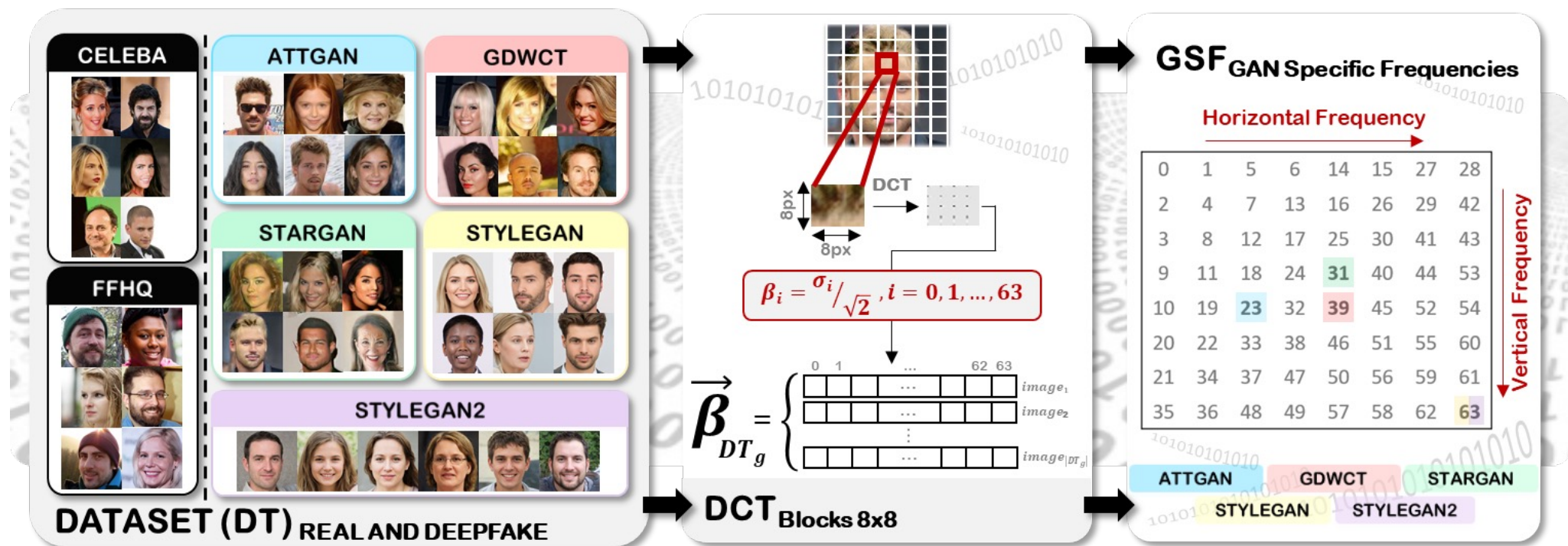


Method	Kernel size of the latest Convolution Layer
GDWCT	4x4
STARGAN	7x7
ATTGAN	4x4
STYLEGAN	3x3
STYLEGAN2	3x3

# Deepfakes - countermeasures

## Forensics

Traces left by GANs during the creation of the Deepfakes can be also detected by analyzing ad-hoc frequencies. The paper by *Giudice et al.* proposed a new pipeline able to detect the so-called **GAN Specific Frequencies (GSF)** representing a unique fingerprint of the different generative architectures.



# Deepfakes - countermeasures

## **Forensics**

Traces left by GANs during the creation of the Deepfakes can be also detected by analyzing ad-hoc frequencies. The paper by *Giudice et al.* proposed a new pipeline able to detect the so-called **GAN Specific Frequencies (GSF)** representing a unique fingerprint of the different generative architectures.

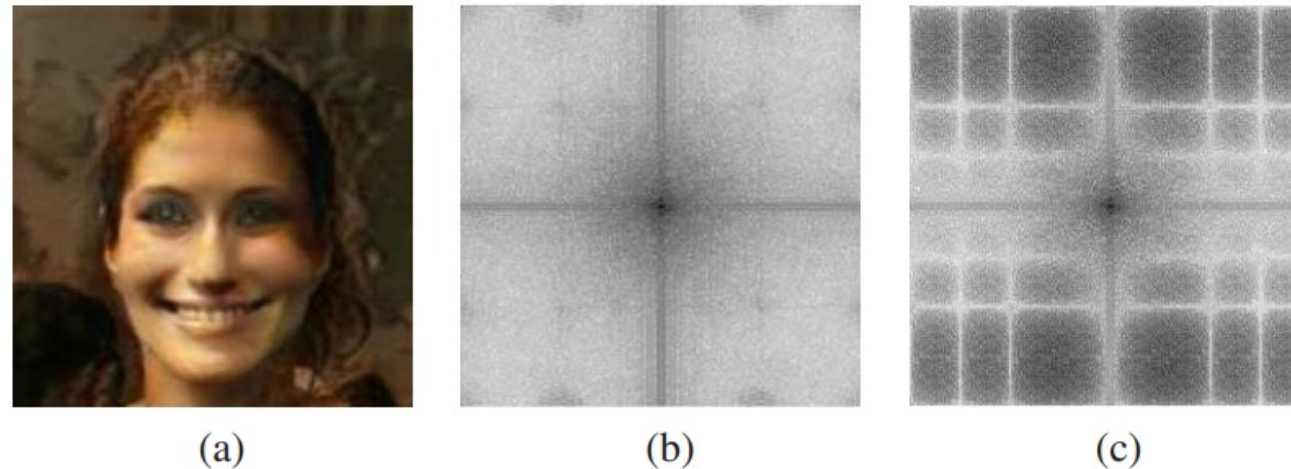


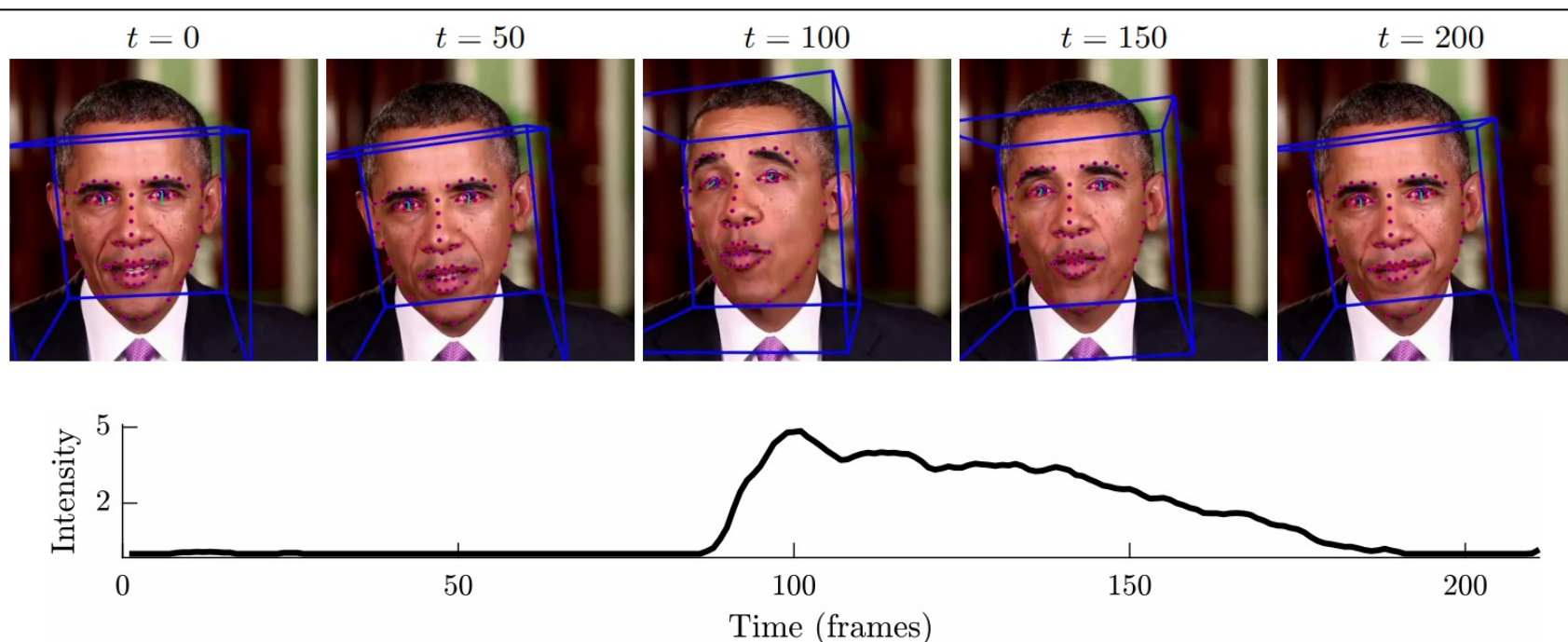
Fig. 8. Abnormal frequencies inspection. (a) Image example from the StarGAN dataset; (b) Fourier Spectra of the input image (a); (c) Abnormal frequency shown by means of *GSF* amplification.



# Deepfakes - countermeasures

## **Behaviour**

Behaviours can be monitored for anomalies. Agarwal et al., hypothesize that as an individuals have distinct facial expressions and movements while speaking. By tracking facial and head movements, the authors defined a detection model that distinguishes an individual from other individuals (1 vs all SVM).

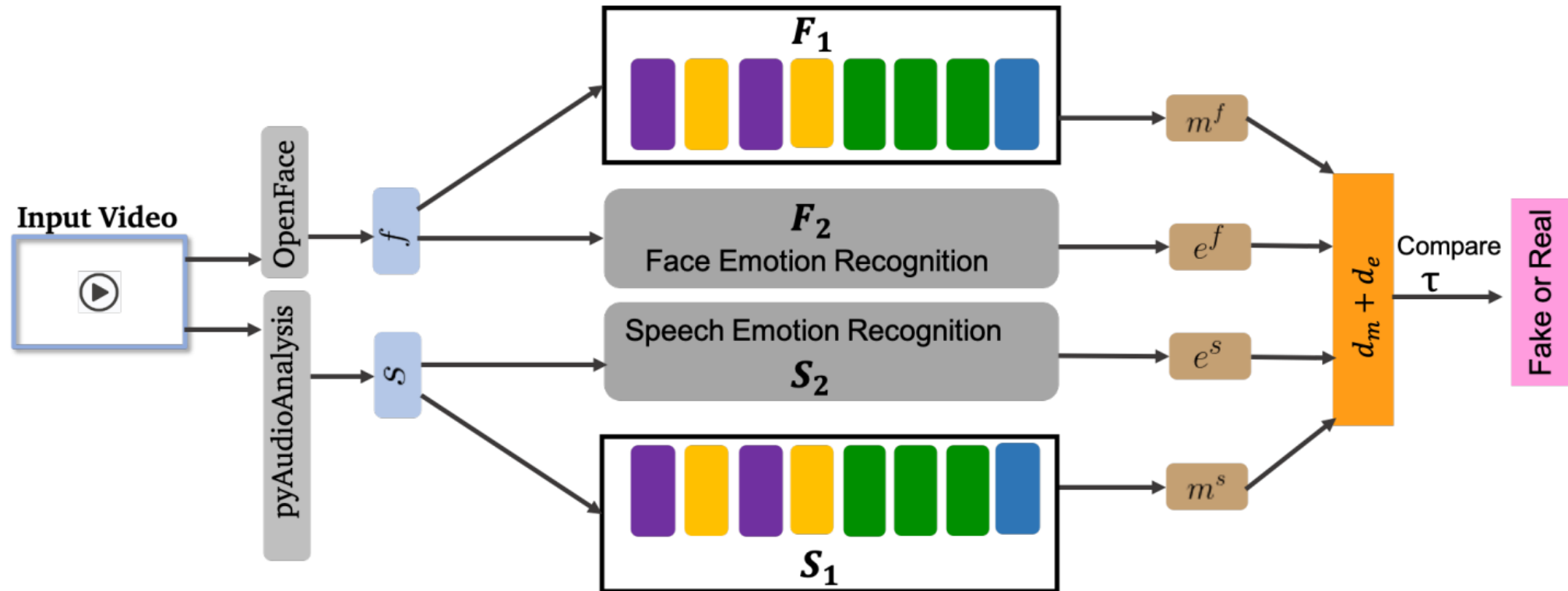


**Obama's  
eye brow  
lift analysis.**

# Deepfakes - countermeasures

## Behaviour

The work by Mittal et al. detects discrepancies in the perceived emotion extracted from the clip's audio and video content. This approach doesn't require a reference footage of the target.

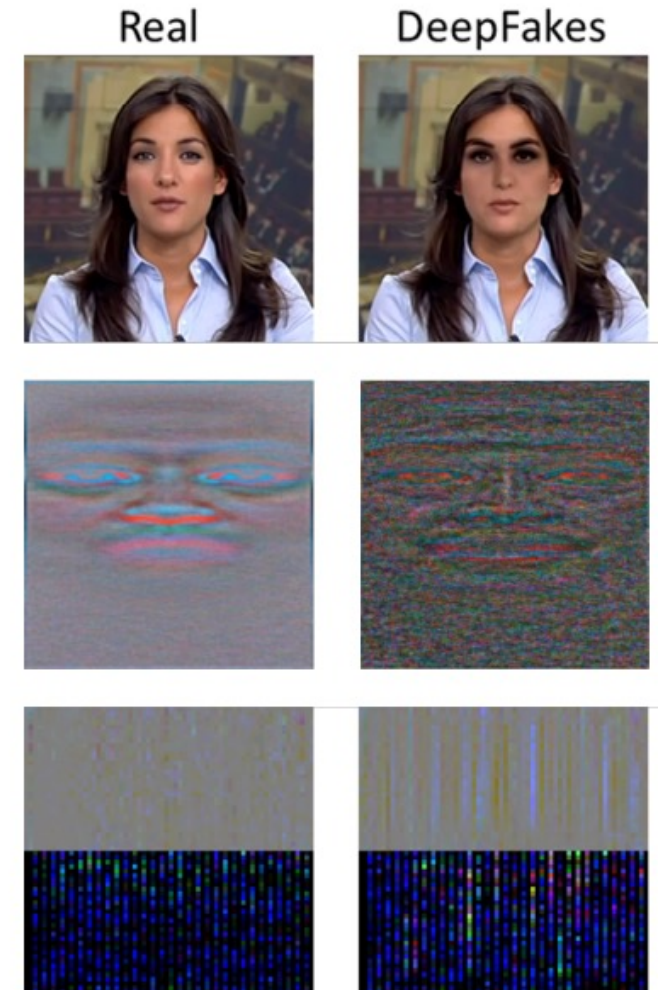
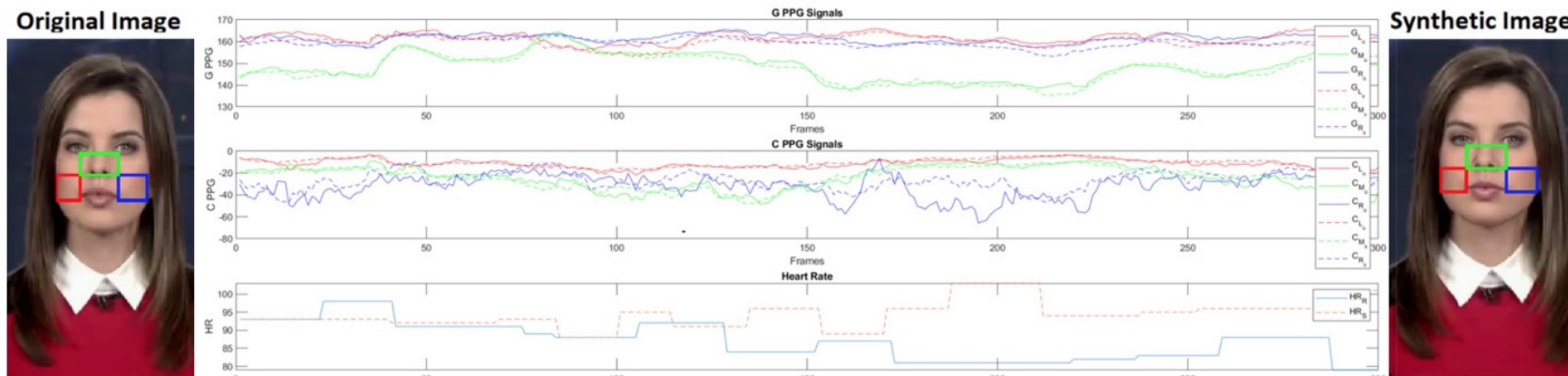


# Deepfakes - countermeasures

**Physiology** Deepfake generated contents lack physiological signals.

Anatomical actions create subtle changes that are not visible to the eye but still detectable computationally.

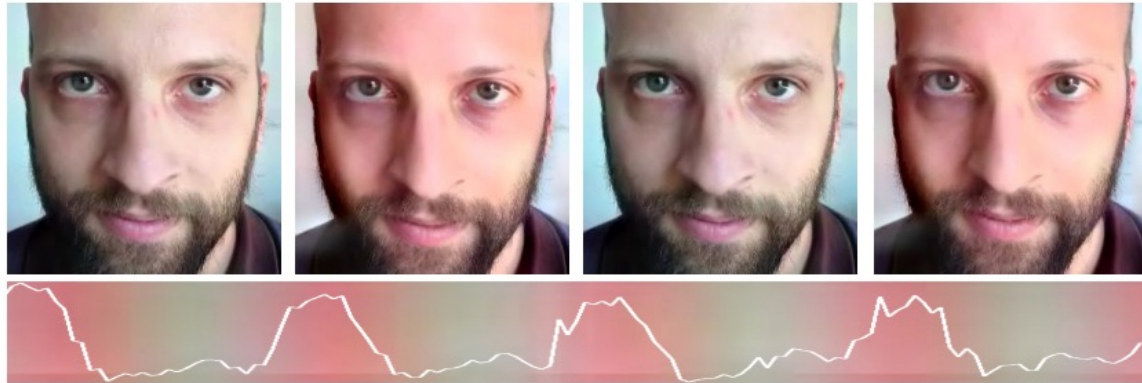
Approaches to extract photoplethysmography (PPG) signals are developed to recognize such changes by image processing techniques.



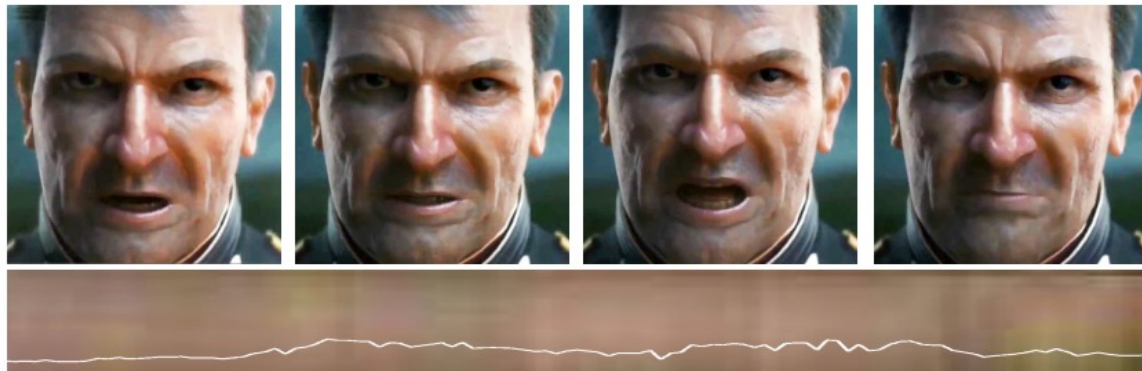
# Deepfakes - countermeasures

**Physiology** Deepfake generated contents lack physiological signals.

**Real**



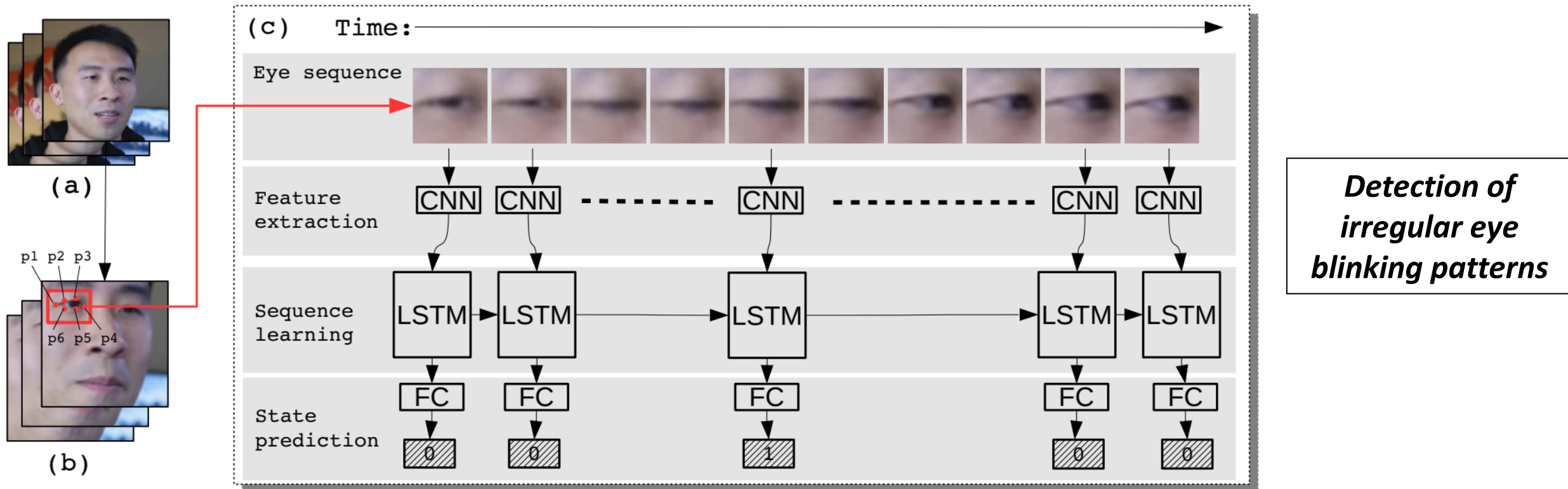
**CG**



**Blood flow  
changes detected  
from the skin.**

# Deepfakes - countermeasures

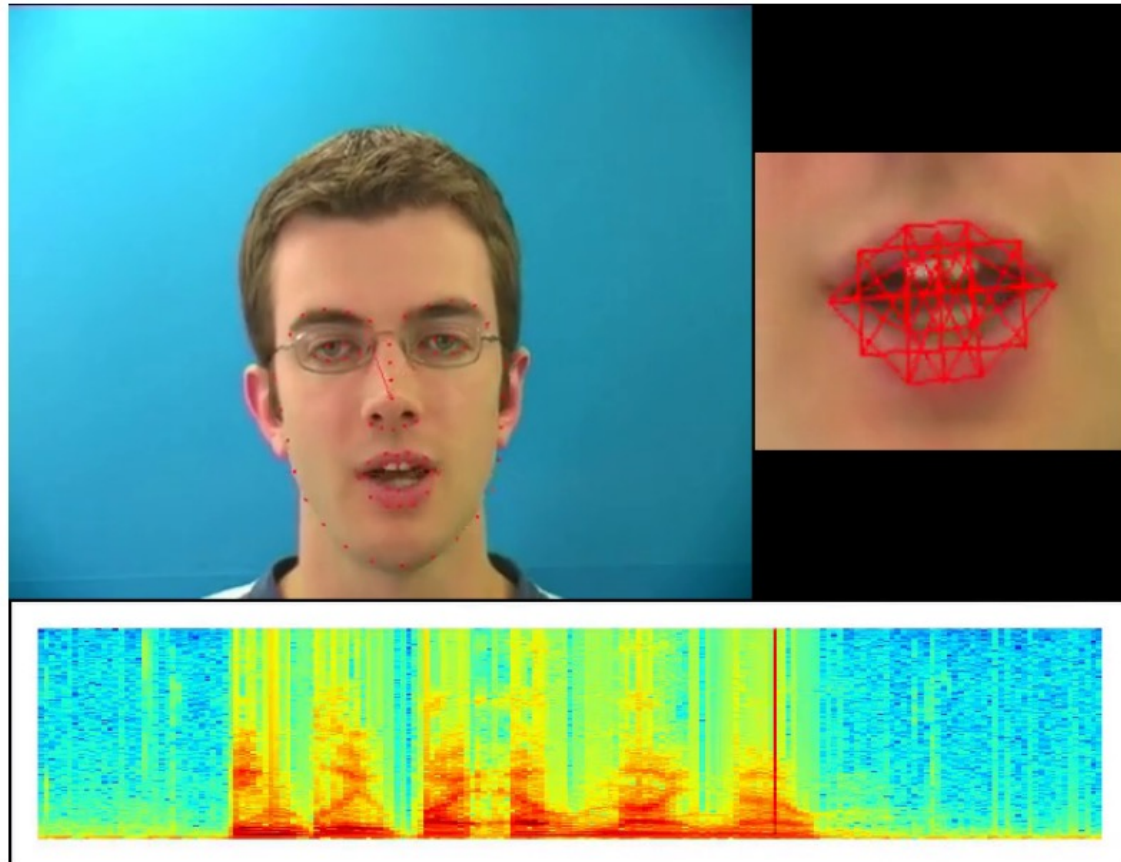
**Physiology** Deepfake generated contents lack physiological signals.



# Deepfakes - countermeasures

## ***Synchronization***

Sync inconsistencies between speech and landmarks around the mouth.



***Comparison  
between lips  
shape and  
spectrogram***

# Deepfakes - countermeasures

Generated images sometimes have typical defects such as shiny blobs or coherence issues (e.g., just one earring, missing parts of objects, background, etc.)

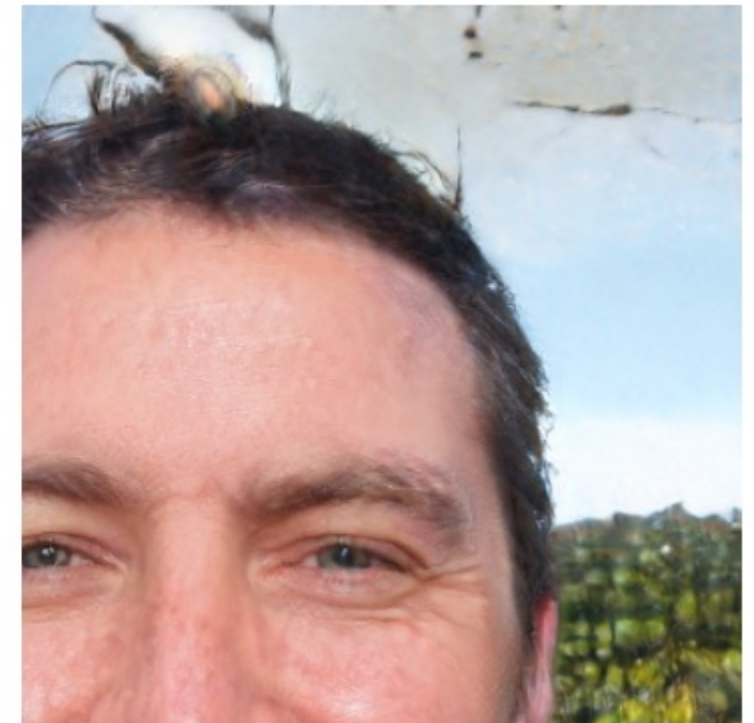
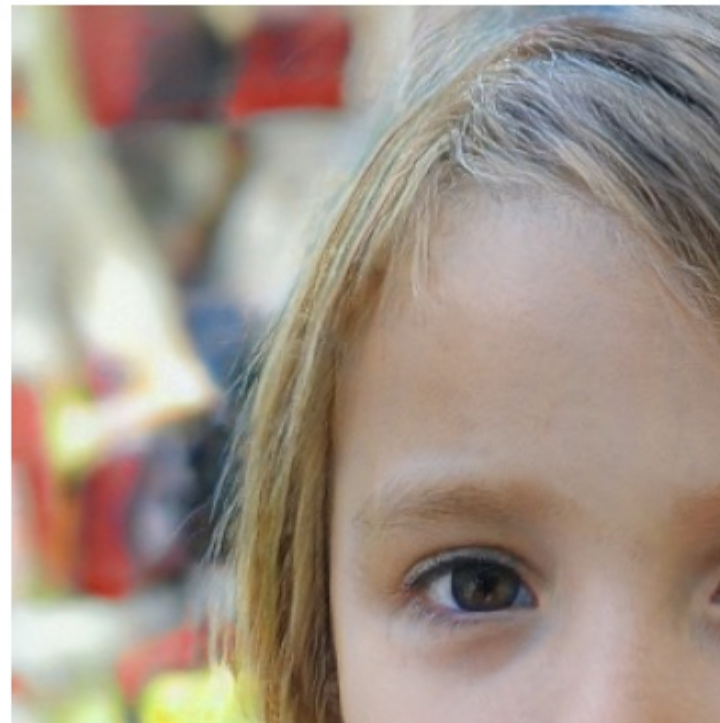
***Shiny blobs***



# Deepfakes - countermeasures

Generated images sometimes have typical defects such as shiny blobs or coherence issues (e.g., just one earring, missing parts of objects, background, etc.)

***Background  
problems***





# Deepfakes - countermeasures

Generated images sometimes have typical defects such as shiny blobs or coherence issues (e.g., just one earring, missing parts of objects, background, etc.)

***Eyeglasses and other symmetry problems***



# Deepfakes - countermeasures

Generated images sometimes have typical defects such as shiny blobs or coherence issues (e.g., just one earring, missing parts of objects, background, etc.)

## *Fluorescent bleed*



# Deepfakes - countermeasures

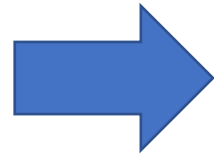


***Due to DS bias the model tries to draw earrings.***

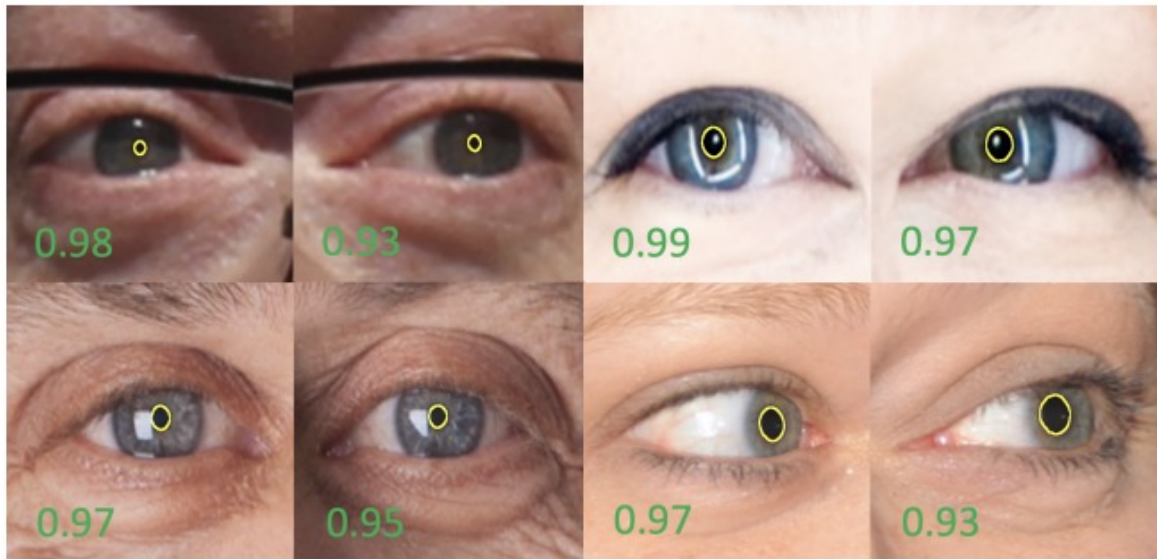
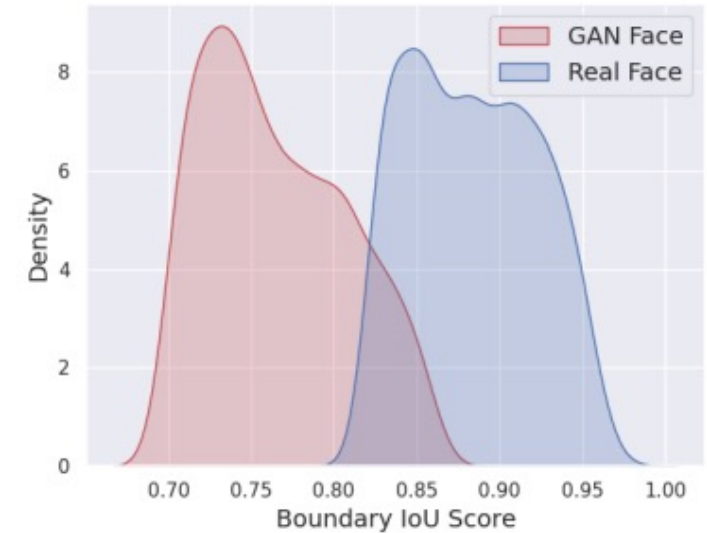
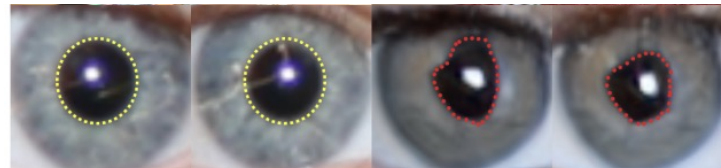


# Deepfakes - countermeasures

*real*                      *fake*



*real*                      *fake*



**Real examples**



**Fake examples**

# Deepfakes - countermeasures

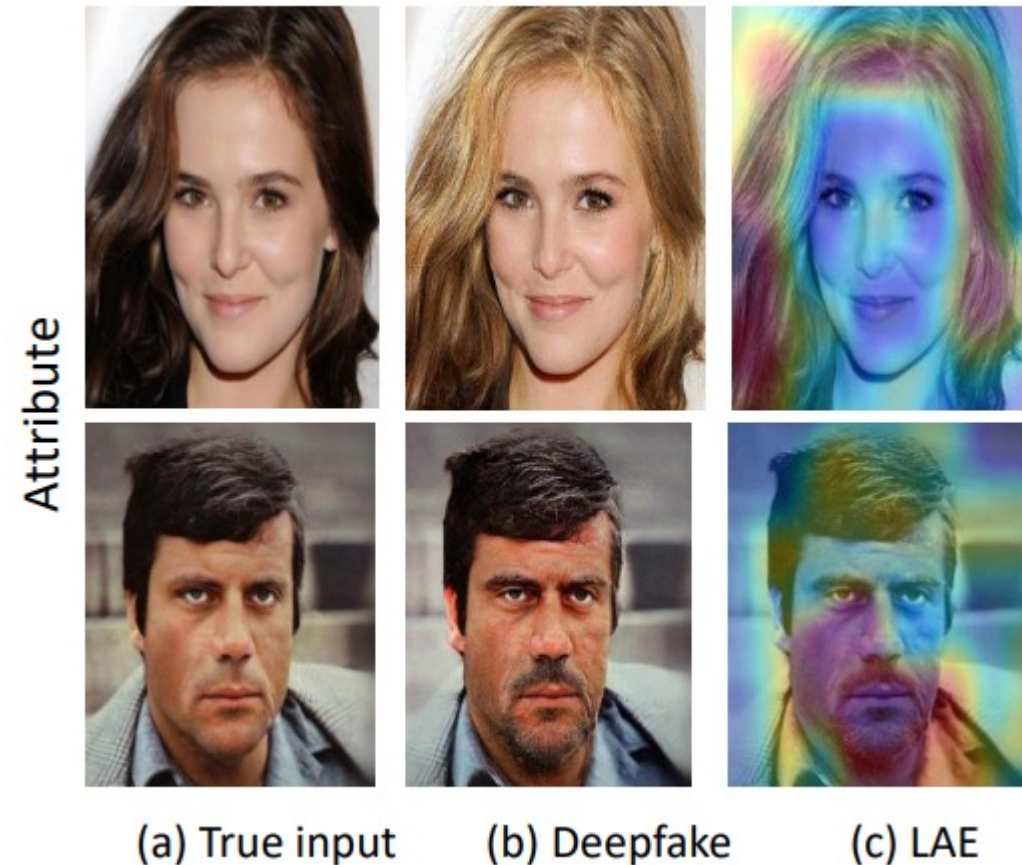
Instead of focusing on artifacts, we can try to train a neural network and let it decide which features to analyse to detect fake contents.

- *Classification*
- *Anomaly Detection*

# Deepfakes - countermeasures

## **Classificaiton**

CNNs can be used to detect and also localize the tampered areas predicting masks learned from GT datasets or by mapping the activations back to the raw image.

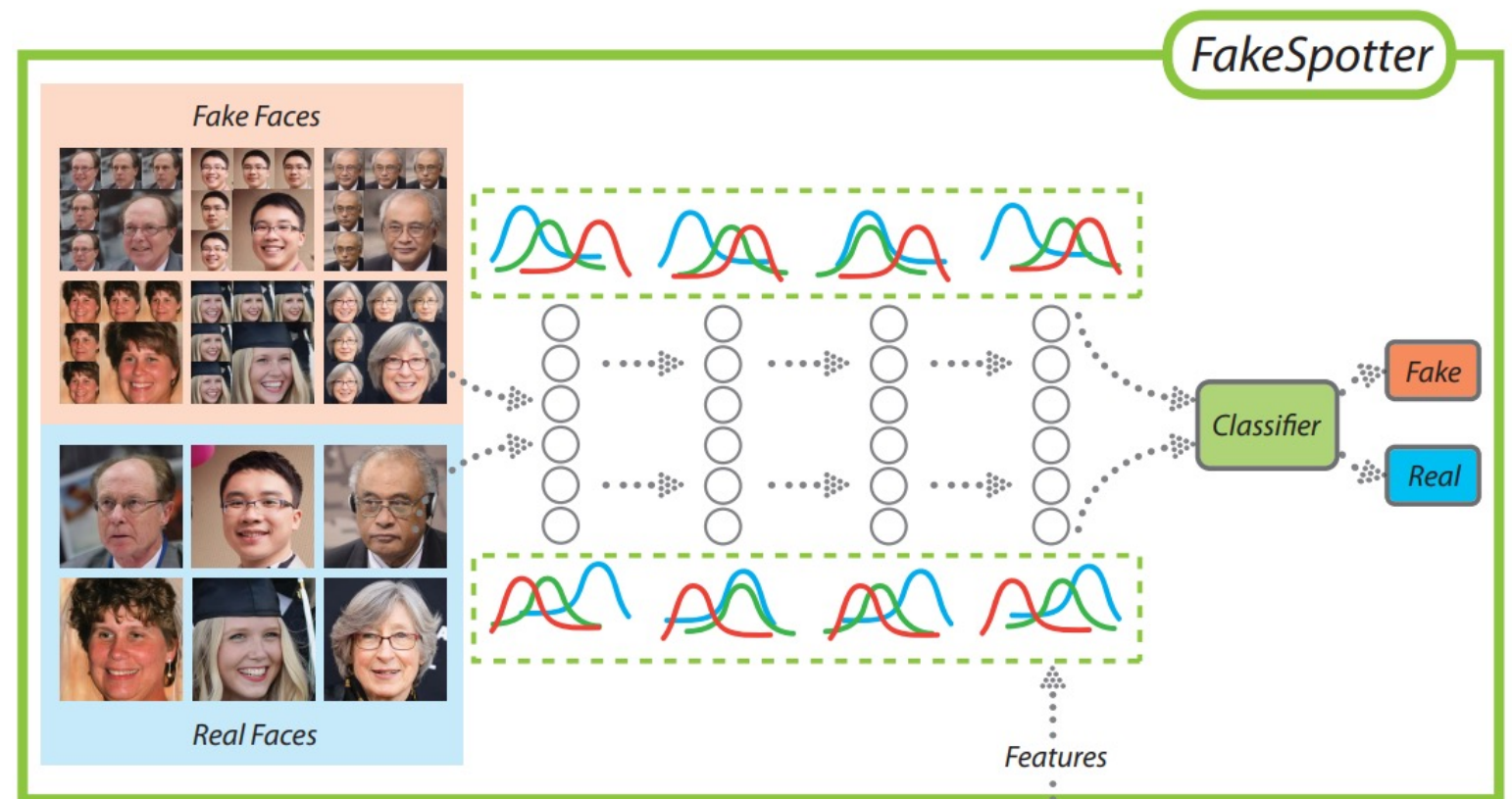


# Deepfakes - countermeasures

## Anomaly Detection

Anomaly detection has been applied on the distributions of the neural activations of a face recognition network.

Distortions are detected by finding anomalies on activations' distributions.



Wang, Run, et al. "Fakespotter: A simple yet robust baseline for spotting ai-synthesized fake faces." arXiv preprint arXiv:1909.06122 (2019).

# Deepfakes - countermeasures

Instead of focusing on artifacts, we can try to train a neural network and let it decide which features to analyse to detect fake contents.

## ***Classification***

Standard CNN can detect deepfakes, but only attacks coming from the models they have been trained on. More robust results can be obtained by ensemble learning.

In general, an attacker can evade detection via adversarial learning (i.e., perturbing the input).

## ***Anomaly Detection***

In general, comparing embedded (and multiple) image representations with the training distributions is more robust than using the raw pixels.



# Deepfakes – Research Trends

Some notable advancements over the last few years include:

- Unpaired training to reduce the amount of training data
- Few-shot learning, which enables identity theft with a single profile picture
- Improvements of quality
- Mitigation of boundary artifacts by using secondary networks to blend composites into seamless imagery

# Deepfakes – Current Limitations

There are a few limitations with the current deepfake technologies:

- Content is always driven and generated with frontal pose (reenactment)
- Reenactment depends on the driver's performances (next generation DF will not use drivers)
- Real-time deepfake is difficult (quality/speed trade-off)
- Difficult to render target's hands (especially when touching the face)
- Other limitations include the coherent rendering of: hair, teeth, tongues and shadows

# Deepfakes – Final Remarks

- Not all deepfakes are malicious
- Attacks are targeting individuals and causing psychological, political, monetary harm
- We expect malicious deepfake will spread to many other modalities
- Deepfake are improving at a rapid rate, it is important that we focus on effective countermeasures

# References (1)

- <https://gizmodo.com/deepfake-lips-are-coming-to-dubbed-films-1846840191>
- <https://www.bbc.com/news/technology-56210053>
- <https://derivative.ca/community-post/deepfake-salvador-dal%C3%AD-interacts-museum-visitors-takes-selfies>
- <https://www.reddit.com/r/SFWdeepfakes/>
- <https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn>
- [https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman?utm\\_source=vicetwitterus](https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman?utm_source=vicetwitterus)
- <https://www.dailymail.co.uk/sciencetech/article-8863233/Disturbing-deepfake-tool-popular-messaging-app-Telegram-forging-NUDE-images-underage-girls.html>
- <https://apnews.com/article/ap-top-news-artificial-intelligence-social-platforms-think-tanks-politics-bc2f19097a4c4fffaa00de6770b8a60d>
- <https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/>  
<https://techmonitor.ai/techonology/cybersecurity/growing-threat-audio-deepfake-scams>
- Mirsky, Yisroel, and Wenke Lee. "The creation and detection of deepfakes: A survey." *ACM Computing Surveys (CSUR)* 54.1 (2021): 1-41.
- <https://github.com/iperov/DeepFaceLab>
- <https://github.com/iperov/DeepFaceLive>
- Kumar, Rithesh, et al. "Obamanet: Photo-realistic lip-sync from text." *arXiv preprint arXiv:1801.01442* (2017).
- Zakharov, Egor, et al. "Few-shot adversarial learning of realistic neural talking head models." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.

# References (2)

- *Li, Lingzhi, et al. "Face x-ray for more general face forgery detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.*
- *Nirkin, Yuval, et al. "DeepFake Detection Based on Discrepancies Between Faces and their Context." IEEE Transactions on Pattern Analysis and Machine Intelligence (2021).*
- *Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "DeepFake Detection by Analyzing Convolutional Traces." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.*
- *Giudice, Oliver, Luca Guarnera, and Sebastiano Battiato. "Fighting deepfakes by detecting GAN DCT anomalies." arXiv preprint arXiv:2101.09781 (2021).*
- *Agarwal, Shruti, et al. "Protecting World Leaders Against Deep Fakes." CVPR workshops. Vol. 1. 2019.*
- *Mittal, Trisha, et al. "Emotions Don't Lie: An Audio-Visual Deepfake Detection Method using Affective Cues." Proceedings of the 28th ACM international conference on multimedia. 2020.*
- *Ciftci, Umur Aybars, Ilke Demir, and Lijun Yin. "How do the hearts of deep fakes beat? Deep fake source detection via interpreting residuals with biological signals." 2020 IEEE International Joint Conference on Biometrics (IJCB). IEEE, 2020.*
- *Conotter, Valentina, et al. "Physiologically-based detection of computer generated faces in video." 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014.*
- *Li, Yuezun, Ming-Ching Chang, and Siwei Lyu. "In ictu oculi: Exposing ai created fake videos by detecting eye blinking." 2018 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, 2018.*
- *Guo, Hui, et al. "Eyes Tell All: Irregular Pupil Shapes Reveal GAN-generated Faces." arXiv preprint arXiv:2109.00162 (2021).*
- *Du, Mengnan, et al. "Towards generalizable deepfake detection with locality-aware autoencoder." Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020.*
- *Wang, Run, et al. "Fakespotter: A simple yet robust baseline for spotting ai-synthesized fake faces." arXiv preprint arXiv:1909.06122 (2019).*