Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Evidence based on Digital Data

Prof. dr. ing. Zeno Geradts

Outline

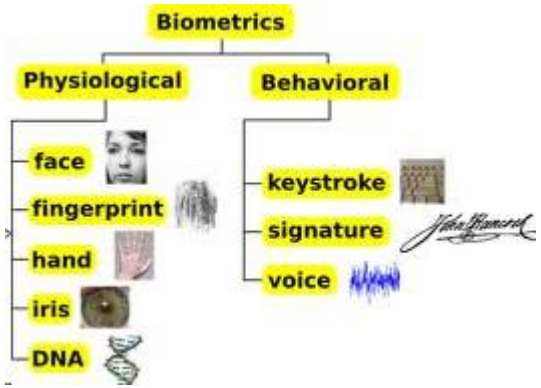
- Introduction
- Daubert and AI
- Explainable AI examples
- Hansken AI
- Deepfakes
- Detecting deepfakes
- Explainable AI
- Conclusions

Casework: Chip cards, Electronic devices
Computers, Networks, phones, cell site analysis

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

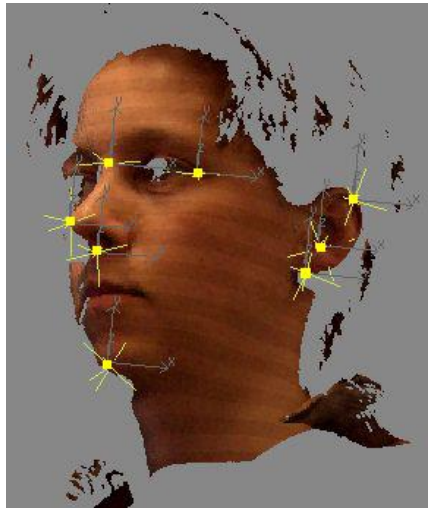
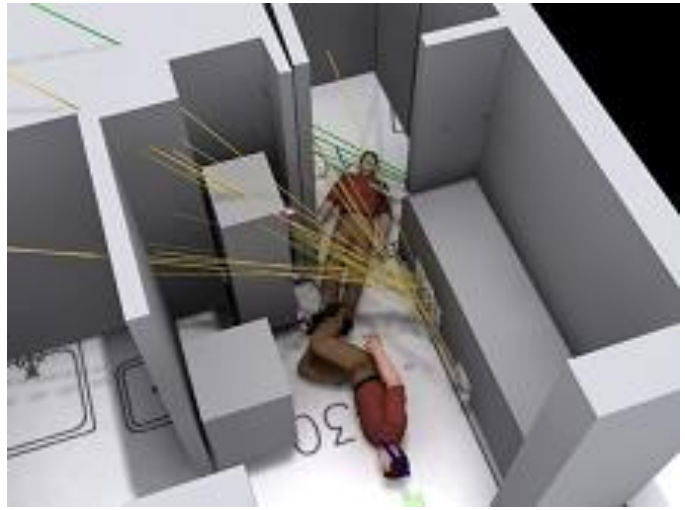
Casework: Traffic and Industrial Accidents

Casework: Audio and speech, Biometric devices



Casework: Image analysis, 3D modeling

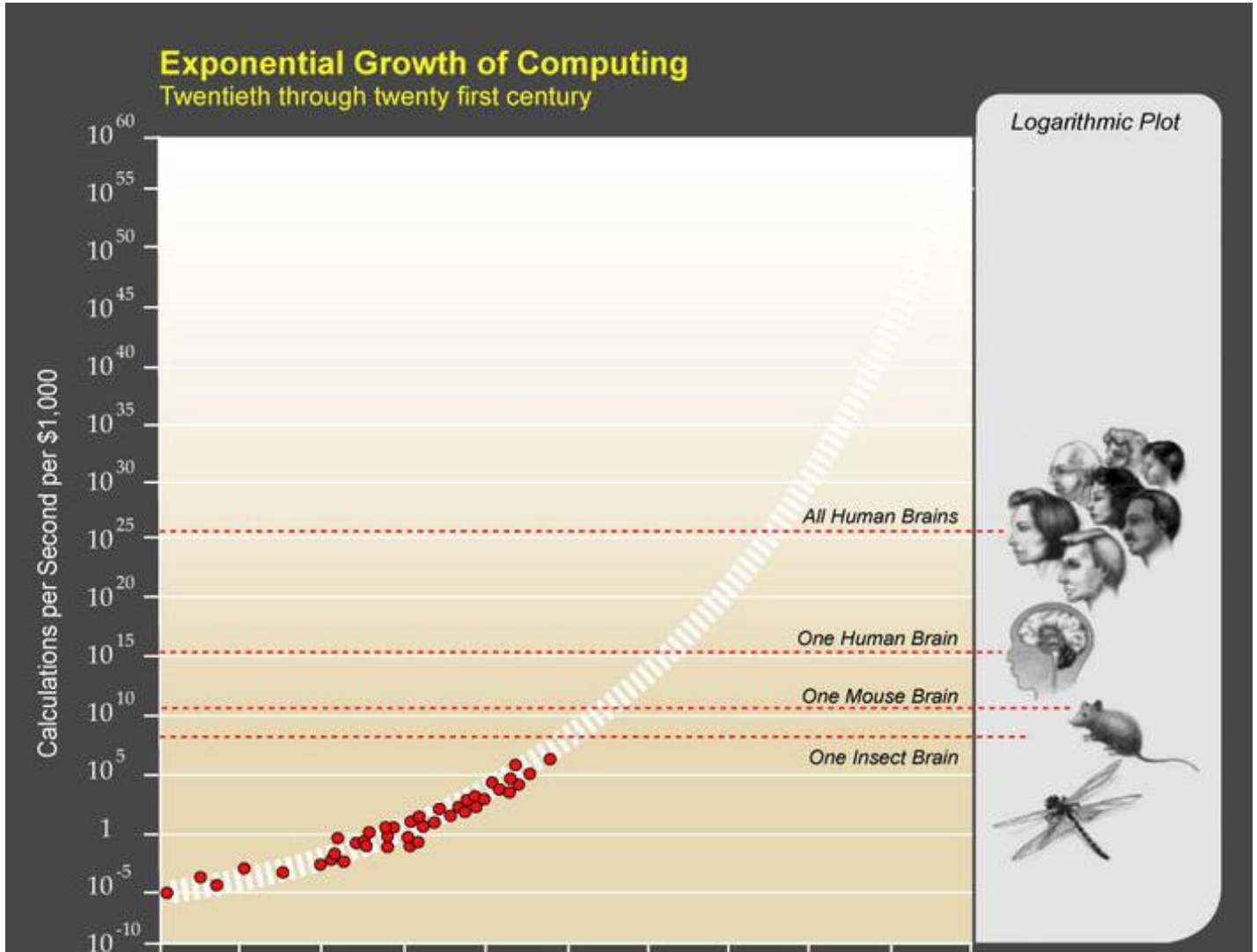


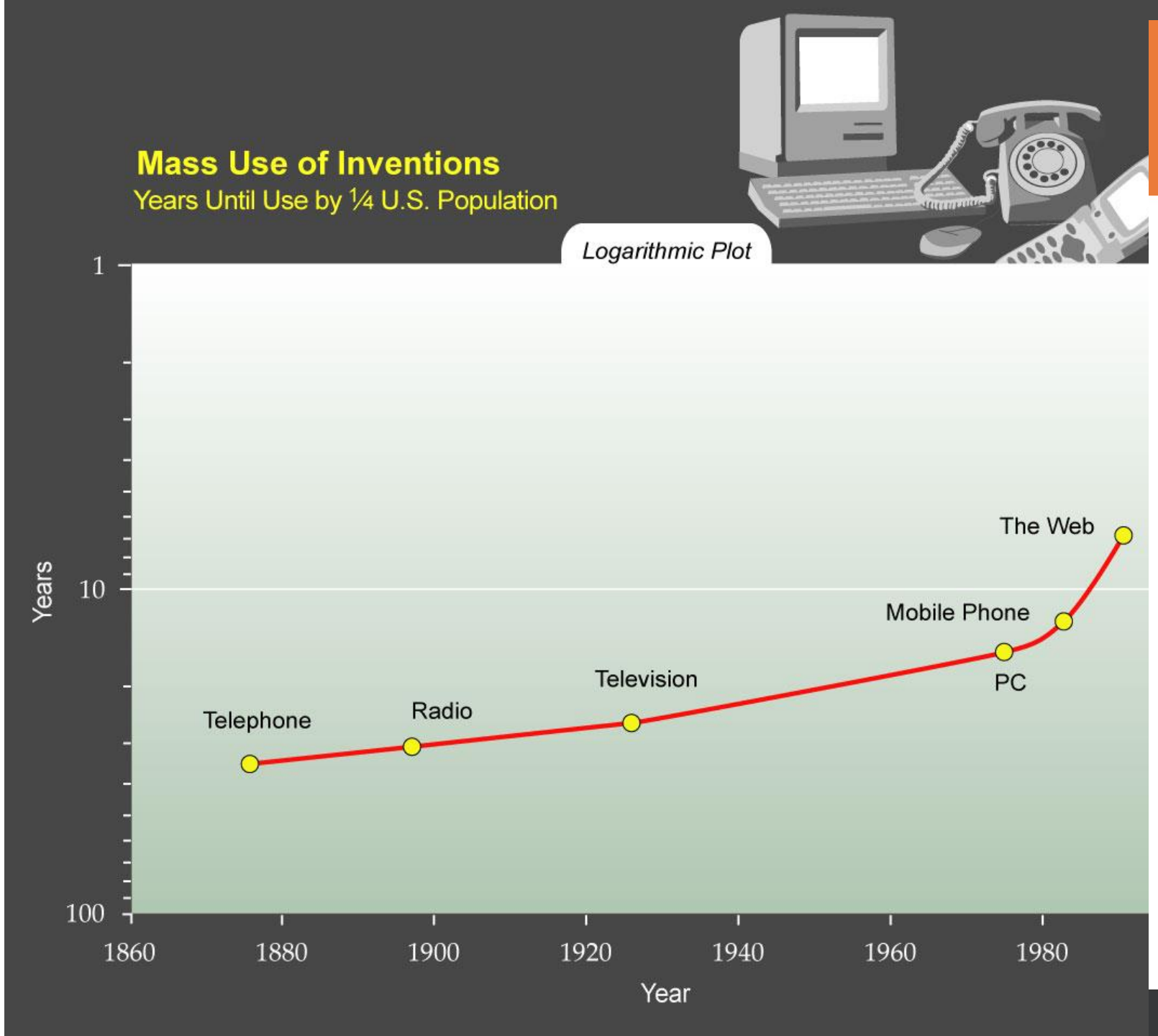




Chair Forensic Data Science

- store and process
- understand and decide
- analyze and model
- Report and visualize
- Higher efficiency
- Data-intensive
- Evidential strength big data ?

Big data


Exponential Growth of Computing
Twentieth through twenty first century
Calculations per Second per $1,000
10^{60}
10^{55}
10^{50}
10^{45}
10^{40}
10^{35}
10^{30}
10^{25}
10^{20}
10^{15}
10^{10}
10^{5}
1
10^{-5}
10^{-10}
All Human Brains
One Human Brain
One Mouse Brain
One Insect Brain
Logarithmic Plot



4 The INDUSTRIAL REVOLUTIONS

1765
THE FIRST INDUSTRIAL REVOLUTION

1870
THE SECOND INDUSTRIAL REVOLUTION

1969
THE THIRD INDUSTRIAL REVOLUTION

NOWADAYS
INDUSTRY 4.0

sentryo

Design by INVOX

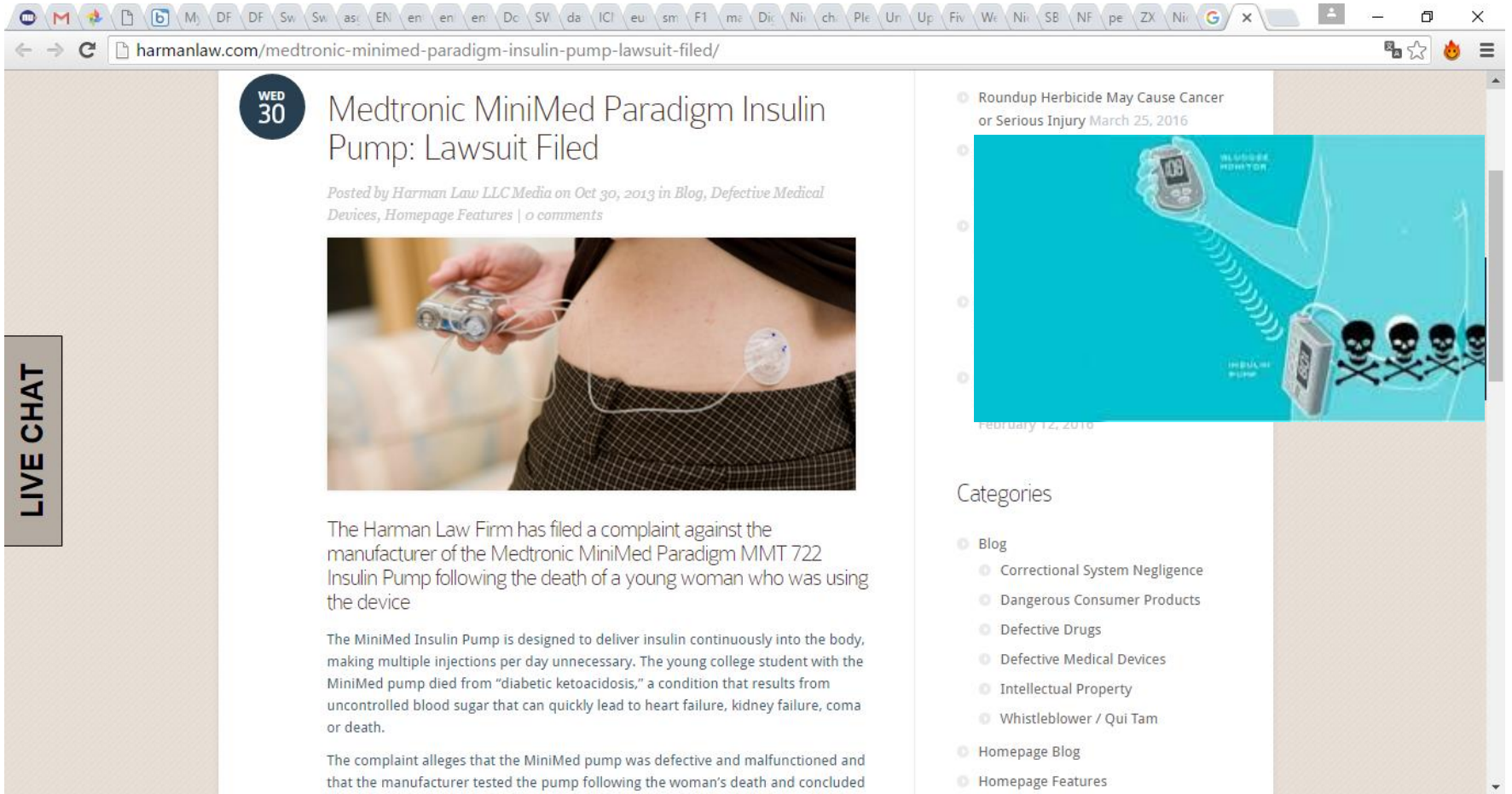
2021 This Is What Happens In An Internet Minute

Challenge:

- Who ? Attribution can be difficult in cyberspace
- What ? Is the data reliable, protocols
- When ? At what time
- Where ? Juridictions etc.
- Combining evidence

Internet of things

Insulin pumps webinterface for hospital remote control



Digital Evidence

Extract data

↓

Make data readable

↓

Organize data

↓

Interpret data

Challenge: many formats, old & new, non-standard

- Tool and library development
- Reverse engineering

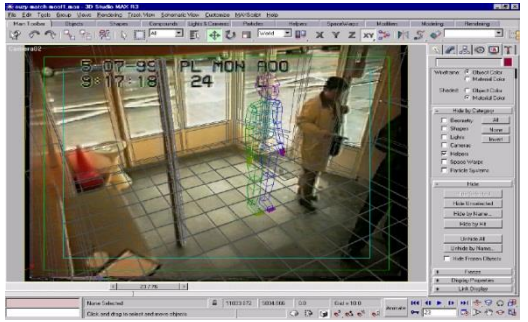
Discover the technological principles of a system (e.g. software or communication protocol) through analysis of its function and operation

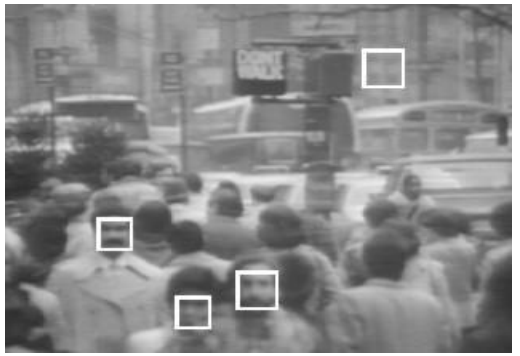
```
000025c0  0e 4a 5b fc 74 d6 21 a2  fb d3 5d bf 59 45 11 9a  |.J[.t.!...].YE..|
000025d0  fd d6 00 28 d6 6f a1 b0  60 59 6c ce c6 d0 4c 5c  |...(.o..`Yl...L\|
000025e0  61 10 8d cf 95 41 c1 3e  b3 f3 62 ff 1b b0 fc dc  |a....A.>..b.....|
000025f0  ea 5b fb 07 95 27 28 59  9a 05 e0 06 27 7b 2a 59  |.[...'(Y....'{*Y|
00002600  0e 43 72 1b ce 4b 1f 59  e2 ce d9 f3 86 34 5e f9  |.Cr..K.Y.....4^.|
00002610  38 d1 4a 0f 06 2e 70 66  c9 49 01 00 7b ca 93 c2  |8.J...pf.I..{...|
00002620  6d 70 02 ab b6 78 90 e1  5b ca 1c 14 29 13 77 93  |mp...x..[...).w.|
00002630  9f 29 a4 d1 1f 1f 3f 20  69 29 c4 ae fd c3 01 bf  |.)....? i)......|
00002640  76 c4 bd a8 cc 99 0b e3  93 74 82 b8 1e cc 2e da  |v........t......|
00002650  64 eb 74 64 5c 6c d7 91  78 5a 58 5b 59 c5 9a 82  |d.td\l..xZX[Y...|
00002660  4d e0 2c 58 1b 5c 83 c7  7e 98 3e 37 b2 93 99 90  |M.,X.\..~.>7....|
00002670  fd 00 e0 3a 8e 4f 13 e5  1f 23 bb b5 f8 b0 a3 85  |...:.O...#......|
00002680  86 74 b9 1b 18 b7 5f 03  4b a1 6a c5 7c c4 46 1e  |.t...._.K.j.|.F.|
00002690  6b 09 51 77 6b 3b 0d 9c  17 36 31 71 07 f4 9a bb  |k.Qwk;...61q....|
```

Challenge: data is not self-explaining

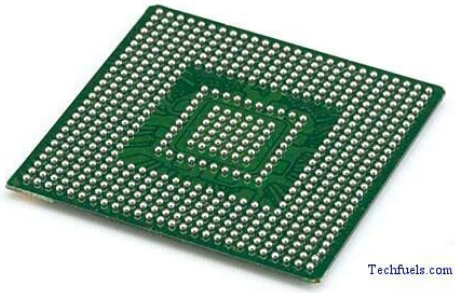
Add models and analysis to support interpretation

- Scenario analysis
- Timeline analysis
- Geographical models: e.g. location of cell phones
- Analysis of images / video / audio
  - Size
  - Speed
  - Face recognition
  - Speech recognition
- Author recognition





Chip forensics: Memory Toolkit II



- Universal forensic solution to read memory chips
- Copy of all data in a memory chip, including information from spare areas, bad blocks etc.
- Memory chips from password locked devices can also be read
- Even memory chips from non functional target devices (damaged by heat, water or force) can be read
- No data is changed

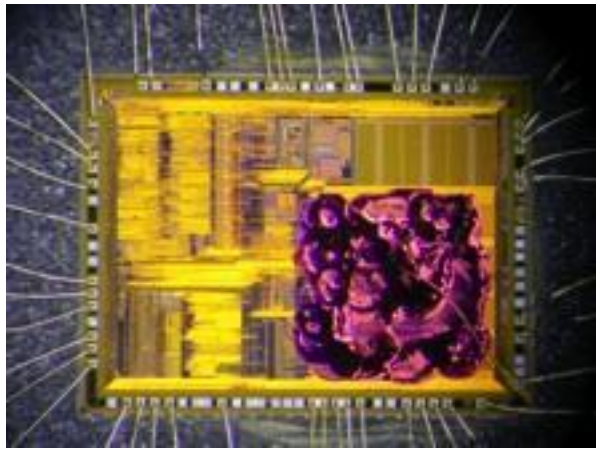
Silicon forensics

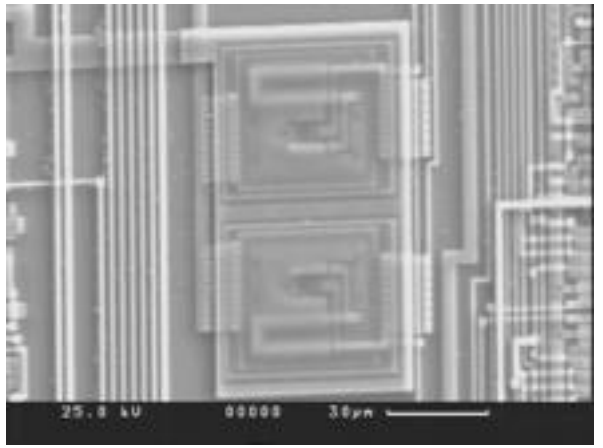
Dual Beam SEM

IMAM - Silicon hacking project

Invasive Memory Access Methods

- Goal: development of methods for getting access to the silicon of live chips and changing their behavior in order to readout their memory that is not accessible via the normal contacts.
- Format: a “toolbox” containing equipment and methods.
- Process: no standard process is offered, since each chip is different.





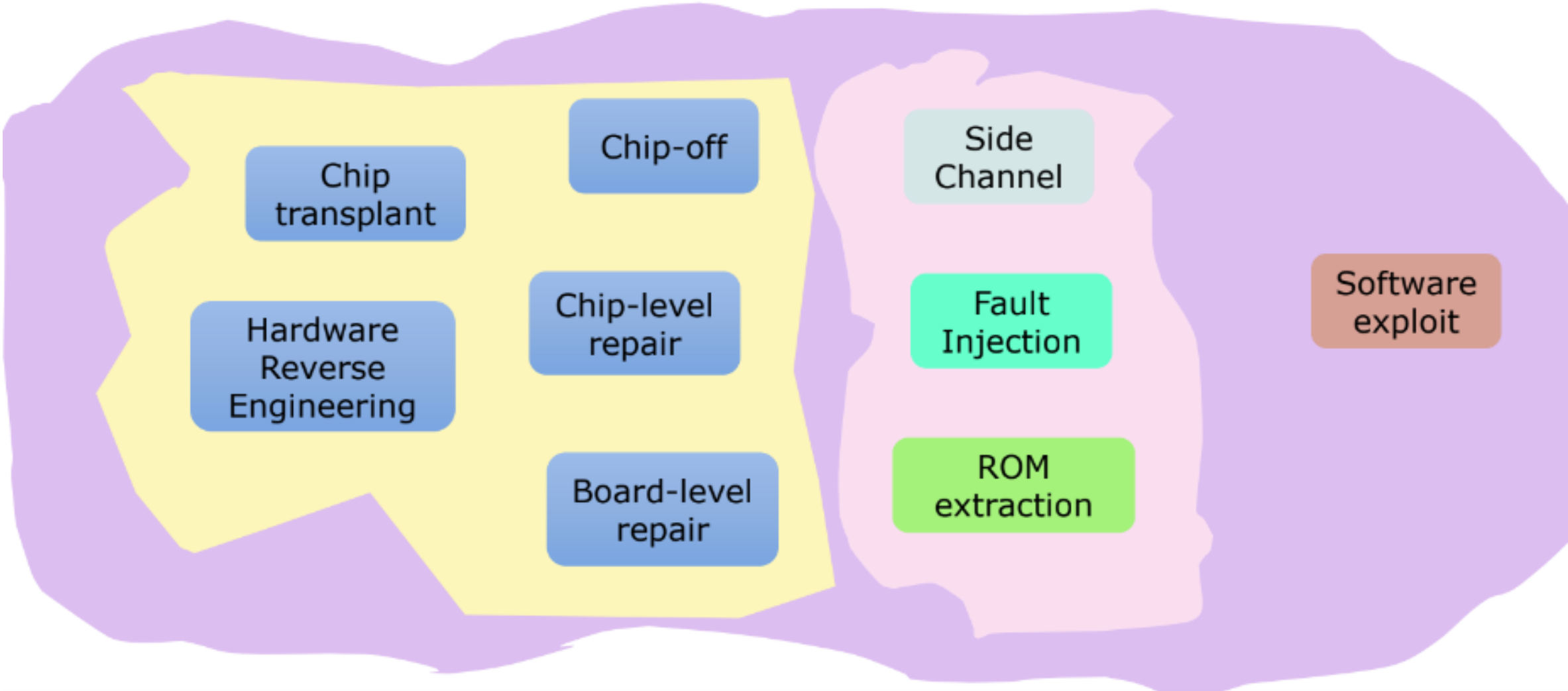
Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

40 kilometers queue of trucks filled with paper!!!

8 Terabyte?

1600 hours HD Video

Hardware-Based Digital Forensic Procedures



Quality Assurance of AI ?

- ISO 17025 / 17020 / 9002 for forensic labs
- CE ?
- Proficiency testing on different labs (how well do they perform with test cases)
- Regularity framework proposal AI from the EU Commission

Daubert and Artificial Intelligence

- This standard comes from the Supreme Court case, *Daubert v. Merrell Dow Pharmaceuticals Inc., 509 U.S. 579 (1993)*.
- Under the Daubert standard, the factors that may be considered in determining whether the methodology is valid are:
- (1) whether the theory or technique in question can be and has been tested;
- (2) whether it has been subjected to peer review and publication;
- (3) its known or potential error rate;
- (4)the existence and maintenance of standards controlling its operation; and
- (5) whether it has attracted widespread acceptance within a relevant scientific community.

AI and Daubert

- (1) whether the theory or technique in question can be and has been tested;
- (2) whether it has been subjected to peer review and publication;
- (3) its known or potential error rate;
- (4)the existence and maintenance of standards controlling its operation; and
- (5) whether it has attracted widespread acceptance within a relevant scientific community.

Much AI already used

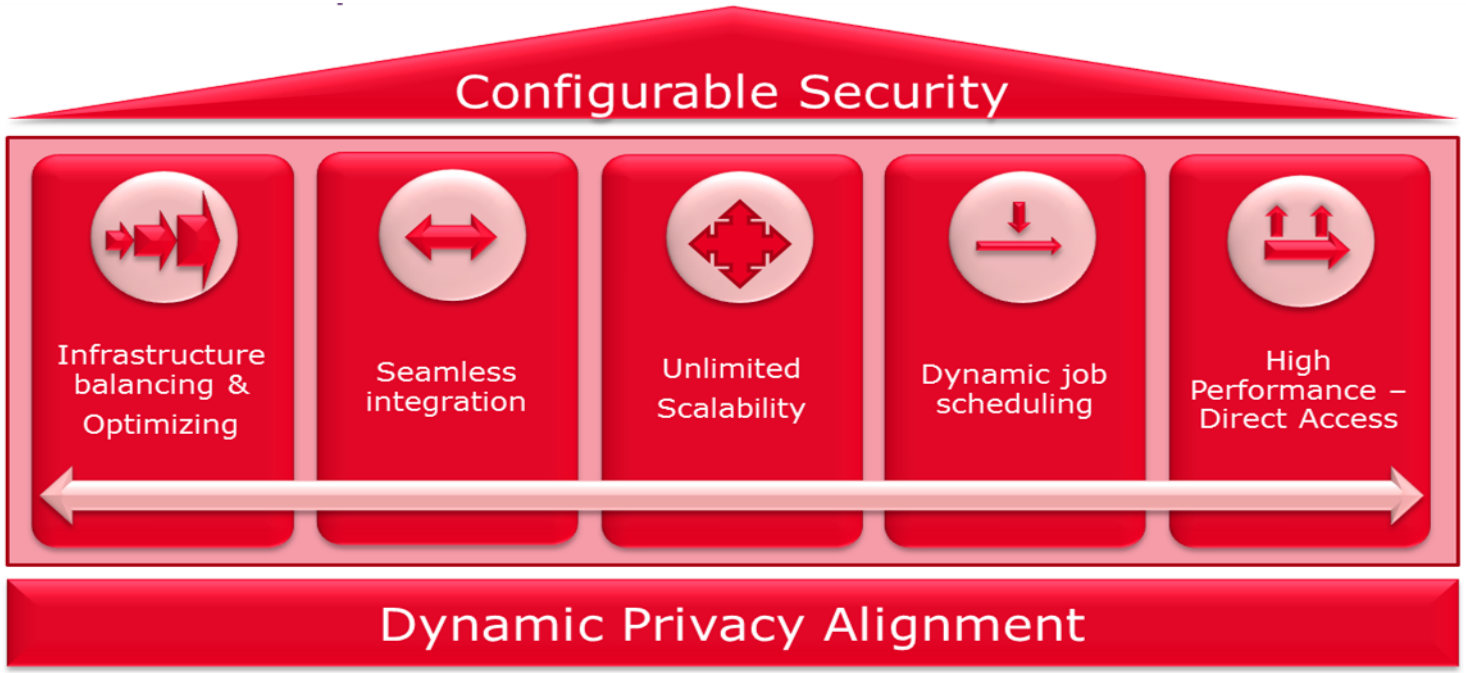
- Fingerprint system AFIS
- Speaker analysis
- Face comparison systems
- Digital Forensic Software
- Chemistry equipment deep learning spectral analysis
- Machine learning for DNA analysis
- …..

Future of digital investigation: HANSKEN

CONFISCATION | **SECURE** | **ENABLE ACCESS / ENRICH** | **REPORT** | **ANALYSE**

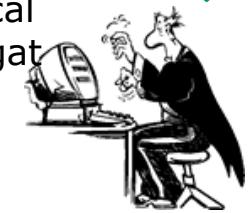
Datastream X

Datastream Y



analyst

technical investigator



tactical investigator

Some hours (1Tb/20 min) – direct results at start

Evolution forensic analysis – automation, speed & coverage

manual import and manual processing

Conventional: throughput months

50% 50%

manual import and automated processing

XIRAF: throughput weeks

70% 30%

automated import and automated massive-parallel processing

HANSKEN:throughput hours

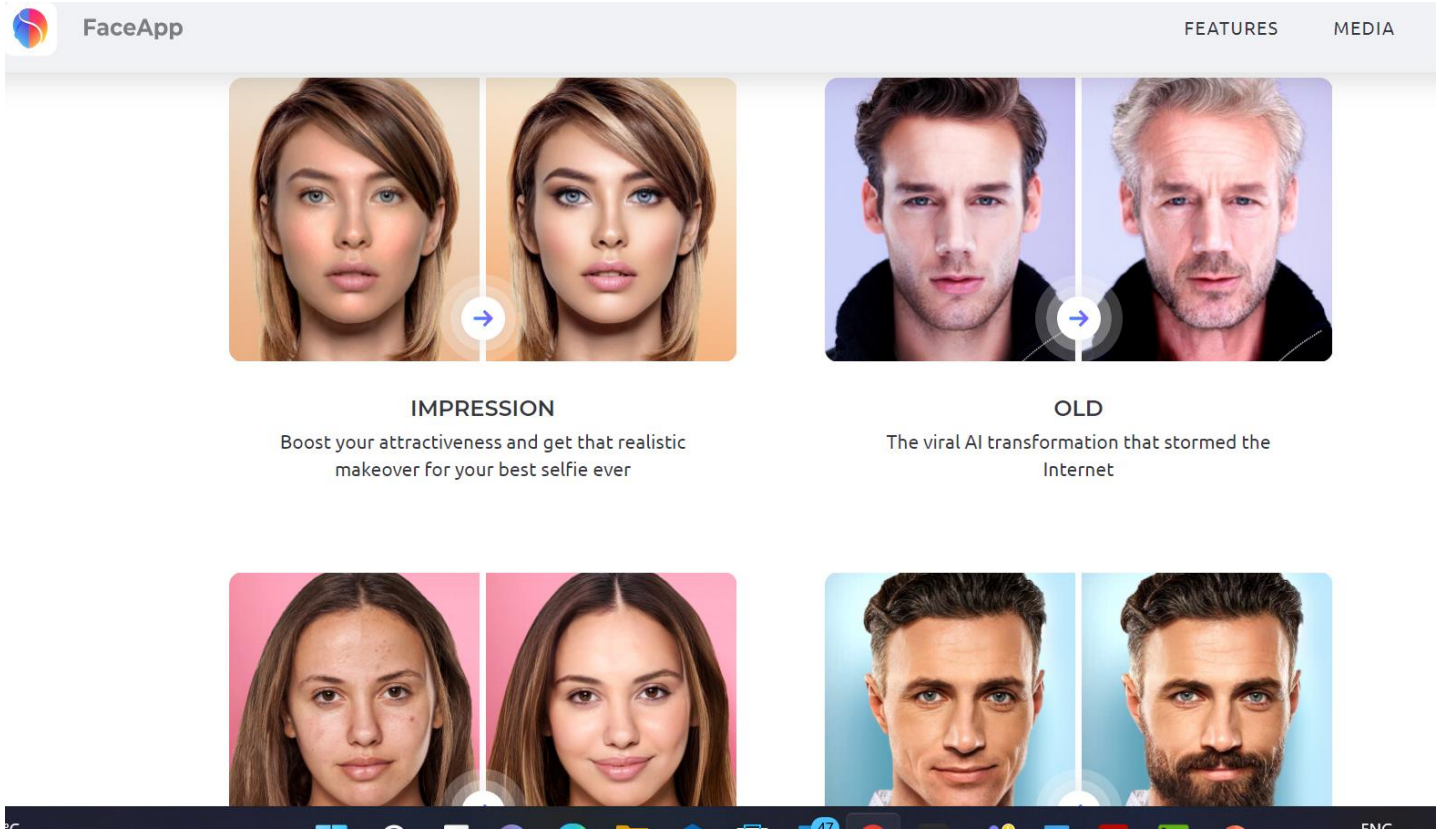
85% 15%

- AI is used in many products
- Black box, can we explain it in court ?
- Use it in interaction with the expert
- Faster forensic science
- Helping the police to make decisions
- Finding deviations in the forensic process / court decisions

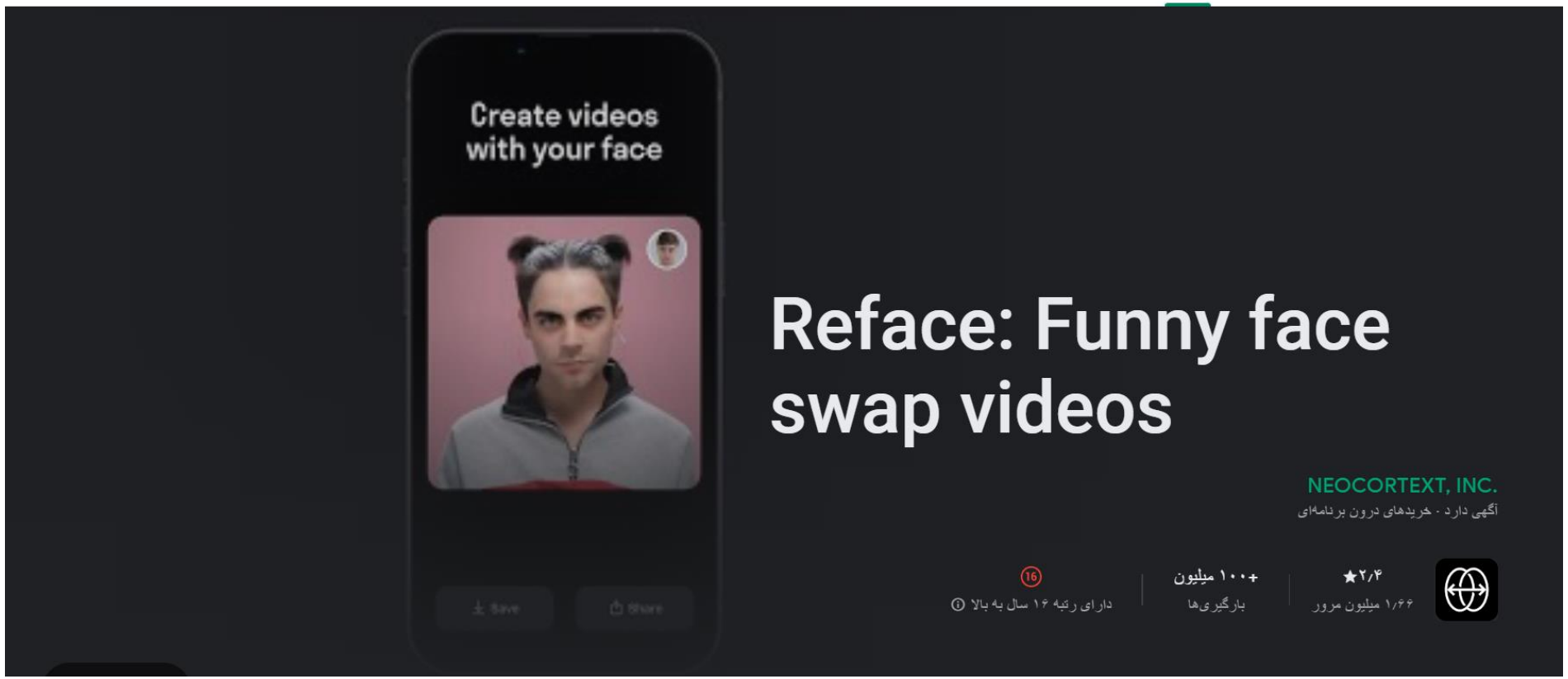
Court proof

- Can we explain it in court
- Does Explainable AI work for this ?
- Example searching in Encrochat or Sky Ecc data, lots of communication Natural Language Processing

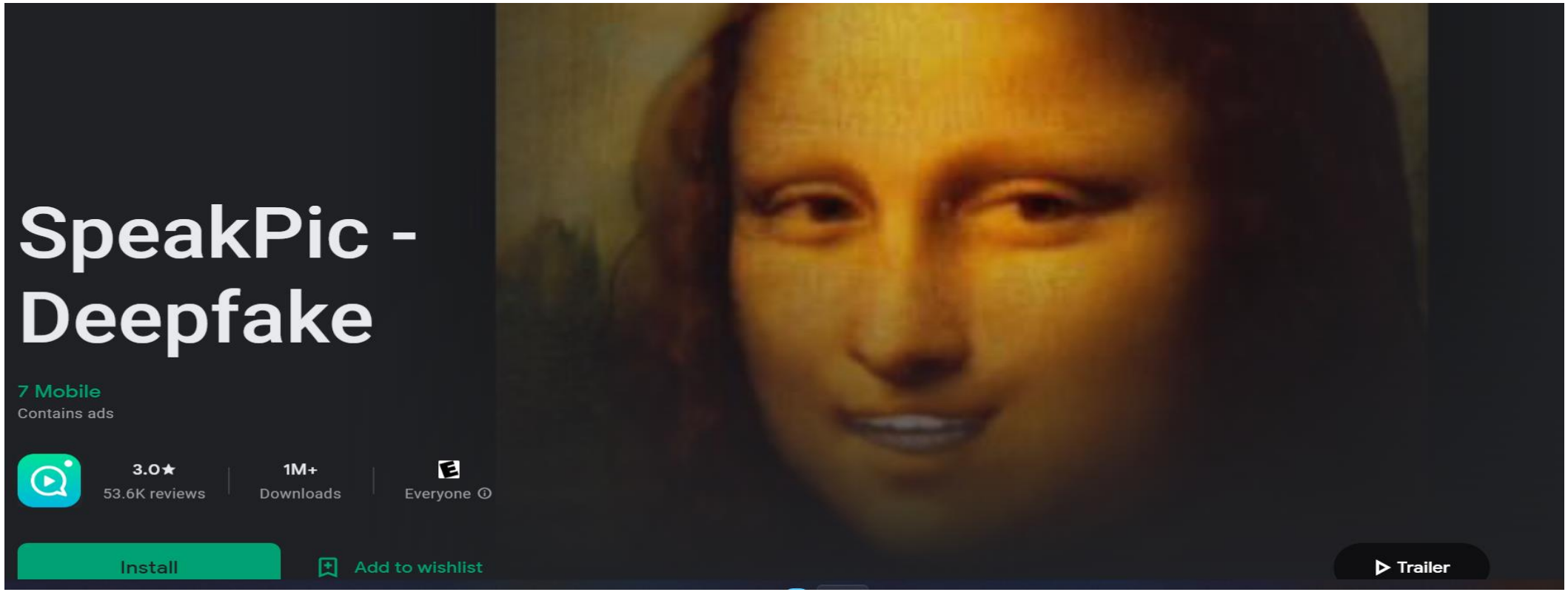
Faceapp


FaceApp
FEATURES
MEDIA
IMPRESSION
Boost your attractiveness and get that realistic makeover for your best selfie ever
OLD
The viral AI transformation that stormed the Internet

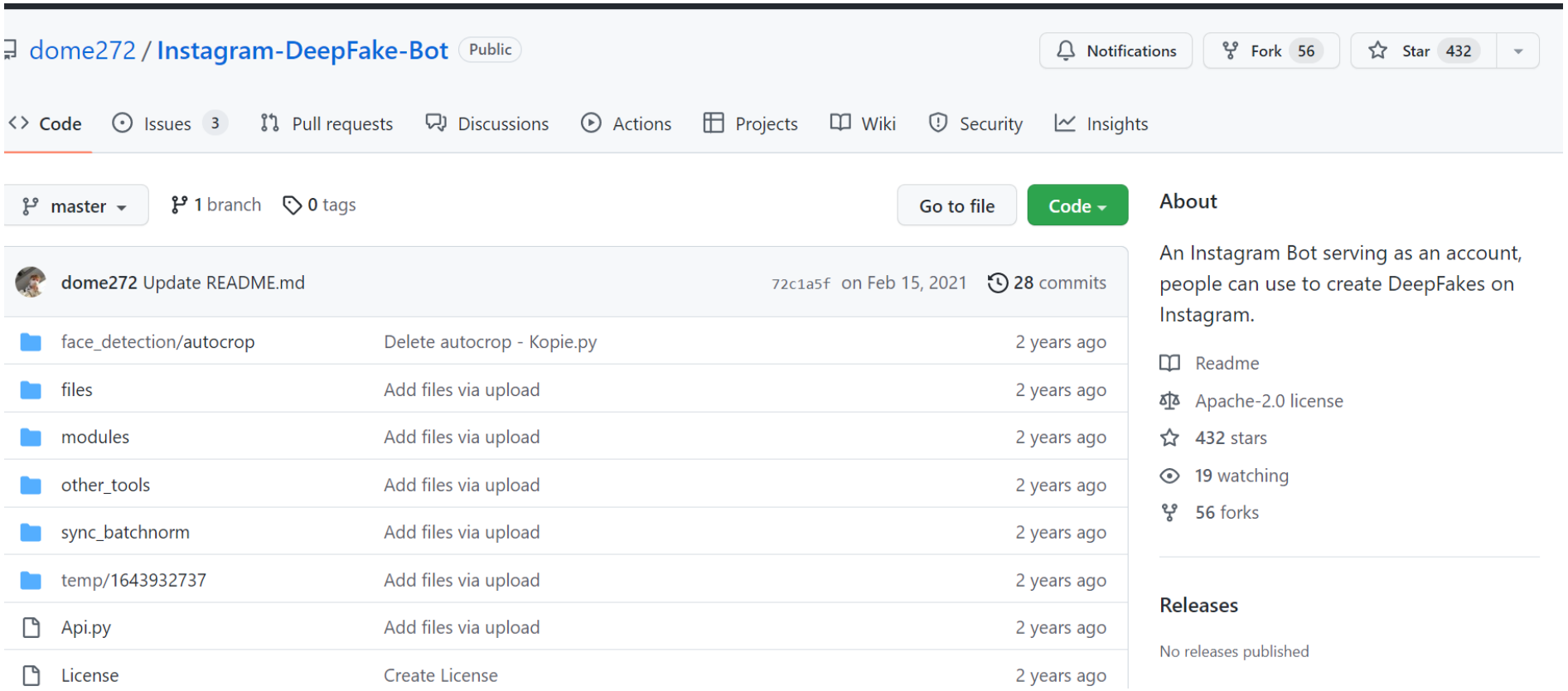
Reface


Create videos
with your face
Reface: Funny face
swap videos
NEOCORTEXT, INC.
آگهی دارد · خریدهای درون برنامه‌ای
۱۰۰+ میلیون
بارگیری‌ها
★۲٫۴
۱٫۶۶ میلیون مرور
16
دارای رتبه ۱۶ سال به بالا

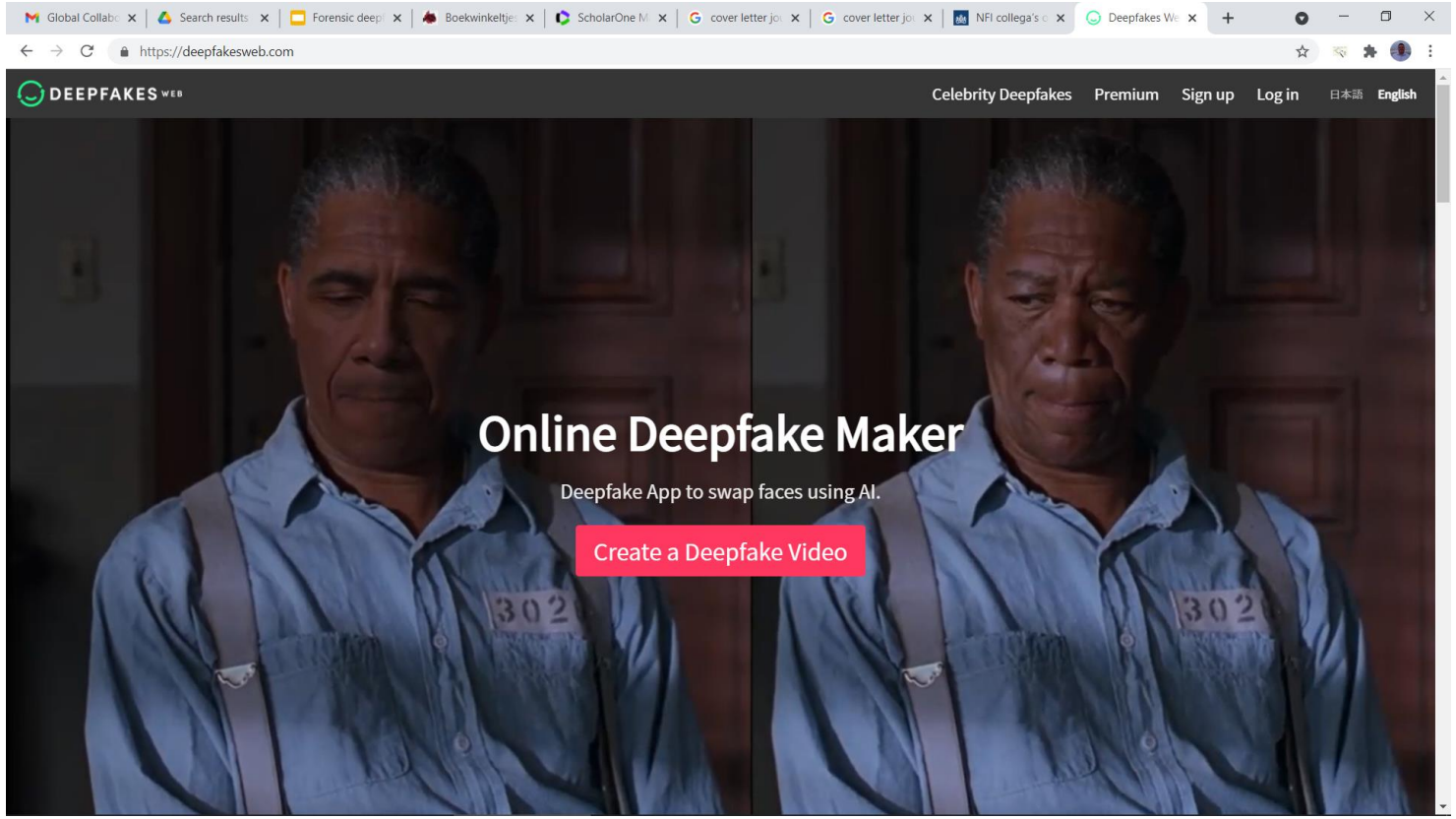
speakpic


SpeakPic - Deepfake
7 Mobile
Contains ads
3.0★
53.6K reviews
1M+
Downloads
Everyone
Install
Add to wishlist
Trailer

Deepfake Bot

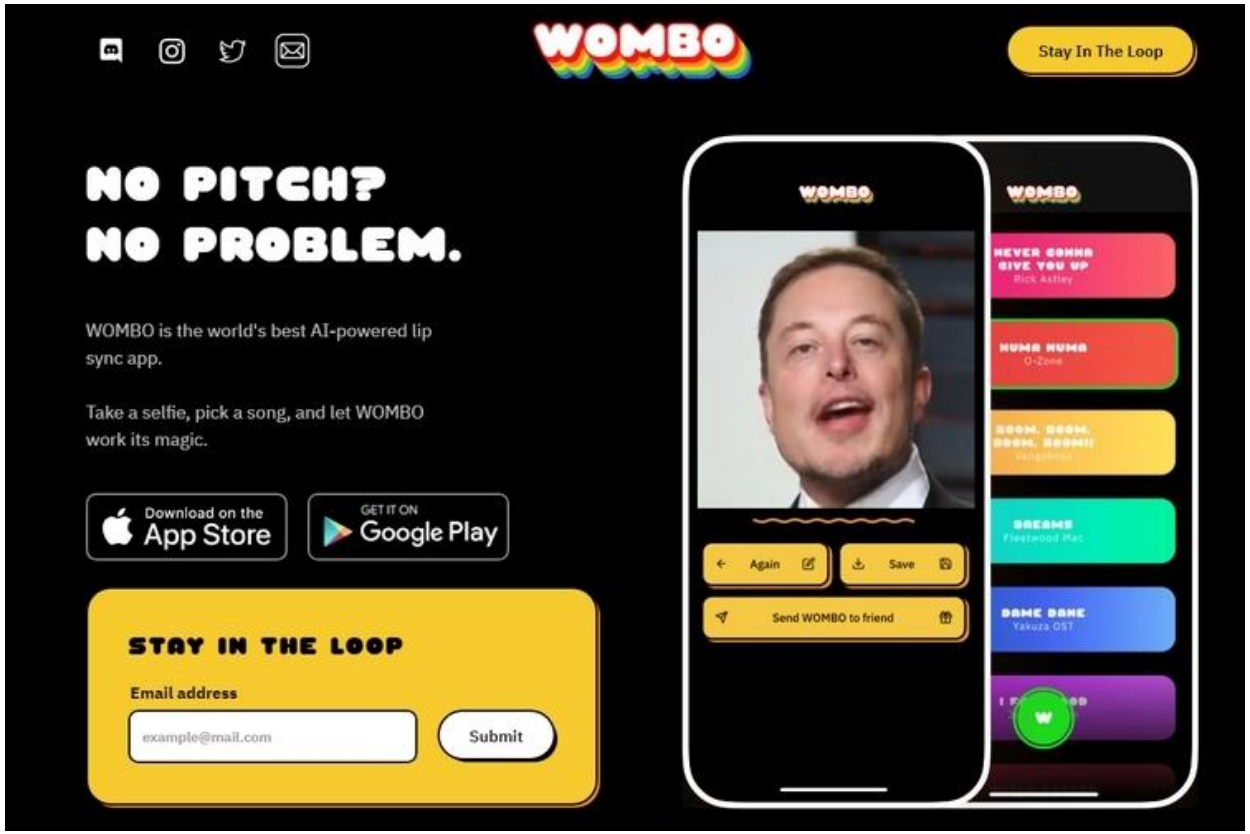


On line deepfake maker



Faceswap booth

Lip syncing app



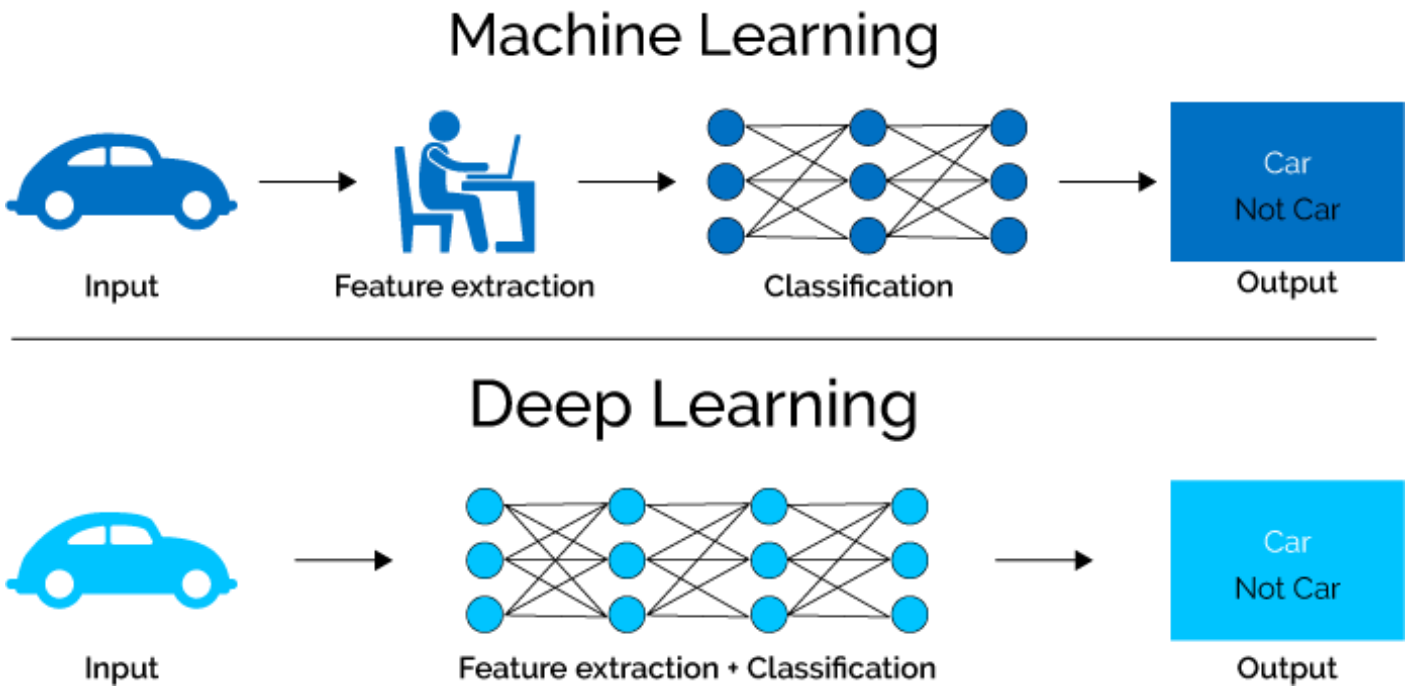
Everybody can deepfake Deepfacelab

Make persons that do not exist thispersondoesnotexist.com

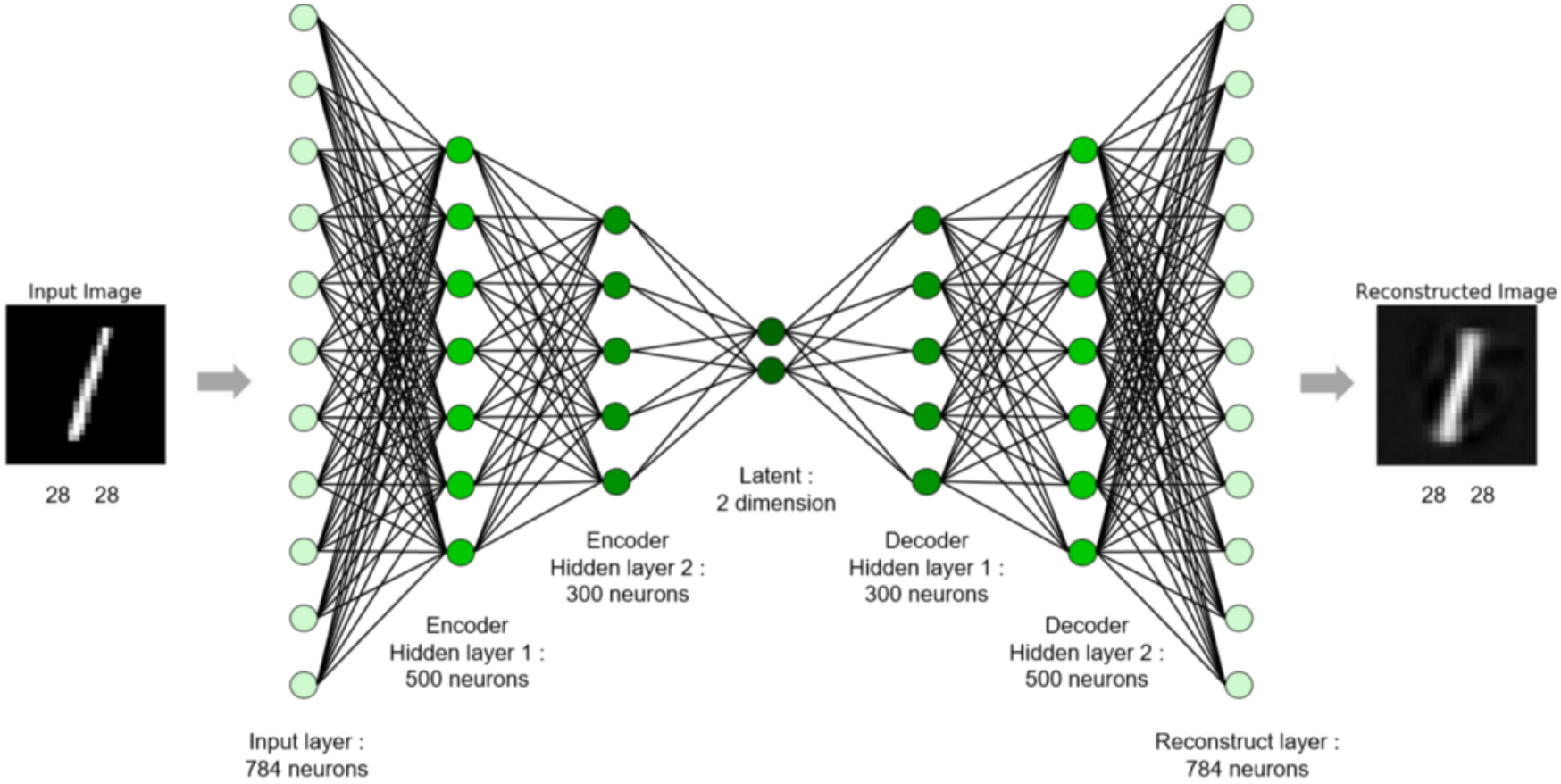
Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Basic principles

Machine learning vs deep learning

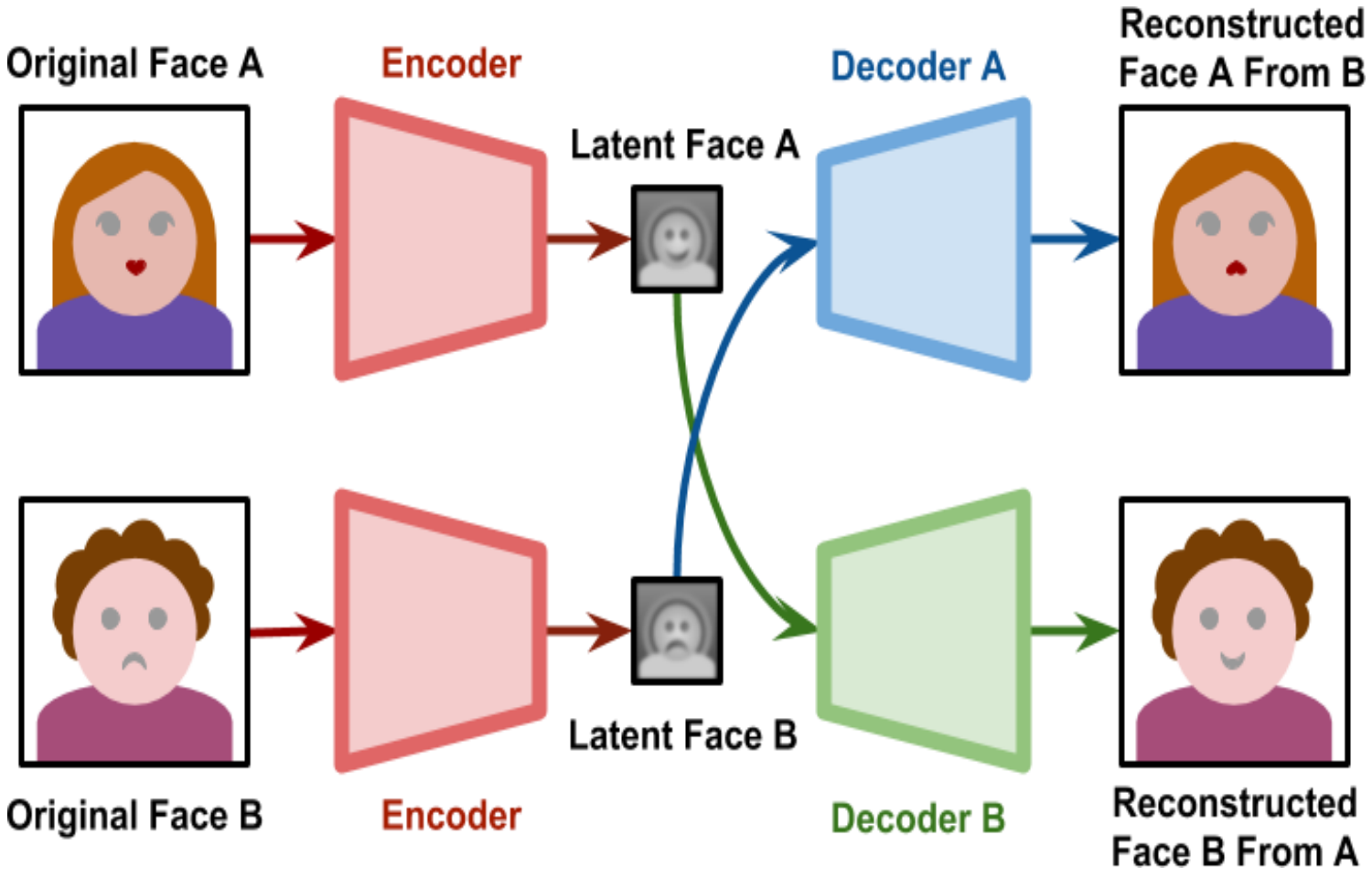


Recap: Autoencoder



Basic approach

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid



Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

It was real

https://www.youtube.com/watch?v=dnZNrp0-mF8

It was a deepfake

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

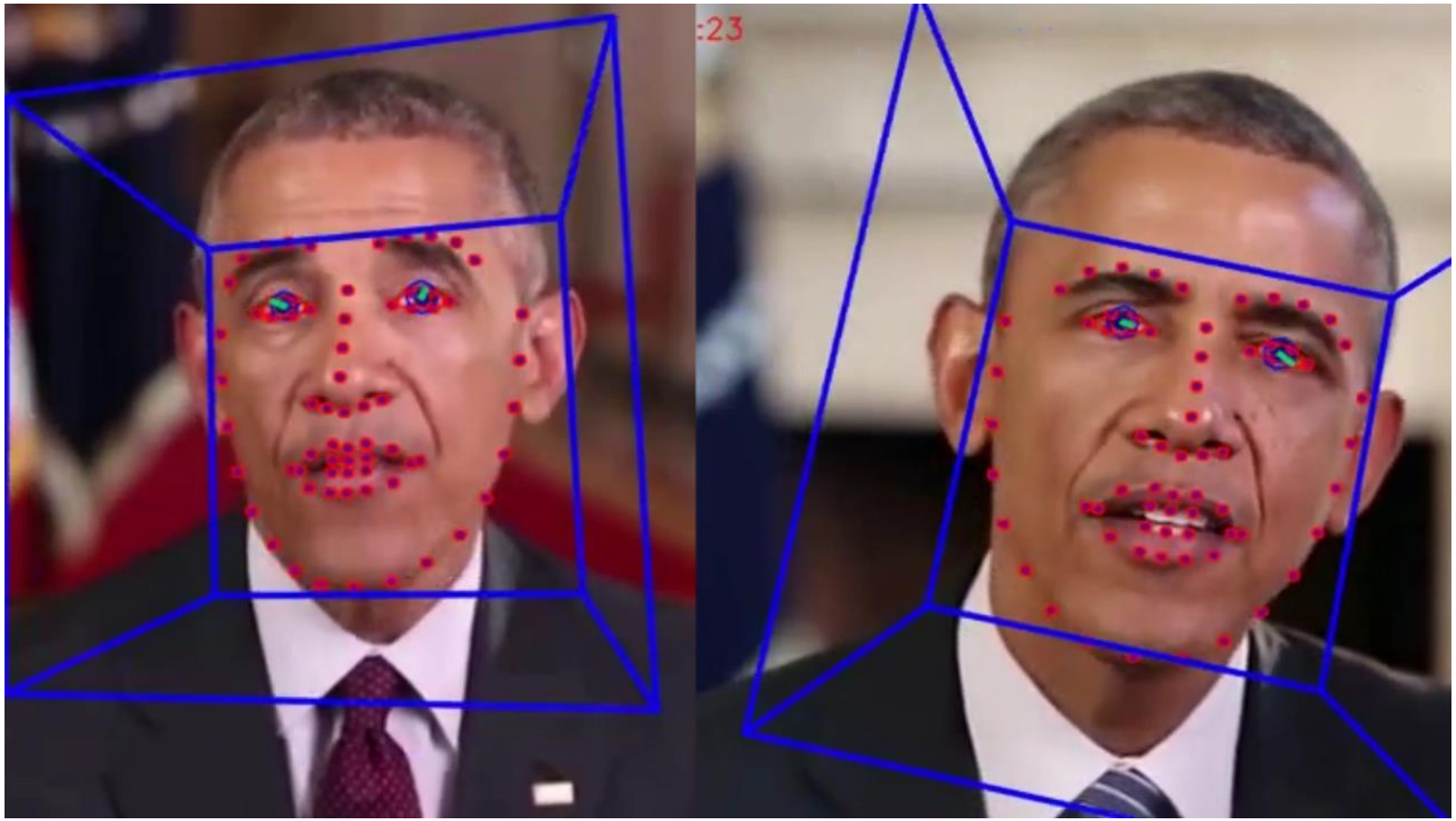
Some issues

- the face is vague, especially at the edges
- look to teeth, mouth and hair they are not detailed
- audio is not synchronized with video
- the body language is different from the way one should expect to speak
- the light source is different

Manipulations of faces

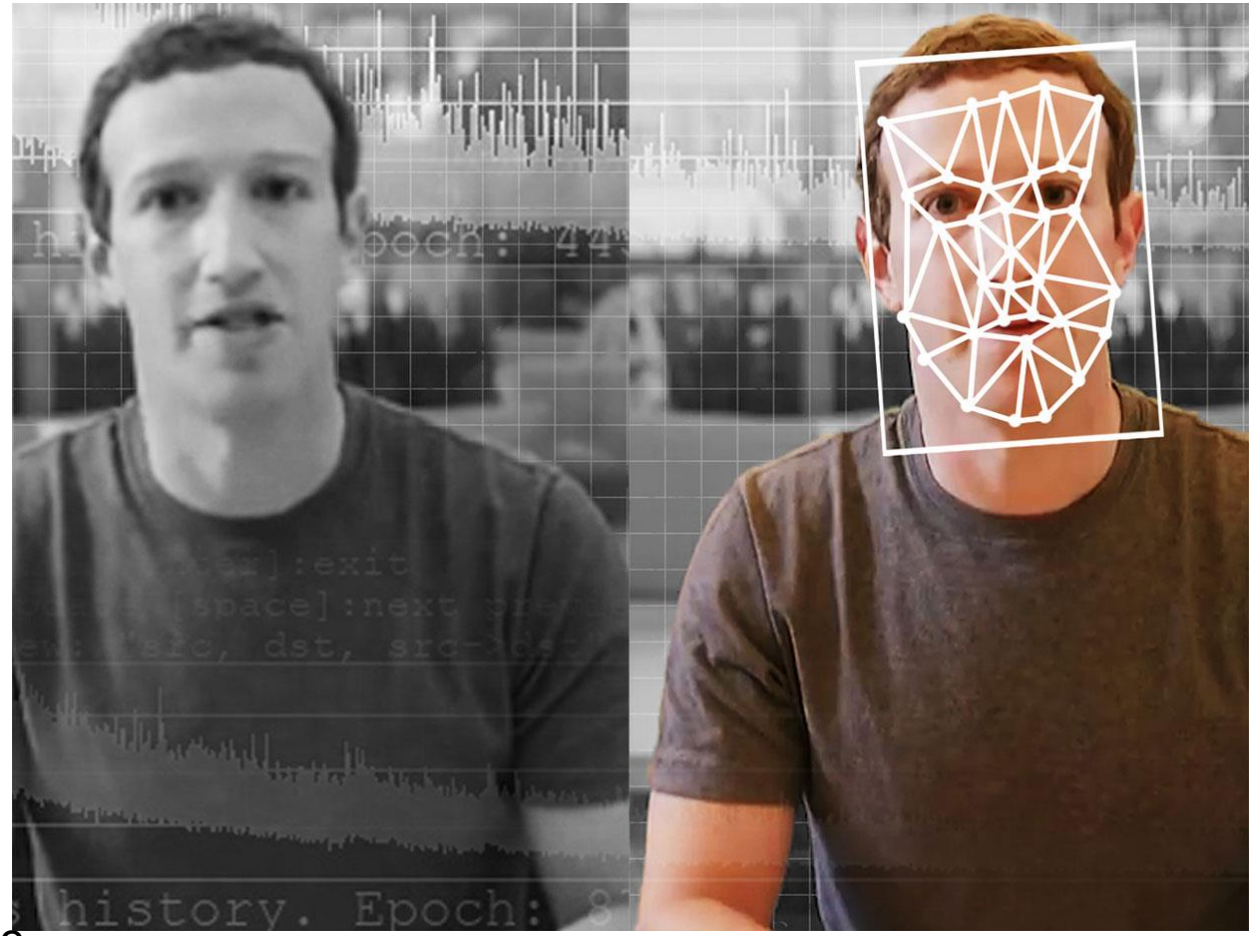
- face swap
- face attributes
- face expressions

lip sync deepfake https://news.berkeley.edu/2019/06/18/researchers-use-facial-quirks-to-unmask-deepfakes/

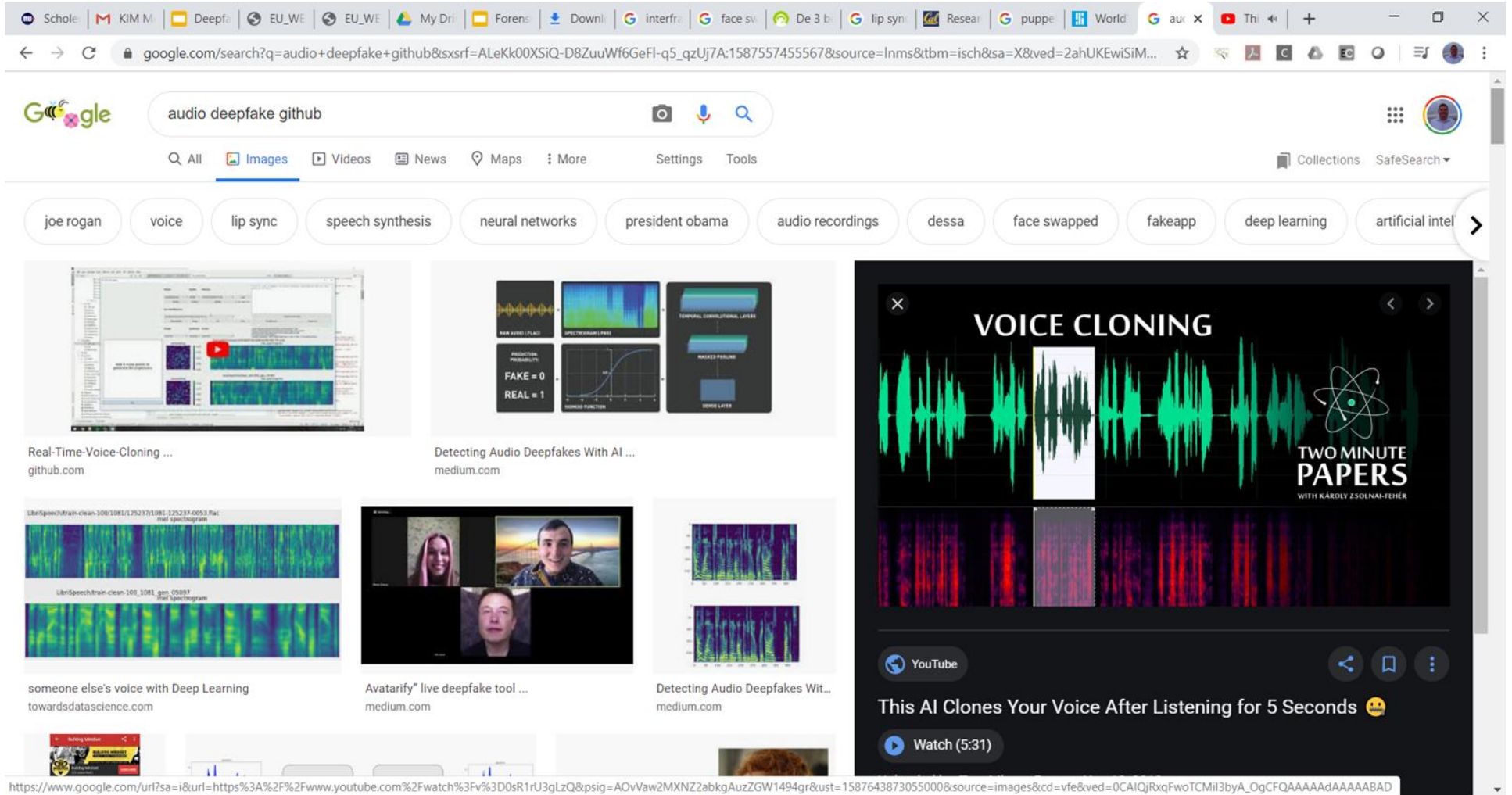


Puppet master

https://spectrum.ieee.org/tech-talk/computing/software/the-worlds-first-audit-of-deepfake-videos-and-tools-on-the-open-web

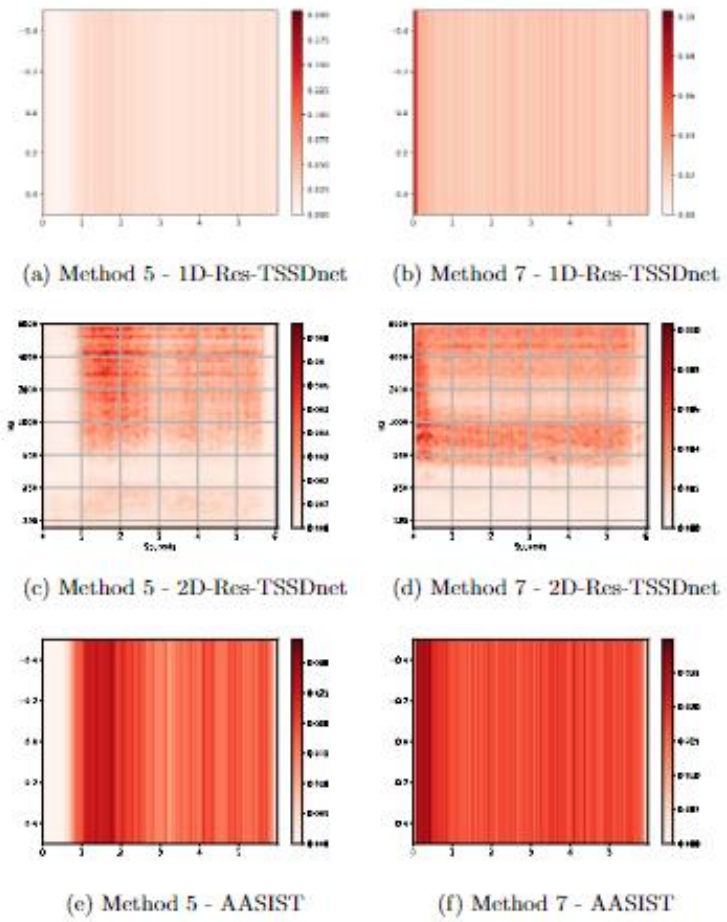


Audio deepfake look at github



Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Explainable AI for voice Thesis Jeroen Bergers EAFS poster award



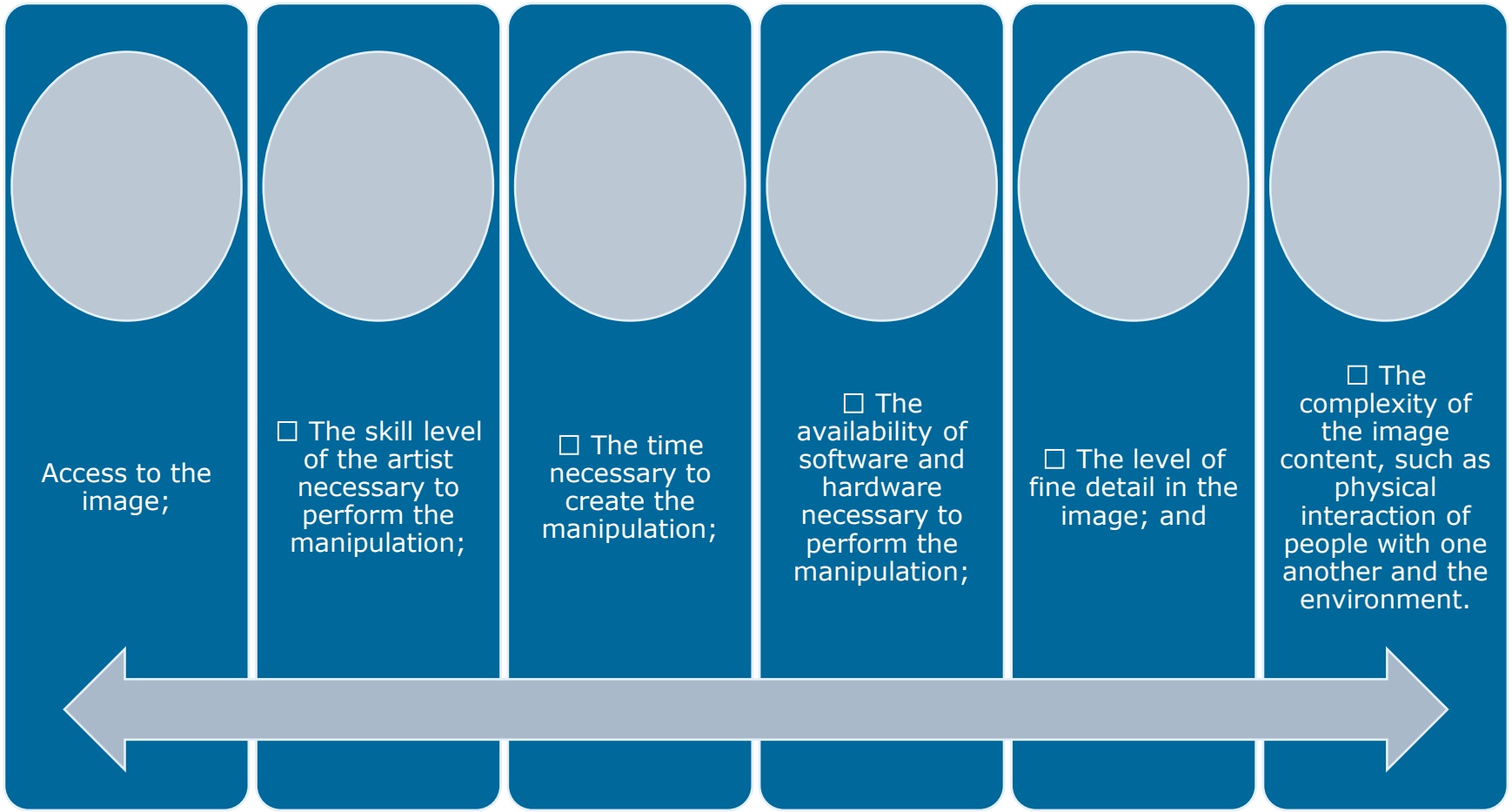
(a) Method 5 - 1D-Res-TSSDnet (b) Method 7 - 1D-Res-TSSDnet

(c) Method 5 - 2D-Res-TSSDnet (d) Method 7 - 2D-Res-TSSDnet

(e) Method 5 - AASIST (f) Method 7 - AASIST

Figure 12: The sum of SHAP explanations from two different synthetic voice methods and three different detection models

Practical issues



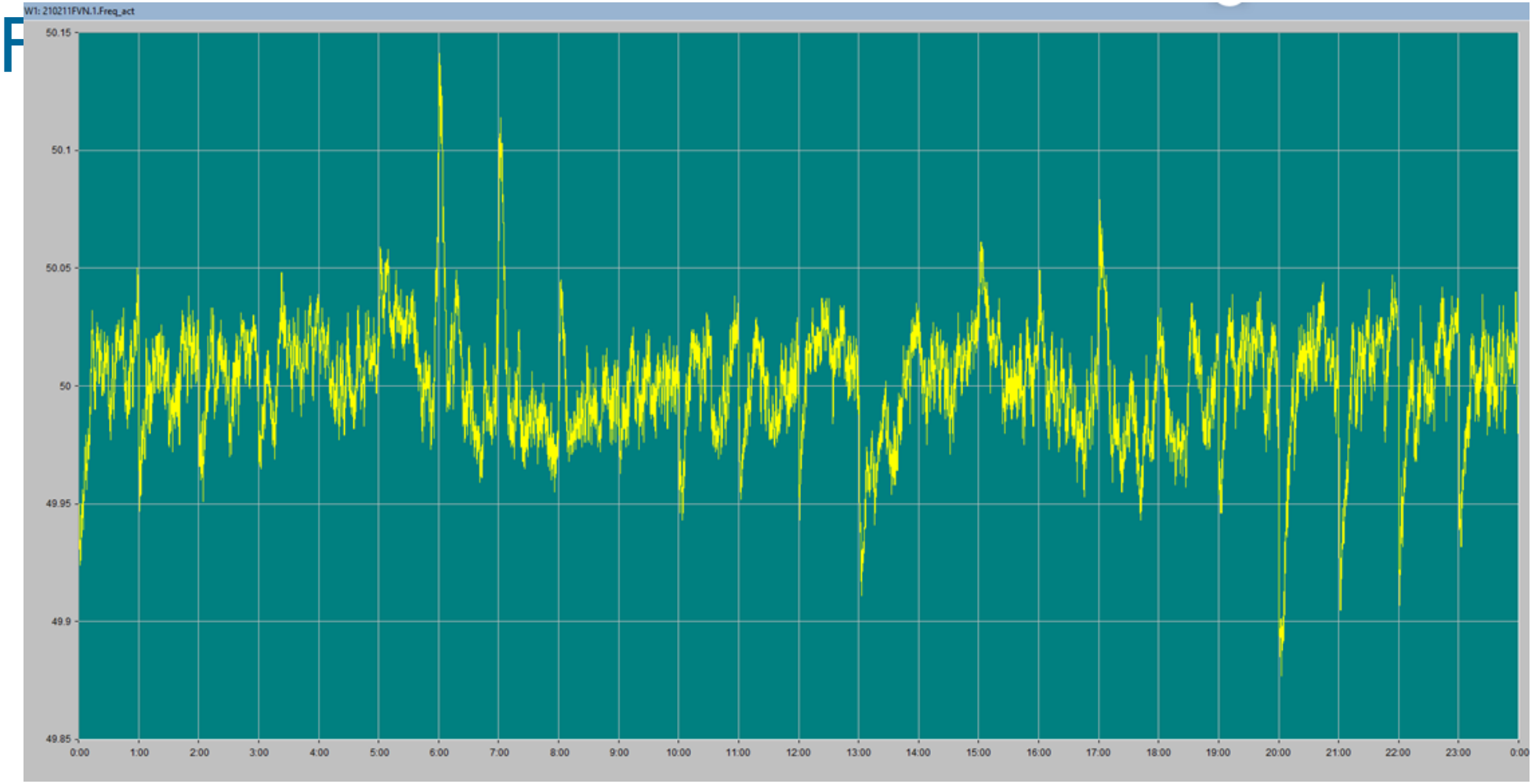
Best practices

- Skin tones & textures
- Skeletal structure
- Flesh & muscle movement
- Body-to-object contact
- Skin-to-skin contact
- Skin creases
- Hair
- Ears
- Eyes
- Reaction of subjects/objects to gravity and physics.
- Continuity Issues
-

Video detection of forgeries

- Same as images, also to metadata
- Look to changes of scenes, detect differences
- Time of recording ENF

ENF



https://www.tennet

Relevance

Verification

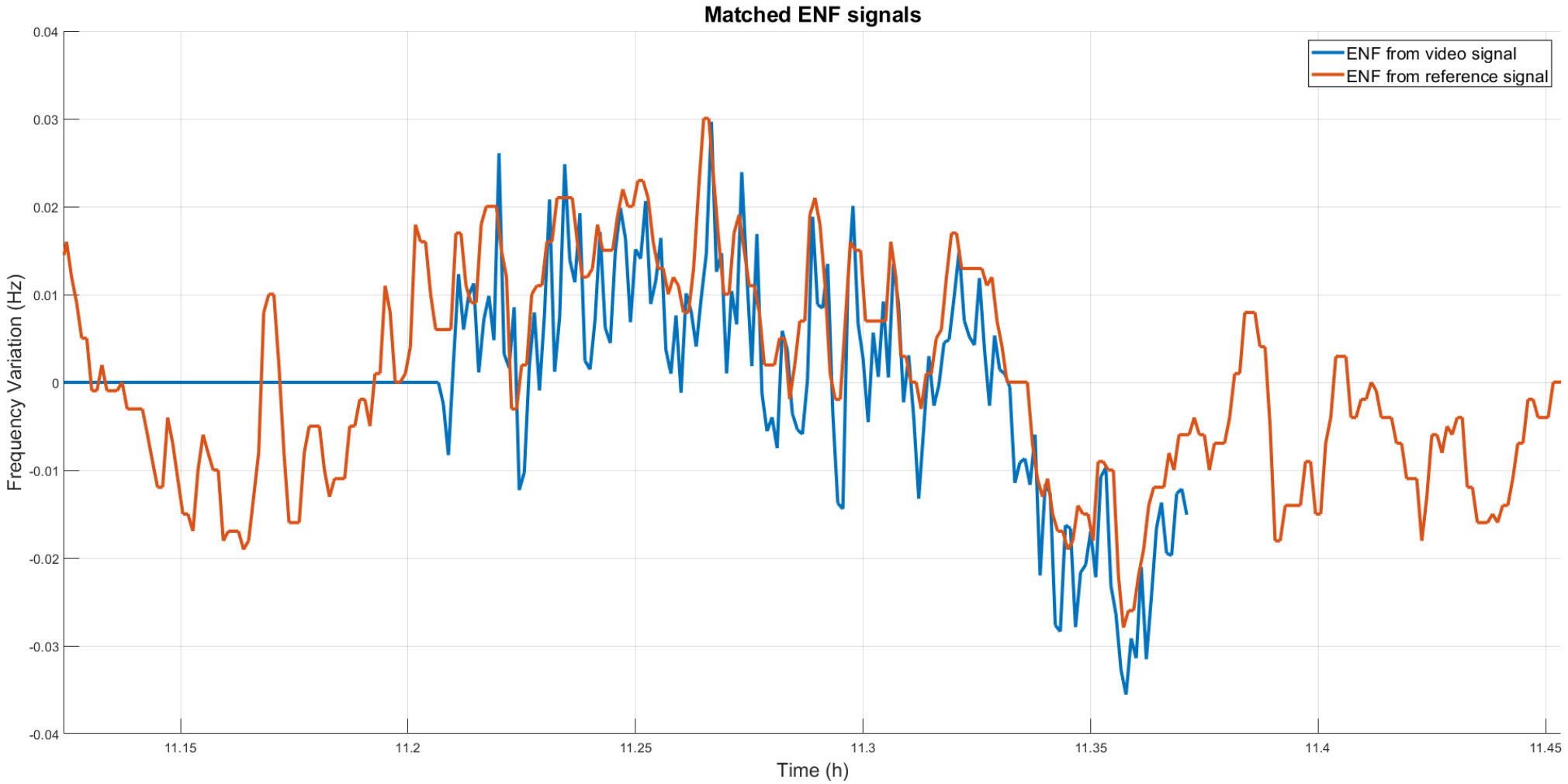
Time estimation

Localization

Focus

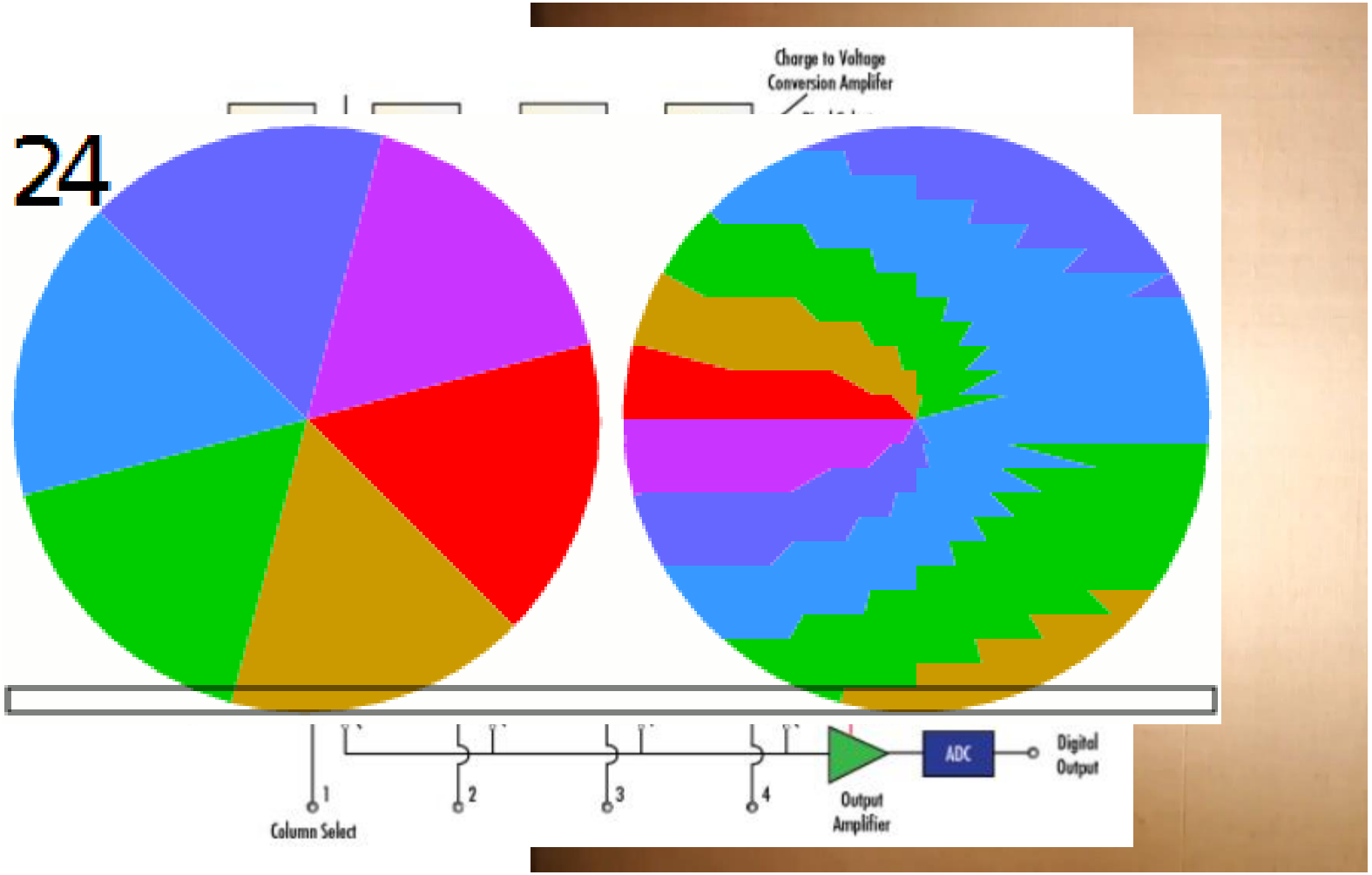
Can the ENF signal be extracted from videos made with smartphones to be used in a forensic context for time estimation?

Co


Matched ENF signals
ENF from video signal
ENF from reference signal
Frequency Variation (Hz)
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
11.15
11.2
11.25
11.3
11.35
11.4
11.45
Time (h)

Videos

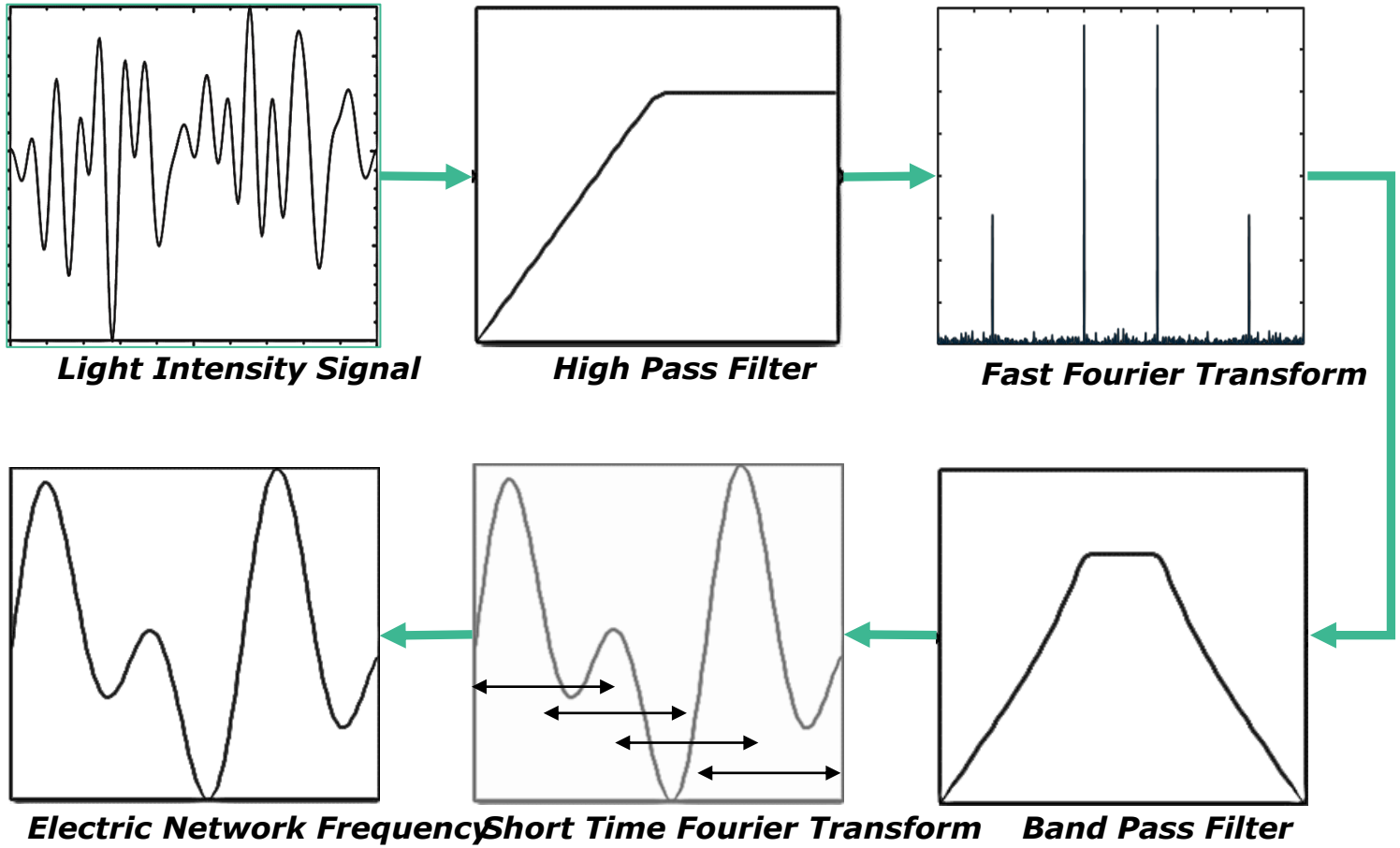
- Smartphones
- CMOS sensor
- Rolling shutter



https://upload.wikimedia.org/wikipedia/commons/1/15/Rolling_shutter_effect_animation.gif

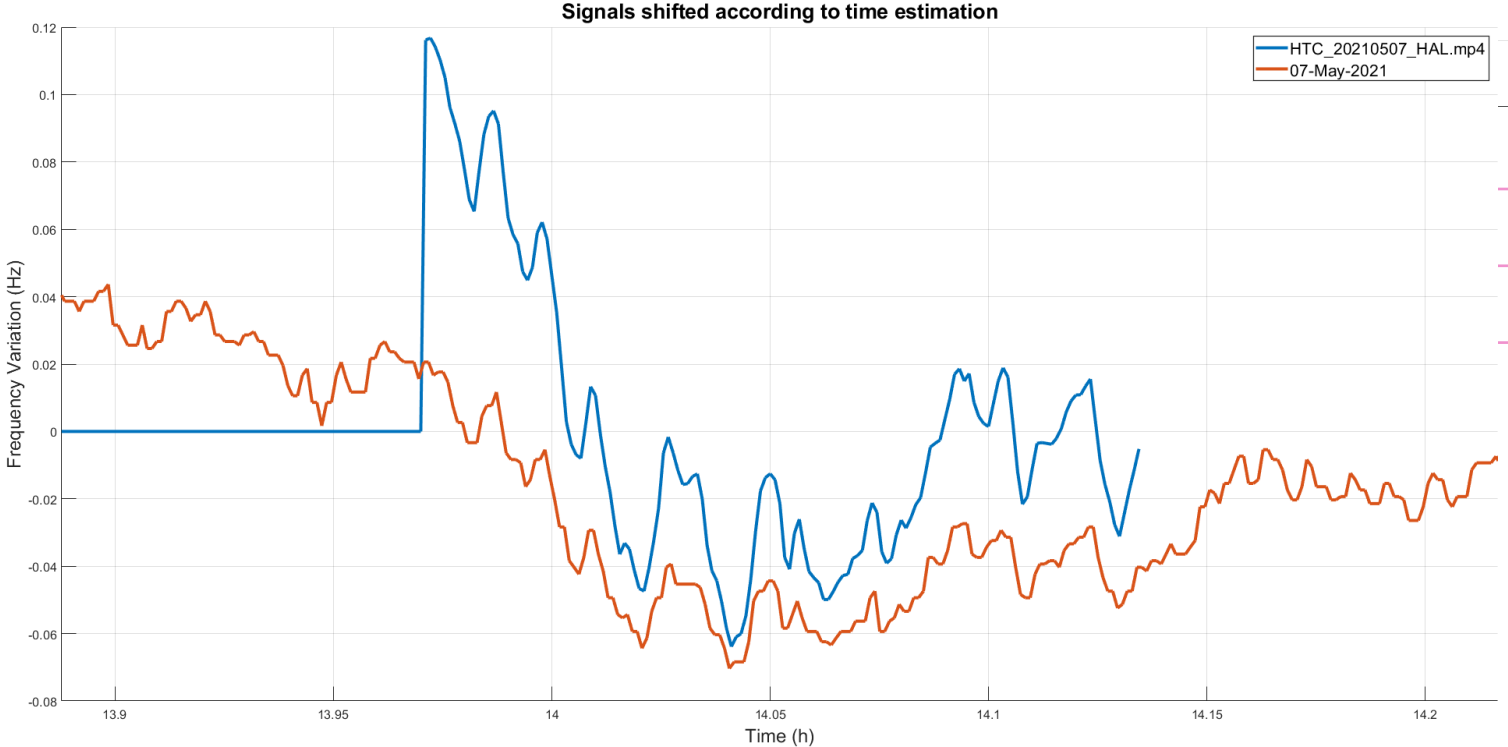
https://www.edmundoptics.com/knowledge-center/application-notes/imaging/understanding-camera-sensors-for-machine-vision-applications/

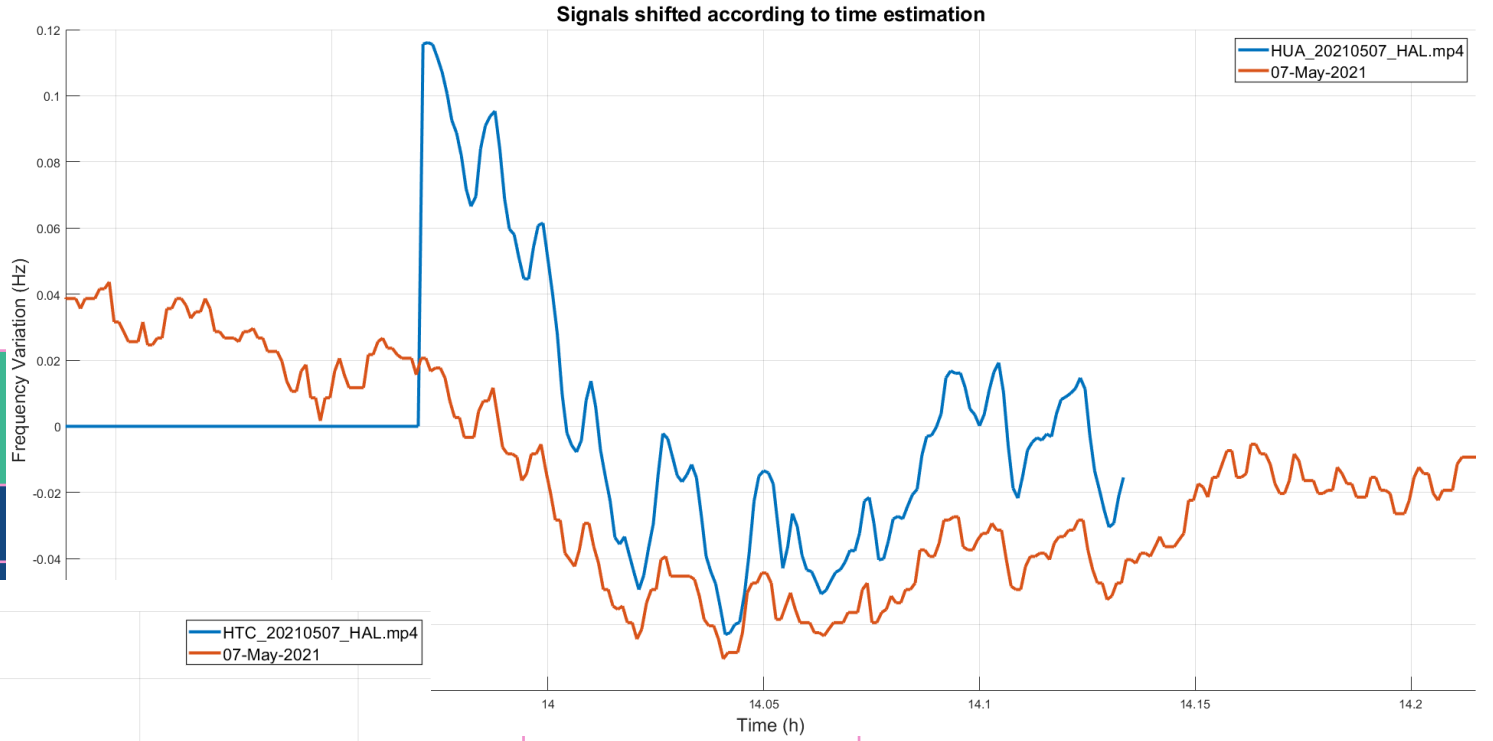
Analysis


Light Intensity Signal
High Pass Filter
Fast Fourier Transform
Band Pass Filter
Short Time Fourier Transform
Electric Network Frequency

Different Phones

Phone	
iPhone 7	12s
	12s





Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Received: 13 August 2021 | Revised: 5 January 2022 | Accepted: 18 January 2022

DOI: 10.1111/1556-4029.15003

JOURNAL OF FORENSIC SCIENCES AAFS

PAPER

Digital & Multimedia Sciences

Use of electric network frequency presence in video material for time estimation

Guus Frijters MSc[1,2] | Zeno J. M. H. Geradts PhD[1,2]

[1]Netherlands Forensic Institute, The Hague, The Netherlands

[2]University of Amsterdam, Amsterdam, The Netherlands

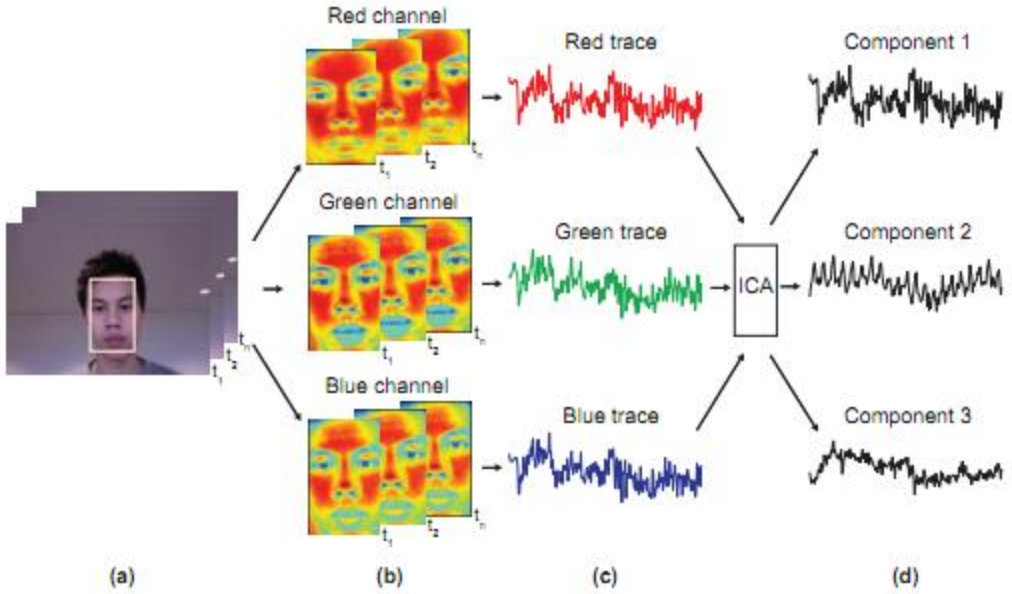
Correspondence
Zeno J. M. H. Geradts, Netherlands Forensic Institute, The Hague, Netherlands.
Email: z.geradts@nfi.nl

Abstract

In this research, the possibility of estimating the time a video was recorded at through electric network frequency is explored by examining various light sources in differentiating circumstances. This research focuses on videos made with smartphones. The smartphone cameras make use of an integrated complementary metal oxide semiconductor sensor. The filmed videos are analyzed using software, which employs a small electric network frequency (ENF) database to determine the time of recording of a video made in experi-

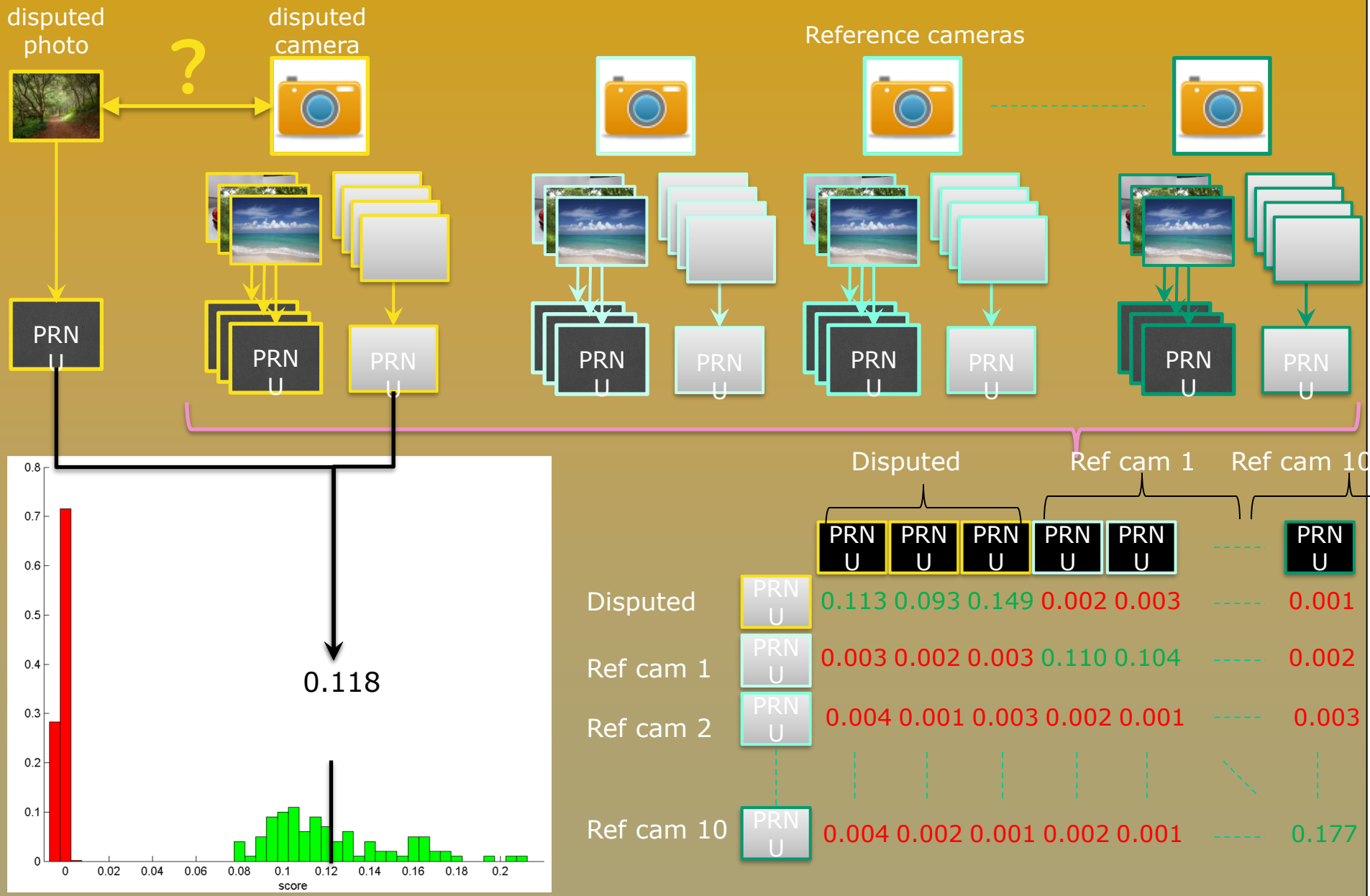
Biological signals heartbeat

Heart Beat



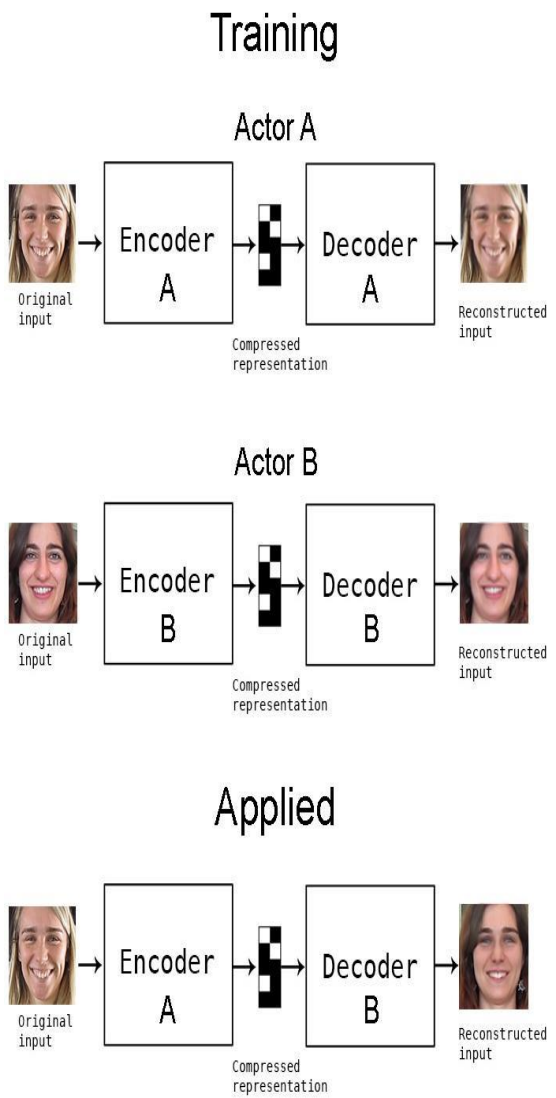
"Cardiocam: Technology for Non-Contact Multi-Parameter Physiologic Measurements". by Ming-Zher Optics Express, Vol. 18, Issue 10, pp. 10762-10774 (2010) doi:10.1364/OE.18.010762

PRNU - Camera identification research



How are Deepfakes made?

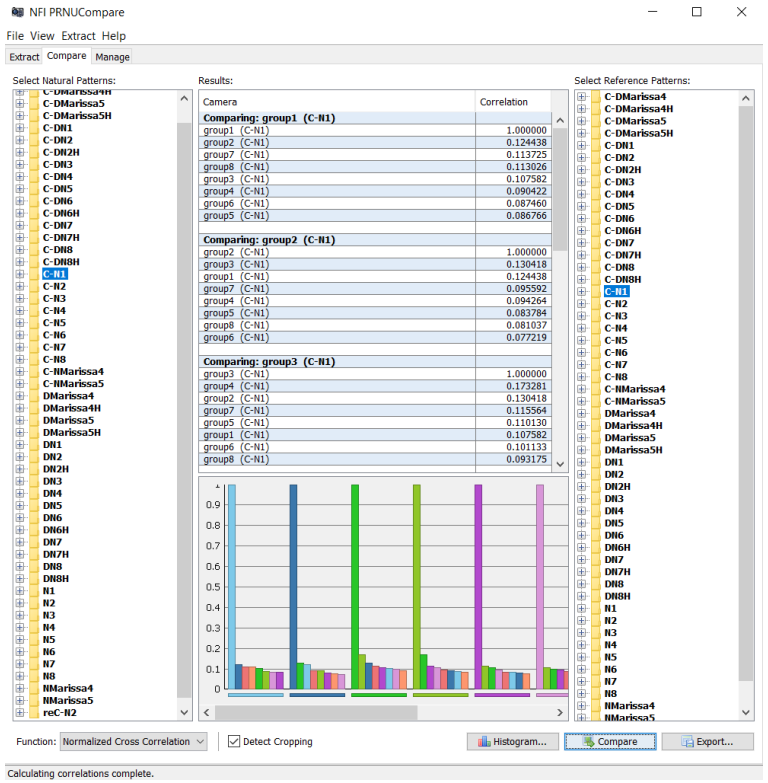
- Thousands of images of actor A and of actor B are needed for good results.
- Two autoencoders (A and B) are trained on these images.
- Autoencoder AB puts face of A on body of B



Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Photo Response Non Uniformity (PRNU) analysis: Method

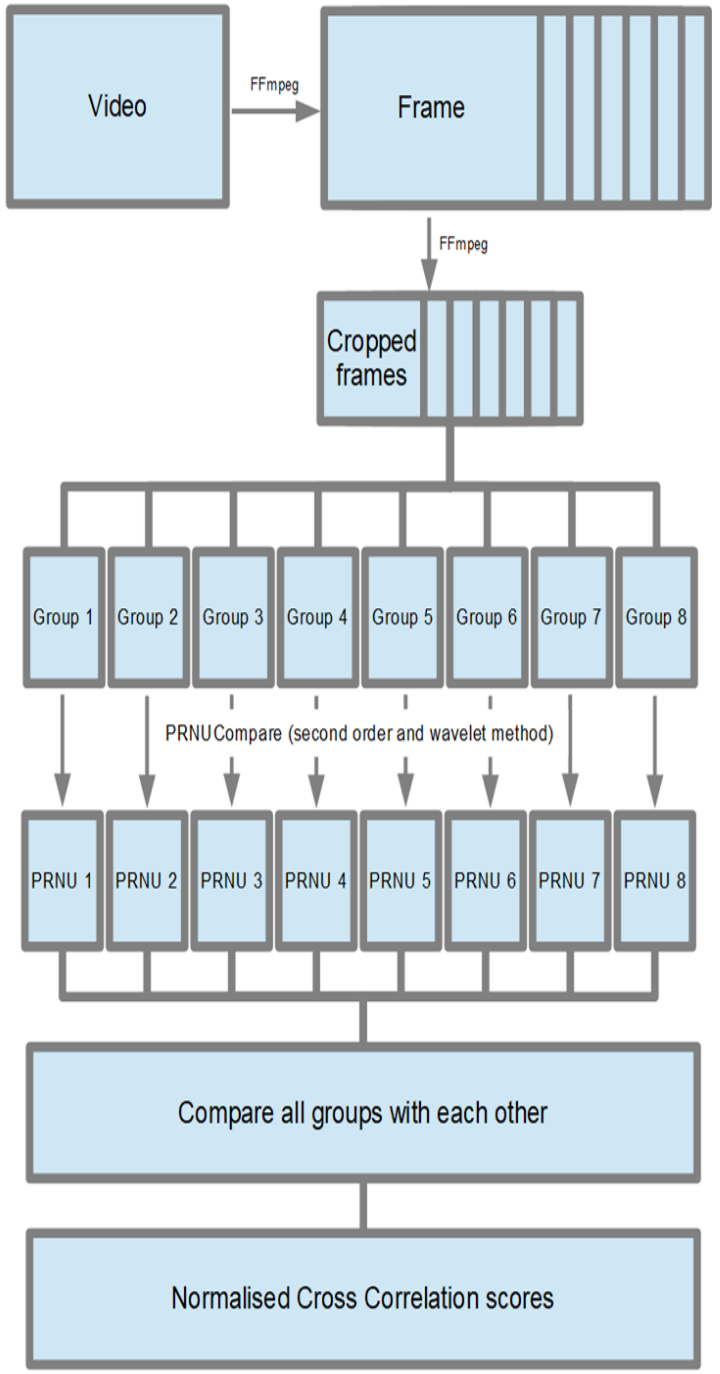
- ‘The fingerprint of the digital camera’
- Manipulation can alter PRNU pattern
- **Deepfake PRNU pattern less consistent throughout video compared to Authentic?**
- Second order and Wavelet method
  - Second wave = faster
  - Wavelet = more reliable
- Cropped and uncropped
  - Increase % variability of PRNU

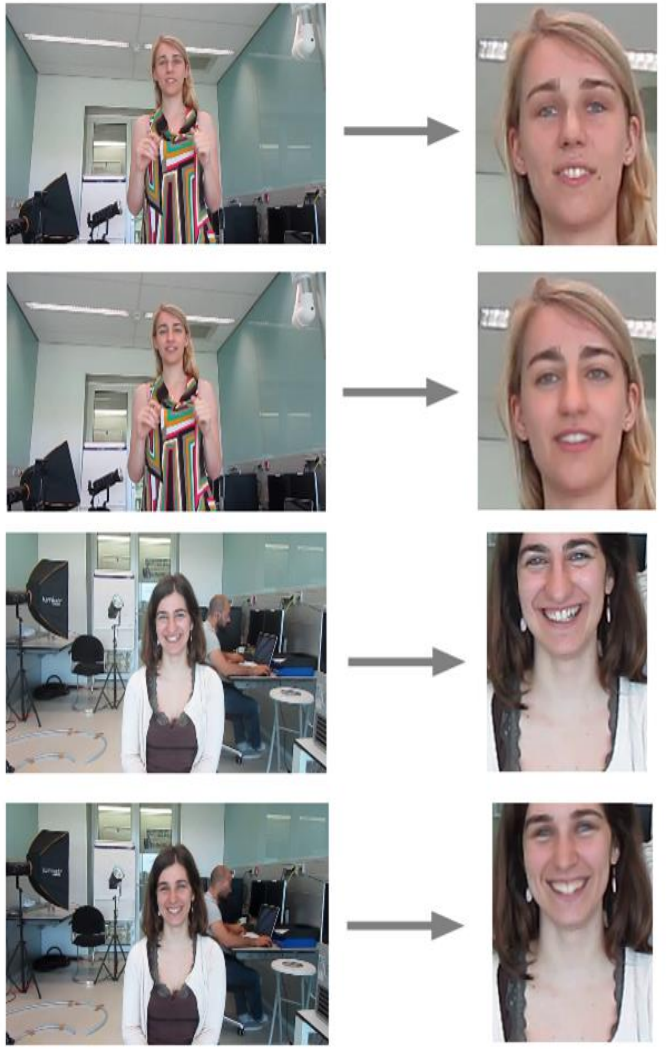


PRNU analysis:

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Method





PRNU analysis: Conclusions

Experiment	P-value variance in normalised cross correlation scores	P-value mean normalised cross correlation scores
Second order cropped	0.593	$5.21 * 10^{-5}$
Second order uncropped	0.303	0.002
Wavelet cropped	0.041	0.188
Wavelet uncropped	0.852	$3.23 * 10^{-4}$

- Second order > Wavelet method
  - Takes less time
  - Stronger correlation
  - More reliable correlation
- Second order cropped > Second order uncropped
  - Stronger correlation

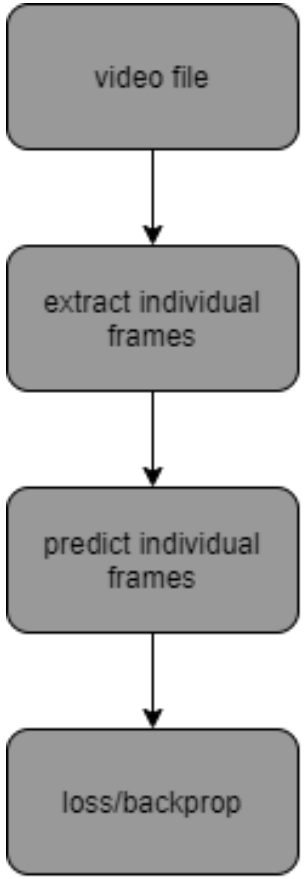
Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Detection methods

Image based

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Extract and annotate each frame individually, then use a feedforward CNN to classify each frame.



Training

calculate/backpropagate each frame-wise loss

Inference

predict each frame, then aggregate over all frames to generate video-level prediction

> detect manipulations by consulting spatial features

Video based

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Extract and annotate frame windows, then use spatio-temporal model to predict frame windows.

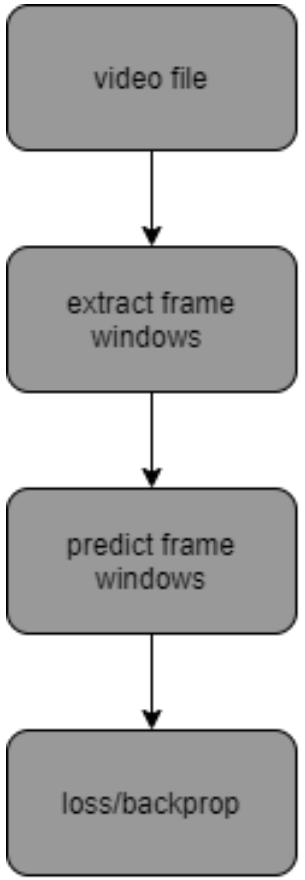
Training

calculate/backpropagate loss associated with each frame-window

Inference

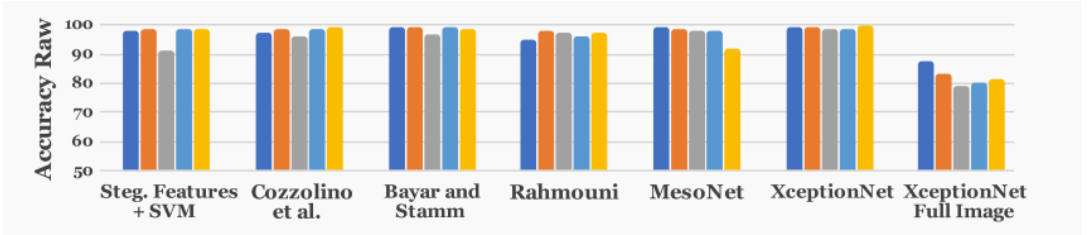
predict each frame window, then aggregate over all windows to generate video-level prediction

> detect manipulations by consulting spatio-temporal features



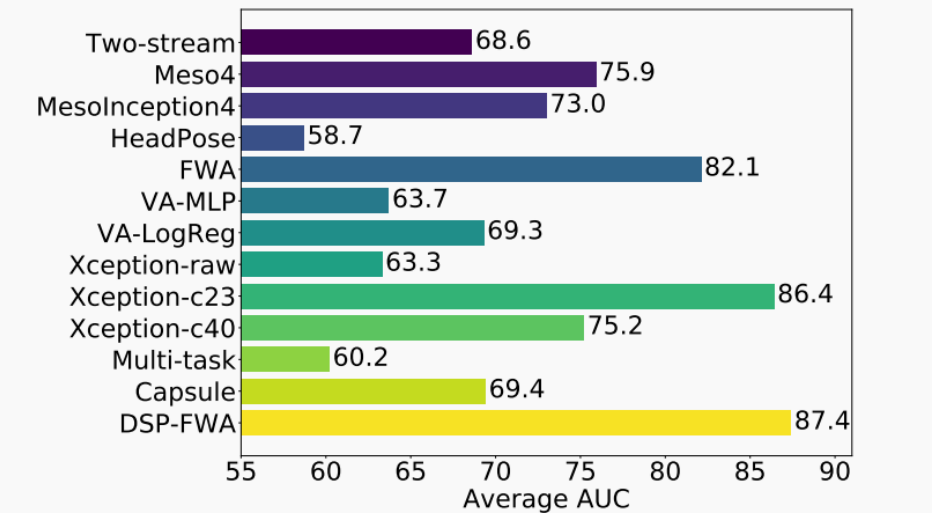
Why use video-based modeling?

Image based methods work well on in-dataset evaluations



Rossler et al. (2019)

But: cross-manipulation accuracy is low!
A good detection method should be able to consider features that underlie all manipulation methods.



Li et al. (2019)

Arms race

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

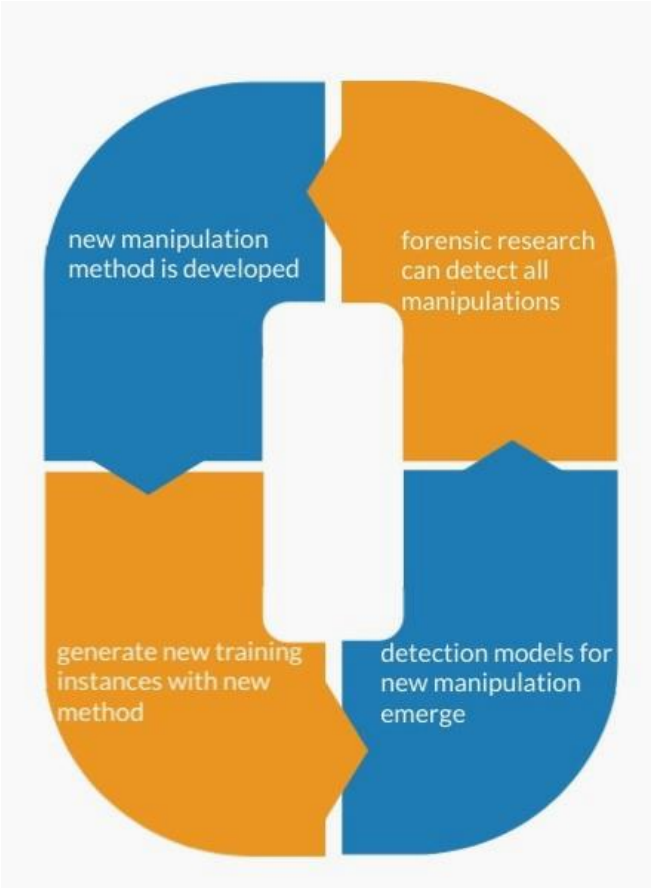
Forensic research on deepfakes has a major disadvantage:

Deep learning is data hungry

> but we cannot wait for availability of training data for each manipulation method

Need to find underlying weakness shared by all manipulations:

> until now: all deepfake generators are image-based!
> frame-to-frame consistencies are not enforced



Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Data

Celeb-DF

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

- 5.639 high resolution deepfakes
- generated by a single manipulation method
- use kalman filter to increase between-frame correlations of facial landmarks
- currently the most challenging set of Deepfakes available

Li et al. (2019)

DFDC - Deepfake Detection Challenge

- 470GB of Deepfake material
- multiple manipulation methods
- diverse in backgrounds, lighting, ethnicity, gender and number of actors (with a subset or all actors manipulated)
- unfortunately, most clips far from real-world quality

Dolhansky et al. (2019)

Real-world data

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

- collected from the web
- multiple, unknown manipulation methods
- subject to pre- and postprocessing
- handpicked, highest quality Deepfakes available

Facebook 100.000 samples deepfakes / real

https://www.technologyreview.com/2020/06/12/1003475/facebooks-deepfake-detection-challenge-neural-network-ai/

MIT Technology Review

Featured Topics Newsletters Events Podcasts Sign in Subscribe

FACEBOOK

Artificial intelligence / Machine learning

Facebook just released a database of 100,000 deepfakes to teach AI how to spot them

The videos are designed to help improve AI's performance—as even the best methods are still not accurate enough.

by **Will Douglas Heaven** June 12, 2020

2 free stories remaining Sign in Subscribe

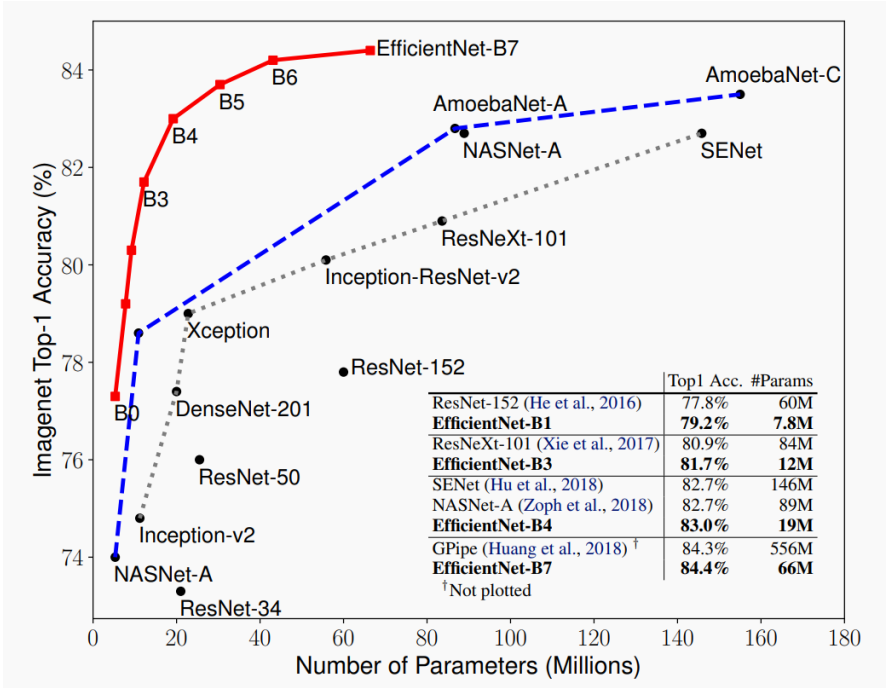
Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Models

EfficientNet

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

- high performance on ImageNet & co
- computationally manageable
- available pre-trained on ImageNet
- most frequently used model in high ranking teams DFDC



	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**79.2%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.7%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**83.0%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.4%**	**66M**

†Not plotted

Tan & Le (2019)

Watch out for database that is used

https://openaccess.thecvf.com/content/CVPR2021W/WMF/papers/Neekhara_Adversarial_Threats_to_DeepFake_Detection_A_Practical_Perspective_CVPRW_2021_paper.pdf

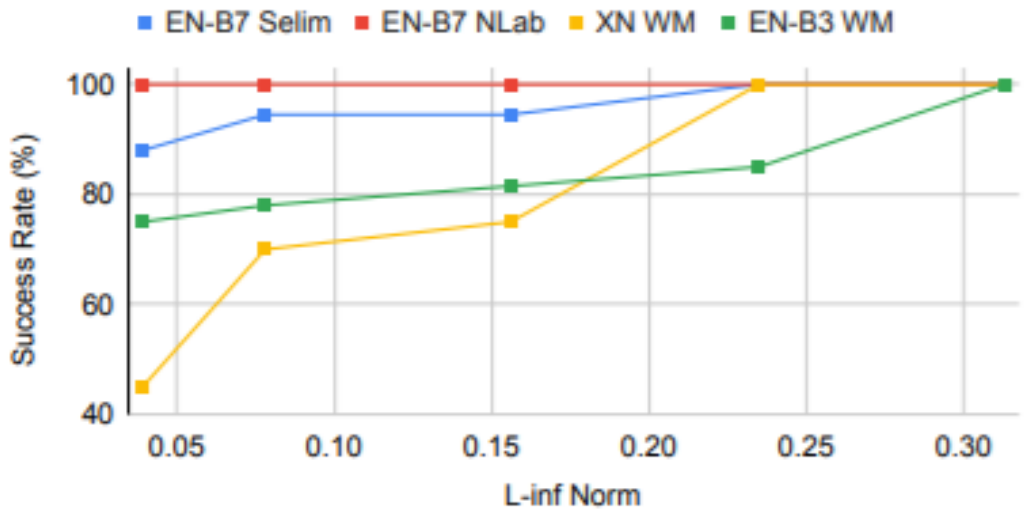
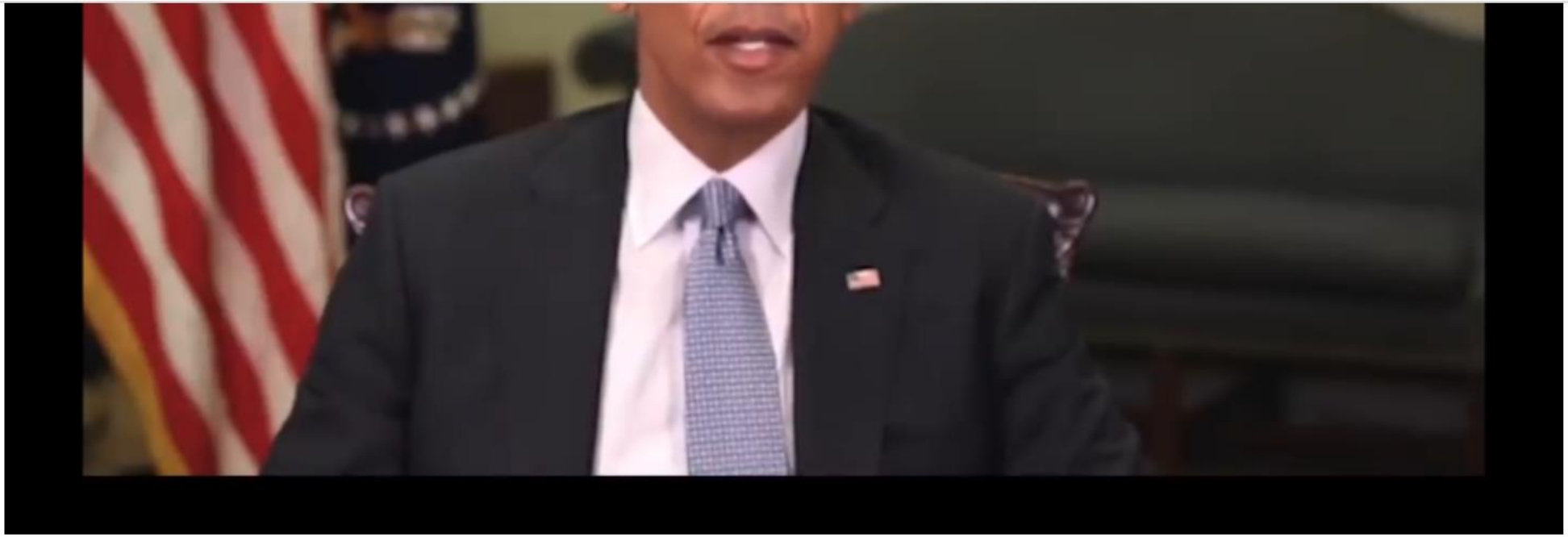


Figure 6. Attack success rate on unseen detectors for a universal perturbation trained on the *EN-B7 NLab* detector at different levels of the L_∞ norm of the perturbation.

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Making your own deepfakes

Try deepfacelab also https://github.com/iperov/DeepFaceLab for Windows



Model Results

Analyst:	DEEPFAKE DETECTED
Avatarify:	NO DEEPFAKE DETECTED(12%)
Deepware:	NO DEEPFAKE DETECTED(0%)
Seferbekov:	SUSPICIOUS(55%)
Ensemble:	NO DEEPFAKE DETECTED(15%)

Video

Duration:	14 sec
Resolution:	1280 x 720
Frame Rate:	23.97 fps
Codec:	h264

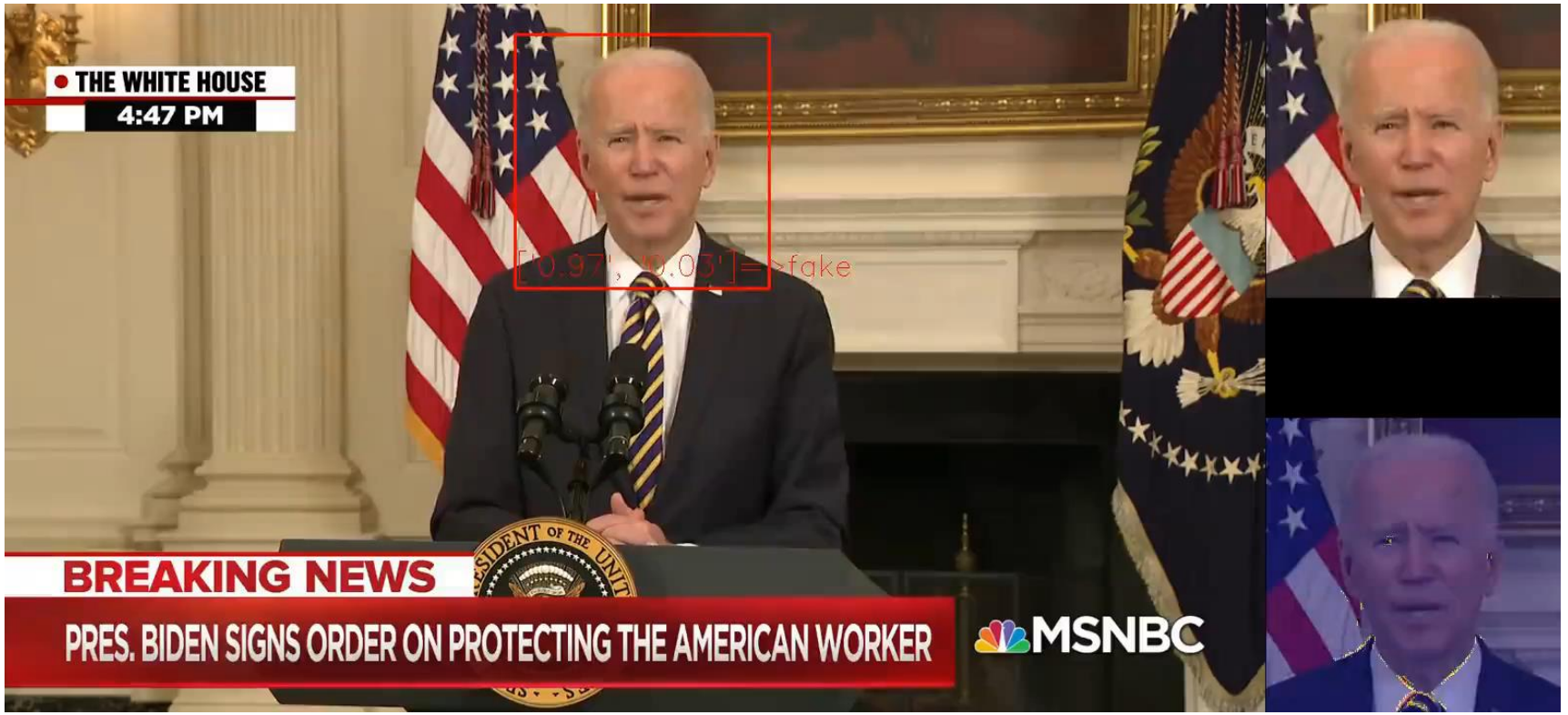
Audio

Duration:	14 sec
Channel:	stereo
Sample Rate:	44 khz
Codec:	aac

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Duckduckgoose

- Deedetector



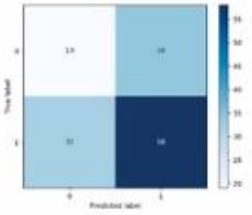
DProtoAnalyser | Prototype Refinement | Video Preprocessing | Deepfake Detection

Model

Select Model: net_003_n

Name	net_003_n
Acuracy:	0.9131
AUC:	0.9513
Num Epochs:	31
Num Prototypes:	20

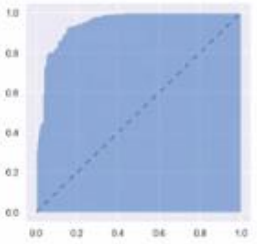
Confusion Matrix



Labels: (0) pristine
(1) manipulated

True Positives:	58
True Negatives:	19
False Positives:	34
False Negatives:	32
Precision:	0.6304
Recall:	0.6444

ROC Curve



Prototypes

Prototype Detailed View

Prototype id: 2 | Image file: 938-294-frame-1.gif | Category: Pristine | Resolution: 176 x 87 pixels | Number of Frames: 10 | Activation Map Color Code:

Prototype Vector

Prototype Viewer

Prototype Scaling: Relative to Image Source | Fill area

Prototype Representation: Animated GIF | Still Image (PNG

Prototype Sequence

Prototype Sequence Representation: RGB Sequence | Optical Flow Sequence

Prototype Deletion

Calculate Impact on Model Performance | Confirm Prototype Deletion | Proceed

Prototype Replacement

Find Alternative Prototype: Nearest Neighours | Farthest Neighours | Random | Find

Results:

Calculate Impact on Model Performance | Confirm Prototype Replacement | Proceed

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Summary

- ICAI lab AI4Forensics
- AI used a lot in digital forensic products
- AI for making deepfakes and detecting them ?
- Use explainable AI : does the court understand what we are doing ?
- Multidisciplinary approach combining different method way forward
- Forensic aspects in court
  - Defense : this evidence was fake produced by AI

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Summary warnings

- Expert / human in the loop
- Does the court really use the evidence ?
- Good Training is required of all users
- Do we mean the same with validation ?

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Questions
z.geradts@nfi.nl