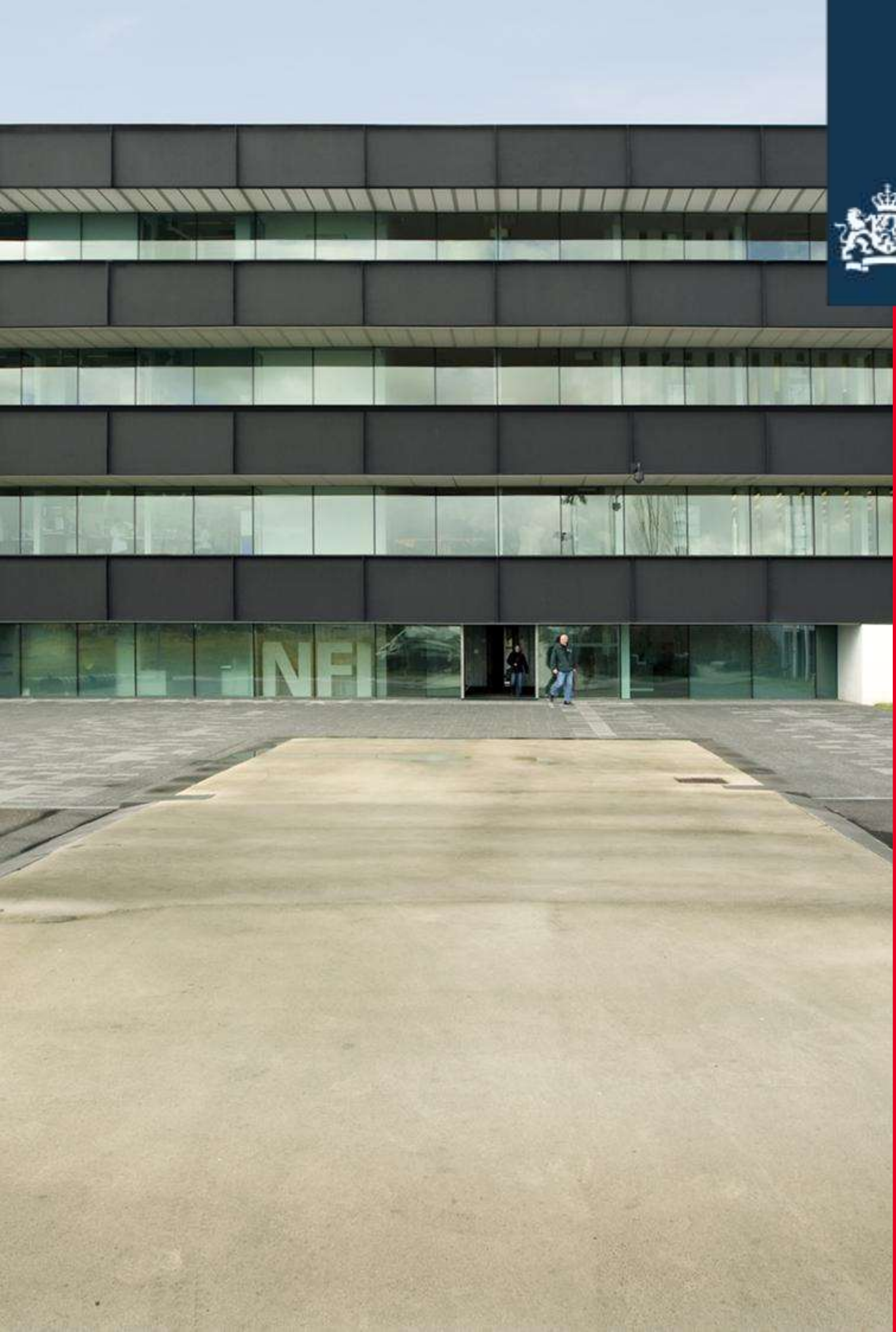


Netherlands Forensic Institute
Ministry of Security and Justice

Forensic Intelligence workshop

prof. dr. ing. Zeno Geradts
Senior forensic scientist /
Special Chair Forensic Data Science
Digital Technology and Biometrics /
University of Amsterdam

DFRWS Oslo 2019

COST Project DigForAsp

DigForAsp (Digital forensics: evidence analysis via intelligent systems and practices) – CA17124 is funded by the European Cooperation in Science and Technology (COST). DigForAsp activities were launched on 10th September 2018 for 4 years.

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme
of the European Union

Outline

- Introduction
- Deep learning and neural networks
- Examples deepfakes
- Issues
- Outlook and conclusion

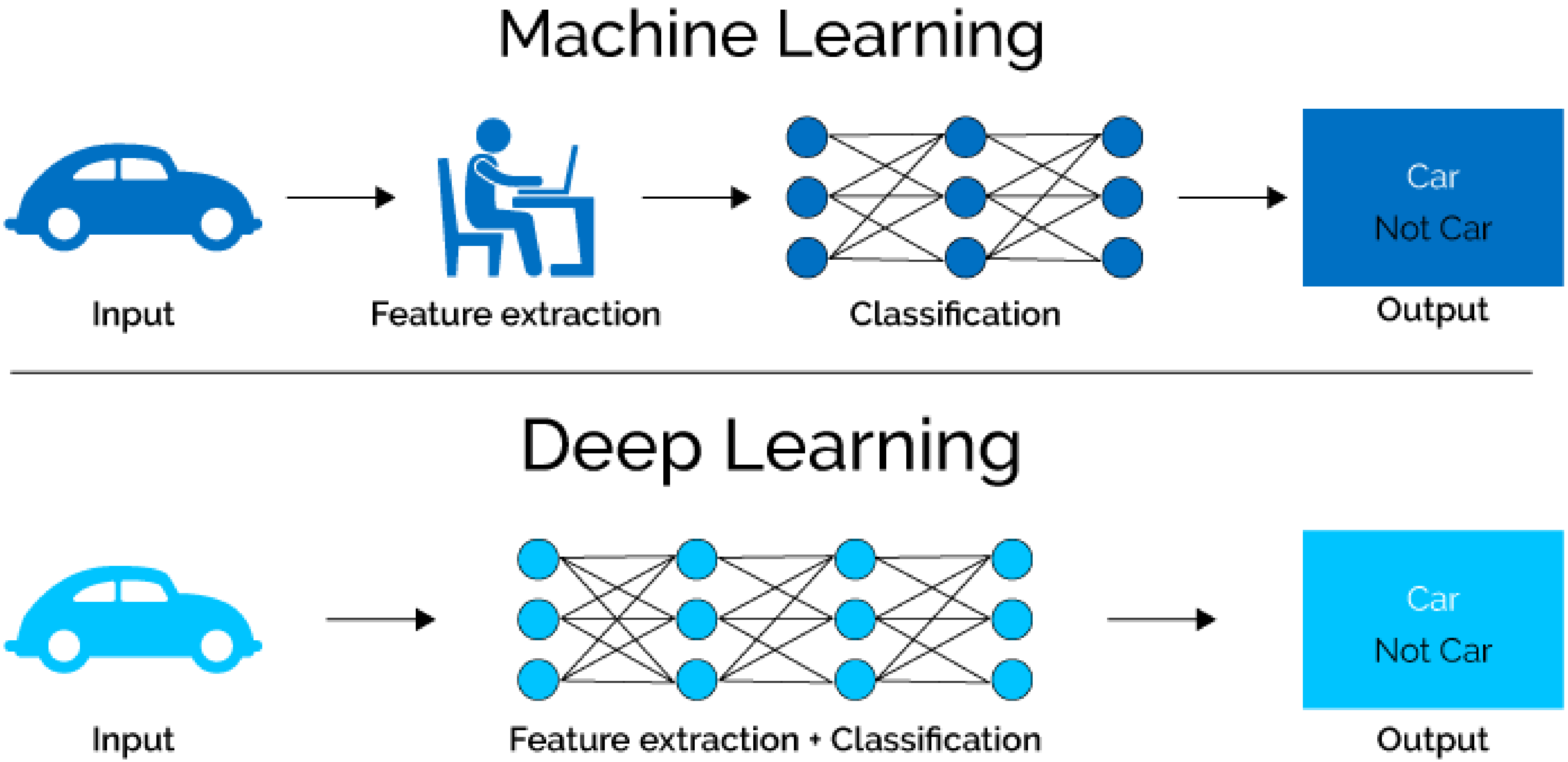
Netherlands Forensic Institute

University of Amsterdam
Chair Forensic Data Science

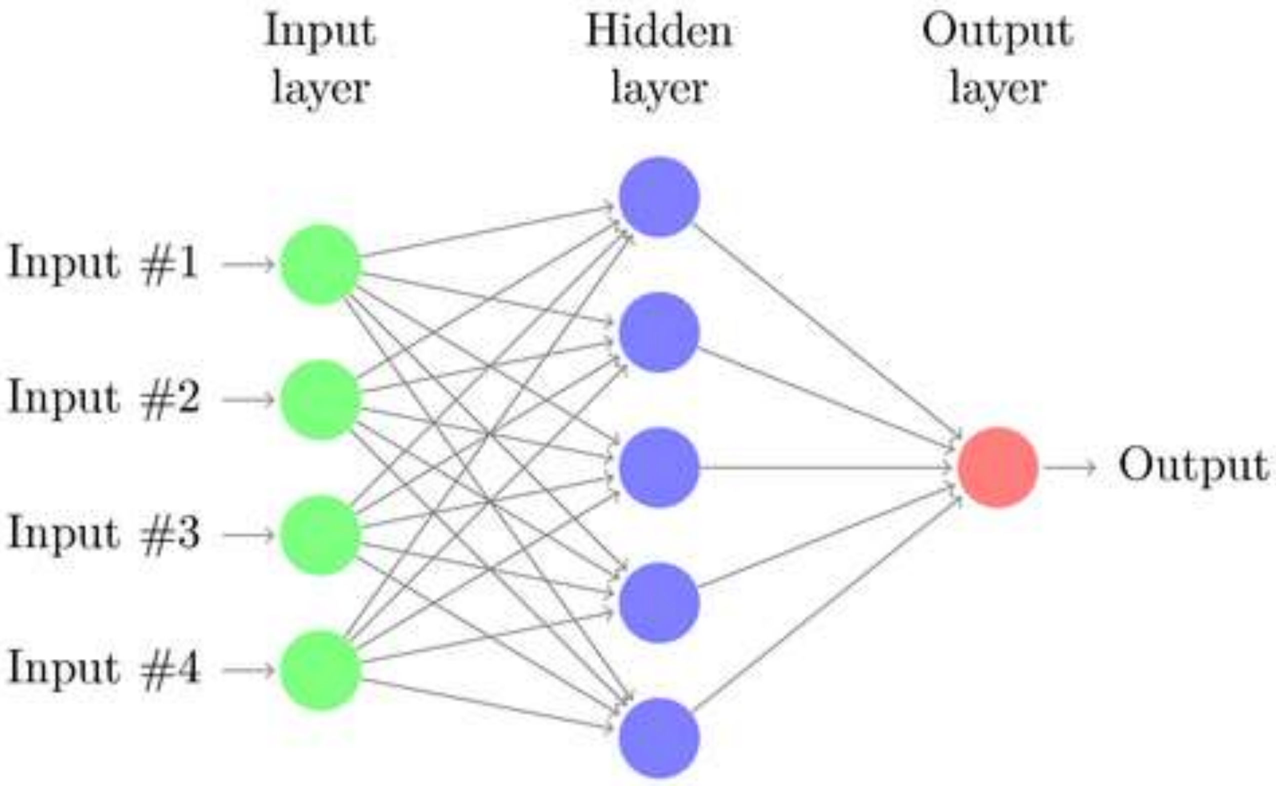
- store and process
- understand and decide
- analyse and model
- Report and visualize
- Higher efficiency
- Data-intensive
- Evidential strength big data

Third Hype in history in AI

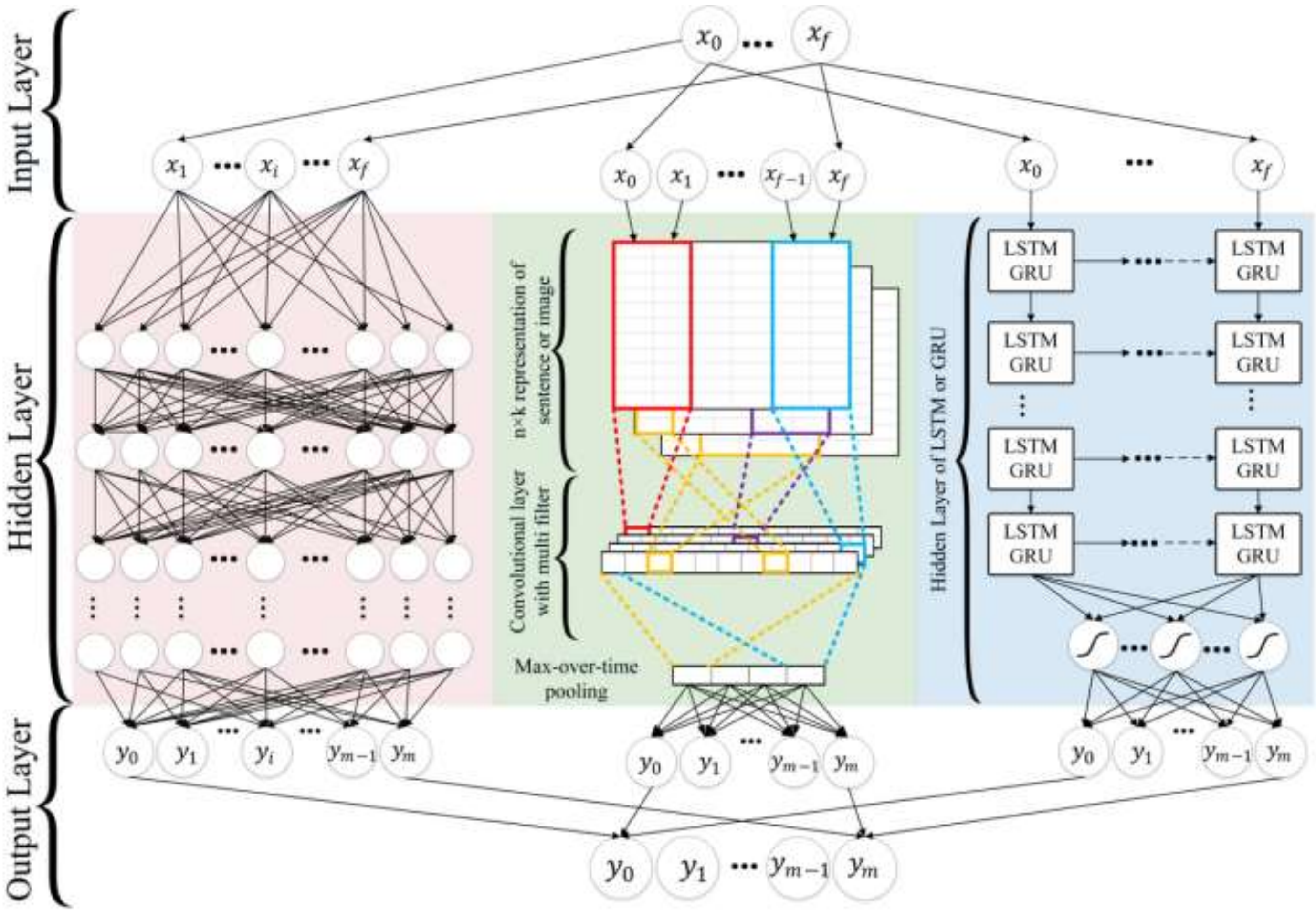
Machine learning vs deep learning



neural network



Neural network multilayer



Calculation speed with Digital Evidence

Digital Evidence

NFI Focus on data

Extract data

Make data readable

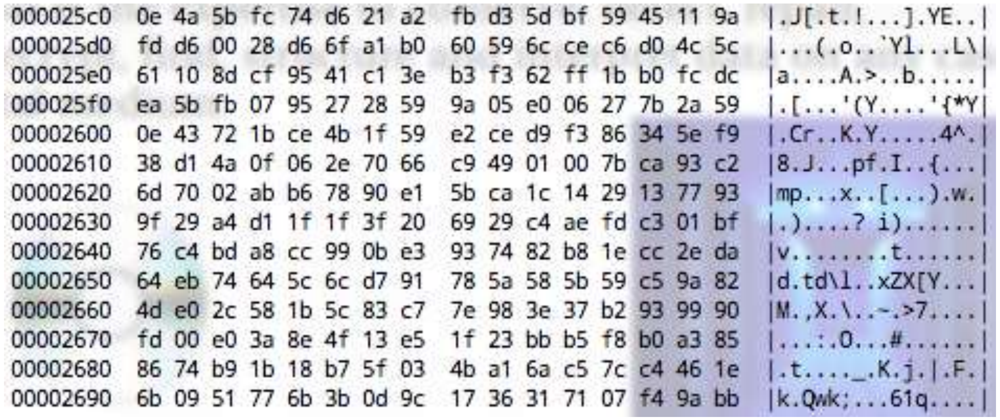
Organize data

Interpret data

Make data readable

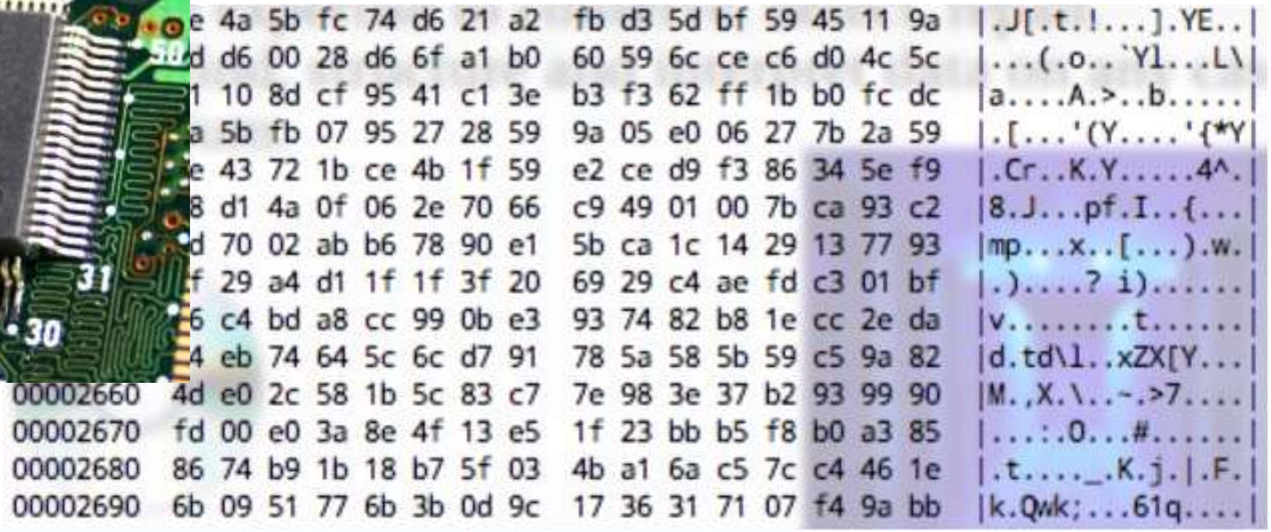
Challenge: many formats, old & new, non-standard

- 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



Trace Recovery & Analysis

Trace-analysis is the expertise to conserve, detect, repair, undelete, decrypt, find, structure and interpret data and traces on any case related digital medium.



TECHNOLOGY ROADMAP: THE INTERNET OF THINGS

Technology Reach

rapid / short development cycles

increasing streaming data volume

time spend online

consumer prices for devices+data drop rapidly

fast global expension of bandwidth

57% per year

digital behaviour

Software agents and advanced sensor fusion

Miniaturization, power-efficient electronics, and available spectrum

Teleoperation and telepresence: Ability to monitor and control distant objects

Physical-World Web

Ability of devices located indoors to receive geolocation signals

Locating people and everyday objects

Ubiquitous Positioning

Cost reduction leading to diffusion into 2nd wave of applications

Surveillance, security, healthcare, transport, food safety, document management

Vertical-Market Applications

Demand for expedited logistics

RFID tags for facilitating routing, inventorying, and loss prevention

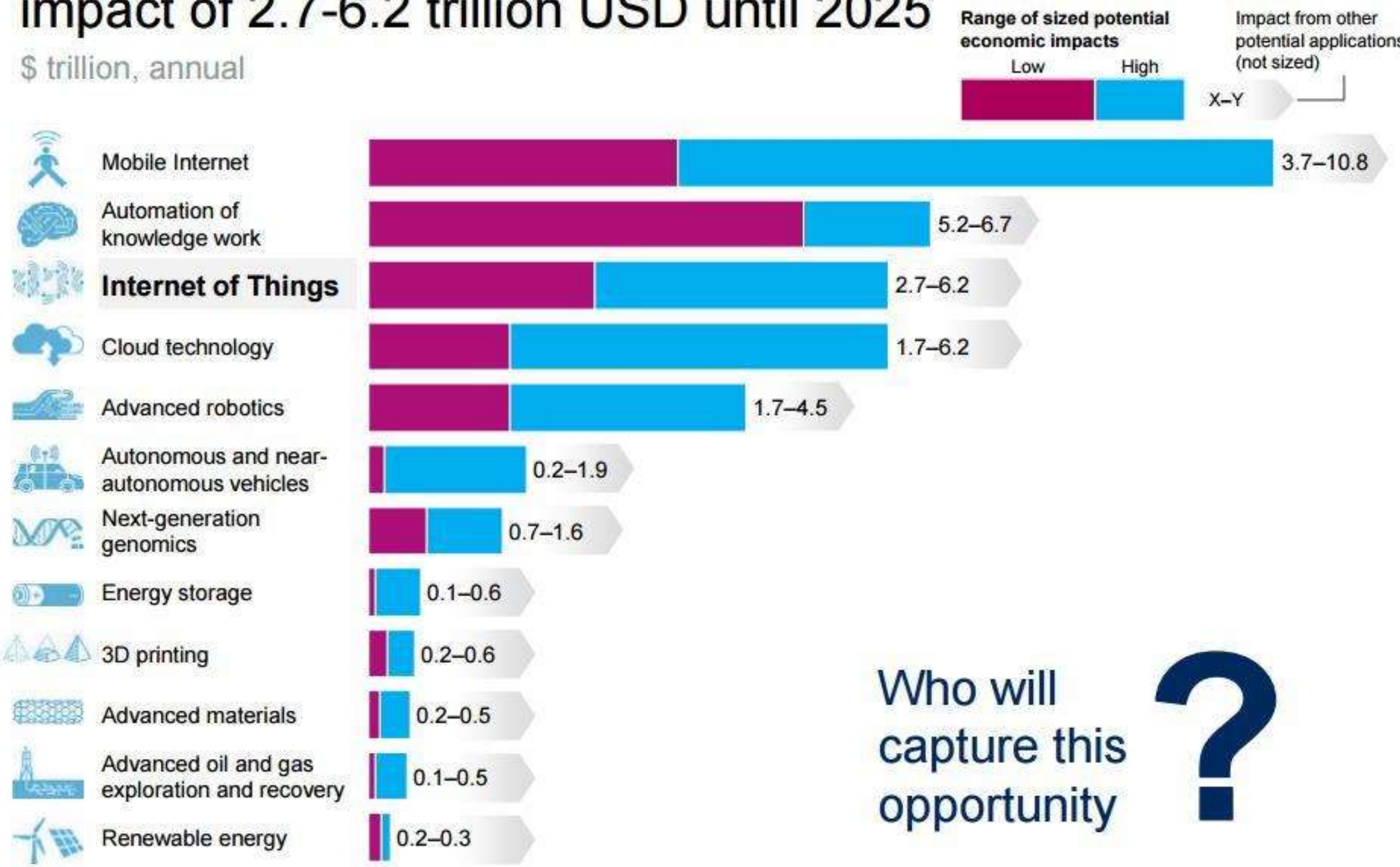
Supply-Chain Helpers

2000 2010 2020 Time

Source: SRI Consulting Business Intelligence

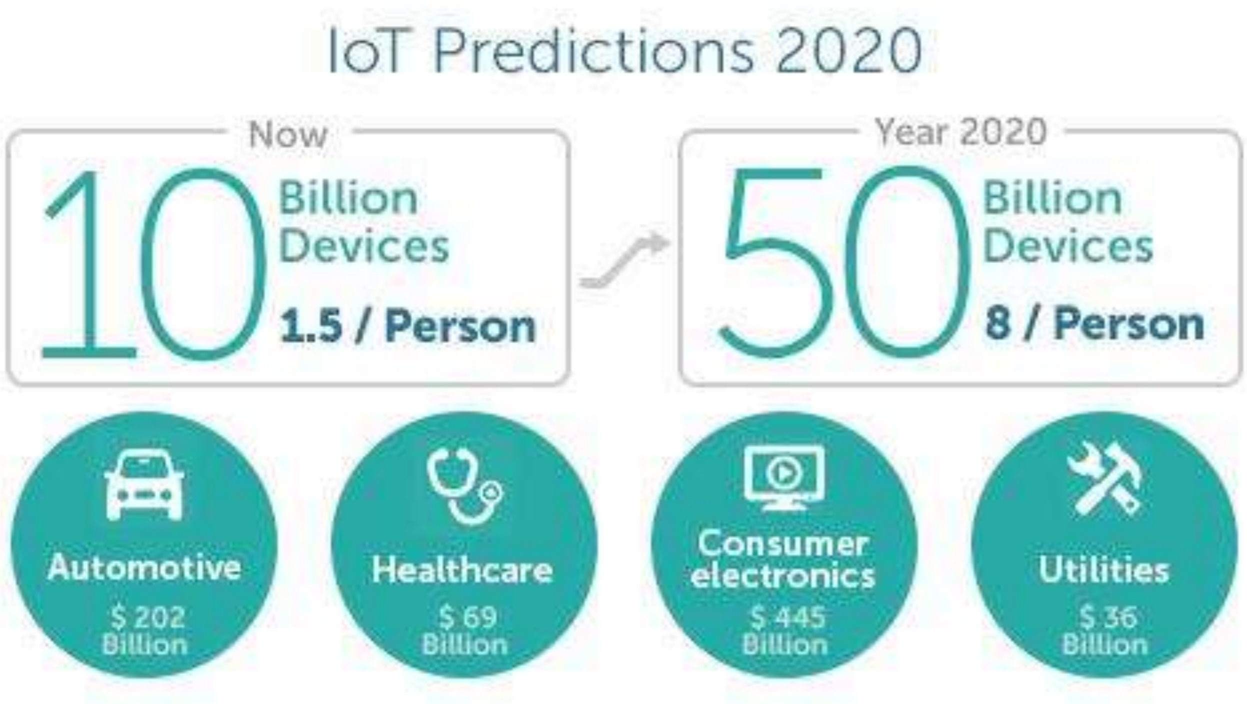
THE IoT PLATFORM OPPORTUNITY

The Internet of Things (IoT) has a potential economic impact of 2.7-6.2 trillion USD until 2025



SOURCE: McKinsey Global Institute analysis

Internet of things 2020 Gartner



5G antenna and fiber boom

Big Data issues

news.berkeley.edu/2018/06/18/big-data-flaws/

UC Berkeley

Berkeley News

Research ▾ People ▾ Campus & communit

MIND & BODY, RESEARCH, TECHNOLOGY & ENGINEERING

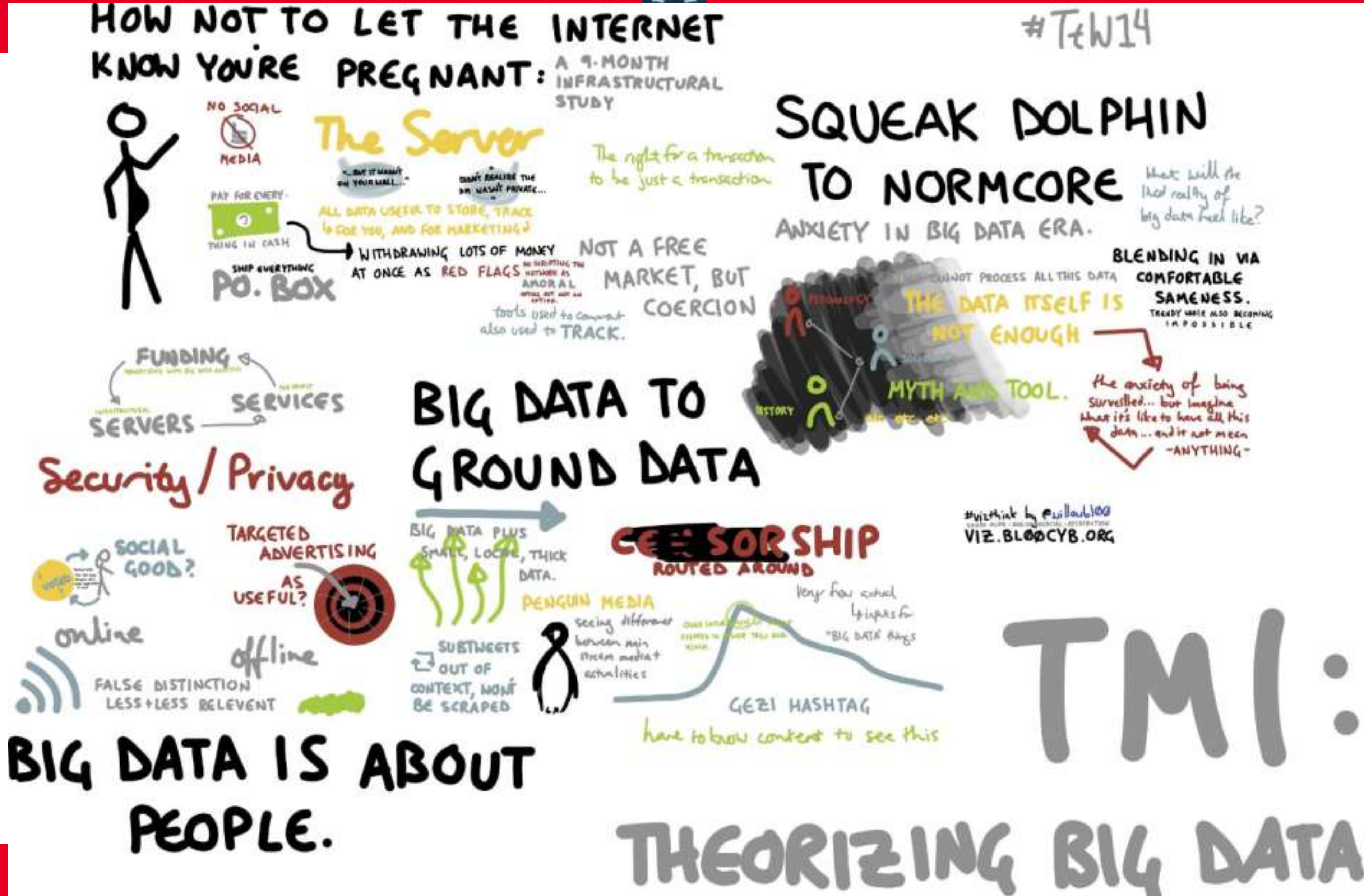
Everything big data claims to know about you could be wrong

By Yasmin Anwar, Media Relations | JUNE 18, 2018

445

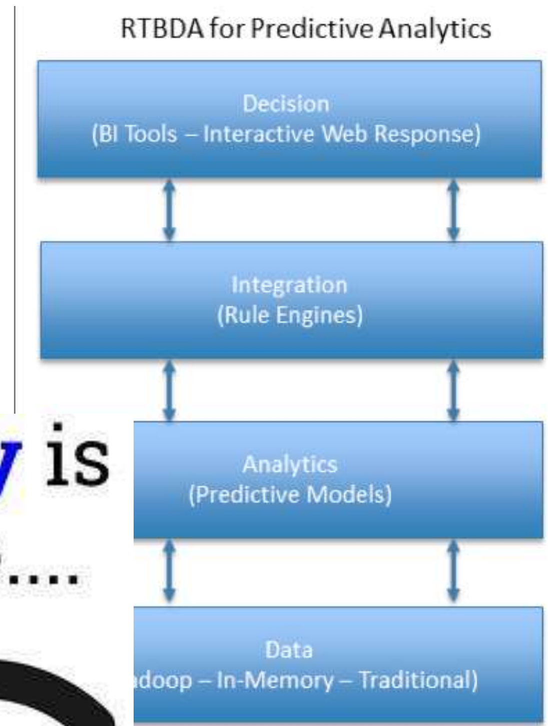
When it comes to understanding what makes people tick — and get sick — medical science has long assumed that the bigger the sample of human subjects, the better. But new research led by UC Berkeley suggests this big-data approach may be wildly off the mark.

That's largely because emotions, behavior and physiology vary markedly from one person to


HOW NOT TO LET THE INTERNET KNOW YOU'RE PREGNANT: A 9-MONTH INFRASTRUCTURAL STUDY
#TtW14
NO SOCIAL MEDIA
The Server
PAY FOR EVERYTHING IN CASH
SHIP EVERYTHING
PO. BOX
WITHDRAWING LOTS OF MONEY AT ONCE AS RED FLAGS
The right for a transaction to be just a transaction
NOT A FREE MARKET, BUT COERCION
tools used to connect also used to TRACK.
SQUEAK DOLPHIN TO NORMCORE
ANXIETY IN BIG DATA ERA.
What will the lived reality of big data feel like?
BLENDING IN VIA COMFORTABLE SAMENESS.
THE DATA ITSELF IS NOT ENOUGH
MYTH AND TOOL.
HISTORY
the anxiety of being surveilled... but imagine what it's like to have all this data... and it not mean -ANYTHING-
FUNDING
SERVICES
SERVERS
BIG DATA TO GROUND DATA
Security / Privacy
SOCIAL GOOD?
TARGETED ADVERTISING AS USEFUL?
online
offline
FALSE DISTINCTION LESS+LESS RELEVENT
BIG DATA PLUS SMALL, LOCAL, THICK DATA.
PENGUIN MEDIA
SUBTWEETS OUT OF CONTEXT, WON'T BE SCRAPED
CENSORSHIP ROUTED AROUND
GEZI HASHTAG
have to know context to see this
#vizthink by @willoublee
VIZ.BLOOCYB.ORG
BIG DATA IS ABOUT PEOPLE.
TMI:
THEORIZING BIG DATA

The good news : many examples were it works well credit card fraud detection and casework

VISA states they save billions of euros a year

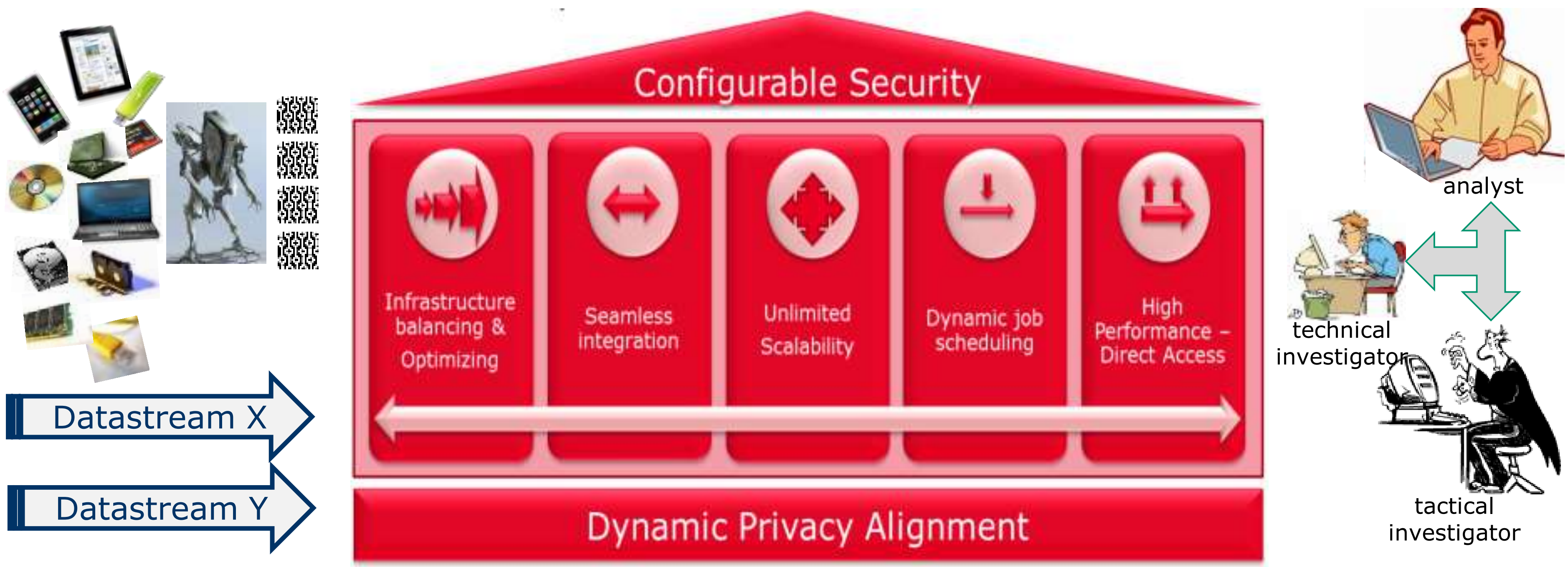


Big Data at NFI

- Text Mining
- Data Profiling
- Financial Data Analysis
- Social Network Analysis
- video and images
- using big data analysis in forensic science

Future of digital investigation: HANSKEN

CONFISCATION › SECURE › ENABLE ACCESS / ENRICH › REPORT › ANALYSE



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%

Many new techniques

https://www.dfrws.org/2014eu/proceedings/DFRWS-EU-2014-7.pdf

DFRWS-EU-2014-7.pdf 1 / 9

https://www.dfrws.org/2014eu/proceedings/DFRWS-EU-2014-7.pdf

Digital Investigation 11 (2014) S54–S62

Contents lists available at ScienceDirect

ELSEVIER

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

Digital Forensics as a Service: A game changer

CrossMark

R.B. van Baar*, H.M.A. van Beek, E.J. van Eijk

Netherlands Forensics Institute, Laan van Ypenburg 6, 2497 GB The Hague, The Netherlands

Keywords:
Digital forensics
DFaaS

A B S T R A C T

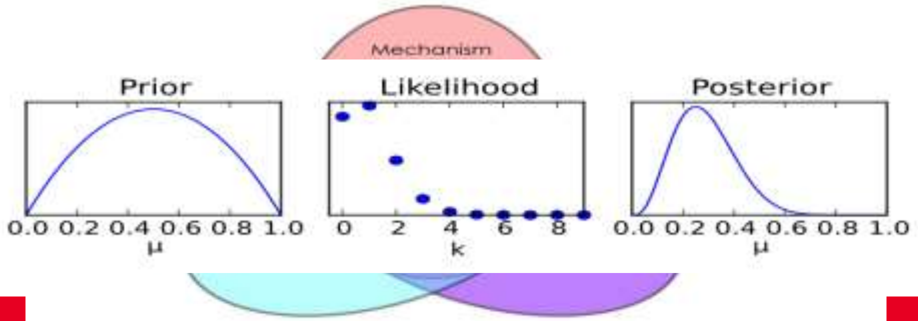
How is it that digital investigators are always busy and still never have enough time to actually dig deep into digital evidence? In this paper we will explore the current implementation of the digital forensic process and analyze factors that impact the efficiency of

is processed and of the manner in which the traces collected by this processing is analyzed

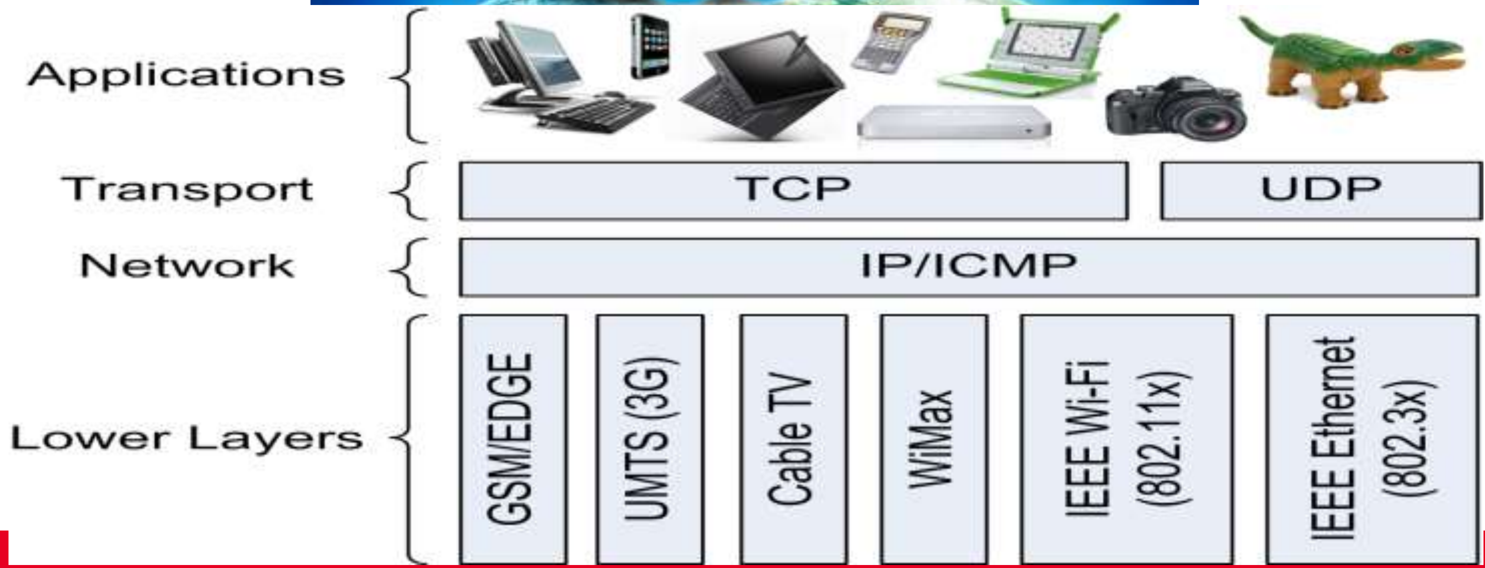
Related work

Examples hypotheses in digital forensic science

- has the computer been hacked or not ?
- has the email been send or not ?
- has the USB been plugged in or not ?
- was the phone in this location or at the location presented by the defence ?
- has the child pornography been send by the computer of the suspect or not ?
- is the child porn photographed with this camera or another camera ?



Lawful internet interception



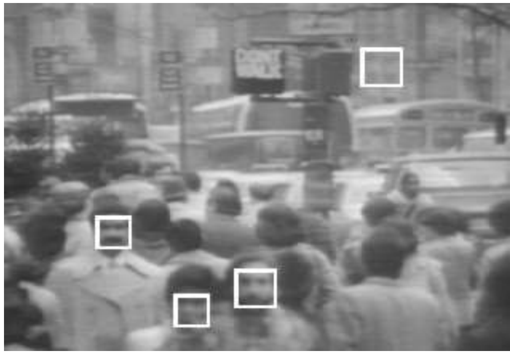
IMSI catcher : privacy by design ?

Interpret data

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



About Forensic Big Data Analysis

- Our data come from confiscated phones, hard drives, licence plate cameras, telephone providers, and so on…



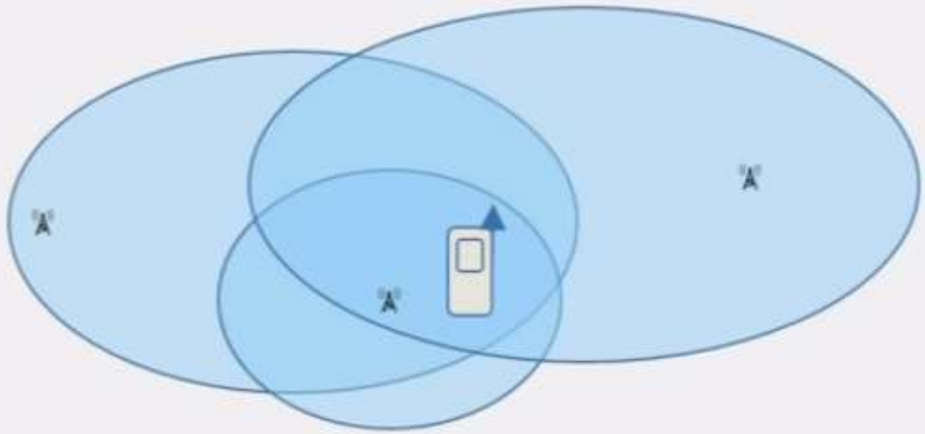
What if…

- An ATM machine is blown up
- A prepaid cell phone is found on the scene
- The police have their eyes on a suspect
- **Research question: is the suspect the user of the prepaid phone?**

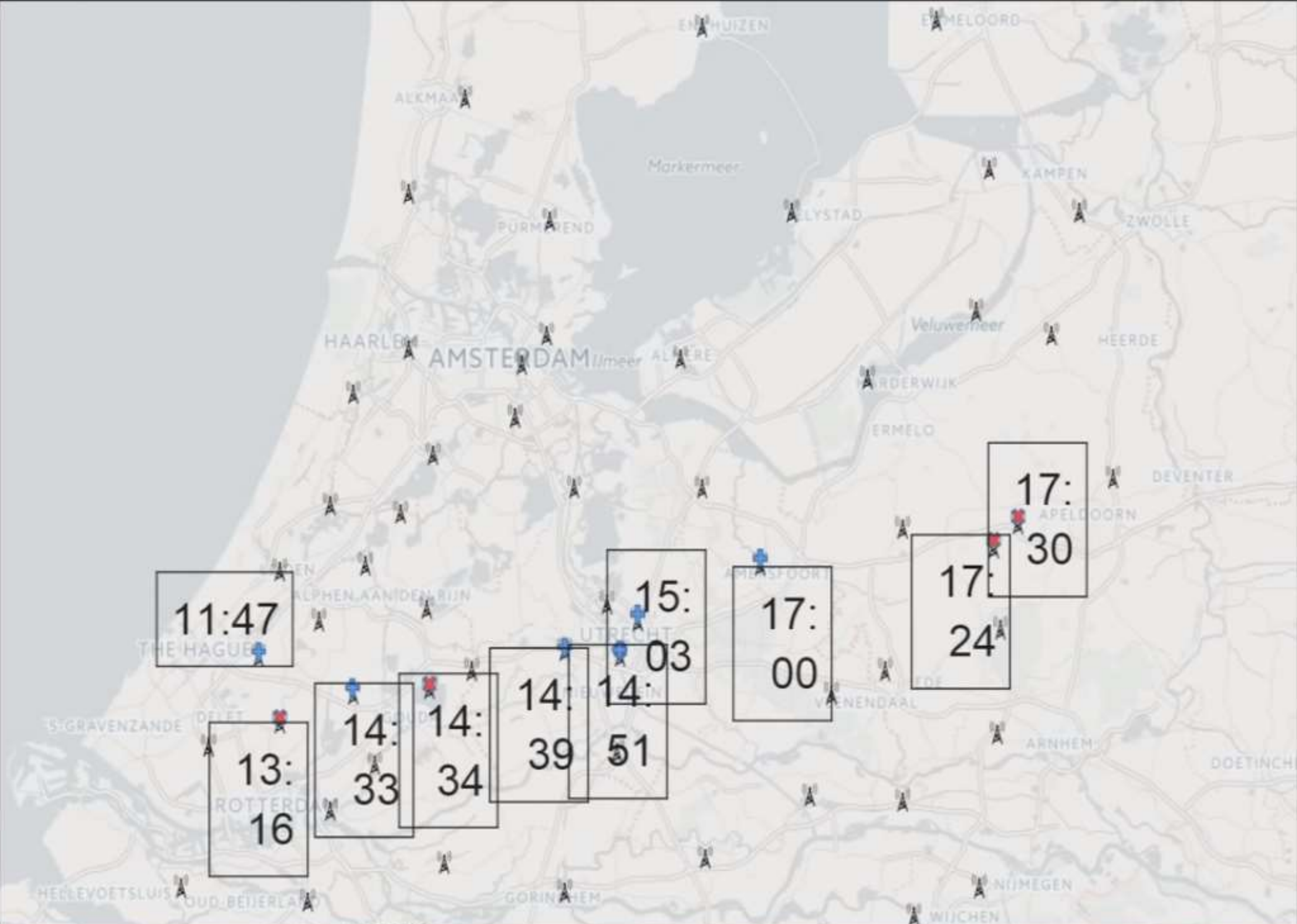
What information do we have?

- You know the phone number of the prepaid phone and that of the suspect's private phone.
- The telephone provider provides the police with usage data for both phones.
- Every time a phone connects to a cell tower, you know when it happened.
- You know the location of each cell tower.

Problem 1: cell tower location data are not precise and depend on…



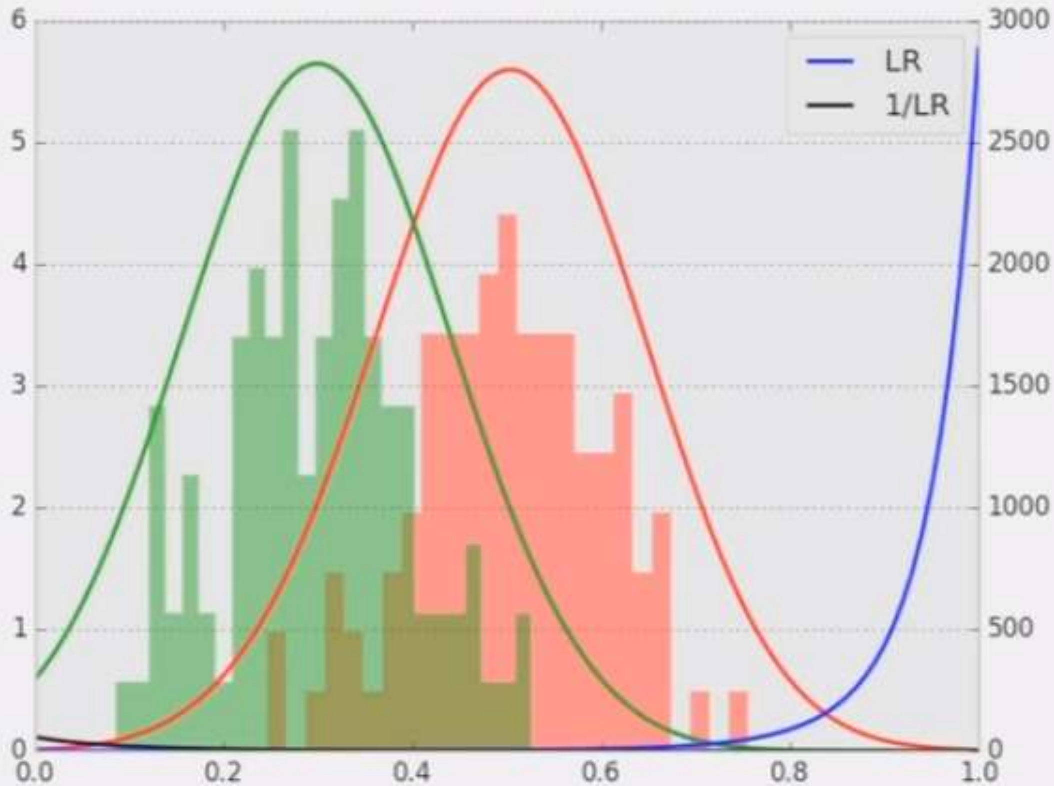
- Theoretical range: 35km
- Direction of transmission
- Distance
- Obstacles (tall buildings)
- Weather conditions
- Network load


ALKMAAR
PURMEREND
Markermeer
LELYSTAD
KAMPEN
ZWOLLE
HAARLEM
AMSTERDAM
Almere
Veluwemeer
HEERDE
HARDERWIJK
ERMELO
DEVENTER
APELDOORN
AMERSFOORT
ALPHEN AAN DEN RIJN
UTRECHT
THE HAGUE
DELFT
'S-GRAVENZANDE
ROTTERDAM
VEENENDAAL
ARNHEM
DOETINCHEM
NIJMEGEN
WIJCHEN
HELLEVOETSLUIS
OUD-BEIJERLAND
GORINCHEM
11:47
13:16
14:33
14:34
14:39
14:51
15:03
17:00
17:24
17:30

To summarize…

- We want to know if the suspect is the user of a prepaid phone that can be linked to a crime.
- We know when and where the prepaid phone was used.
- We know when and where the suspect's phone was used.
- But our data are sparse and imprecise…

Likelihood Ratio

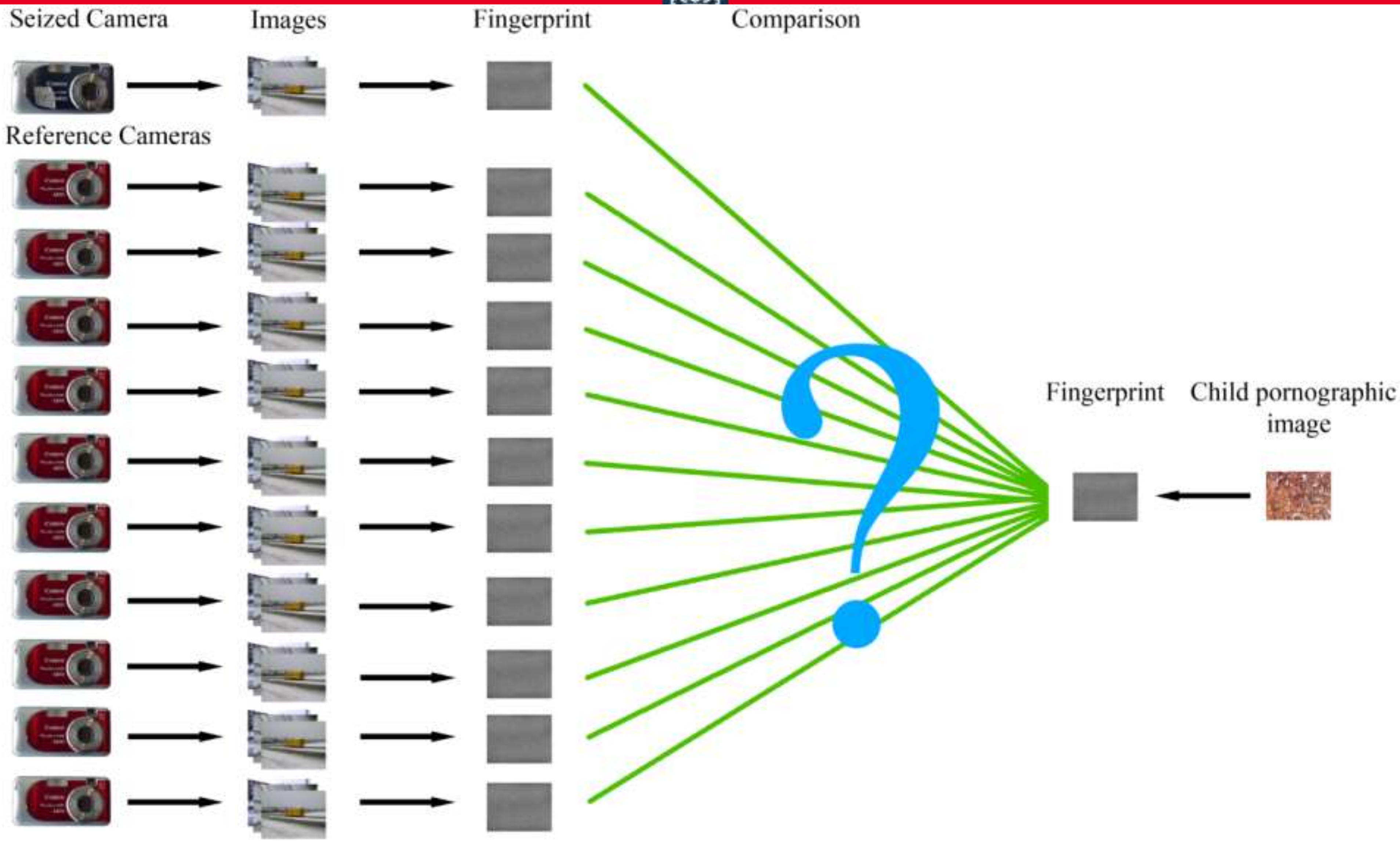


Digital Camera Identification

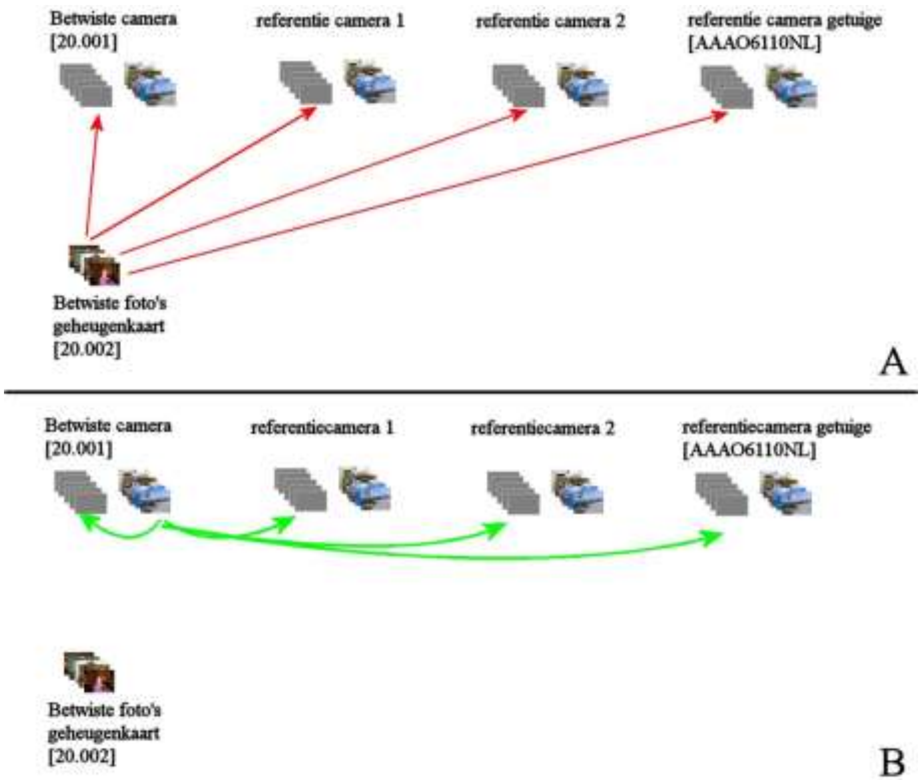
The process of

Linking images to the source camera

Linking images to images in a database
to determine a common source

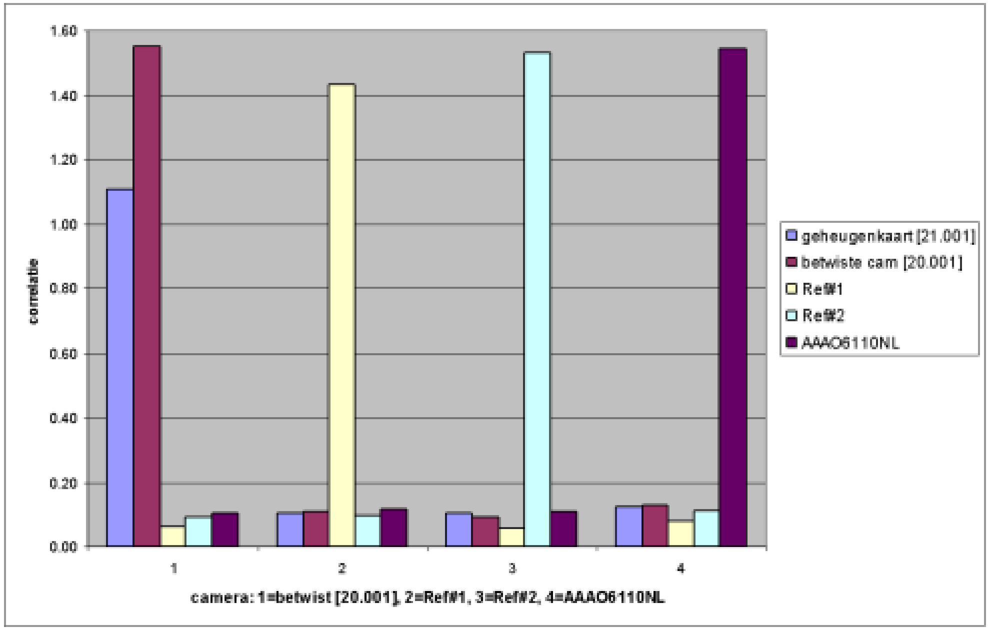

Seized Camera
Images
Fingerprint
Comparison
Reference Cameras
Fingerprint
Child pornographic image

Casework links

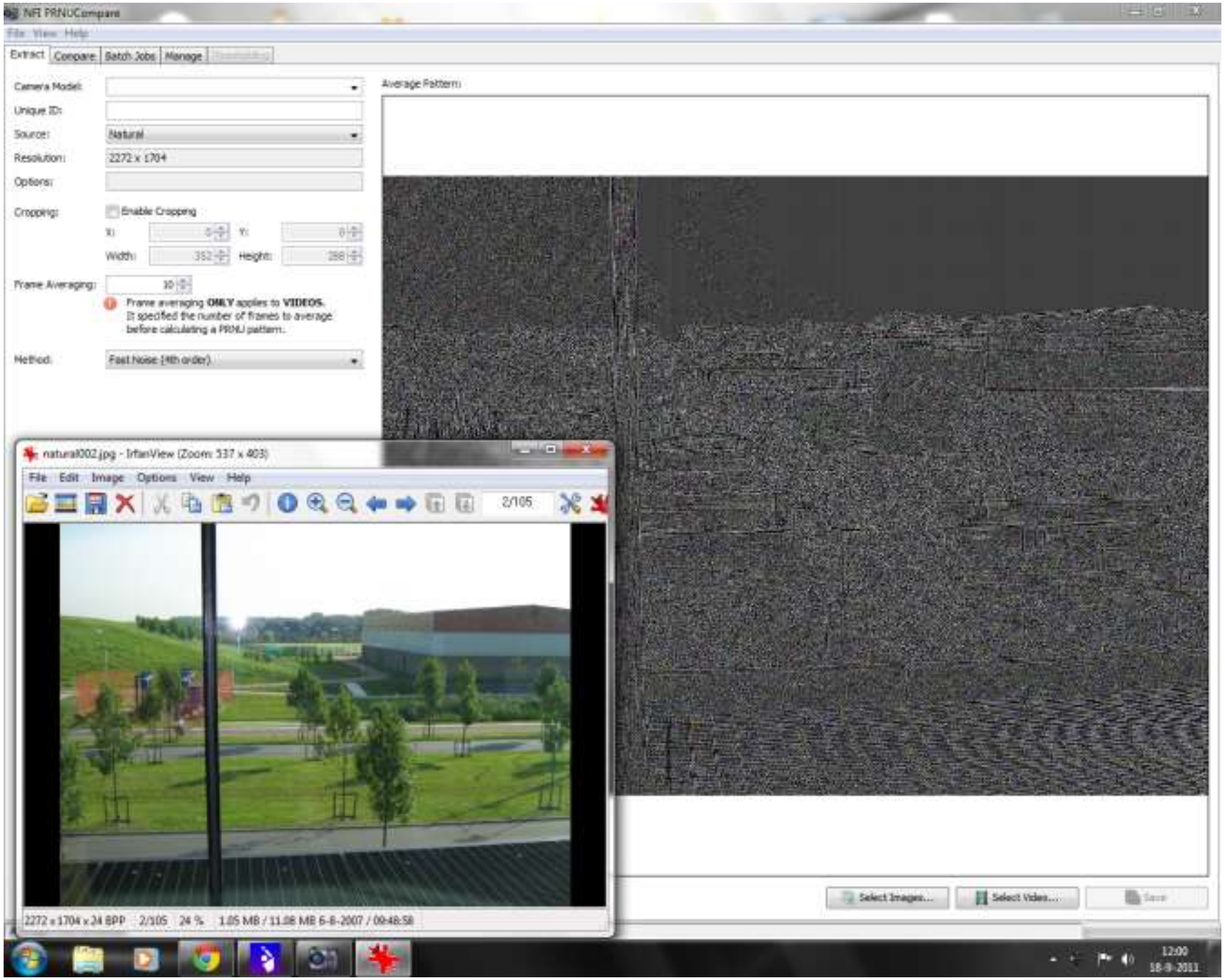


Casework

- Example where it worked



PRNU Compare



Bayesian

Question: were the images made with the seized camera?

Conclusion

The findings of the investigation are:

Equally likely

Somewhat more likely

More likely

Much more likely

Very much more likely

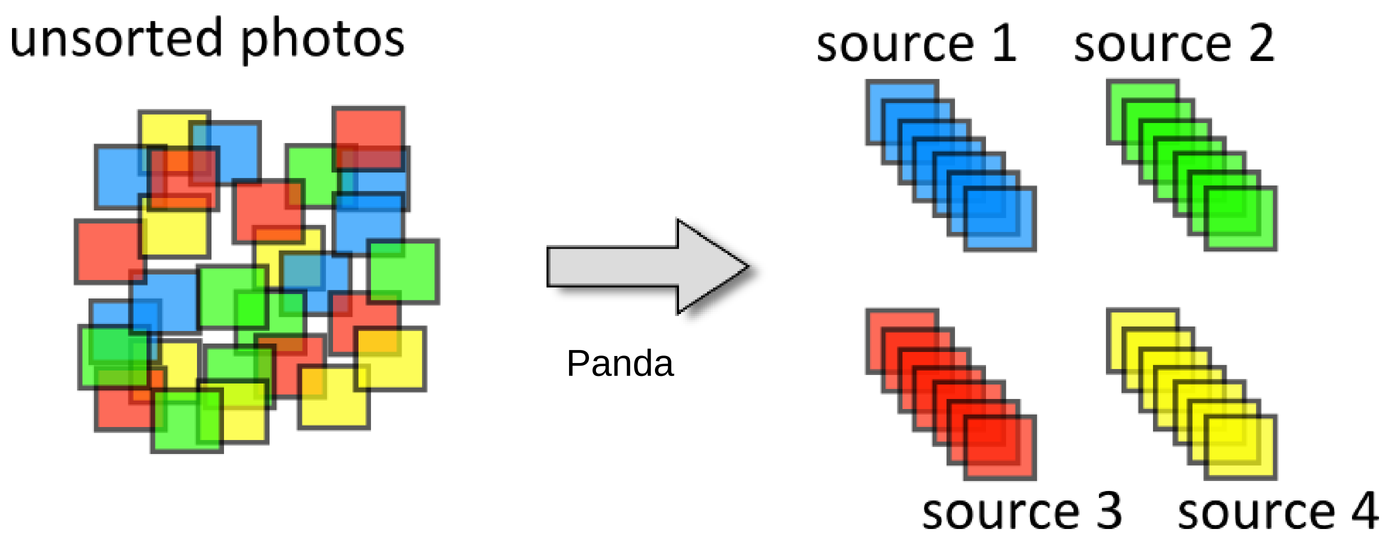
if H1 is true, than if H2 is true.

The findings are very much more likely if the Seized Camera took the child pornographic image, than if another camera took the image.

Camera	...	...	...	Sum
Comparing: Foto in kwestie (Onbekend)				
Verdachte Camera reference (Canon PowerShot)	..	..	...	0.131330
Camera 1 reference (Canon PowerShot)	..	..	...	0.008054
Camera 4 reference (Canon PowerShot)	..	..	...	0.007700
Camera 2 reference (Canon PowerShot)	..	..	...	0.007022
Camera 3 reference (Canon PowerShot)	..	..	...	0.006287

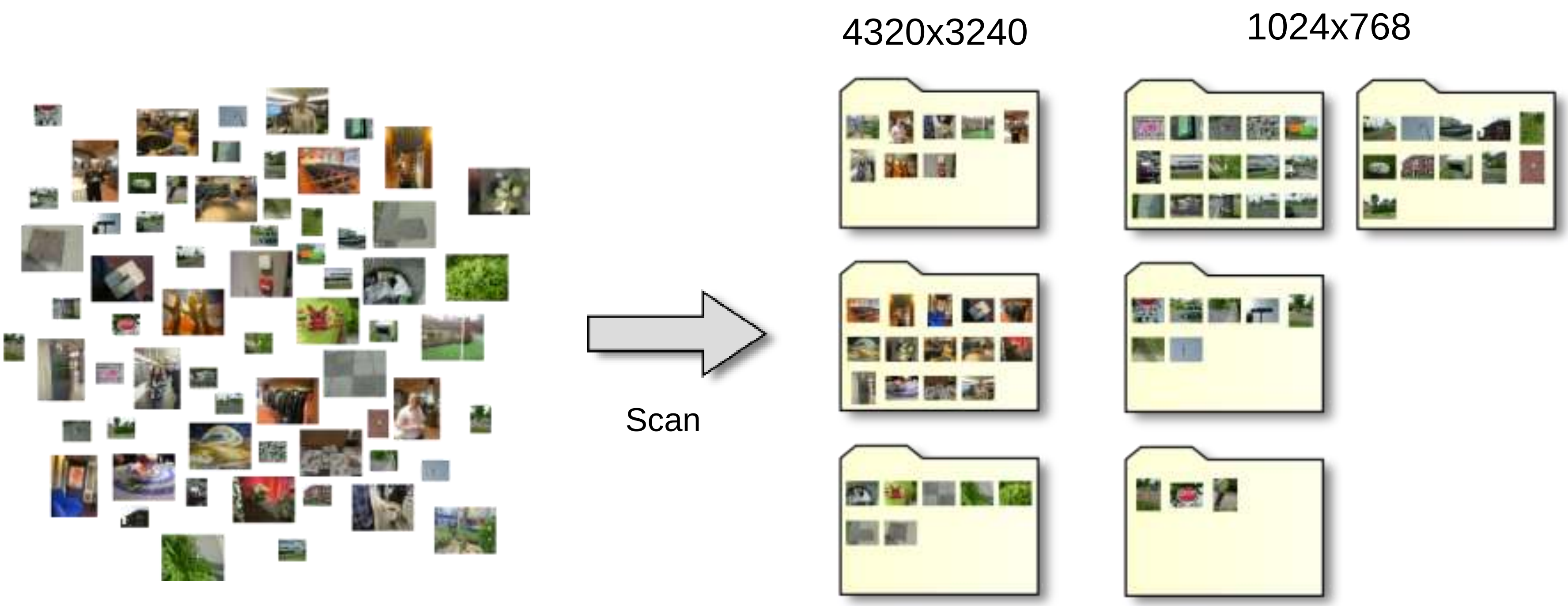
Large Scale Camera Identification

- Sorting photos by source
- Identify photos from the same source (camera)
- New valuable information and insight



Sorting Images by Source

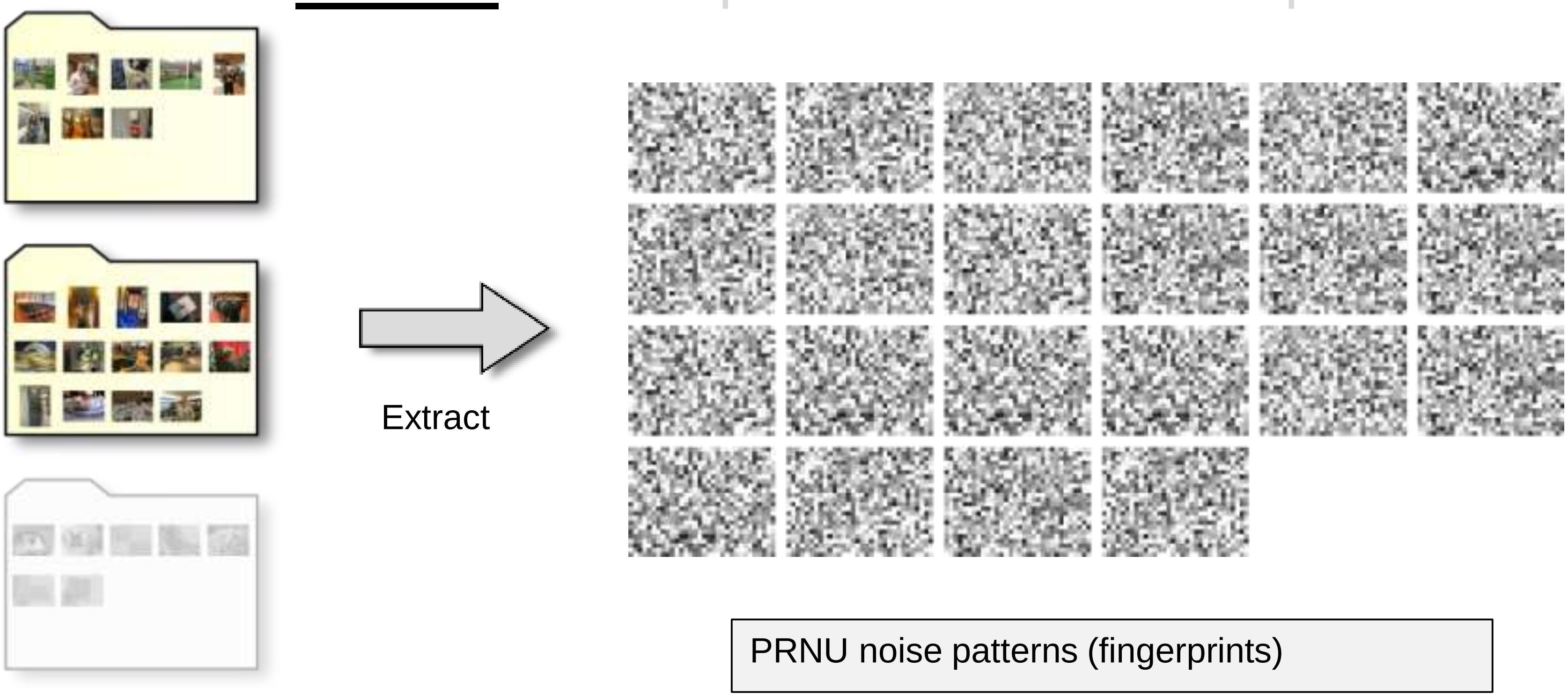
Scan→Extract → Compare → Cluster → Explore



Sorted by resolution and directory

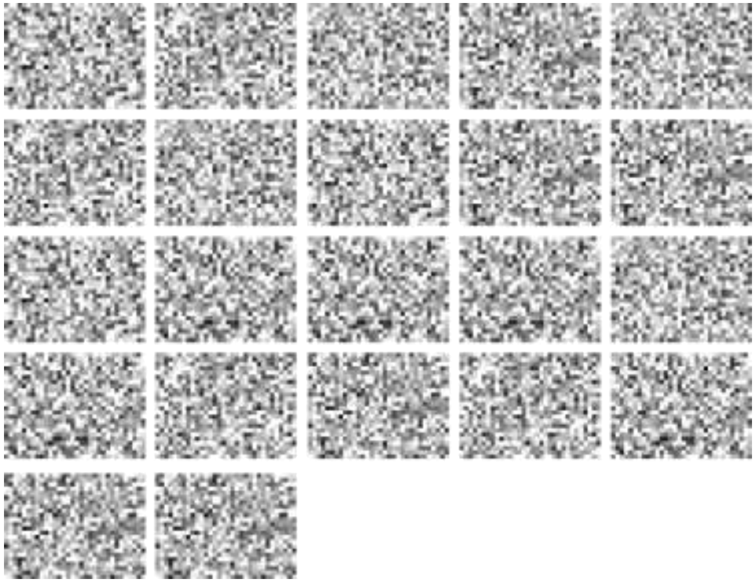
Sorting Images by Source

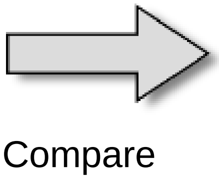
Scan→**Extract** → Compare → Cluster → Explore



Sorting Images by Source

Scan→Extract → **Compare** → Cluster → Explore



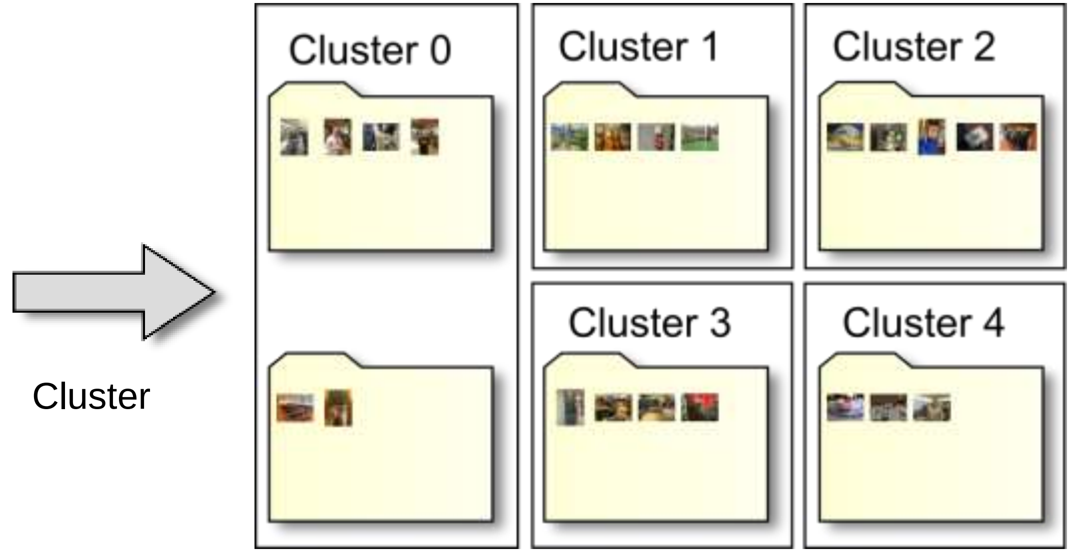


Images compared to all images

Sorting Images by Source

Scan→Extract → Compare → **Cluster** → Explore

threshold = 0.001



Images grouped by source

Sorting Images by Source also GPU / social networks also deep learning applied

Scan→Extract → Compare → Cluster → **Explore**

GP

Facial comparison

NO

YES

NIST test of faces in the wild

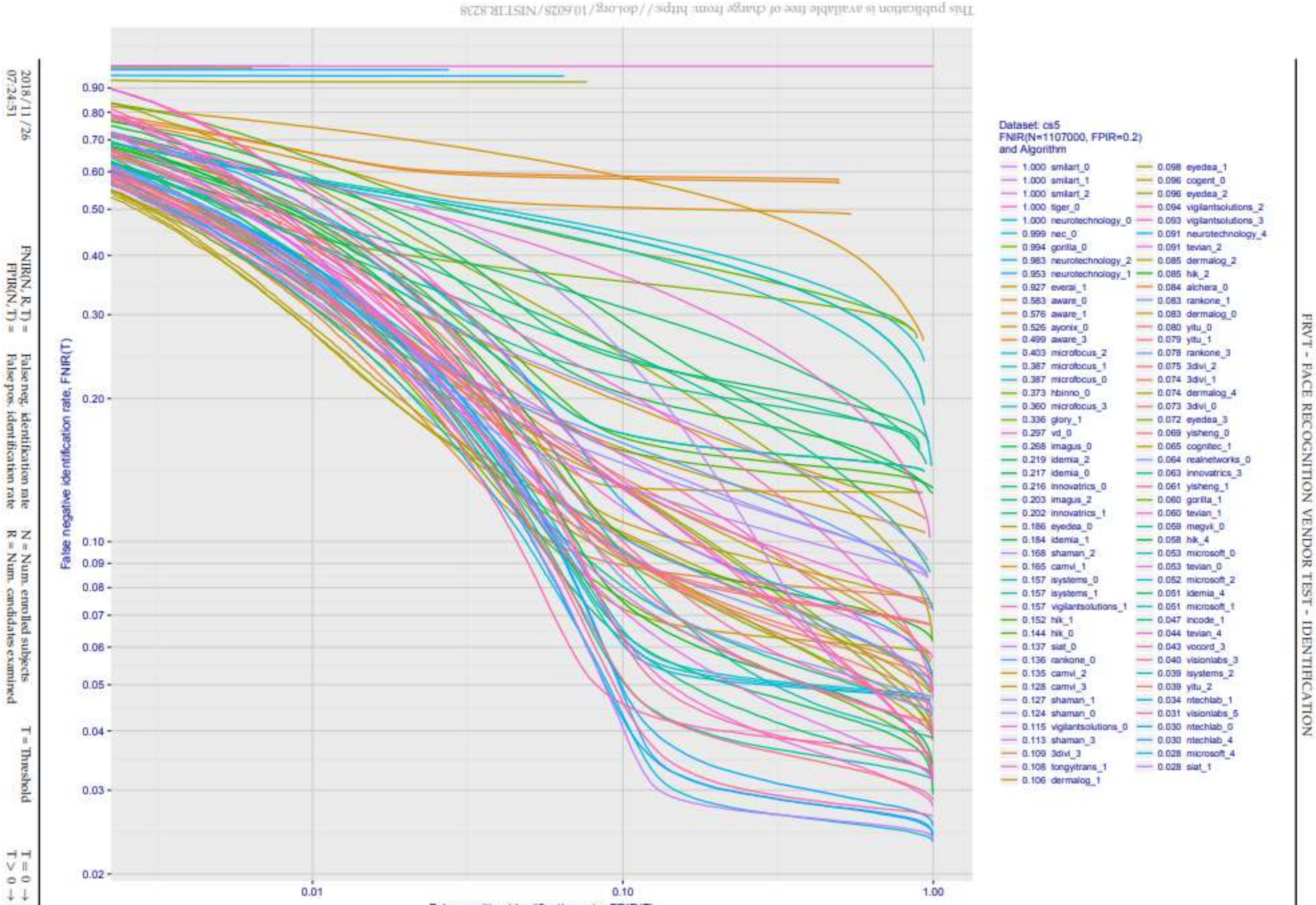


Figure 99: **[Wild Dataset] Identification miss rates vs. false positive rates.** *The figure shows accuracy of algorithms on wild images searched against wild images of*

Other examples of deep learning

- manipulation detection
- face morphing / deepfakes
- court findings finding irregularities

Nederlands Forensisch Instituut
Ministerie van Veiligheid en Justitie

Detecting face morphing in video and documents

30 August 2018

Ilias Batskos
Andrea Macarulla Rodriguez
Marissa Koopman
Zeno Geradts

EAFS 2018 Lyon

Contents

- Face Morphing
- Deepface
- Conclusions

Definition and problem statement

- Morph: A novel photo created by blending the photos of two different individuals

Face **a**

Morph of
a and **b**

Face **b**

- Problem: 2 individuals, one being the criminal and the other the accomplice, can use the same travel document

56

Security issues

Detection can be performed in 2 stages by humans, computers or both

- Issuing stage: Seemingly flawless morphs can be accepted as genuine photos, ~50% FAR when unaware and ~20% FAR when aware (Issuing officer) [1]
- ABC stage: Morphs can bypass Automatic Border Control (Face recognition systems),FaceVACS, FAR = 83.62% [2]

In real case scenarios of both stages, there are two photos that could be compared:

- Issuing stage : Photo from previous ID/Passport and presented new photo
- ABC stage : Passport photo and probe photo

[1] Robertson et al.(2017). Fraudulent ID using face morphs:Experiments on human and automatic recognition
[2] Raghavendra et al.(2017). Face Morphing Versus Face Averaging: Vulnerability and Detection

Example 2

- Genuine or morph ?

Probe

e-Pass

Example 3

- Genuine or morph ?

Probe

e-Pass

Example 4

- Genuine or morph ?

Probe

e-Pass

Example 5

- Genuine or morph ?

Probe

e-Pass

Example 7

- Genuine or morph ?

Probe

e-Pass

Problem/Vulnerability

The possibility to provide printed photographs creates a vulnerability in the system, which can be exploited by criminals with face morphing skills.

Solution?

- Live photograph enrollment at an authorized facility
- Secure Police web application for enrollment
- Additional biometric features (fingerprint, iris) on the chip
- Or better detection

State of the Art detection

- Texture based features using Local Binary Patterns (LBP), Binarised Statistical Image Features (BSIF), Image Gradient magnitude (IG) , Local Phase Quantitation (LPQ), blind/referenceless image spatial quality evaluator (BRISQUE)
- Double jpeg compression detection: Benford features , DCT coefficients of JPEG compressed face images
- Neural networks

Examples

SoA limitations

- Vulnerable to image processing and print & scan process
- Ghost artifacts, interpolation effects, morphing traces can be mitigated by image processing.(If the feature space of a detector is known it can be bypassed)
- Crucial pixel information is lost during print & scan, resulting to significantly increased errors of SoA detection methods

Ghost artifacts

Processed

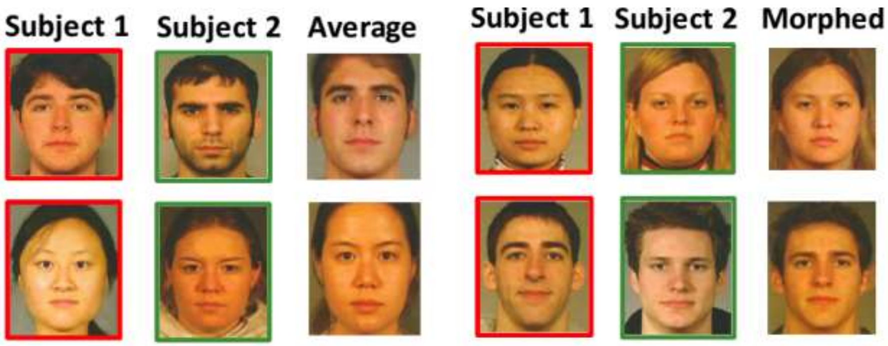
Printed and Scanned

Digital photo

Candidates selection

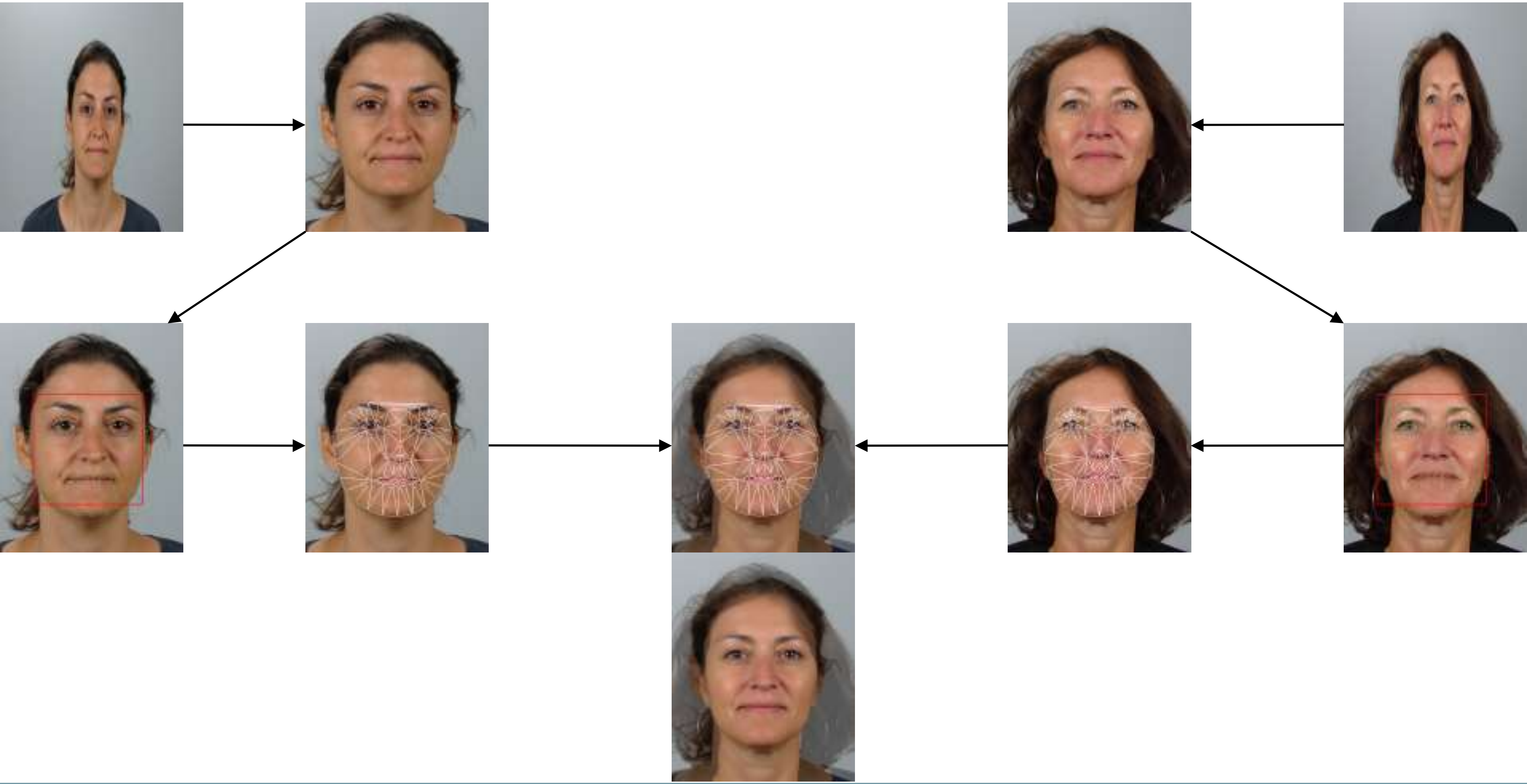
- Face encodings(128d) were extracted from all candidates.(Resnet34, Dlib)
- Using the Euclidean distance, a list of distance scores was calculated for each candidate.
- The person whose face encoding was closer to the candidate's encoding was selected for morphing

Automatically selected candidates
according to similarity



Morphing candidates found in the literature

Morphing pipeline



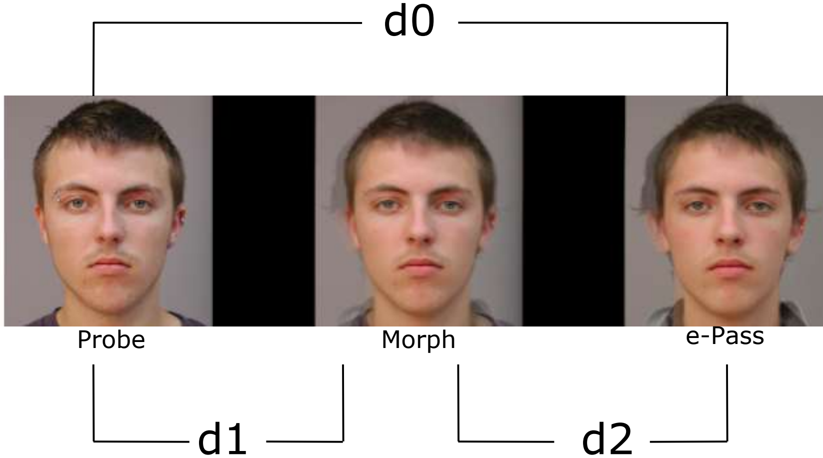
Experimental method

Criminal case

In a criminal scenario the photograph intended for e-Pass use already contains 50% of another individual, thus the morph will still contain 25% of that other individual.

After creating the morph which is intended as the e-pass photo the same process as before is followed

Note that the photo used for morphing was taken at a different time than the probe photo to simulate a real case scenario



Experimental method

- Traning set: H0=56, H1=55
- Testing set: H0=163, H1=359
- Total H0=219
- Total H1=414

Experimental results

- True Positives (Genuine photos classified as genuine) = 146
- False Negative (Genuine photos classified as morphs) = 17
- Total Positives (Genuine photos) = 163

- True Negatives (Morphs classified as morphs) = 353
- False Positives (Morphs classified as genuine) =6
- Total Negatives (Morphs)= 359

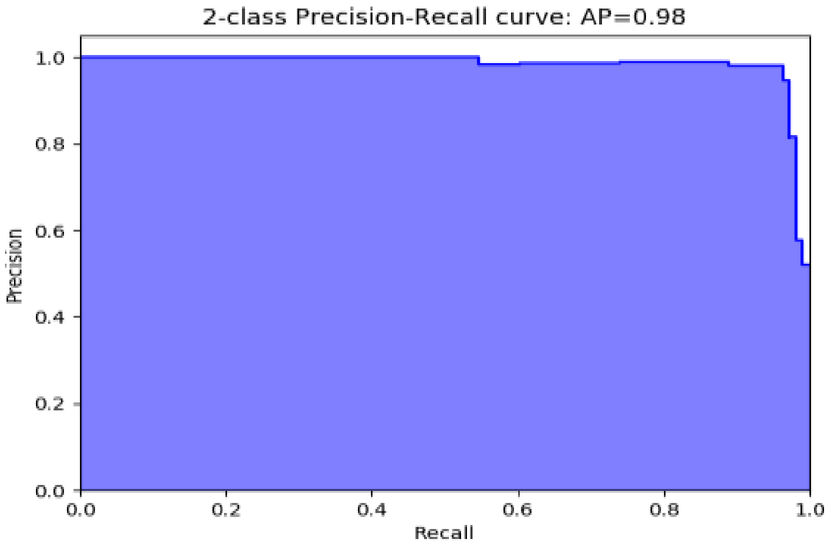
- True Positive Rate (Recall) = 0.8957055214723927
- True Negative Rate = 0.9832869080779945
- False Positive Rate (1-TNR) = 0.016713091922005572
- False Negative Rate (1-TPR) = 0.10429447852760736

- Precision = 0.9605263157894737
- False Discovery Rate (1- Precision) = 0.039473684210526314
-
- Accuracy (proportion of true results) = 0.9559386973180076
- F1 score = 0.926984126984127

Cross-validation

10-fold accuracy scores= [0.98245614, 1, 0.94736842, 0.96491228, 0.96428571, 0.92727273,
 0.96363636, 0.98181818, 0.94545455, 0.98181818]

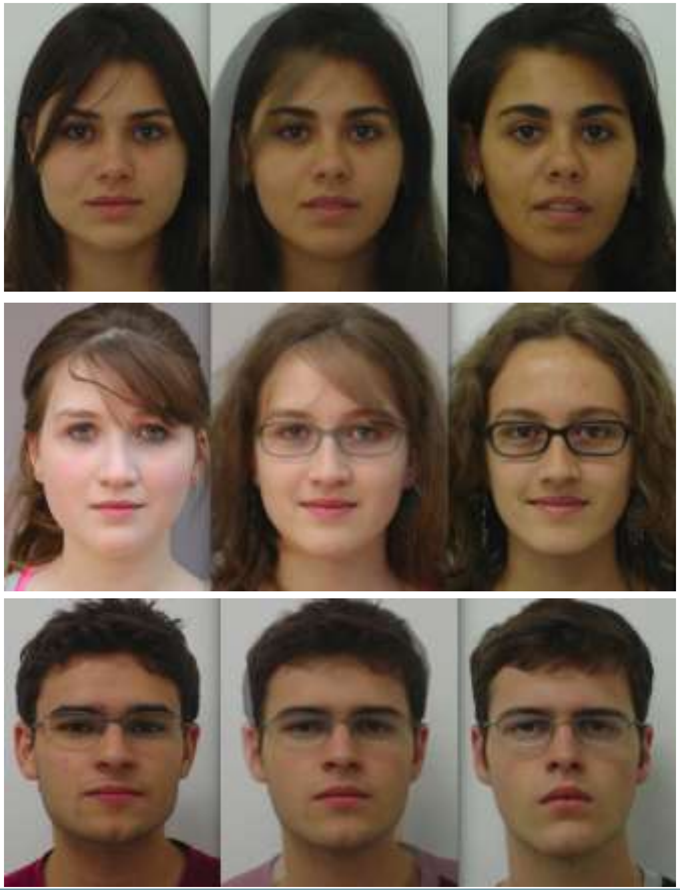
Accuracy: 0.97 (+/- 0.04)
Average precision-recall score: 0.98
auc= 0.989441195582



ROC curve

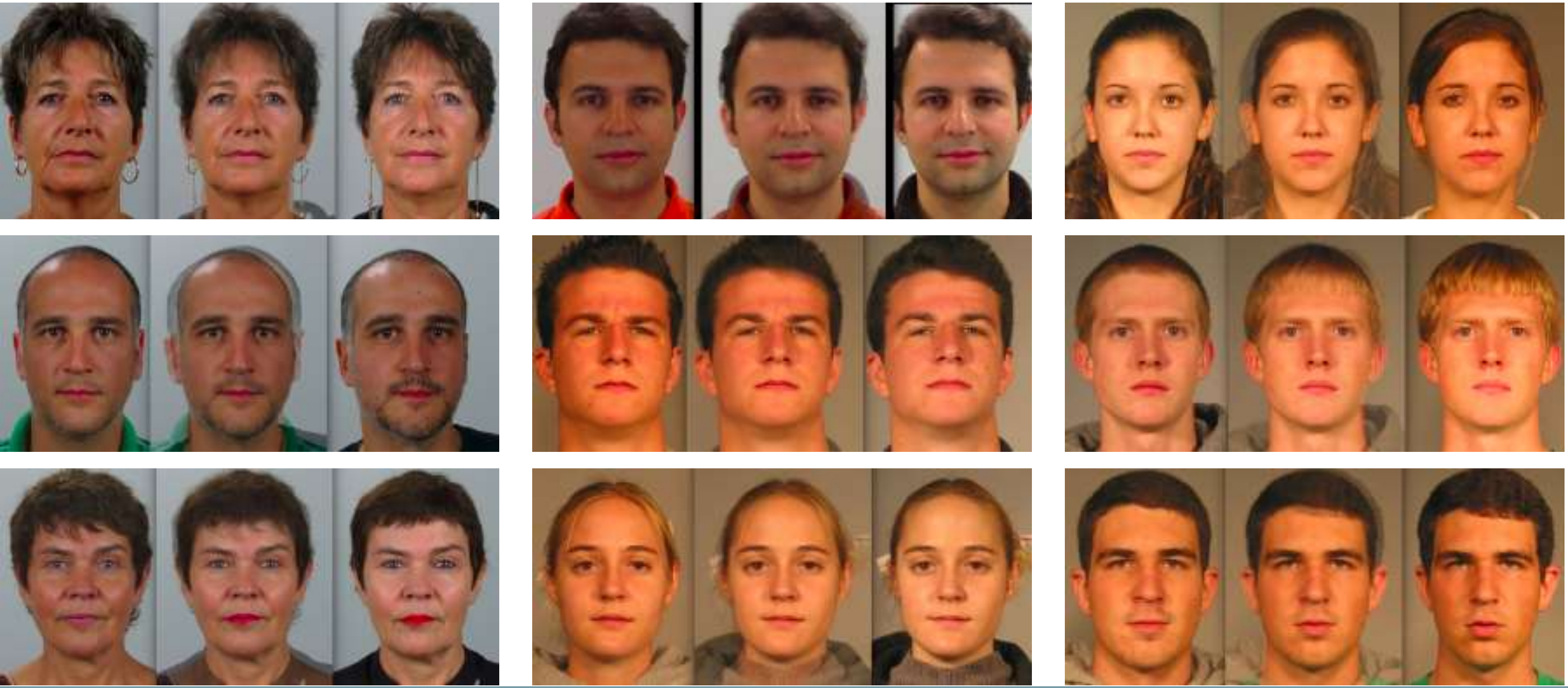
Classification examples

False positives



Classification examples

False negatives



Classification examples

True negatives






Limitations

- Sensitive to extreme photometric variations
- Sensitive to pose variations
- Sensitive to angle variations

Improvements

- Cluster faces by gender and ethnic characteristics for better morphing candidates
- Ratios of distances between pairs of landmarks to create the face descriptor
- Additional landmarks to improve morph quality
- Second morphing, additional discriminating power(?)

Conclusion

- The scores are unaffected by manual morphing(better quality than automatic) and Print and Scan process(?)

- Good experimental results indicate the effectiveness of the proposed method. The method could be implemented in parallel with SoA detection methods as an additional protection layer to counter cases of highly sophisticated and skilled criminals.

Contents

MOTHERBOARD

AI-Assisted Fake Porn Is Here and We're All Fucked

Someone used an algo *** ce of 'Wonder Woman' sta... porn video, and the implications are terrifying.**

THE VERGE TECH SCIENCE CULTURE CARS REVIEWS LONGFORM VIDEO MORE

WEB \ TECH \ ARTIFICIAL INTELLIGENCE

Gfycat starts removing fake AI porn GIFs from its platform

Gfycat is taking a firm stance against permitting 'deepfakes,' or AI-generated fake pornography

By Nick Statt | @nickstatt | Jan 31, 2018, 4:01pm EST

SHARE

Deepfakes: People are now swapping their friends' faces into porn

Deepfakes use facial recognition technology to superimpose faces on to porn stars.

By Kashmira Gander

January 29, 2018 15:30 GMT

SCIENCE CULTURE CARS REVIEWS LONGFORM VIDEO MORE

TECH \ ARTIFICIAL INTELLIGENCE \ SEX

Is it legal to swap someone's face into porn without consent?

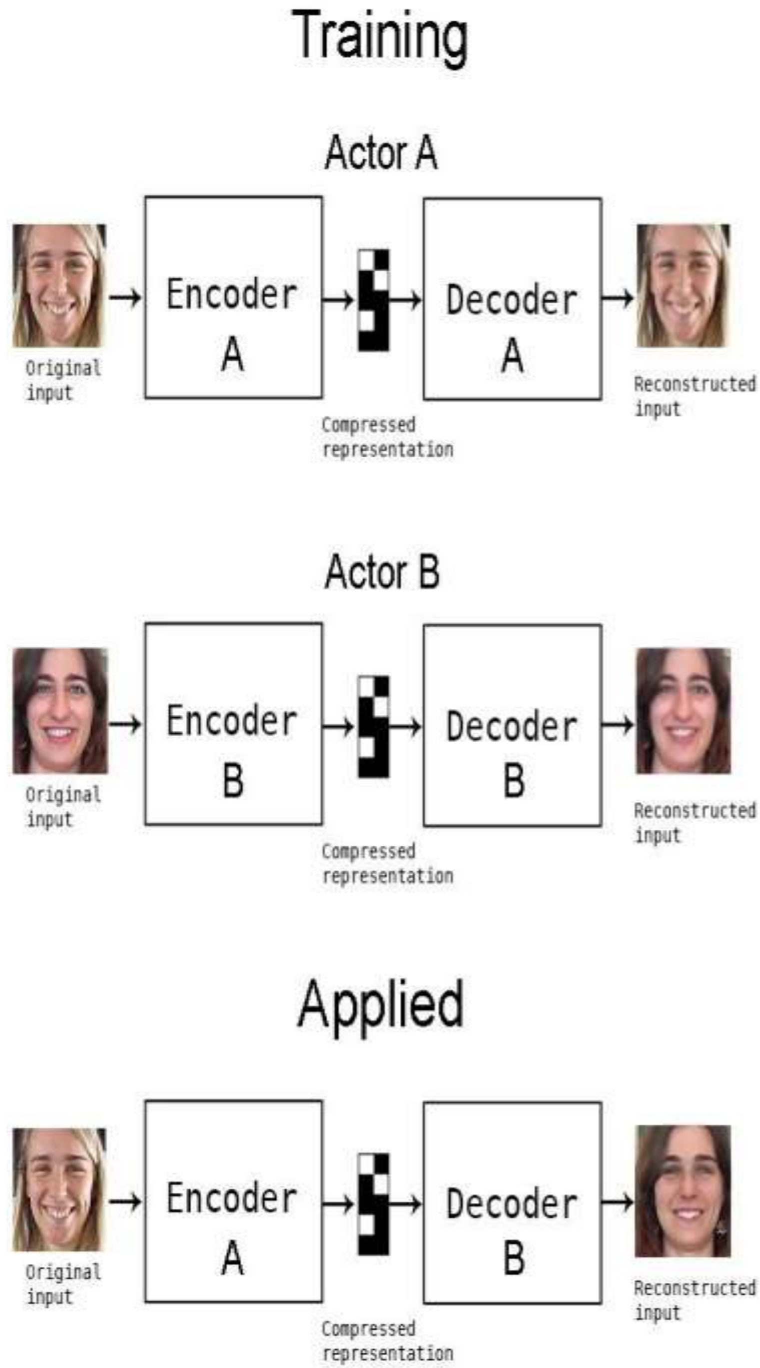
Yes, no, maybe

By Megan Farokhmanesh | @Megan_Nicolett | Jan 30, 2018, 2:30pm EST


DEEP FAKES

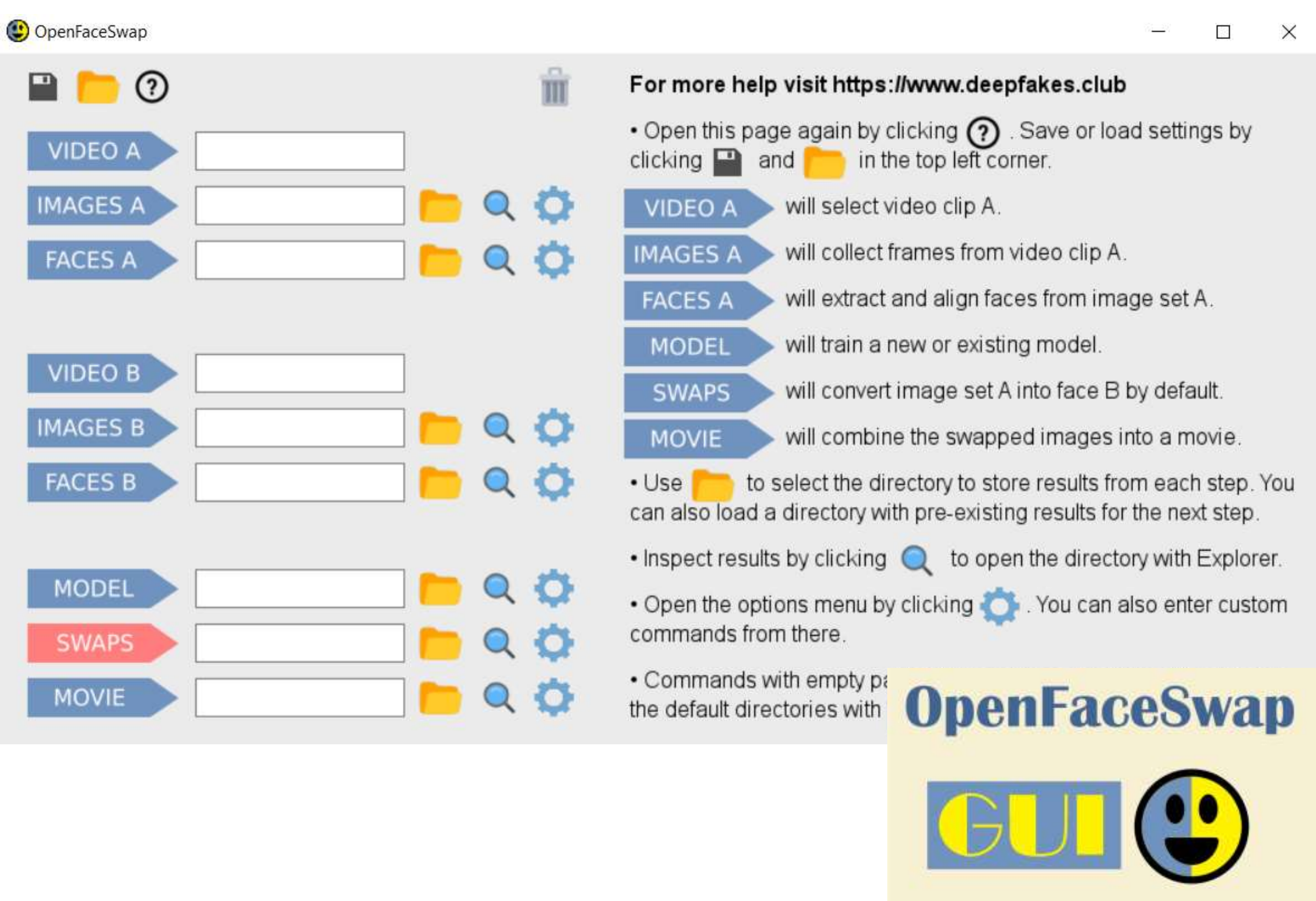
How are Deepfakes made?

- 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



Forensic Relevance

- Authenticate video evidence
- Technology widely accessible on the internet
- Easy to use for everyone through GUIs
- Authentication of videos is also important for journalism, social media, etc.


OpenFaceSwap
VIDEO A
IMAGES A
FACES A
VIDEO B
IMAGES B
FACES B
MODEL
SWAPS
MOVIE
For more help visit https://www.deepfakes.club
• Open this page again by clicking . Save or load settings by clicking and in the top left corner.
VIDEO A will select video clip A.
IMAGES A will collect frames from video clip A.
FACES A will extract and align faces from image set A.
MODEL will train a new or existing model.
SWAPS will convert image set A into face B by default.
MOVIE will combine the swapped images into a movie.
• Use to select the directory to store results from each step. You can also load a directory with pre-existing results for the next step.
• Inspect results by clicking to open the directory with Explorer.
• Open the options menu by clicking . You can also enter custom commands from there.
• Commands with empty pa
the default directories with
OpenFaceSwap
GUI

Research question

Can Deepfakes be distinguished from authentic videos with the use of:

1. Convolutional Neural Networks (CNN)
2. Discrete Cosine Transformation (DCT) coefficients analysis
3. Photo Response Non Uniformity (PRNU) analysis

- A pioneering study into the detection of Deepfakes.

Dataset

- 16 Deepfakes, 10 authentic videos
- Average video length = 29 seconds
- Frames extracted with FFmpeg
- Three actors used to create dataset

FFMPEG

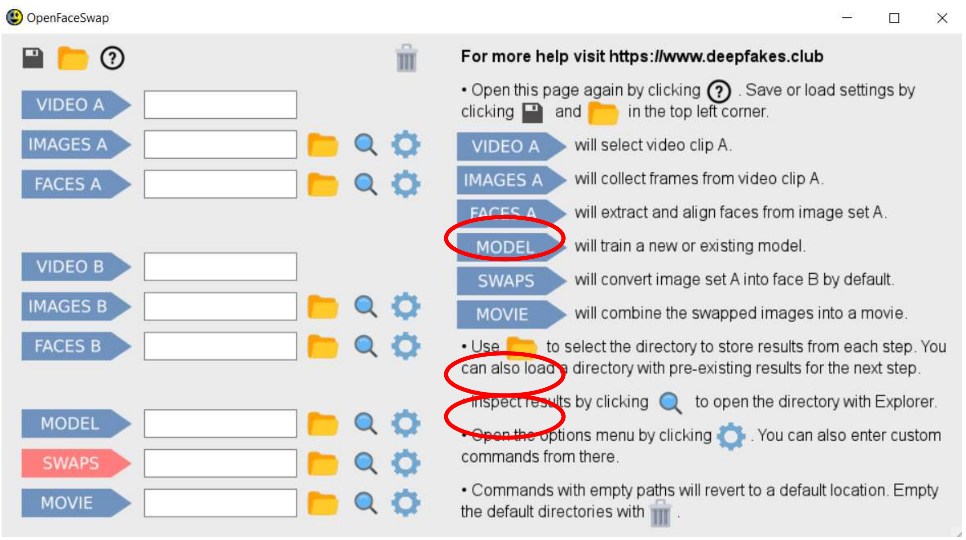


OpenFaceSwap

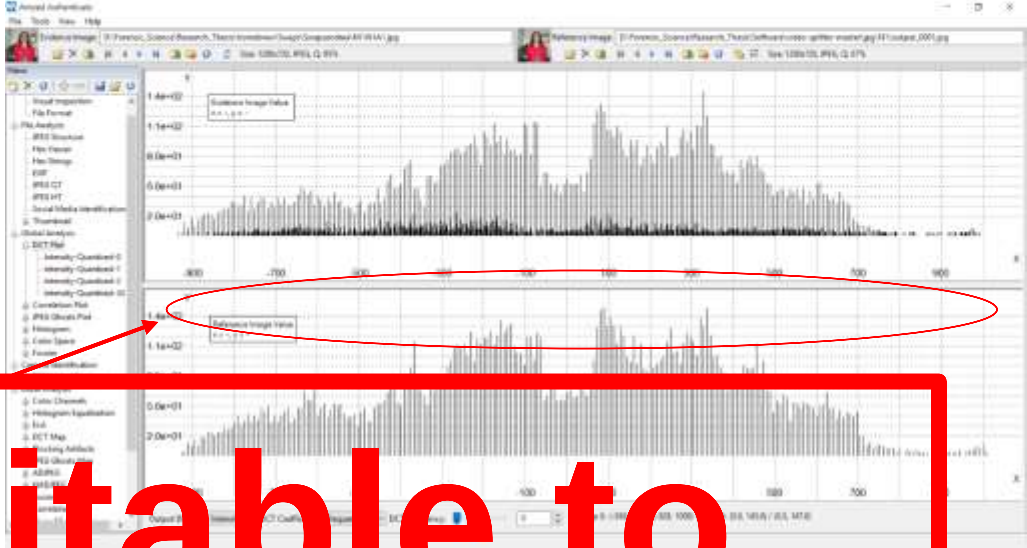
GUI

Discrete Cosine Transformation (DCT): Method

- DCT are used in JPEG compression - can leave traces
    - Can be used to see whether the file has been saved more than once
- Authentic frame vs Swap frame
- Authentic frame vs Deepfake movie frame (extracted by FFmpeg)



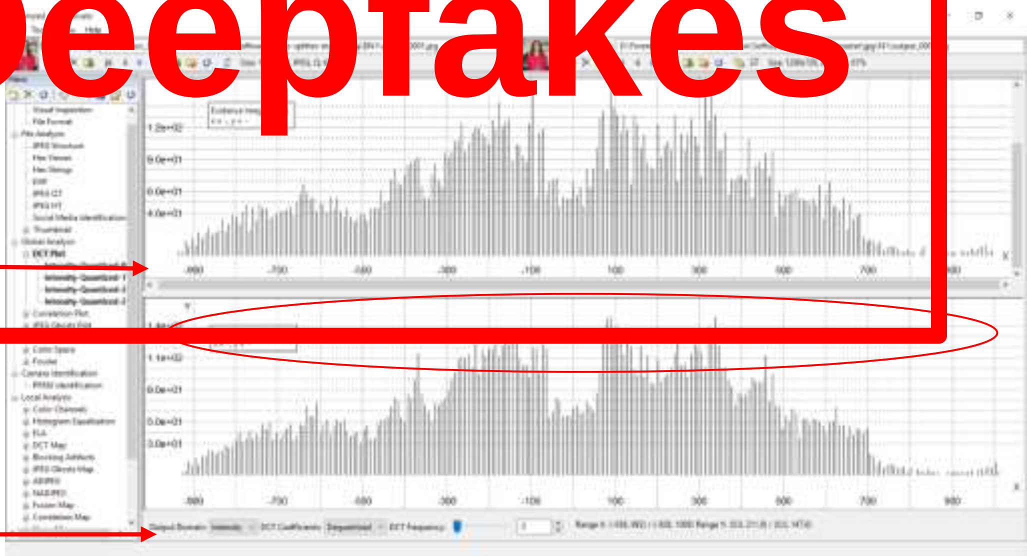
DCT: Results



Deepfake frame before video assembly

Authentic frame

Not suitable to detect Deepfakes



Deepfake frame extracted from video after assembly

No sign of double compression

Authentic frame

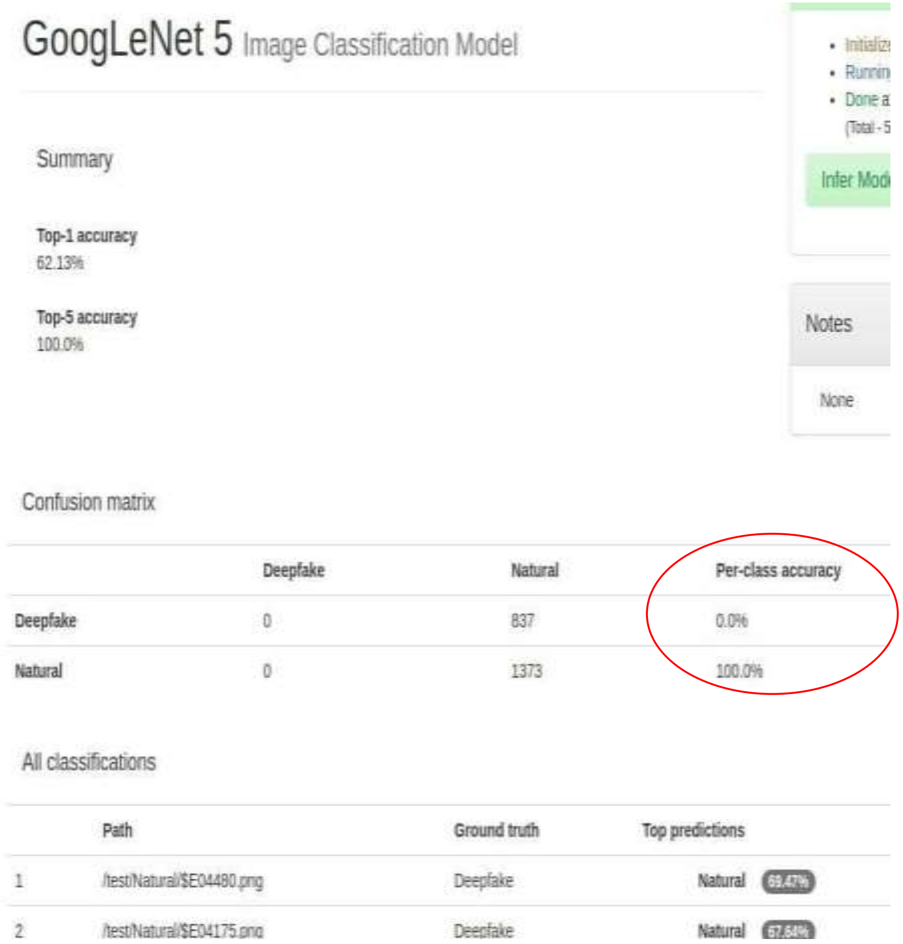
Convolutional Neural Network (CNN): Method

- CNN are AI which take an input and learn to eg. classify it, without a human telling them how to do so.
- CNN based on GoogLeNet
- Classify frames from dataset as ‘Natural’ or ‘Deepfake’
- 60/20/20
- Model trained for max. 100 epochs

NVIDIA DIGITS

CNN: Results

- Classified all frames as 'Natural' (authentic), or all as 'Deepfake'
- Changing regularisation methods had no effect.

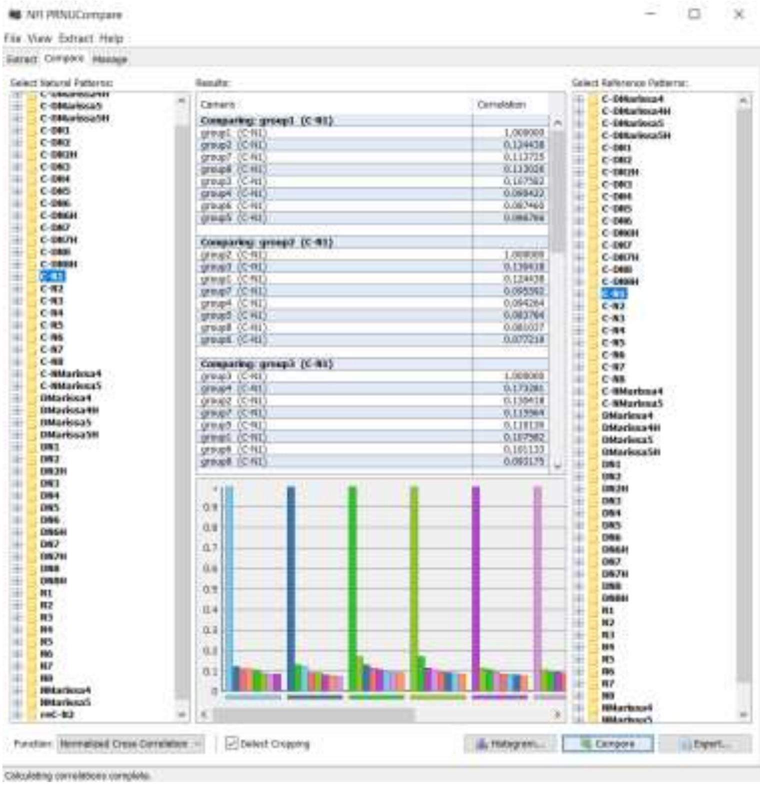


CNN: Conclusions

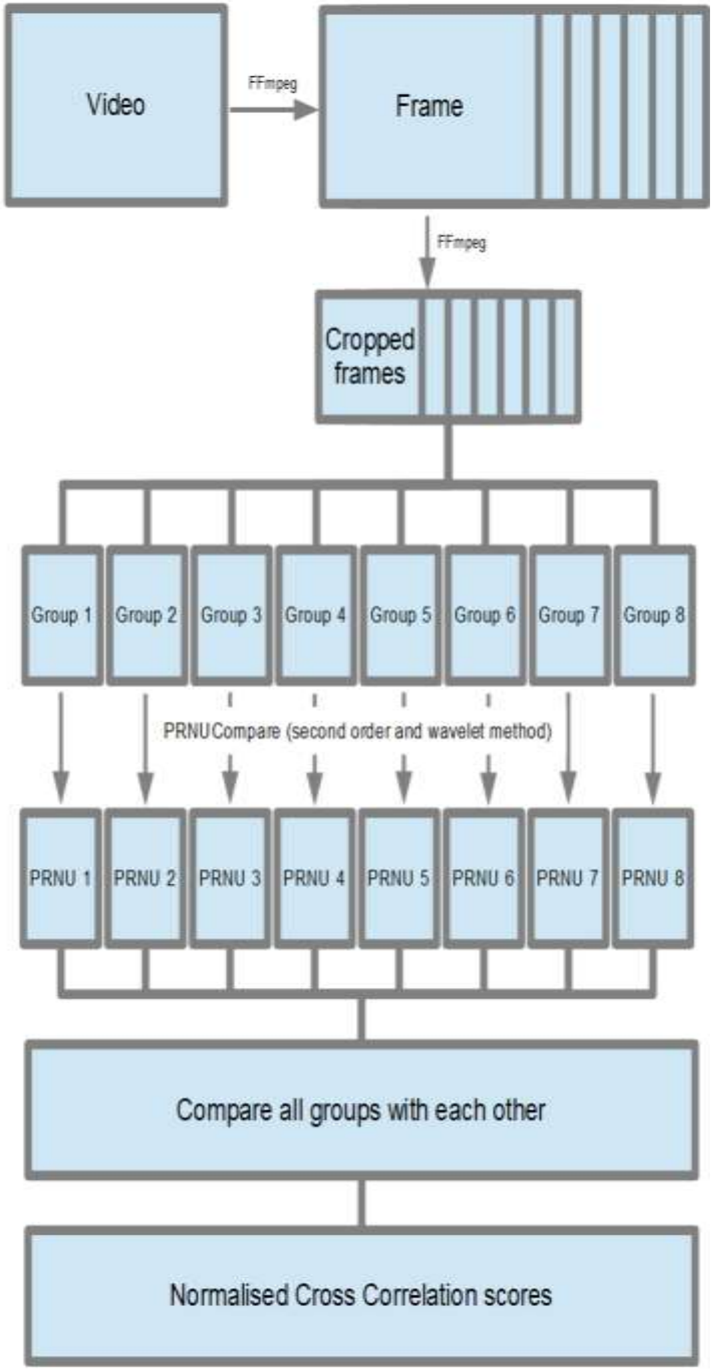
- Persistent overtraining
- Dataset with even number of Deepfake and Natural frames may improve learning
- Cannot conclude that CNNs are unsuitable
- General Adversarial Networks (GANs) may be more suitable.

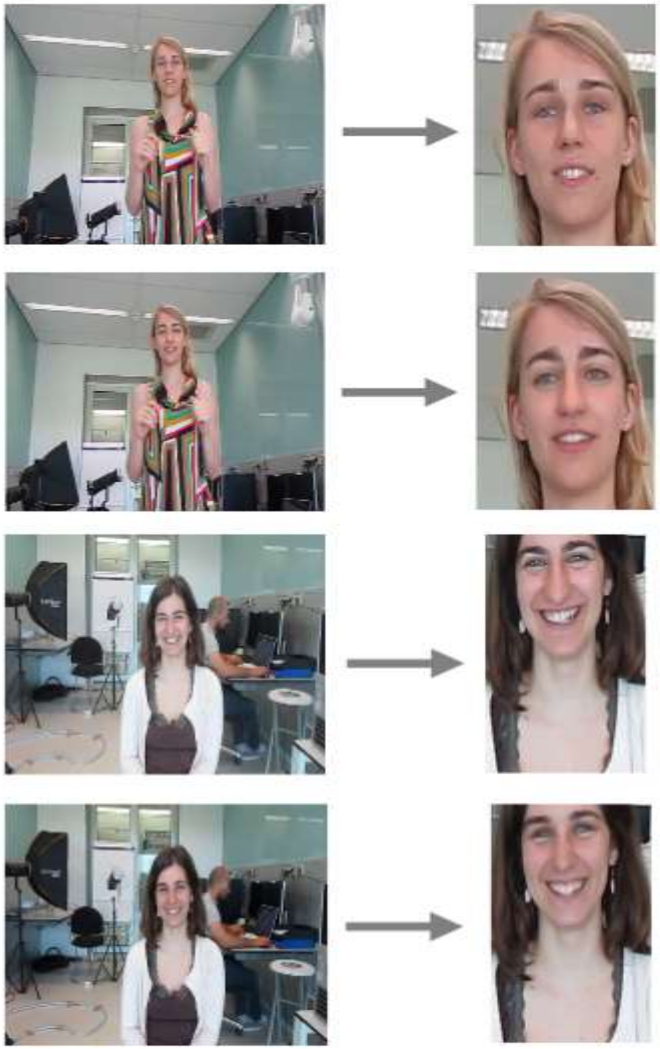
Photo Response Non Uniformity (PRNU) analysis: Method

- '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: 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

Limitations of study

- Dataset
  - Small
  - Imbalanced
  - One camera
- CNN
  - Only CNN based on GoogLeNet
  - Imbalanced dataset
- PRNU analysis
  - Cropping method not suitable for videos with large movements
  - All results from one camera’s PRNU pattern

Further Research

- PRNU
  - Confirm correlation
  - Likelihood ratios
  - Effect different cameras
  - Effect camera software (phone apps etc)
  - Is second order uncropped be sufficient
- CNN
  - General Adversarial Networks (GAN)
  - Balanced dataset

Thank you for your time

Questions zeno@holmes.nl / andrea@holmes.nl

Discussion

Rafferty said: “Cost-cutting and outsourcing has put the administration of justice at risk … I don’t think it’s bad faith by the police. They have been under-resourced. They are swamped. In some of my cases it’s the police who have revealed material that’s helpful to the defence.”

Collie, the head of Discovery Forensics in London who mainly works for defendants, said: “The odds are stacked against the defence in many ways. We rarely get access to the actual piece of equipment. In the past I could go to the police station and see a phone or a computer and physically check it’s the right piece. Now everything comes prepackaged and is handed over on a hard drive or USB stick.”

Collapsed rape prosecutions

December: Liam Allan
The first case to be abandoned due to the failure by police to hand over crucial digital evidence was that of London student Liam Allan, 22, in December. Allan was charged with 12 counts of rape and sexual assault, but his trial was abandoned after police were ordered to hand over phone records that should have already been provided to the defence.

December: Isaac Itiary
Shortly before Christmas, an alleged child rapist, Isaac Itiary, 25, was cleared at Inner London crown court when the prosecution offered no evidence. Material recovered from the phone of the complainant by police was only handed over to defence lawyers shortly before it was due to come to trial.

January: Oliver Mears
In January, Oliver Mears, 19, a student at Oxford University, was

Challenges

- Explain Deep Learning in court
- Bias in Model
- Training of users

Shielding

- Anti forensic software
- Encryption
- Darknets
- crypto currencies

Questions