ACM Multimedia 2020

Deep Learning for Privacy in Multimedia


Andrea Cavallaro


Ali Shahin Shamsabadi


Mohammad Malekzadeh

The Alan Turing Institute

CIS centre for intelligent sensing

Queen Mary University of London

Privacy threats

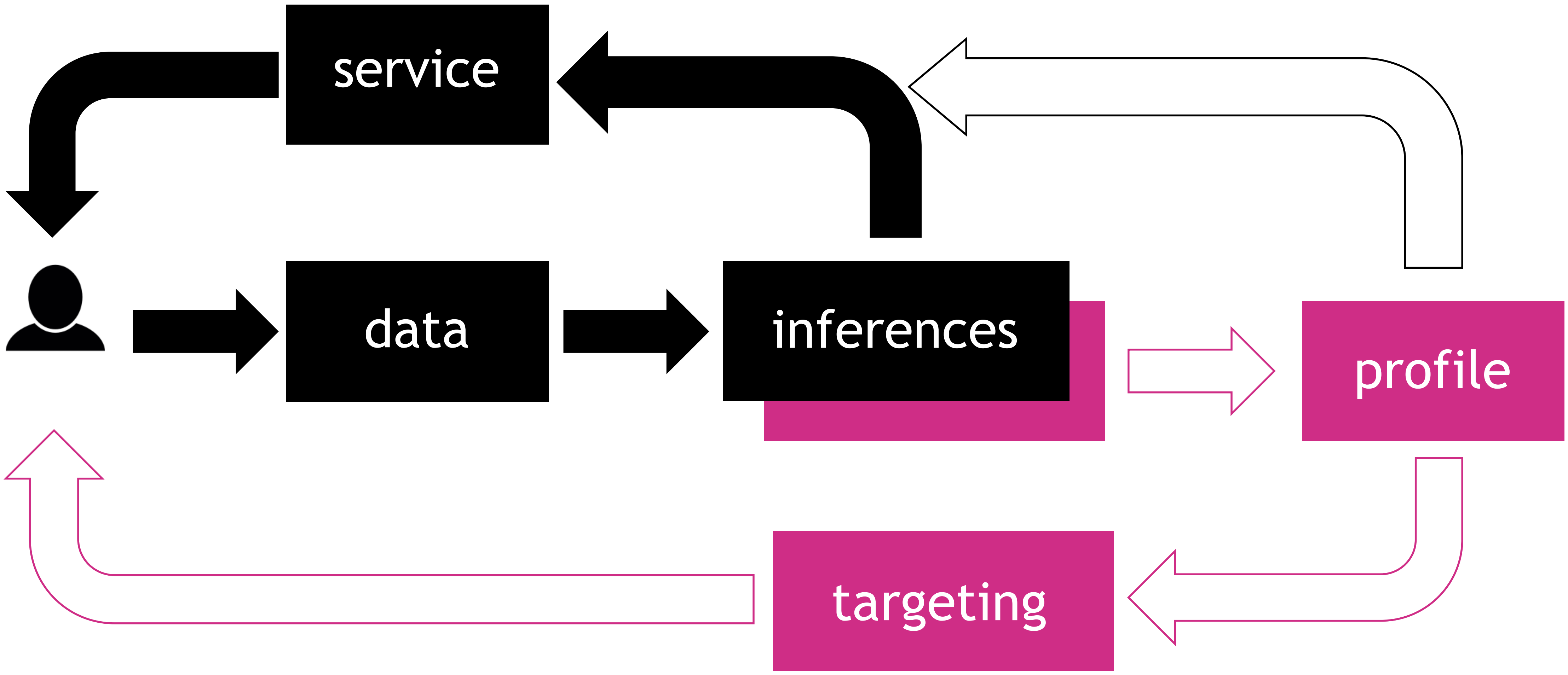
whom or what should information be protected from?

Privacy protection

from unwanted, automatic inferences (AI-powered services)

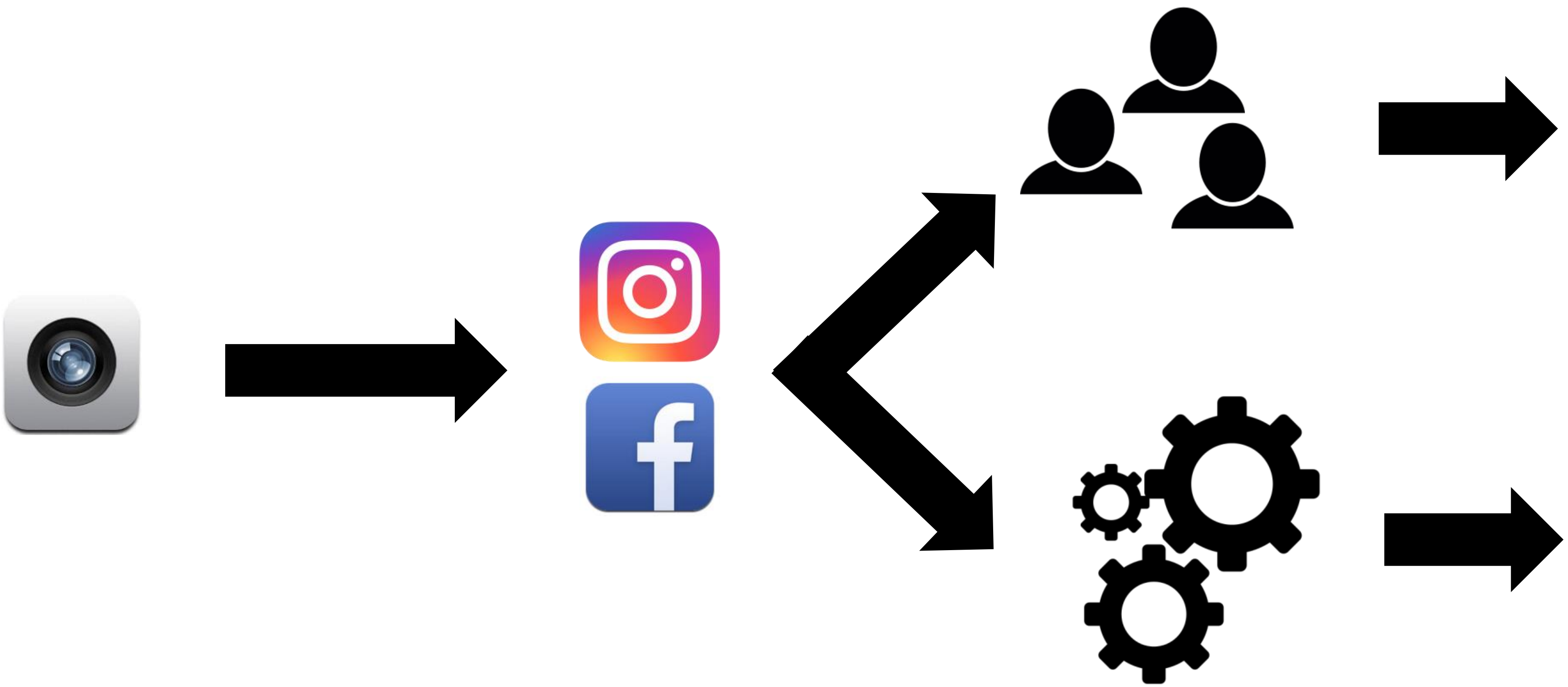
Tools to control the information we share

software distributed as open source



a.cavallaro@qmul.ac.uk

Inferences on images shared on social media

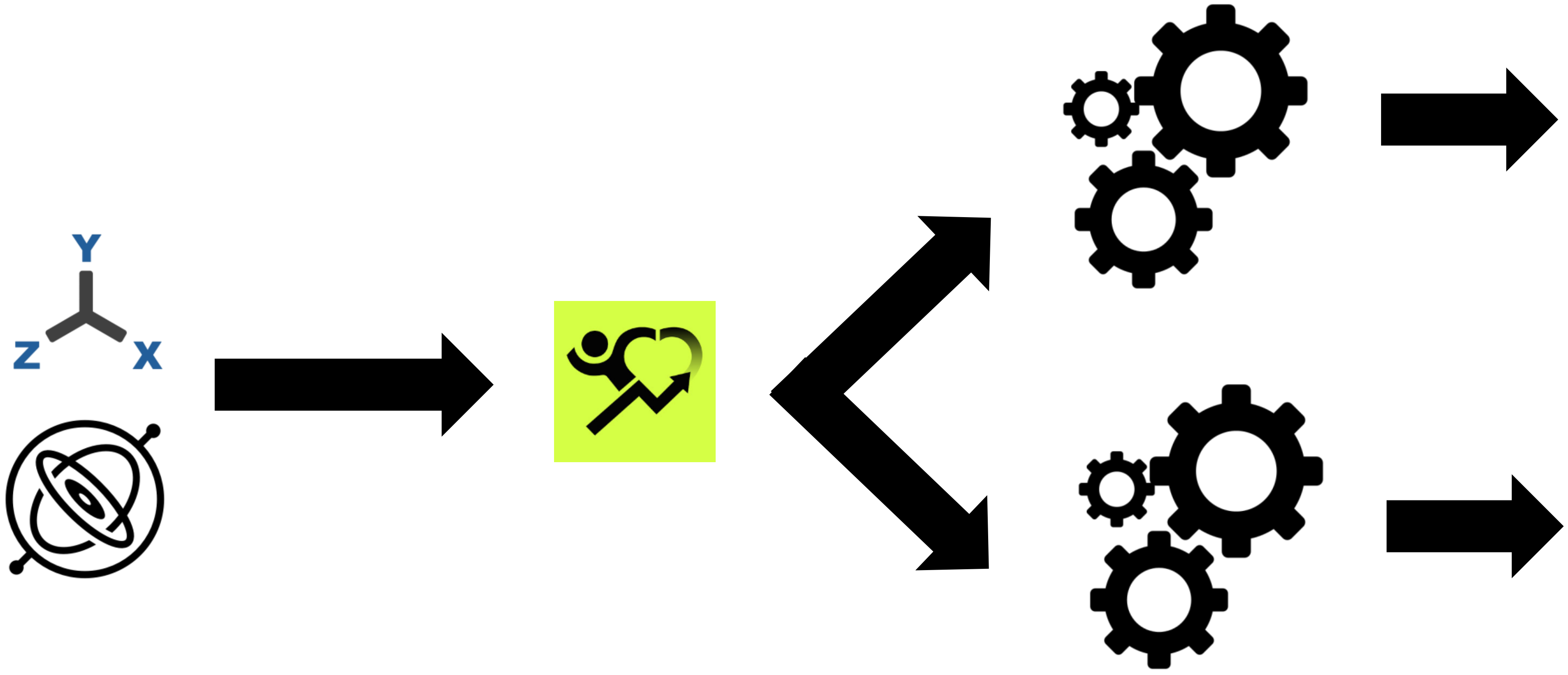


- scene depicted
- presence of people
- their relationship
- age, gender
- emotional state
- clothing style

informed consent?

a.cavallaro@qmul.ac.uk

Inferences on motion data shared with apps



- step count
- activities

- identity
- gender
- height
- weight
- how many pints
- how many cigarettes
- …

informed consent?

a.cavallaro@qmul.ac.uk

Definitions

Consent should be a "*freely given, specific, informed and unambiguous indication of the data subject's wishes*".

Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation), Apr. 2016

Personal data: "*any information relating to an identified or identifiable natural person ('data subject')*"

a.cavallaro@qmul.ac.uk

Definitions

informed consent

data minimization

"Personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed"

vs

data monetization

Definitions

Personal data

Personal information

a.cavallaro@qmul.ac.uk

Motivation

- We generate & share loads of data about ourselves
  - images we post in social media platforms
  - motion data from our wearables we provide to apps
  - audio data captured by smart speakers and digital assistants

- These data also reveal information we might wish to keep private

- Whom or what should this information be protected from?
  - individuals observing the data → access control procedures
  - algorithms extracting personal information → focus of this tutorial

Algorithmic inferences that reveal personal information

- Audio data
  - height and weight
  - emotional state
  - health conditions

Krauss at al. "*Inferring speakers' physical attributes from their voices*"

Trigeorgis et al. "*Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network*"

Schuller et al. "*The INTERSPEECH 2013 computational paralinguistics challenge: Social signals conflict emotion autism*"

- Motion data (wearables)
  - height and weight
  - level of activity
  - changes in behavioral patterns

Masuda and Maekawa, "*Estimating physical characteristics with body-worn accelerometers based on activity similarities*"

Zainudin at al. "*Monitoring daily fitness activity using accelerometer sensor fusion*"

Gruenerbl et al. "*Using smartphone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients*"

Privacy as a feature for body worn cameras
M.S. Cross, A. Cavallaro
Signal Processing Magazine, July 2020

Optimisation for/on the user

User as source and target of a process

Undesirable inferences

Defending from a 'machine'

This tutorial

**Methods that (learn to) alter data,
while maintaining their task-specific quality,
to prevent classifiers from inferring personal information
that is not necessary for the intended task**

Focus on
images & motion data

routinely analyzed by
algorithms for content
annotation and user profiling

fitness trackers infer activities,
calculate energy expenditures
or sleep quality scores, and …

First, let’s look back

The right to privacy

Recent inventions [..] call attention to the next step which must be taken for the protection of the person, and for securing to the individual [..] the right "to be let alone".

Instantaneous photographs [..] have invaded the sacred precincts of private and domestic life; and numerous [..] devices threaten to make good the prediction that "what is whispered in the closet shall be proclaimed from the house-tops. "

Warren and Brandeis, "**The Right to Privacy**"
Harvard Law Review, Vol. IV, No. 5, December 15, **1890**

Privacy

- Privacy
  - *"a state in which one is not observed or disturbed by other people"*
  - *"the state of being free from public attention"*
  - *"the right to select what personal information about me is known to what people"*

- What is new about privacy and multimedia data?

- Who should care about privacy?

- What are the right questions to ask about privacy?

a.cavallaro@qmul.ac.uk

What is the problem today?

- (Personal) data aggregation across systems, inferences from multimedia data
- Internet expanding into the physical world
- Dataveillance
  - the practice of monitoring digital data relating to personal details or online activities
- Difficulty in conceptualising privacy problems
- Private content: direct and indirect

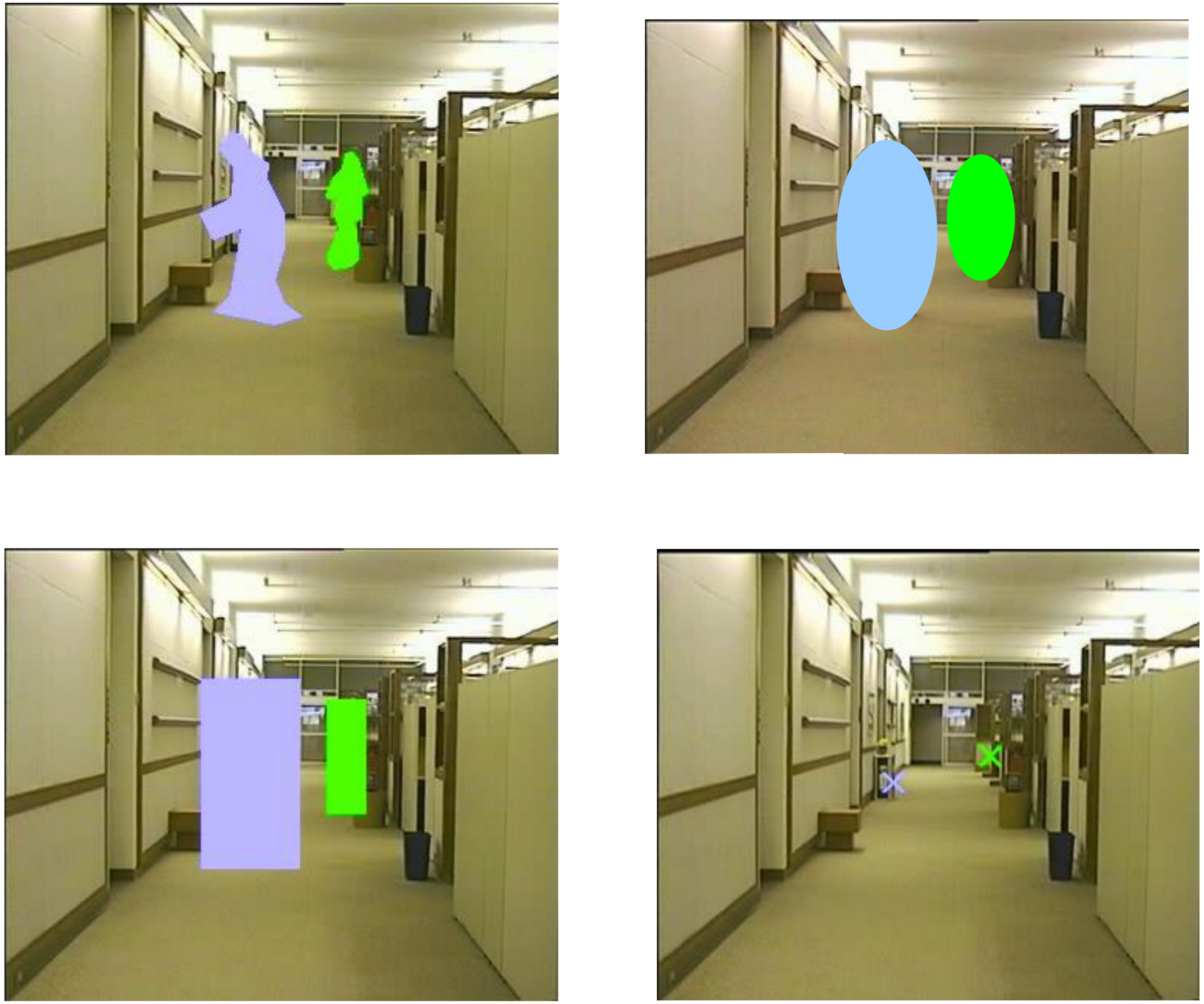
Direct vs indirect personal data

- Direct personal data
  - e.g. in images
    - face
    - body
    - licence plate

- How to reduce the risk of privacy loss?
  - generalisation
  - suppression → redactions, encryption, scrambling
    - during capture, transmission, storage, sharing, visualisation, access
    - multiple redactions
  - data management policies
    - selective archival
    - access control
    - duration

a.cavallaro@qmul.ac.uk

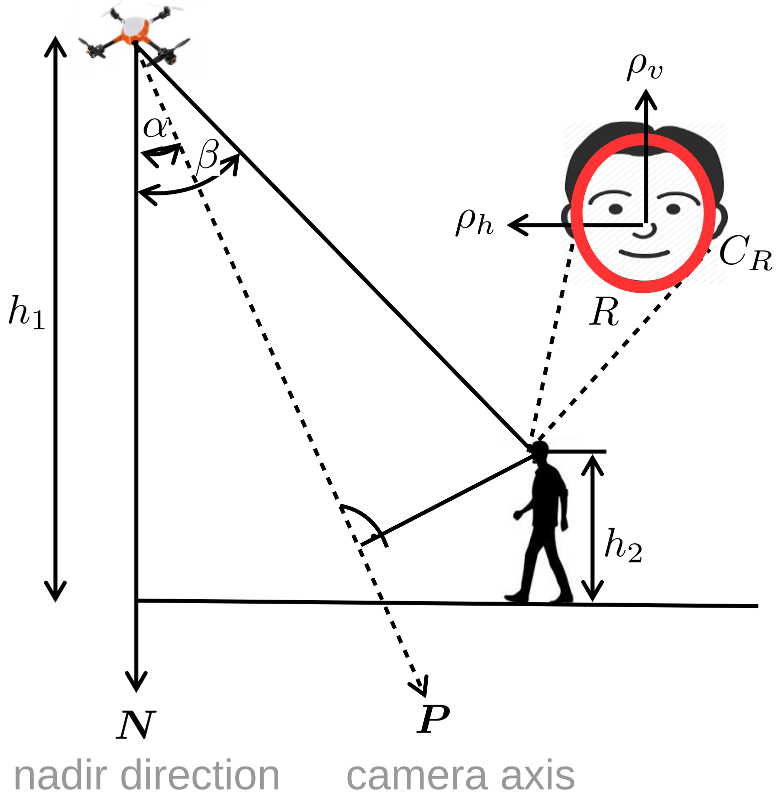
Privacy-related data: redaction



Adding privacy constraints to video-based applications
A. Cavallaro
Proc. of Eur. Workshop on the Integration of
Knowledge, Semantics and Digital Media Technology, 2004

Privacy-preserving drone videography

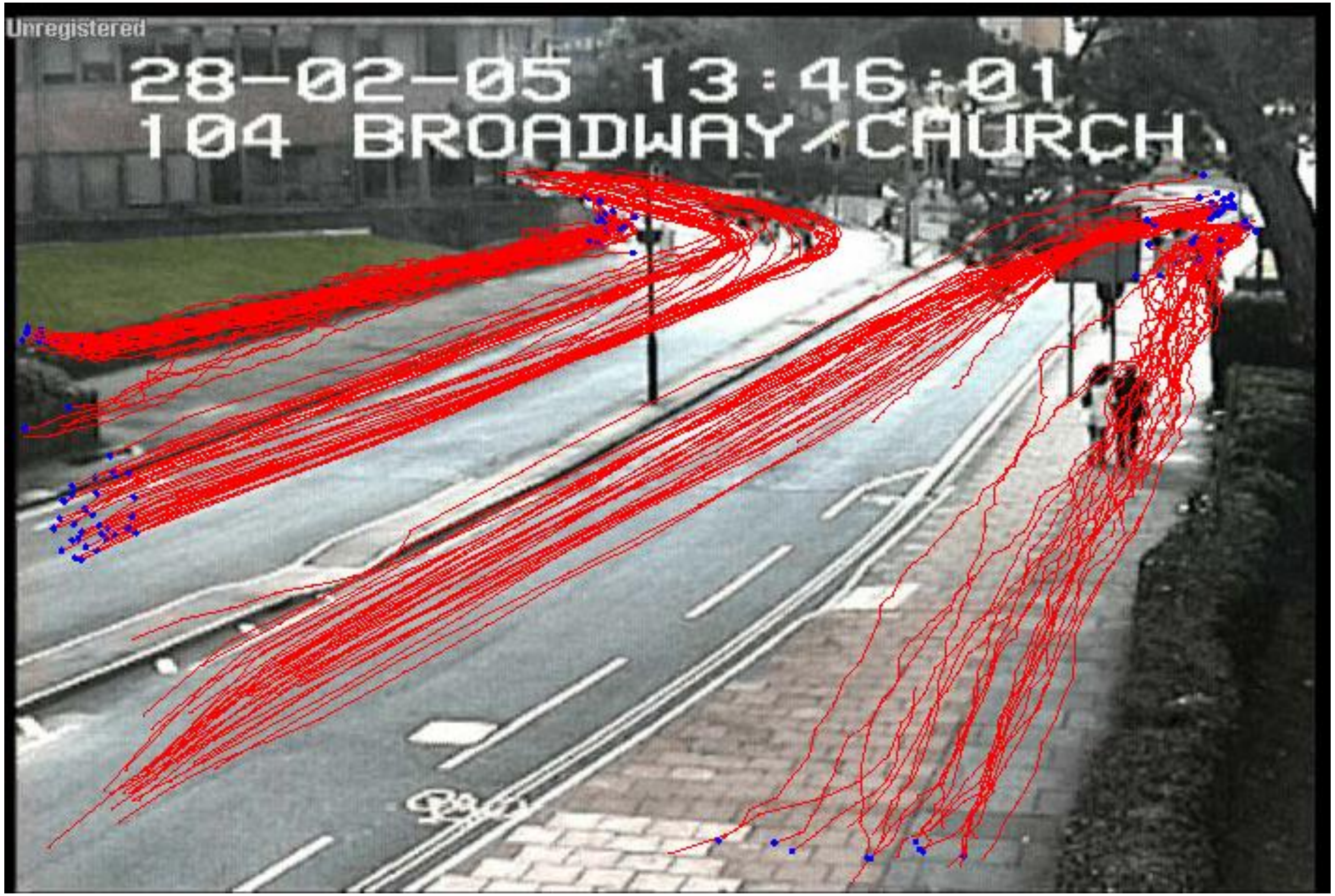
- To appropriately and selectively alter visual data
  - *objective*: to robustly protect the identity, while limiting spatio-temporal distortions
  - adaptively distorts the face appearance as a function of its resolution
  - locally changes parameters values to prevent estimation attacks

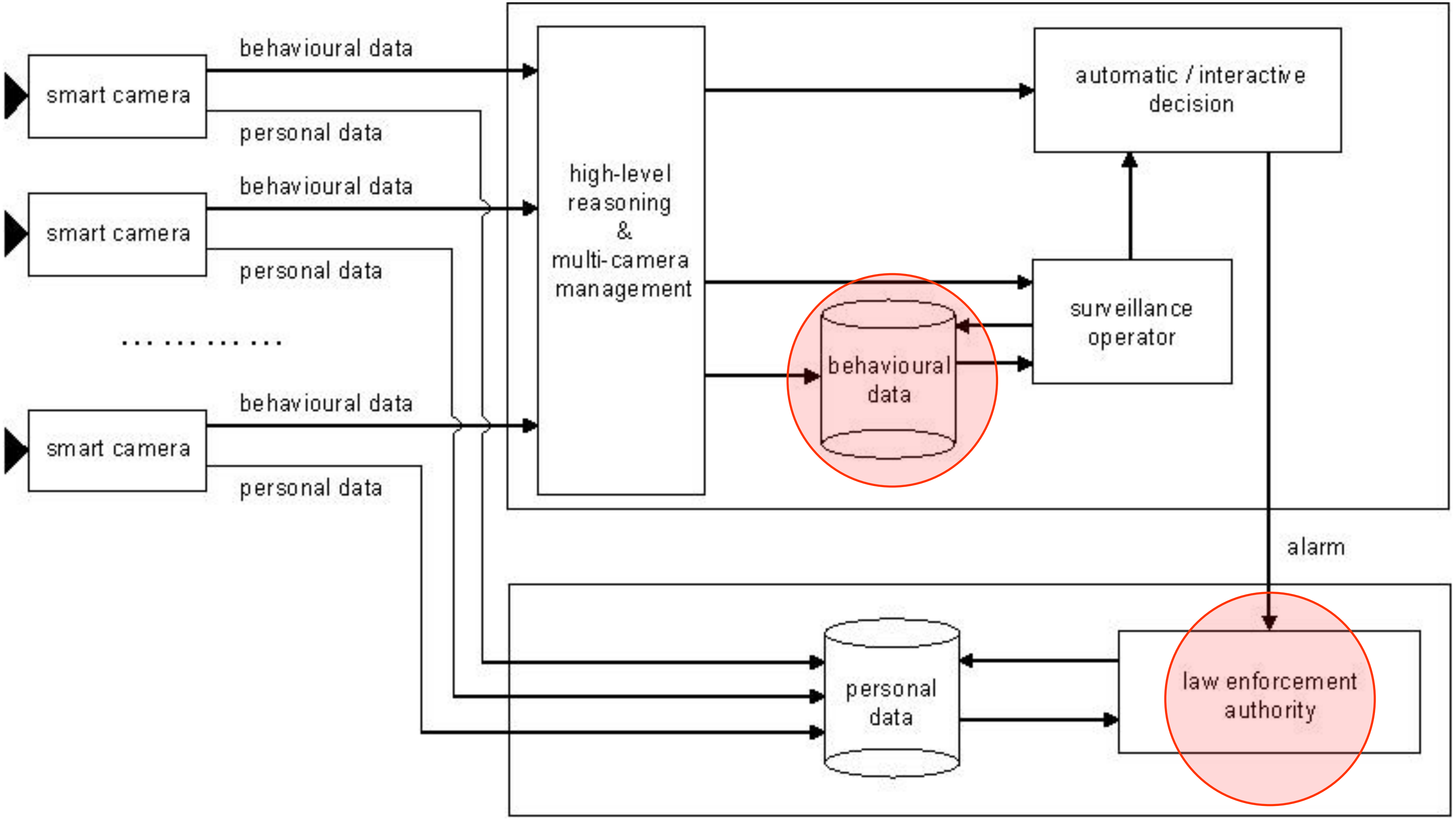


Temporally smooth privacy protected airborne videos
Sarwar, Cavallaro, Rinner
IROS 2018

A privacy-preserving filter for oblique face images based on adaptive hopping Gaussian mixtures
Sarwar, Rinner, Cavallaro
IEEE Access, October 2019

Behavioural data: example





Privacy in video surveillance
A. Cavallaro
IEEE Signal Processing Magazine, Vol. 24, Issue 2, March 2007

Wearable cameras

- Transition from purposive to passive data collection
  - expected shipment: 5+ million units in the next year
  - compounded annual growth rate of 16% in the next five years

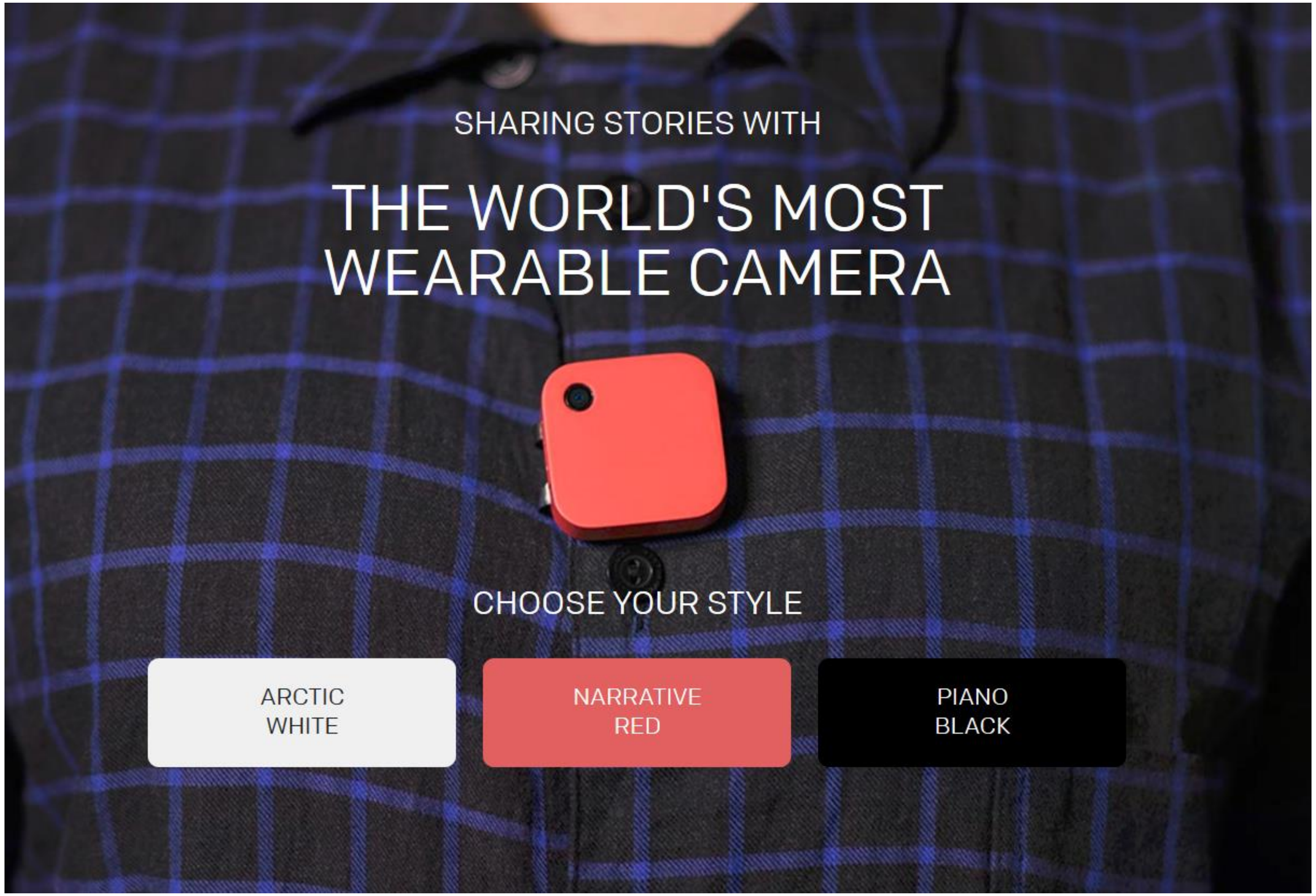
- Collecting personal information of
  - people being imaged
  - person wearing it!

- Always on … info on the person as well
  - location (identify bathroom/bedroom/computer screens)
  - social & affective visual computing
  - extraction of behavioural & health data

Privacy as a feature for body worn cameras
M.S. Cross, A. Cavallaro
Signal Processing Magazine, July 2020

Privacy-aware human activity recognition from a wearable camera
G. Abebe Tadesse, O. Bent, L. Marcenaro, K. Weldemariam, A. Cavallaro
Signal Processing Magazine, May 2020

Example: Narrative



http://getnarrative.com/

Example: smart glasses

https://about.fb.com/realitylabs/projectaria/

Direct vs indirect personal data

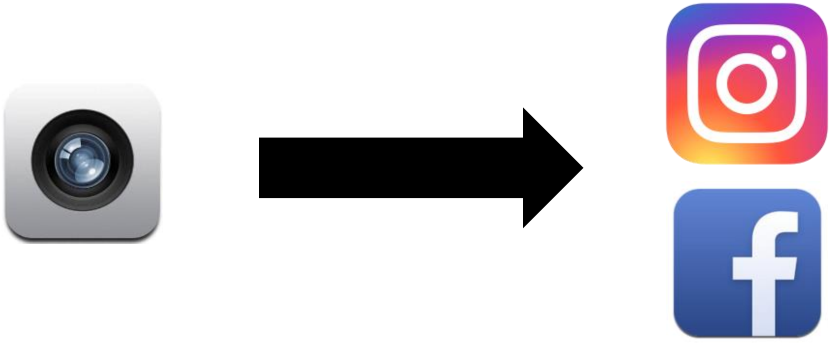
- Indirect personal data
  - location: coordinates or place (e.g. https://demos.algorithmia.com/classify-places)
  - time
  - activity

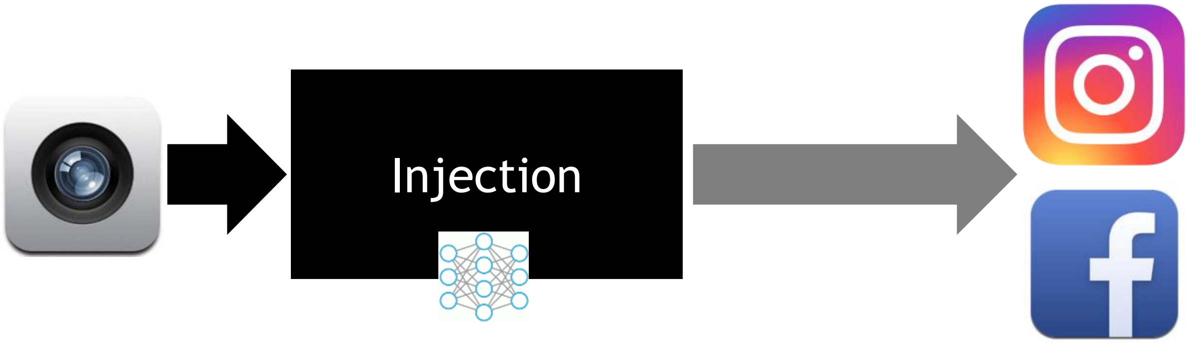
- Derived/secondary personal data (inferred from multimedia data)
  - gender, body shape, age (e.g. https://www.how-old.net)
  - expressions, emotions
  - personality (e.g. https://www.faception.com)
  - mental health

→ Novel privacy implications!

- How to reduce the risk of privacy loss?

Learning data manipulations for privacy protection





to protect the private content of images
(*that a user shares with other users*)
from undesirable automatic inferences

Part 2 of this tutorial

$$y = C(X)$$

training process

training data

a.cavallaro@qmul.ac.uk

Data manipulation: adversarial example

- Intentionally perturbed image that evades one or more classifiers

original (clean) image $\boldsymbol{X}$ $\boldsymbol{y = C(X)}$

class label

classifier

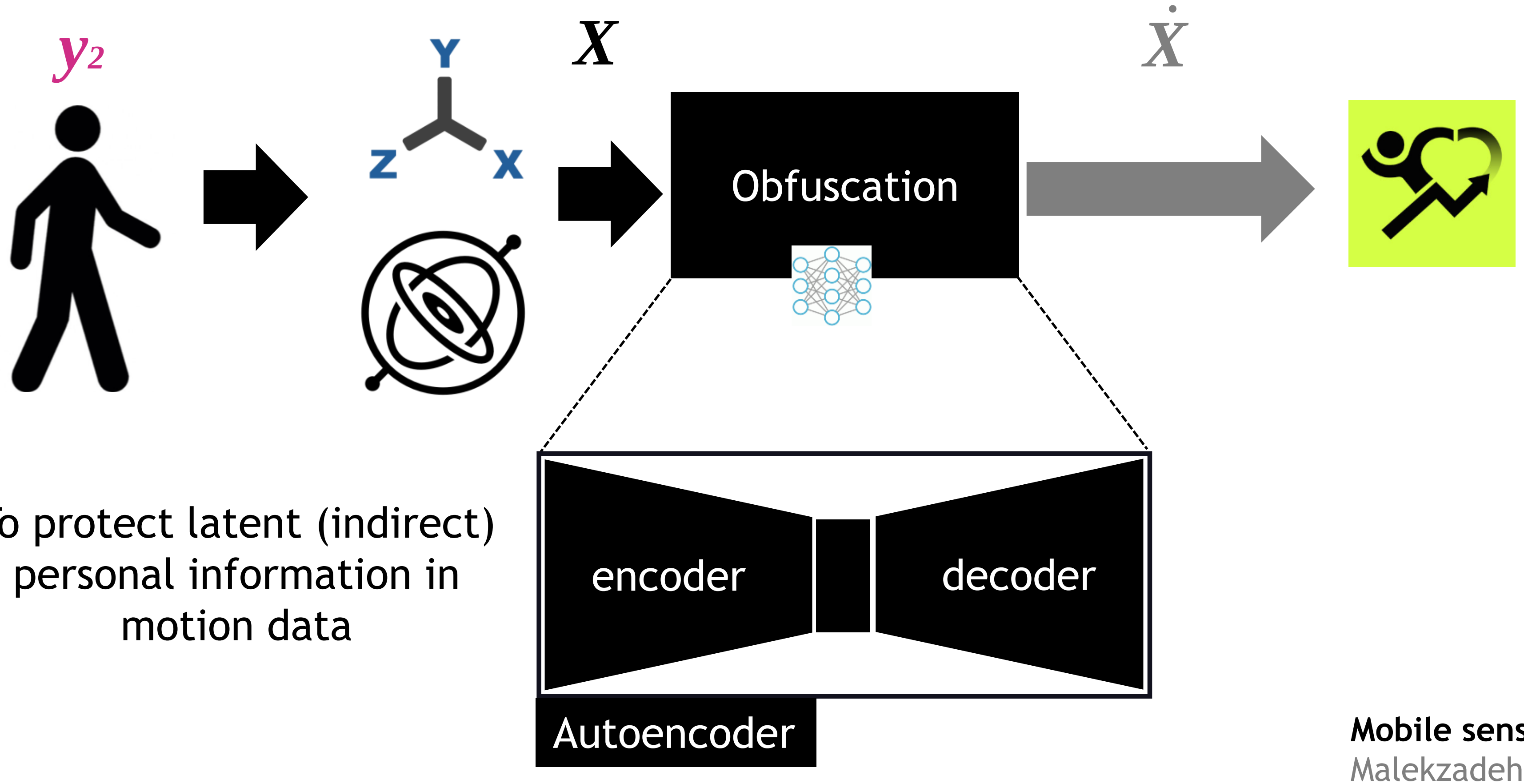
$\downarrow$?

perturbed (adversarial) image $\dot{\boldsymbol{X}}$ $\boldsymbol{y \neq C(\dot{X})}$

a.cavallaro@qmul.ac.uk

Obfuscate data by minimising user-identifiable attributes

Objective: to minimise the amount of information leakage from y_2 to $\dot{X}$

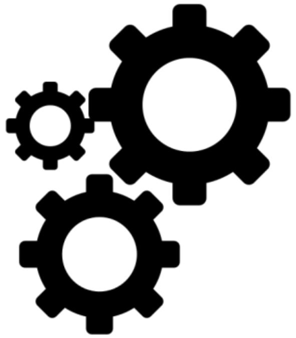


To protect latent (indirect) personal information in motion data

Part 3 of this tutorial

Mobile sensor data anonymization
Malekzadeh, Clegg, Cavallaro, Haddadi
ACM/IEEE IoTDI 2019

Example: MotionSense dataset



activities:
jogging
walking
w. downstairs
w. upstairs

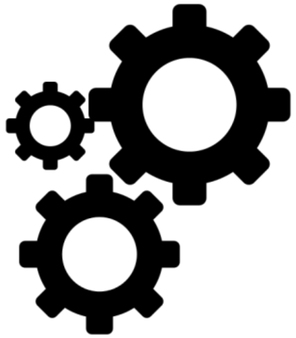
$y_1 = C_1(X)$

93%

$y_1 \approx C_1(\dot{X})$

93%

F1 score



label:
identity

$y_2 = C_2(X)$

96%

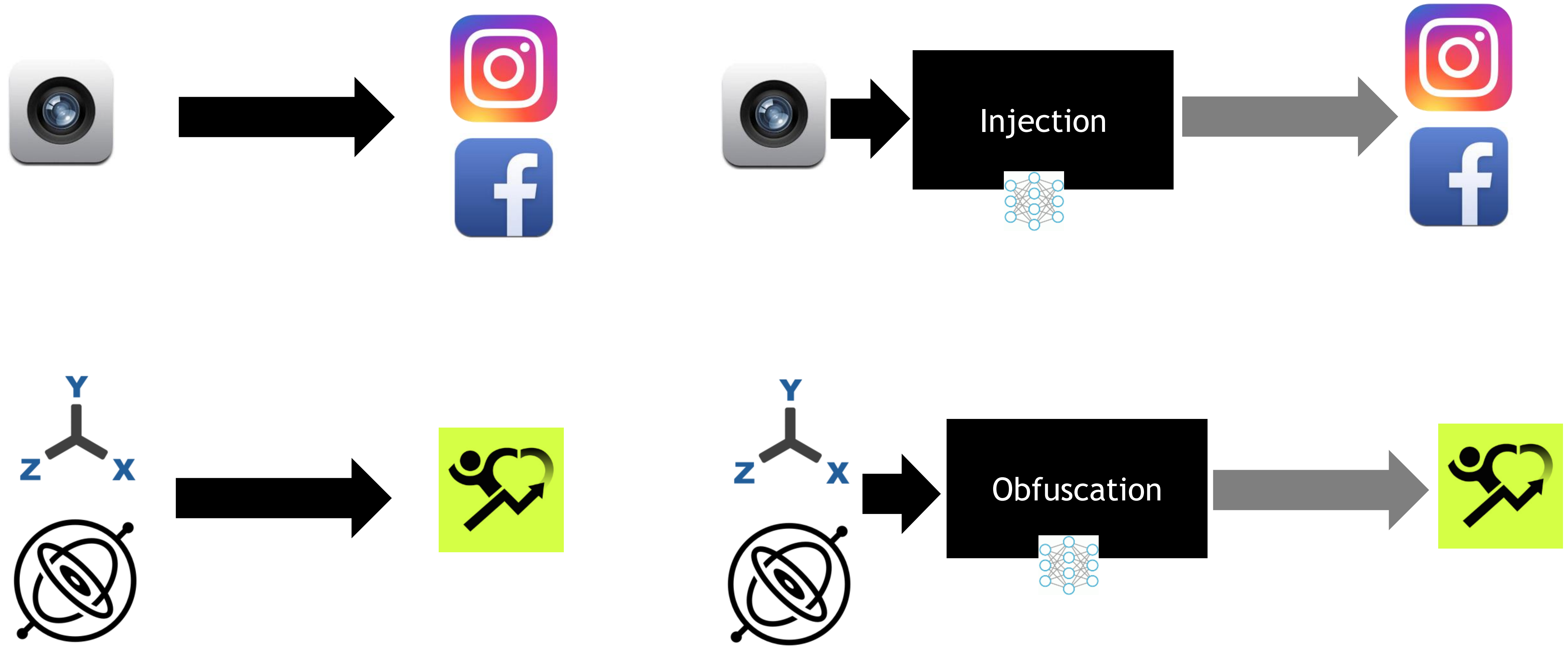
$y_2 \neq R(X) \approx C_2(\dot{X})$

7%

accuracy

Mobile sensor data anonymization
Malekzadeh, Clegg, Cavallaro, Haddadi
ACM/IEEE IoTDI 2019

Learning data manipulations for privacy protection



Privacy and utility preserving sensor-data transformations
M. Malekzadeh, R.G. Clegg, A. Cavallaro, H. Haddadi
Pervasive and Mobile Computing, Vol 63, Article 101132, 2020

Properties of a ‘*good*’ data manipulation

effectiveness

success rate in
evading a classifier

targeted

untargeted

robustness

success rate in
evading a classifier
despite defenses

transferability

success rate in
evading an unseen classifier

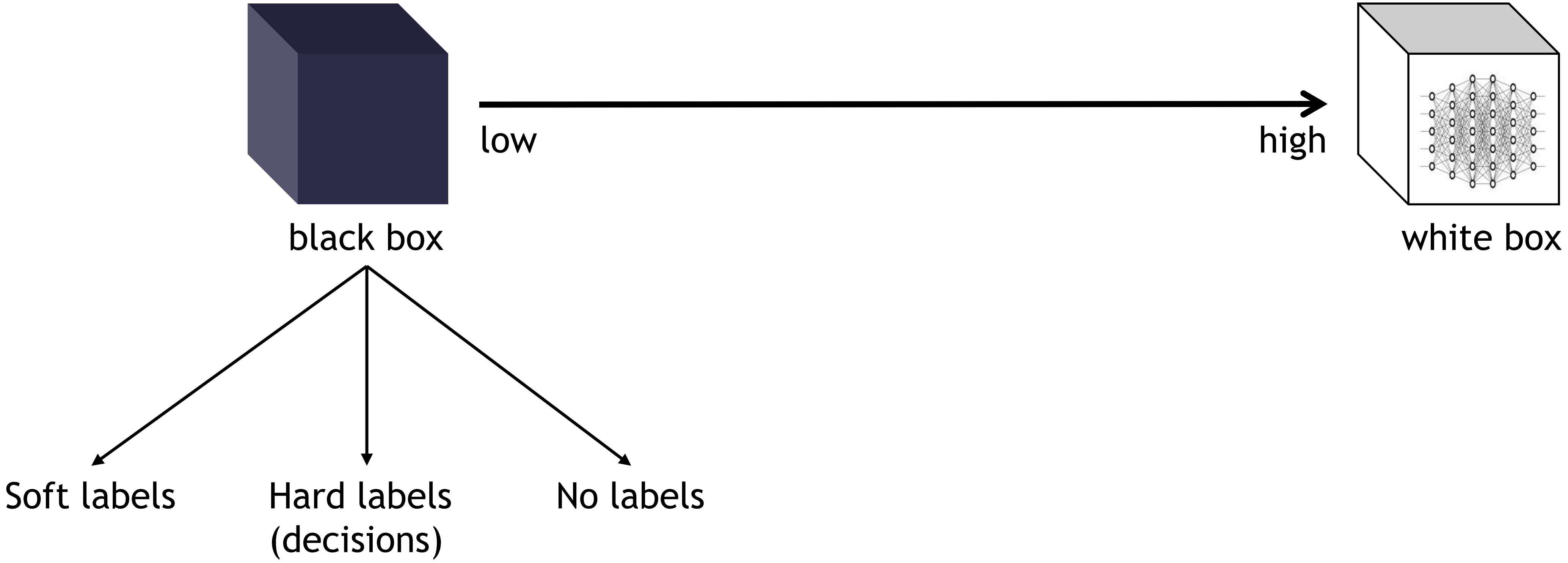
un-detectability

signal “perceived”
as an original one

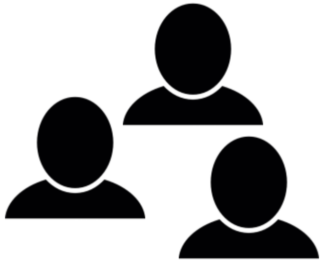
irreversibility

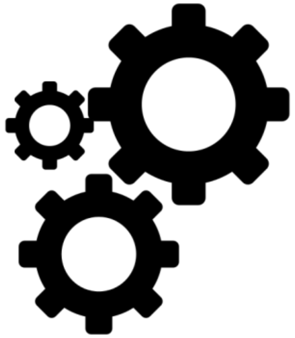
if transformation is detected,
low probability of
recovering the original class

Knowledge about the classifier (or its output)



Example





X

$y = C(X)$

class: *church*

83%

$\dot{X}$

$y \neq C(\dot{X})$

class: *zen garden*

99%

Scene Privacy Protection
Li, Shamsabadi, Sanchez-Matilla, Mazzon, Cavallaro
IEEE ICASSP 2019

Amount of perturbation to create Adversarial Examples

$$\dot{X} = X + \delta$$

constrained

- small perturbations (bounded)
- limited success with unseen classifiers
- vulnerable to defenses
  (high-frequency spatial perturbations are easily defeated by denoising algorithms)

unconstrained

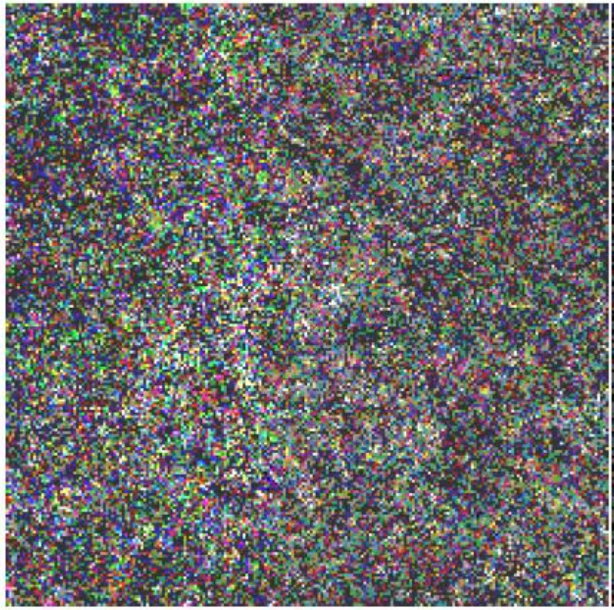
- more transferable to unseen classifiers
- more robust to defenses (less detectable)
- may severely degrade images
  (large perturbations are noticeable)

Constrained perturbations: examples

$\boldsymbol{x}$

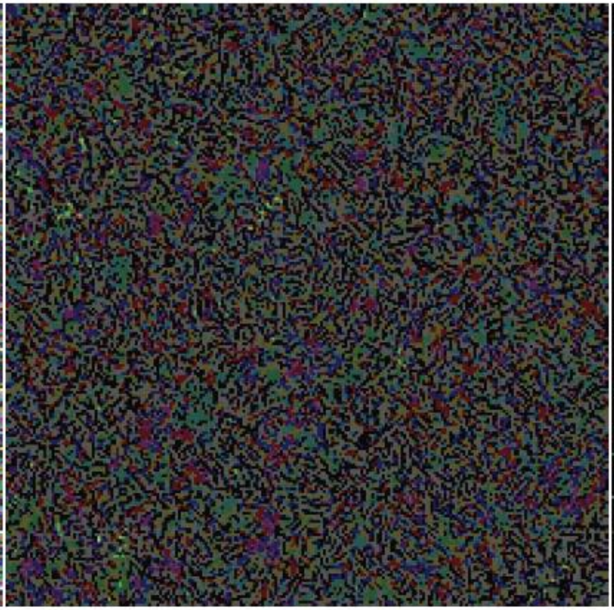
Original

δ



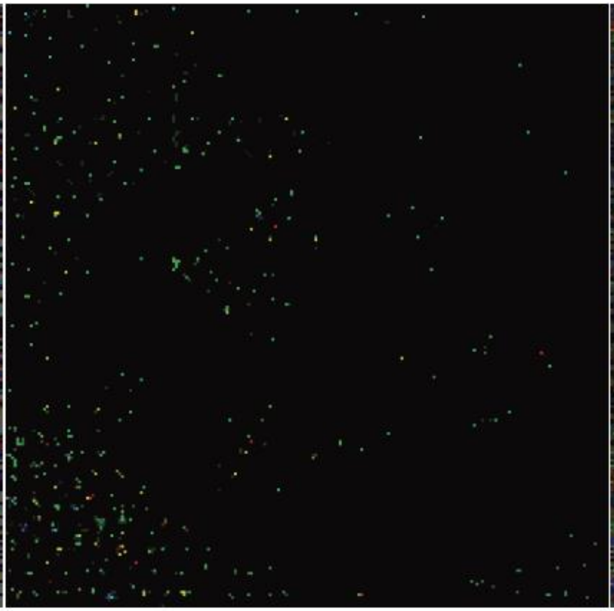
P-FGSM

δ



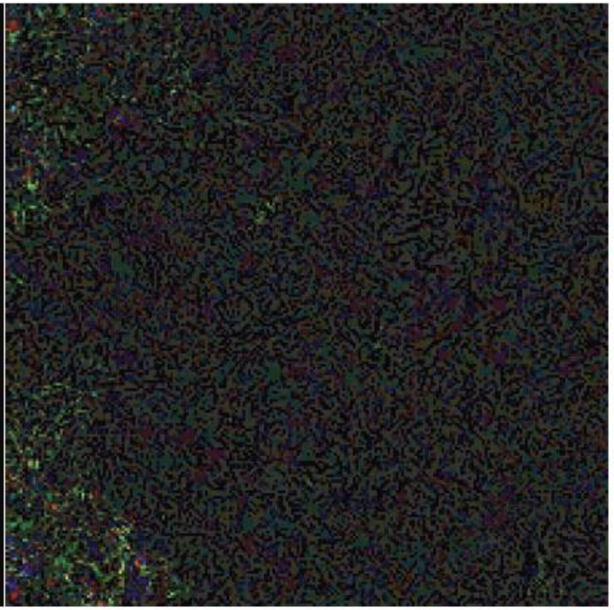
DeepFool

δ



SparseFool

δ



CW

Unconstrained perturbations: approaches

- Shifting randomly hue and saturation values
- Transferring new textures to images
- Colorization

can we do any better?

Part 2 of this tutorial

selectively modifies colors

maintains natural colors

Lab color space

exploits image semantics

object categories

(person, vegetation, sky, water, other)

ColorFool: semantic adversarial colorization

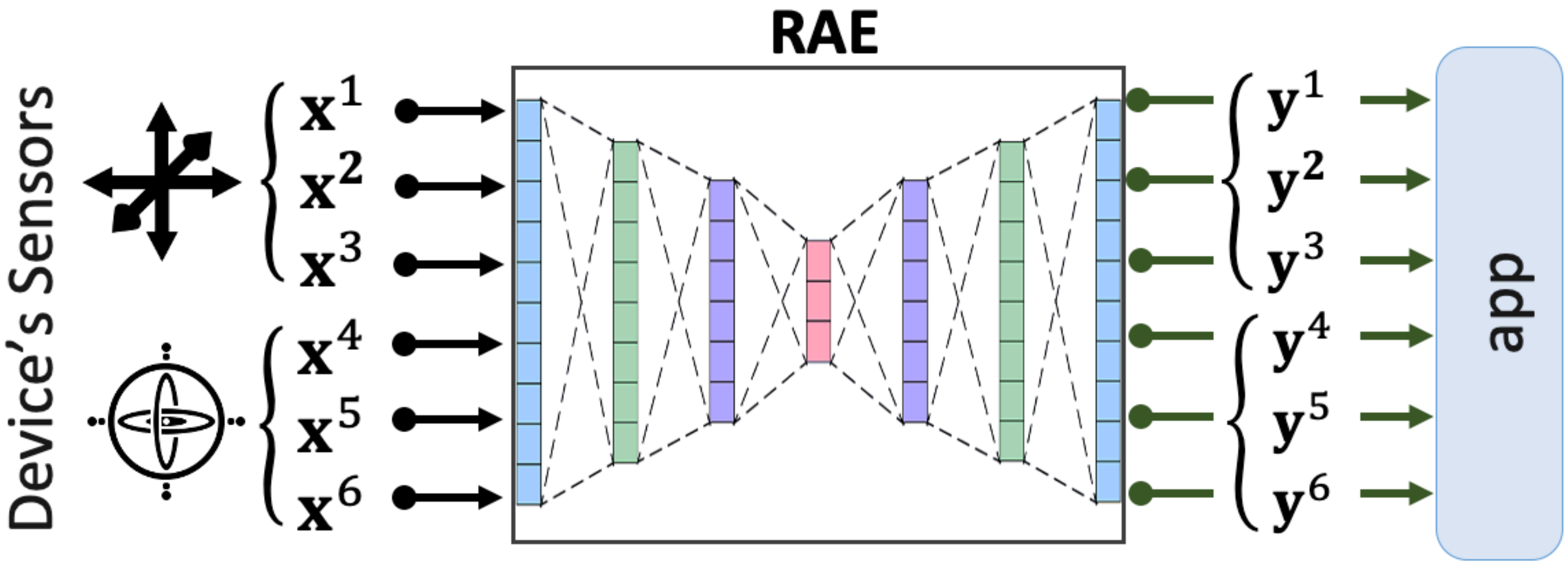
Shamsabadi, Sanchez-Matilla, Cavallaro

CVPR 2020

Injection

structure-aware perturbations
end-to-end training
multi-task loss
- image detail enhancement *objective*
- misleading *objective*

EdgeFool: an adversarial image enhancement filter
Shamsabadi, Oh, Cavallaro
IEEE ICASSP 2020



Privacy and utility preserving sensor-data transformations

M. Malekzadeh, R.G. Clegg, A. Cavallaro, H. Haddadi

Pervasive and Mobile Computing, Vol 63, Article 101132, 2020

References

- M.S. Cross, A. Cavallaro. 2020. **Privacy as a feature for body worn cameras.** *Signal Processing Magazine*, 37 (4), doi: 10.1109/MSP.2020.2989686.
- A.S. Shamsabadi, R.S. Matilla, A. Cavallaro. 2020. **ColorFool: Semantic Adversarial Colorization.** *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, Washington, USA.
- A.S. Shamsabadi, C. Oh, A. Cavallaro. 2020. **EdgeFool: An Adversarial Image Enhancement Filter.** *In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain.
- R. S. Matilla, C. Y. Li, A. S. Shamsabadi, R. Mazzon, A. Cavallaro. 2020. **Exploiting vulnerabilities of deep neural networks for privacy protection**. *IEEE Transactions on Multimedia*, 22 (7), doi: 10.1109/TMM.2020.2987694.
- M. Malekzadeh, R. G. Clegg, A. Cavallaro, H. Haddadi. 2020. **Privacy and Utility Preserving Sensor Data Transformations.** *Pervasive and Mobile Computing*, Volume 63, Article 101132.
- O. Sarwar, B. Rinner, A. Cavallaro. 2019. **A privacy-preserving filter for oblique face images based on adaptive hopping Gaussian mixtures**. *IEEE Access*, 7, doi: 10.1109/ACCESS.2019.2944861.
- C. Y. Li, A. S. Shamsabadi, R. S. Matilla, R. Mazzon, A. Cavallaro. 2019. **Scene Privacy Protection.** *In Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK.
- M. Malekzadeh, R. G. Clegg, A. Cavallaro, H. Haddadi. 2019. **Mobile Sensor Data Anonymization.** *In Proc. of the IEEE/ACM International Conference on Internet of Things Design and Implementation (IoTDI)*, Montreal, Canada.
- M. Malekzadeh, R. G. Clegg, A. Cavallaro, H. Haddadi. 2018. **Protecting Sensory Data Against Sensitive Inferences**. *In Proc. of the 1st ACM Workshop on Privacy by Design in Distributed Systems (W-P2DS)*, Porto, Portugal.
- M. Malekzadeh, R. G. Clegg, H. Haddadi. 2018. **Replacement Autoencoder: A Privacy-Preserving Algorithm for Sensory Data Analysis**. *In Proc. of the IEEE/ACM International Conference on Internet-of-Things Design and Implementation (IoTDI)*, Orlando, Florida, USA.
- O. Sarwar, A. Cavallaro, B. Rinner. 2018. ***Temporally smooth privacy protected airborne videos***. *In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain.

Resources

- Code and data
  - https://github.com/smartcameras/ColorFool
  - https://github.com/smartcameras/EdgeFool
  - https://github.com/smartcameras/P-FGSM

  - https://github.com/mmalekzadeh/motion-sense
  - https://github.com/mmalekzadeh/dana
  - https://github.com/mmalekzadeh/replacement-autoencoder

Part 2

Part 3

Privacy threats

whom or what should information be protected from?

Privacy protection

from unwanted, automatic inferences (AI-powered services)

Tools to control the information we share

software distributed as open source

Part 2

Part 3

ACM multimedia
Seattle, USA Oct 12-16 2020

ACM Multimedia 2020

Deep Learning for Privacy in Multimedia

Andrea Cavallaro

Part 2

Ali Shahin Shamsabadi

Part 3

Mohammad Malekzadeh

Thanks to the Alan Turing Institute (EP/N510129/1), which is funded by the EPSRC, for its support through the project PRIMULA

The Alan Turing Institute

CIS centre for intelligent sensing

Queen Mary University of London