

Discussion to Bart Custer's “*Discrimination, privacy and related issues in predictive big data analytics*”

Law & Robots Workshop, May 16, 2018

Thilo Stadelmann



datalab

www.zhaw.ch/datalab

Observations

From the **talk**:

- The law community **expects** technical **solutions** from us **engineers**, e.g.,
 - privacy-preserving data mining
 - privacy by design/default

From my **practice**:

- Engineers **expect** the **society** and legislator **to define** an appropriate **environment**, e.g.
 - data ethics
 - in what kind of society do we want to live?



Observations

From the **talk**:

- The law community **expects** technical **solutions** from us **engineers**, e.g.,
 - privacy-preserving data mining
 - privacy by design/default

From my **practice**:

- Engineers **expect** the **society** and legislator **to define** an appropriate **environment**, e.g.
 - data ethics
 - in what kind of society do we want to live?



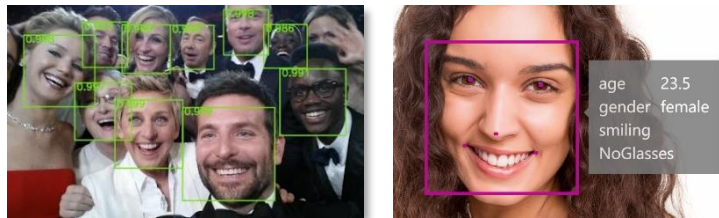
Problem

I like engineering solutions, but... it's difficult

MY RESEARCH FIELD

Machine learning for pattern recognition:

- Object detection in images



- Spoken language processing

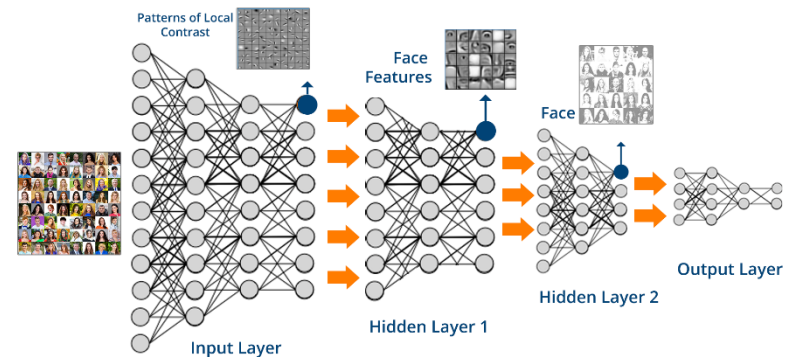


- Natural language processing

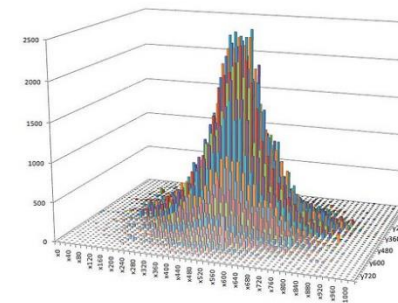


ITS FOUNDATION

- Deep neural networks

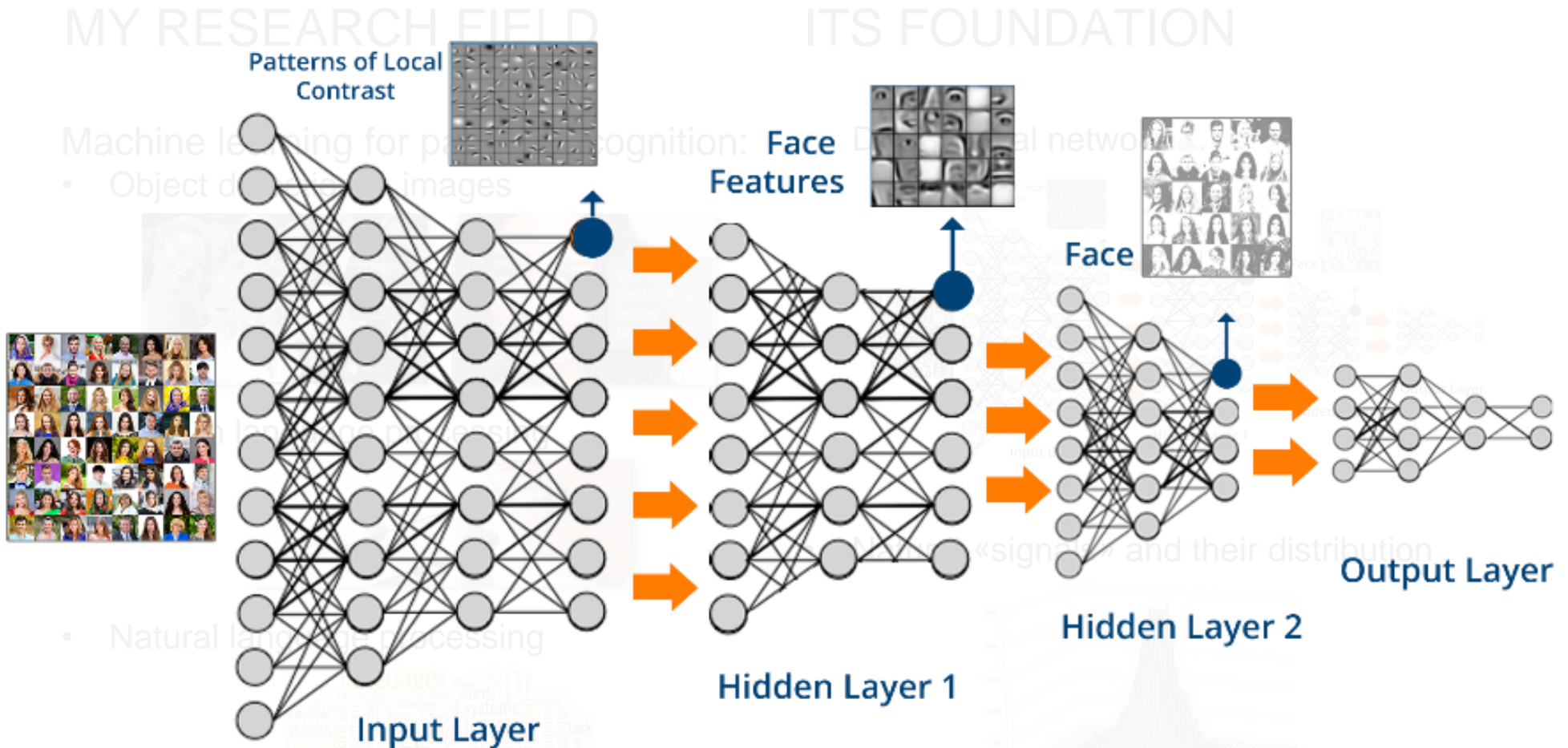


- Natural «signals» and their distribution



Problem

I like engineering solutions, but... it's difficult



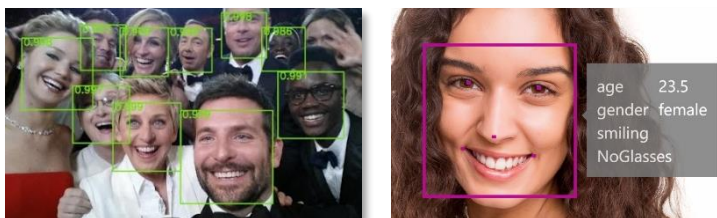
Problem

I like engineering solutions, but... it's difficult

MY RESEARCH FIELD

Machine learning for pattern recognition:

- Object detection in images



- Spoken language processing

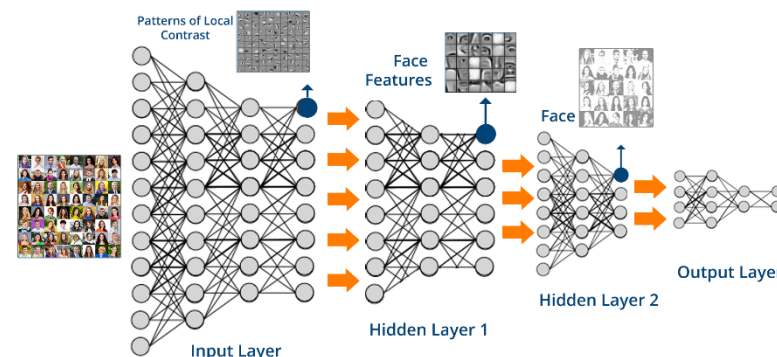


- Natural language processing

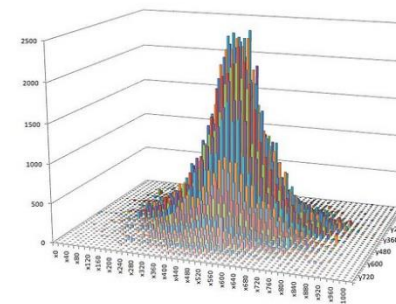


ITS FOUNDATION

- Deep neural networks



- Natural «signals» and their distribution

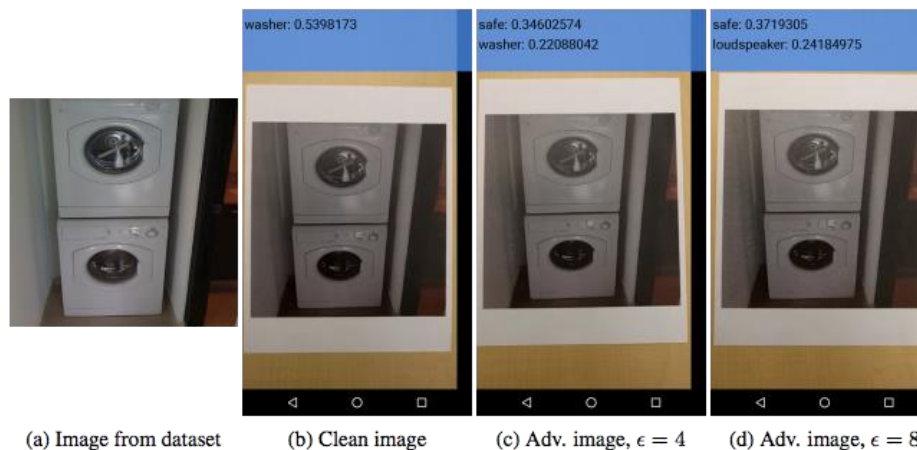


More background

The NN doesn't see as we do – and depends on very subtle patterns



Up: Olah, et al., "Feature Visualization", Distill, 2017, <https://distill.pub/2017/feature-visualization/>; down: <https://blog.openai.com/adversarial-example-research/>



Conclusions– Contributions to the discussion

- Doing **responsible engineering** is **important** and should be stressed more
- Current **approaches** are **very limited** (to specific domains / methods / data types)
- Interdisciplinary (and intercultural) **dialogue is key**
→ proposal: explore to use **unsupervised** machine **learning to detect algorithmic bias** in data

How to get in touch

- Dr. Thilo Stadelmann, ZHAW School of Engineering
- Head of ZHAW Datalab, Vice President SGAICO, Board of the Swiss Alliance for Data-Intensive Services
- thilo.stadelmann@zhaw.ch, @thilo_on_data
- +41 58 934 72 08
- <http://stdm.github.io>



The footer contains three logos. On the left is the 'datalab' logo, which includes a grid of icons (lightbulb, minus, plus, divide, multiply, bar chart) and the text 'datalab' and 'www.zhaw.ch/datalab'. In the center is the 'SGAICO' logo, with the text 'swiss group for artificial intelligence and cognitive science' above it. On the right is the 'data+service' logo, featuring the words 'data' and 'service' in white on overlapping orange and blue circles, with a red circle containing a white plus sign between them. Below this is the text 'Swiss Alliance for Data-Intensive Services'.