Machine Learning and Communication

Thomas Wiegand

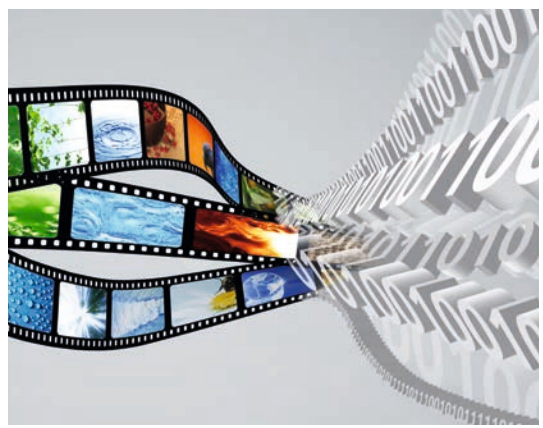
TU berlin

Fraunhofer
Heinrich Hertz Institute

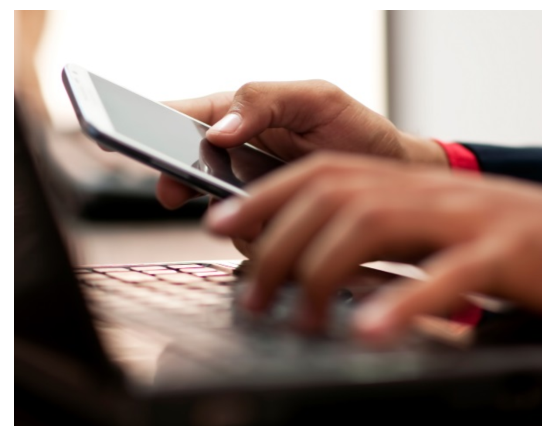
Fraunhofer Heinrich Hertz Institute

- Globally active player in digital infrastructure research
- Annual budget of 50 M€ / 450 Researchers
- Research & Development in Photonics, Video & Wireless Technologies
- **Every second bit on the internet touches Video or Photonic technology invented or made by Fraunhofer HHI**

10^0 – 10^2 – 10^4 Gbps



H.264 – H.265 – H.266



3G – 4G – 5G

Outline

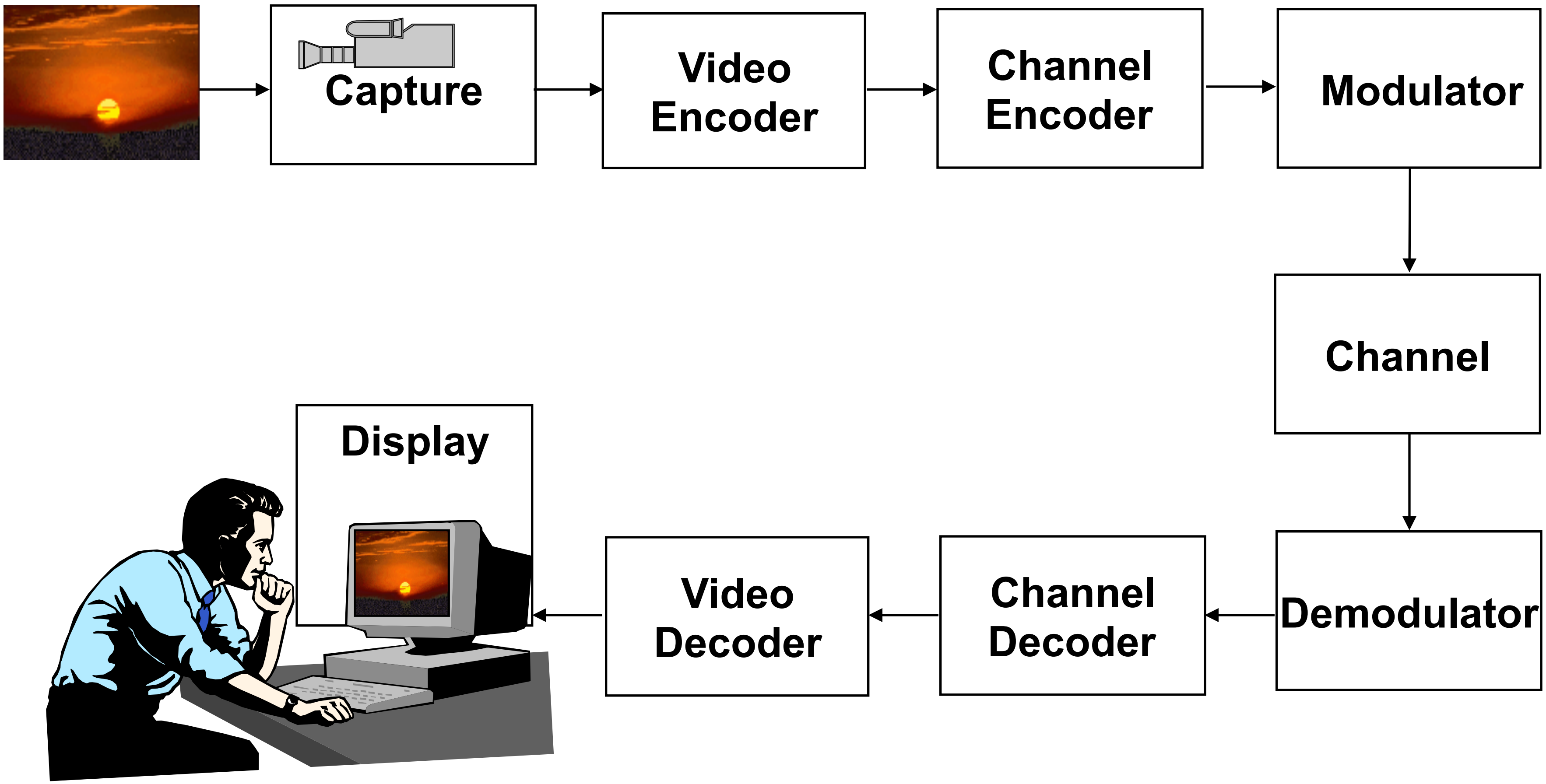
Machine Learning and

- Video Coding Standards
- Data Communication
- Decision Making Explained

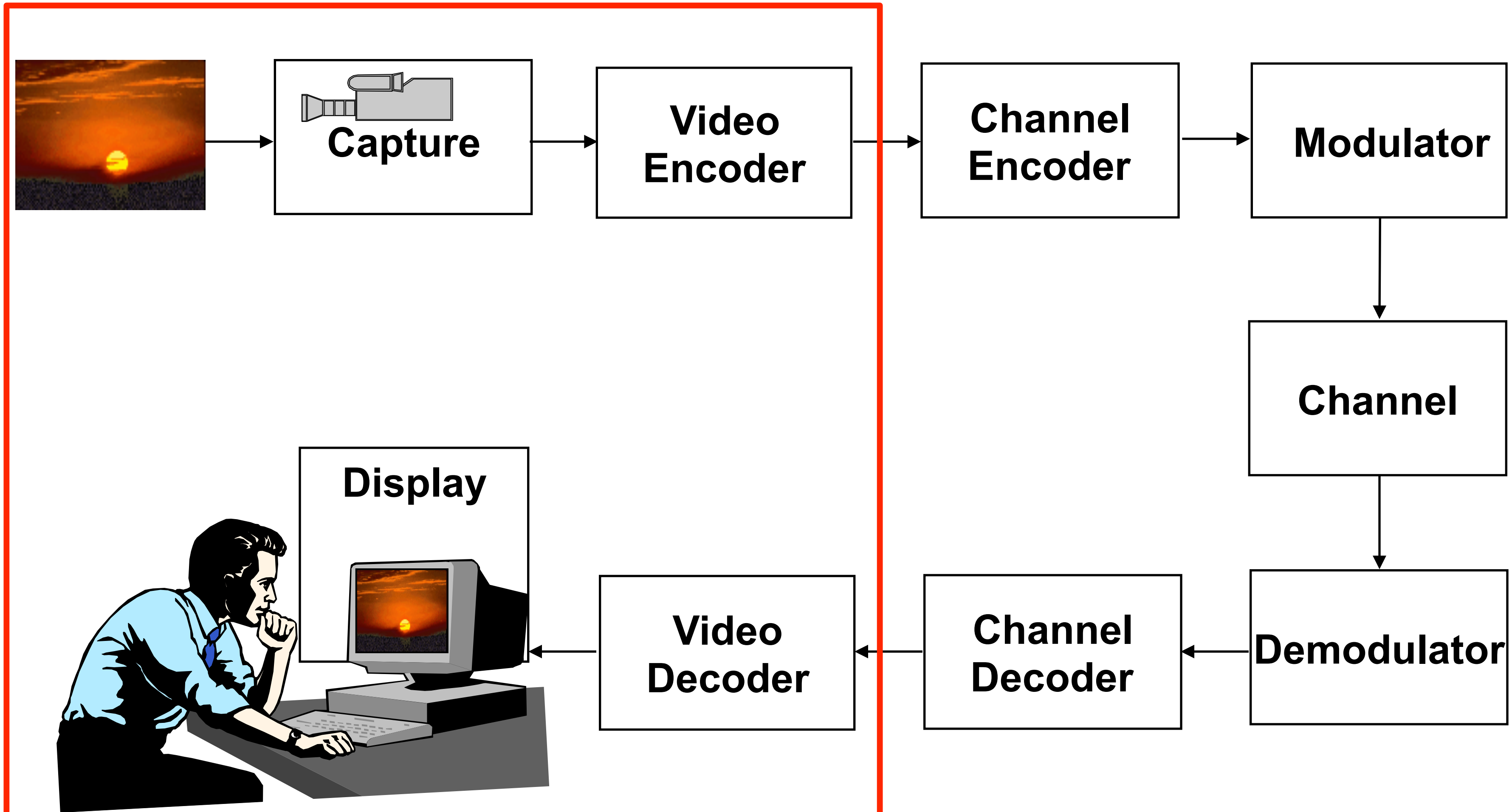
Machine Learning and Video Coding Standards

Fraunhofer
Heinrich Hertz Institute

Visual Communication Systems



Visual Communication Systems



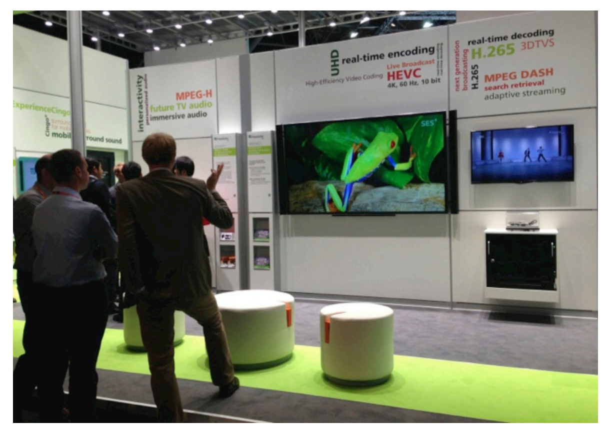
Video Coding Standards

International standardization of video coding:

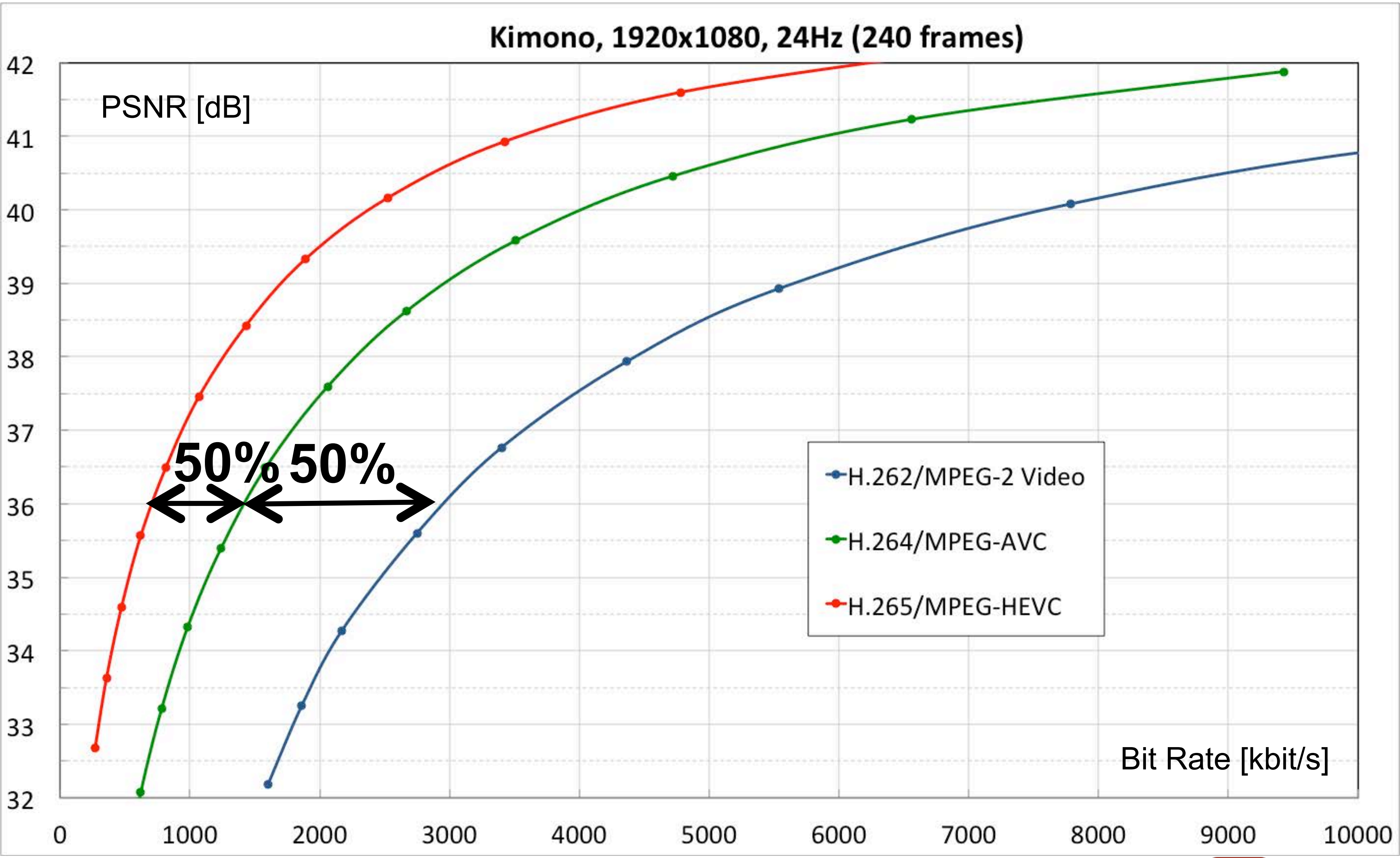
- Every 2nd bit on the Internet is H.264
- H.265 is starting to become relevant (12/2016: about 1 Billion devices)
- H.266 is in future planning stage

Implementations of video coding standards:

- Only decoder is specified
- Real-time video encoding is developed by manufacturers



Performance of Video Standards



Machine Learning

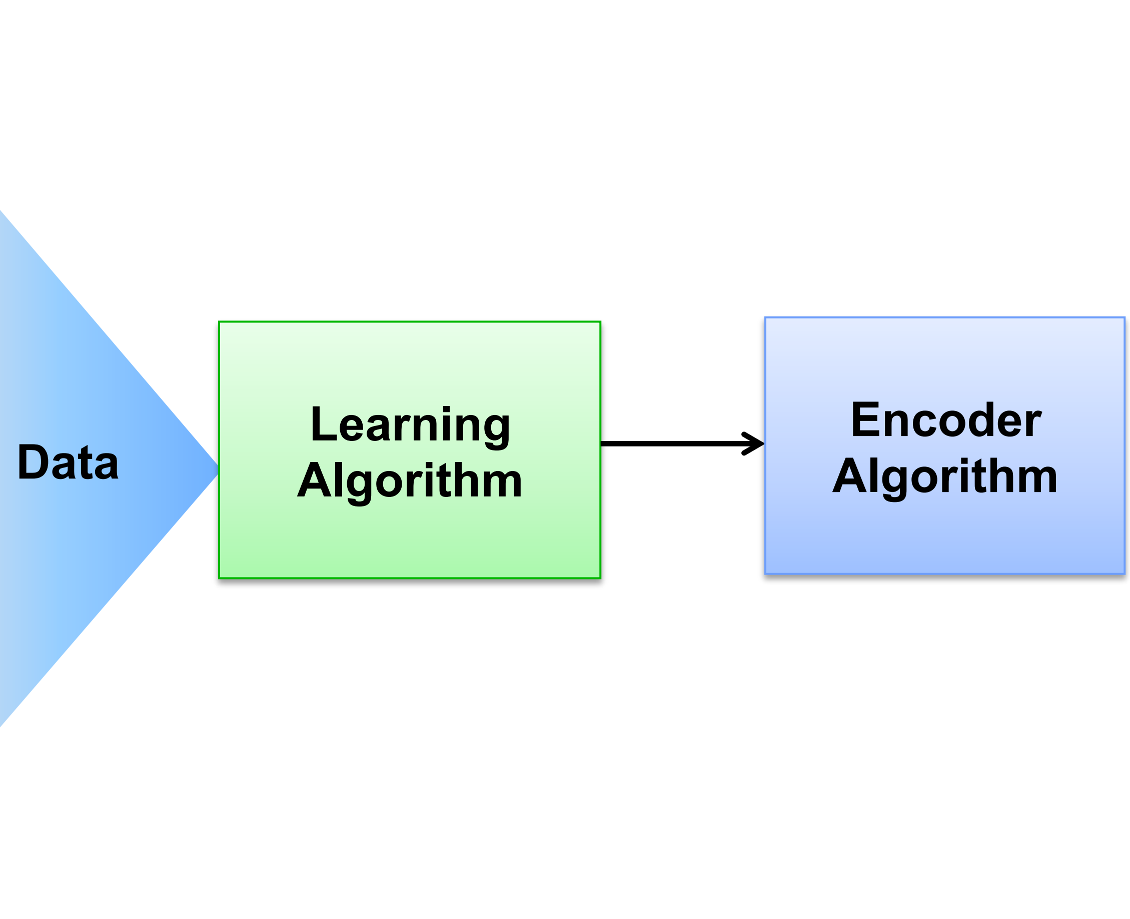
- Natural video

- H.265/MPEG-HEVC

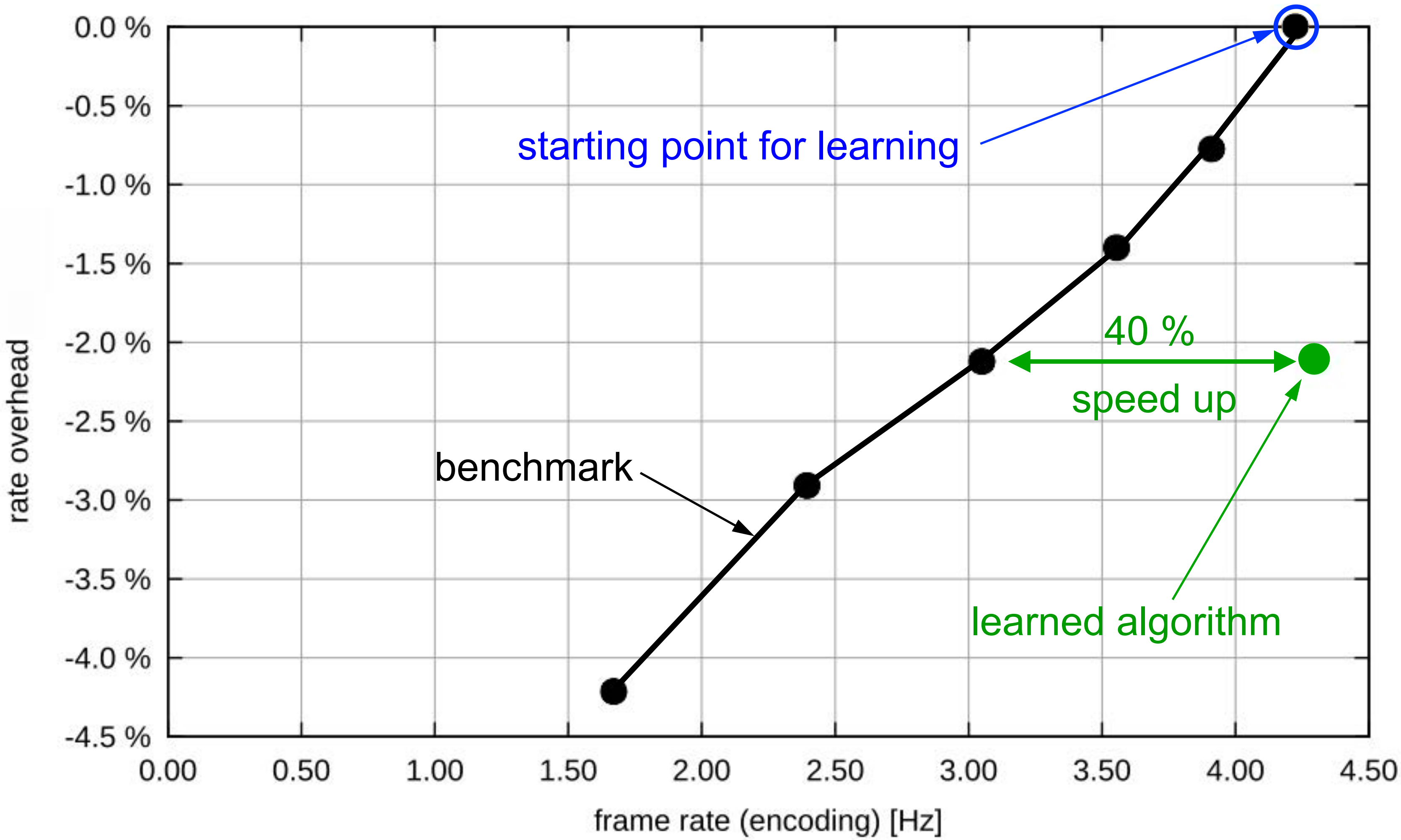
- Boundary conditions

  Rate <= R,

  Time <= T, ...



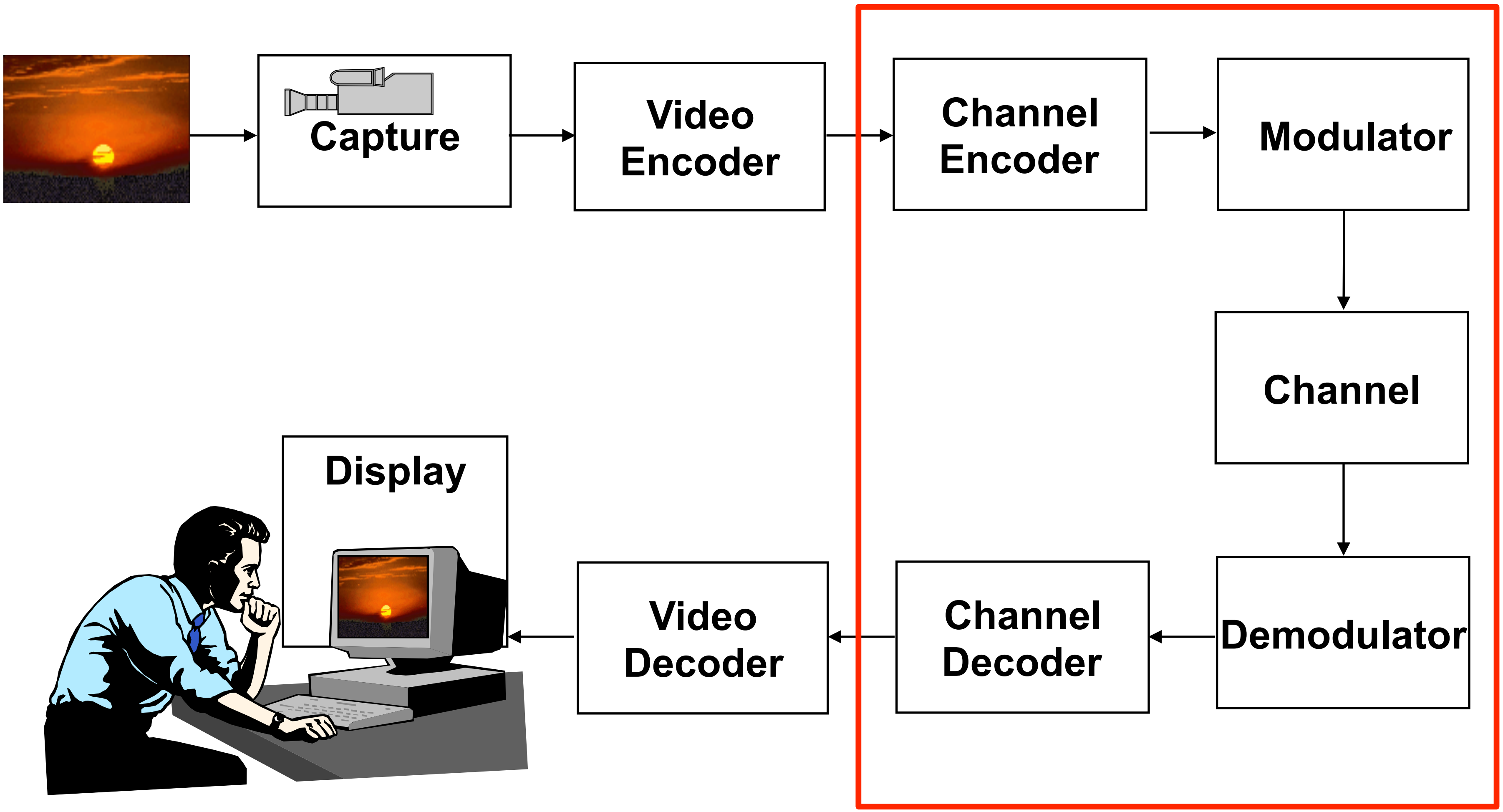
First Results: Fraunhofer HHI H.265 Encoder



Machine Learning and Data Communication

Fraunhofer
Heinrich Hertz Institute

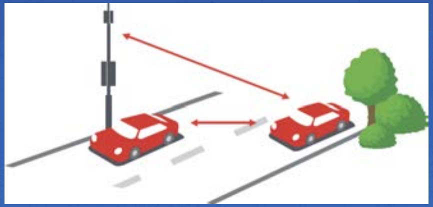
Visual Communication Systems



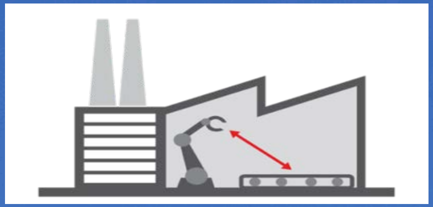
The Next Generation: 5G Network

5G Berlin

Mobile High Speed Internet

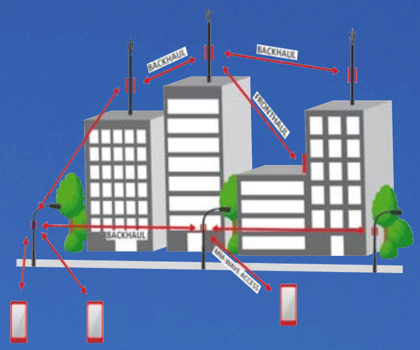


Car2Car & Car2X Communications



Industrial Wireless

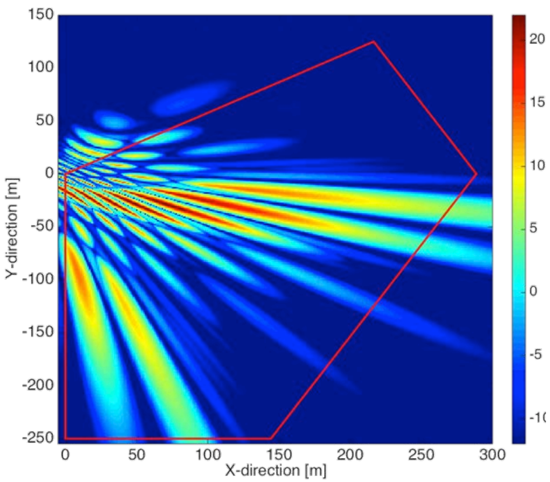
Requirements
- 1000 x throughput
- 100 x devices
- 10 x battery life
- 1 ms latency

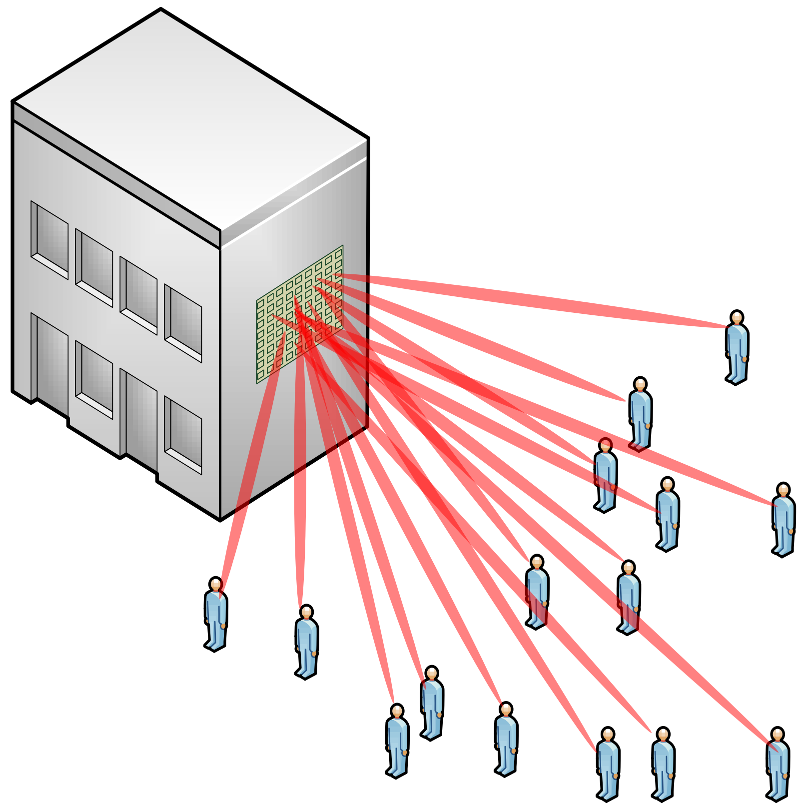


Technology
- DSL boxes and street lights become senders
- Optical fiber

Wireless Fiber and Location Sensing

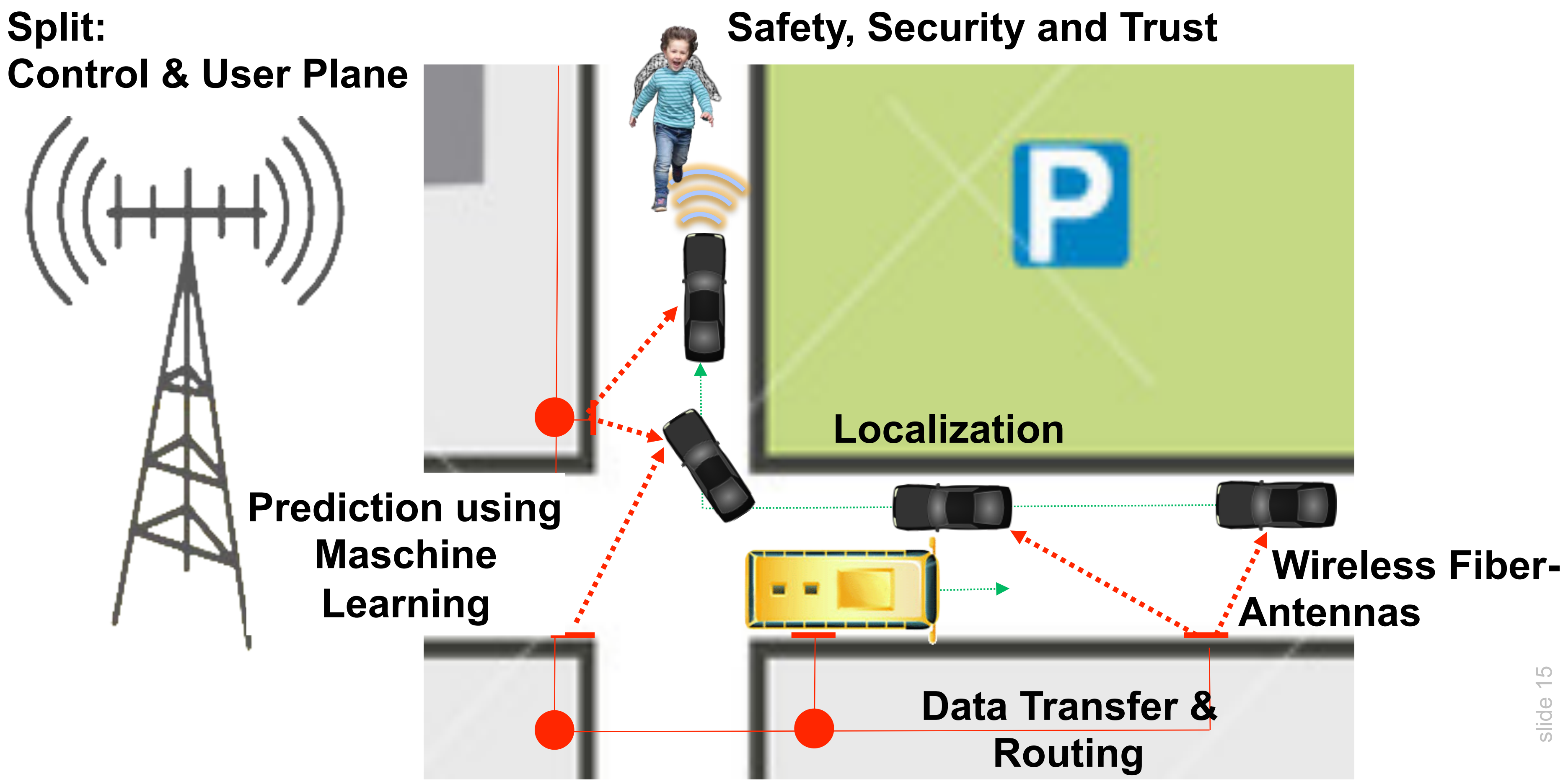
- 3D beamforming with MIMO Antennas



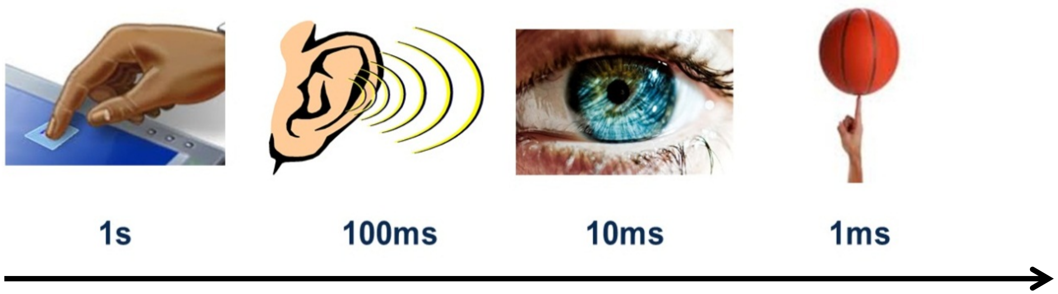


- Location of users via sensors

Future Mobile Digital Infrastructure Example: Networked Autonomous Driving



The Tactile Internet



Human reactions times

source: ITU TechWatch Report: The Tactile Internet

source: https://netzoekonom.de

- Very low end-to-end latencies (1ms)
- Ultra high reliability
- Can be realised as part of WiFi, 5G or fixed networks

Collaborative Driving

Source: ITU TechWatch Report: The Tactile Internet

Driver assistance with AR of potentially dangerous objects and situations

Interpretable Machine Learning

Fraunhofer
Heinrich Hertz Institute

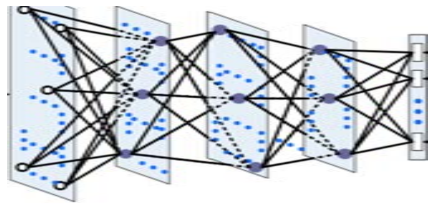
Classification using Machine Learning

Big Data

14.2 Million images, 22.000 classes

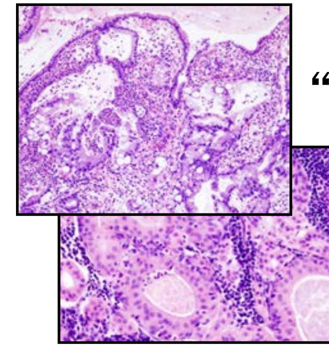
+

Machine Learning



=

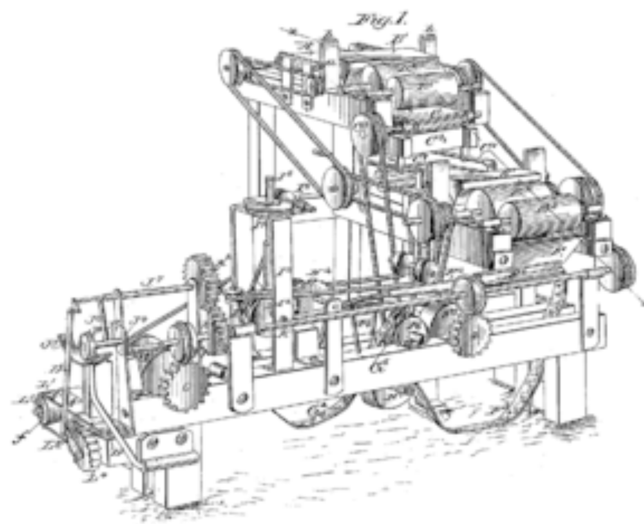
Automatic Annotation



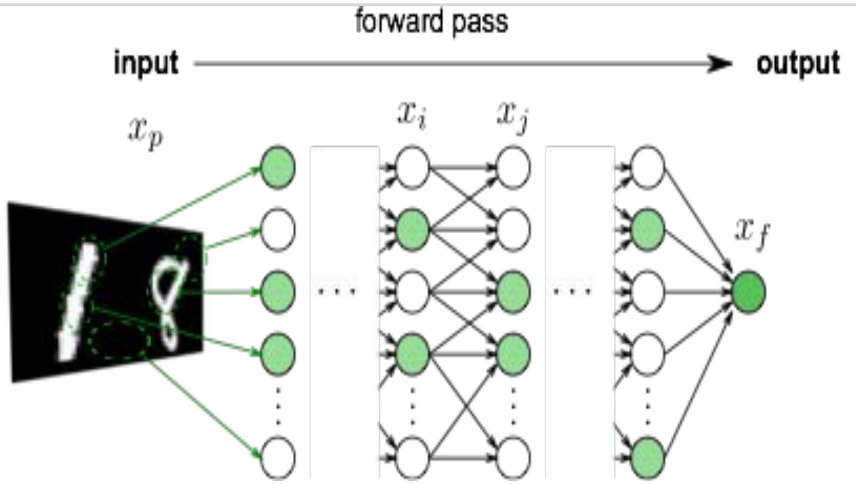
"no cancer"

"cancer"

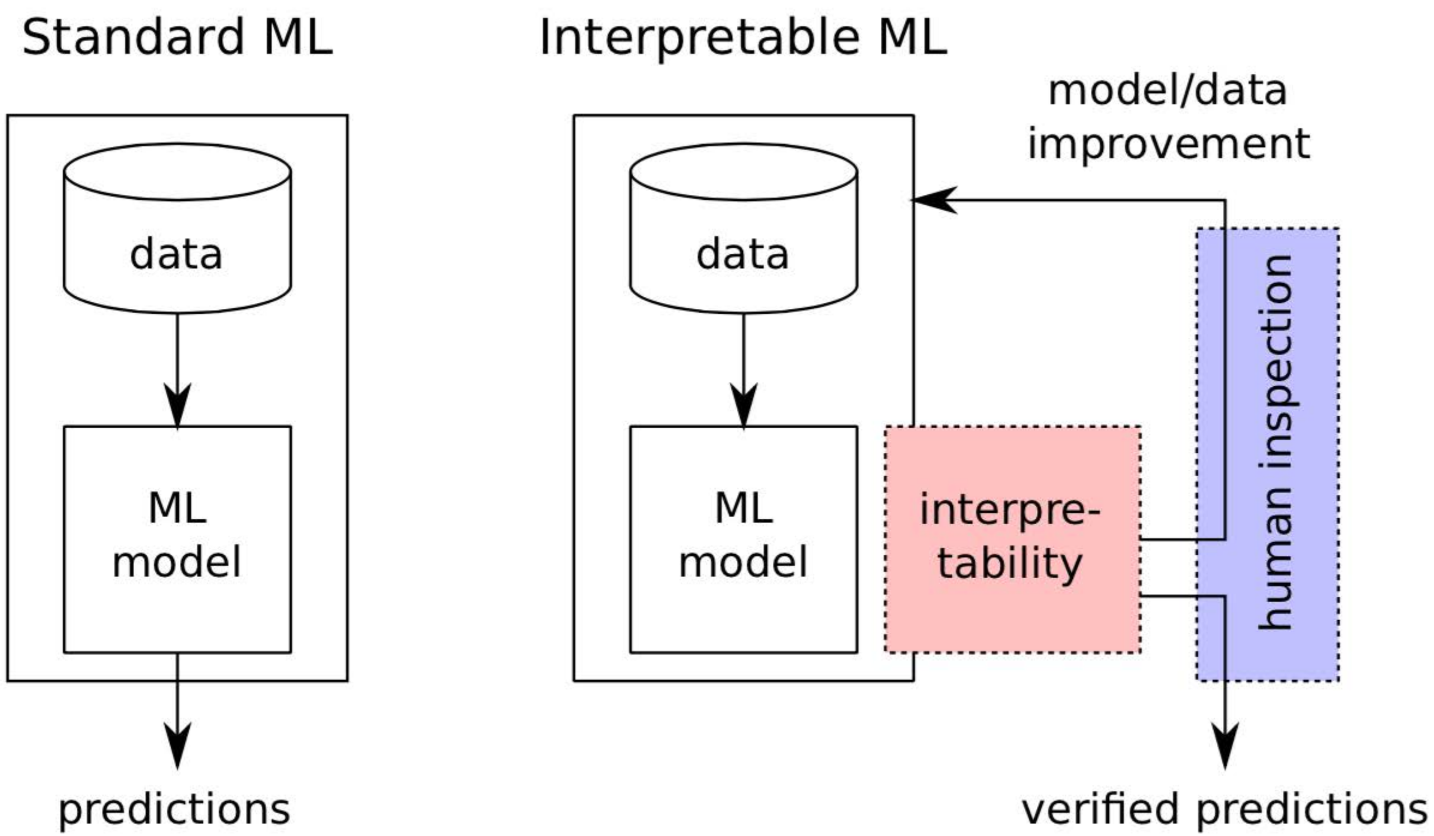
Do we trust the machine ???



Revert the Deep Neural Network



Interpretability of Machine Learning

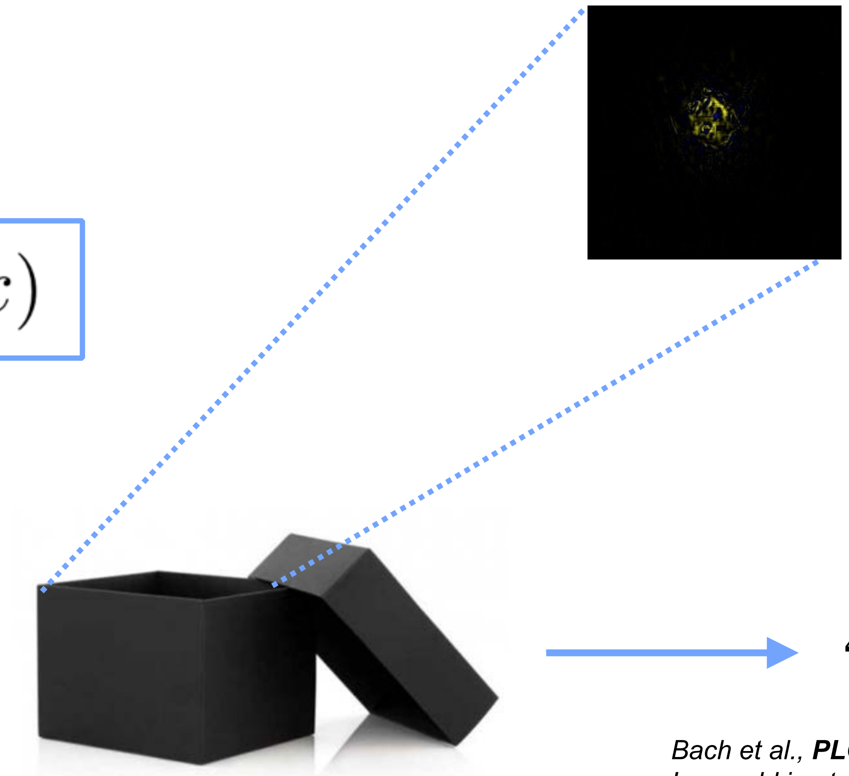


Interpretability is first step towards making sure (i.e. verifying) that ML algorithms do the right thing !

Idea for Interpretable Machine Learning

W. Samek, K.-R. Müller et al.:
general method to explain **individual** classification decisions.

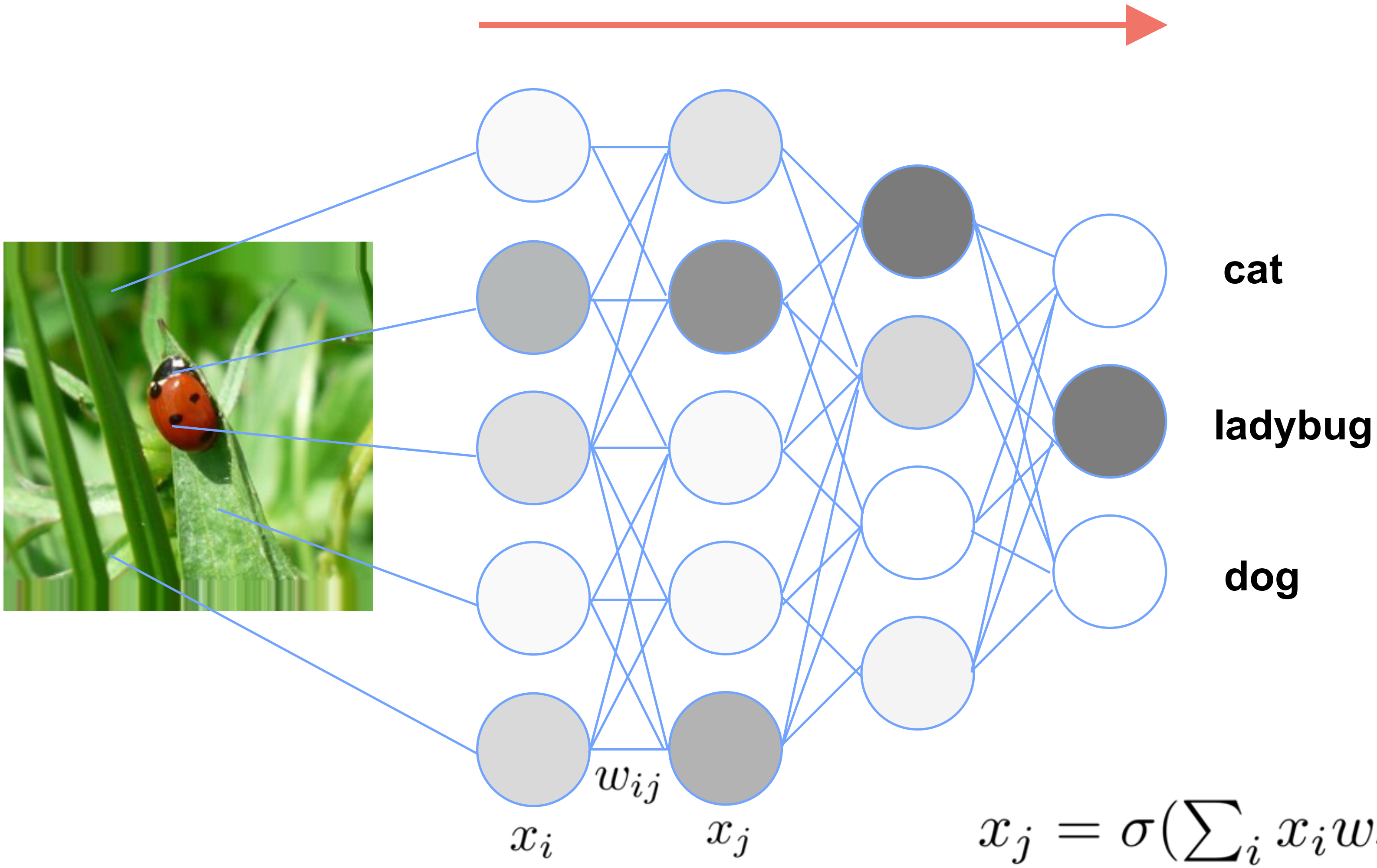
Main idea: $\sum_p r_p = f(x)$



"ladybug"

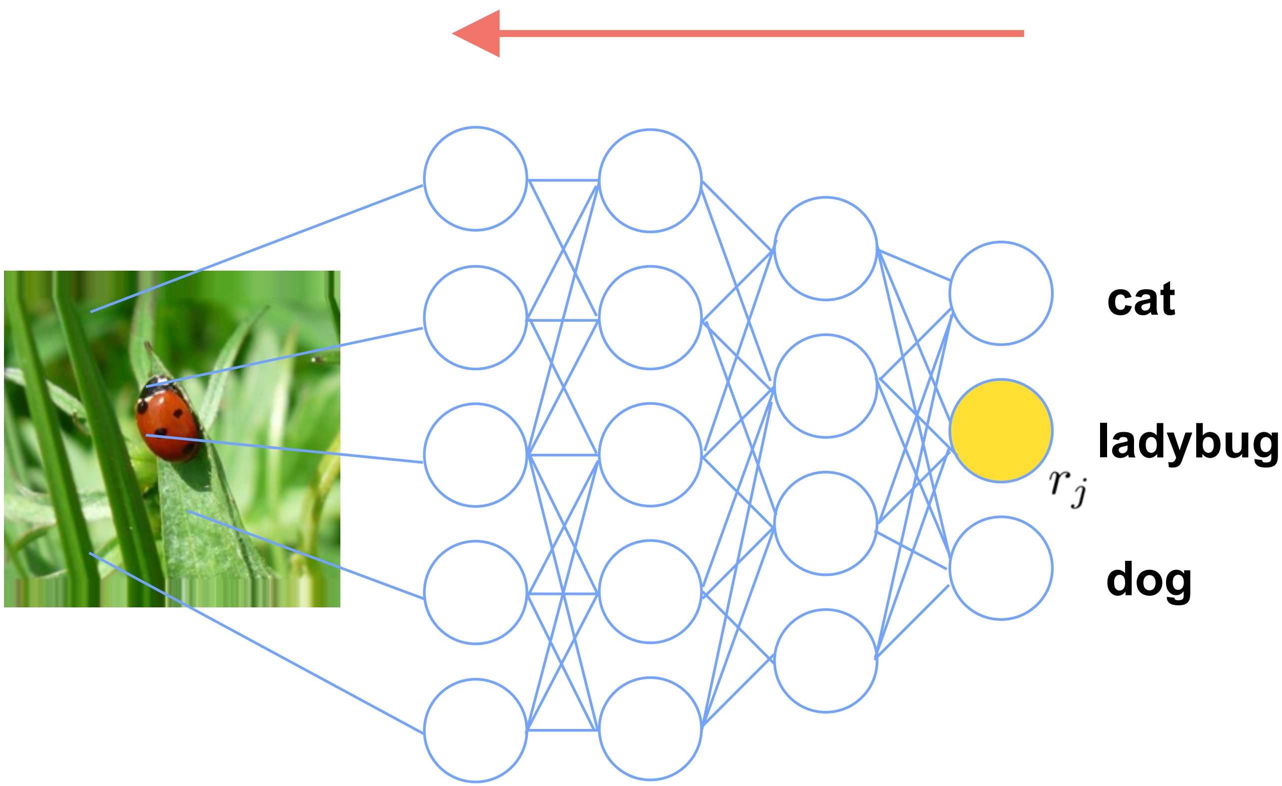
*Bach et al., **PLOS ONE**, 2015*
*Lapuschkin et al., **CVPR**, 2016*
*Samek et al., **TNNLS**, 2016*
….

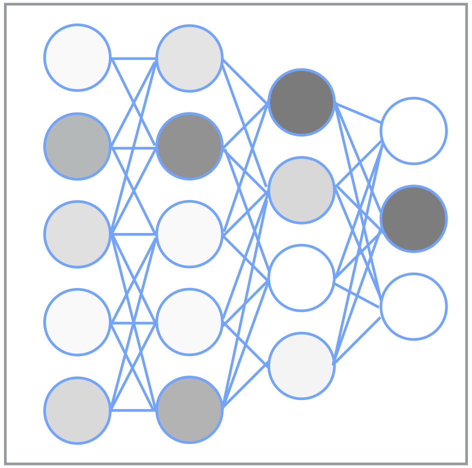
Classification



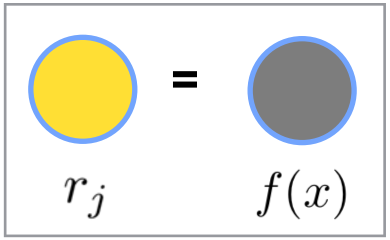
$$x_j = \sigma(\sum_i x_i w_{ij} + b_j)$$

Explanation

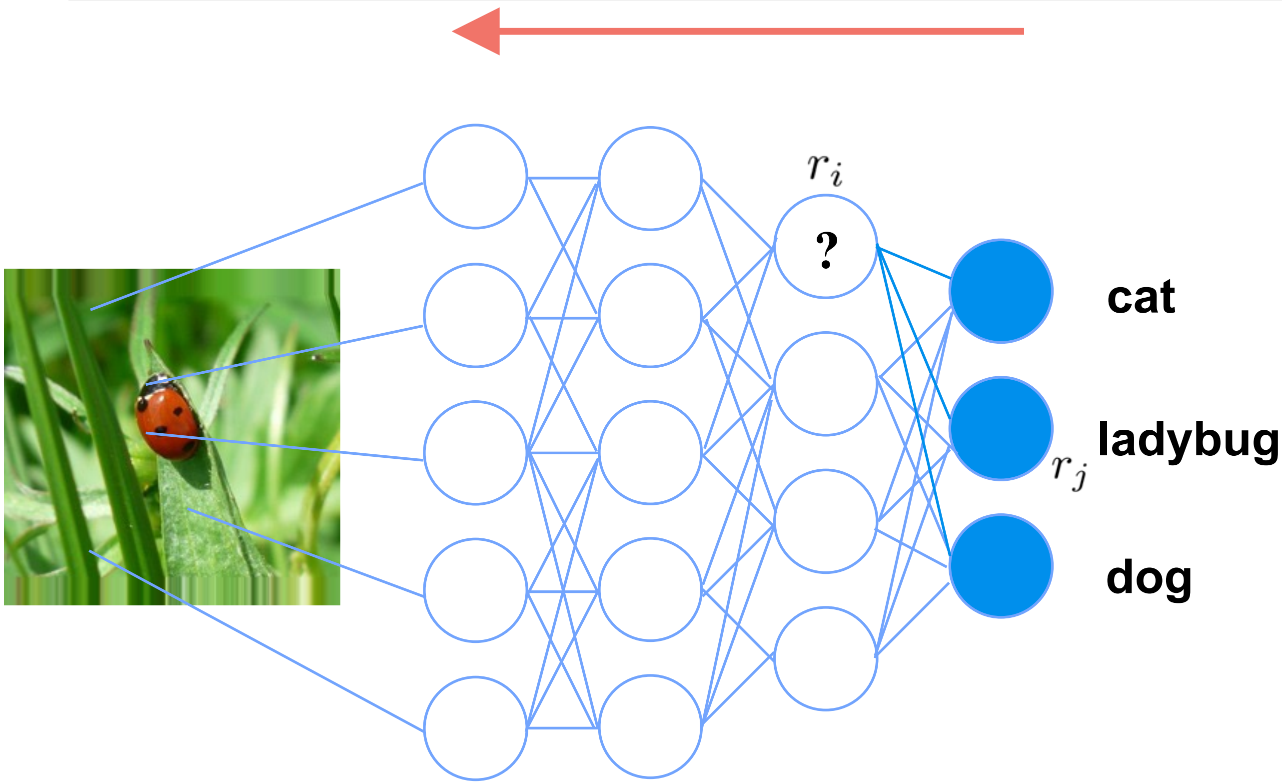


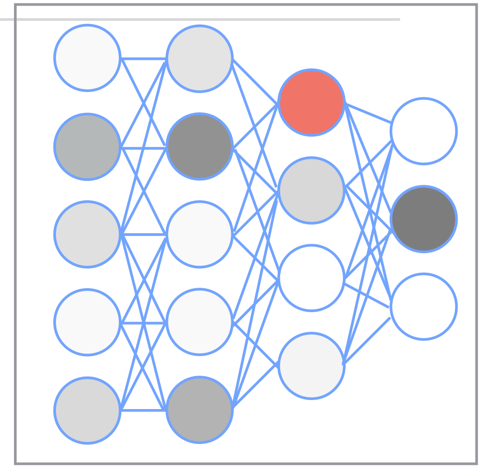


Initialization



Relevance Propagation





$$r_i = x_i \sum_j \frac{w_{ij} r_j}{\sum_i x_i w_{ij}} = x_i c_i$$

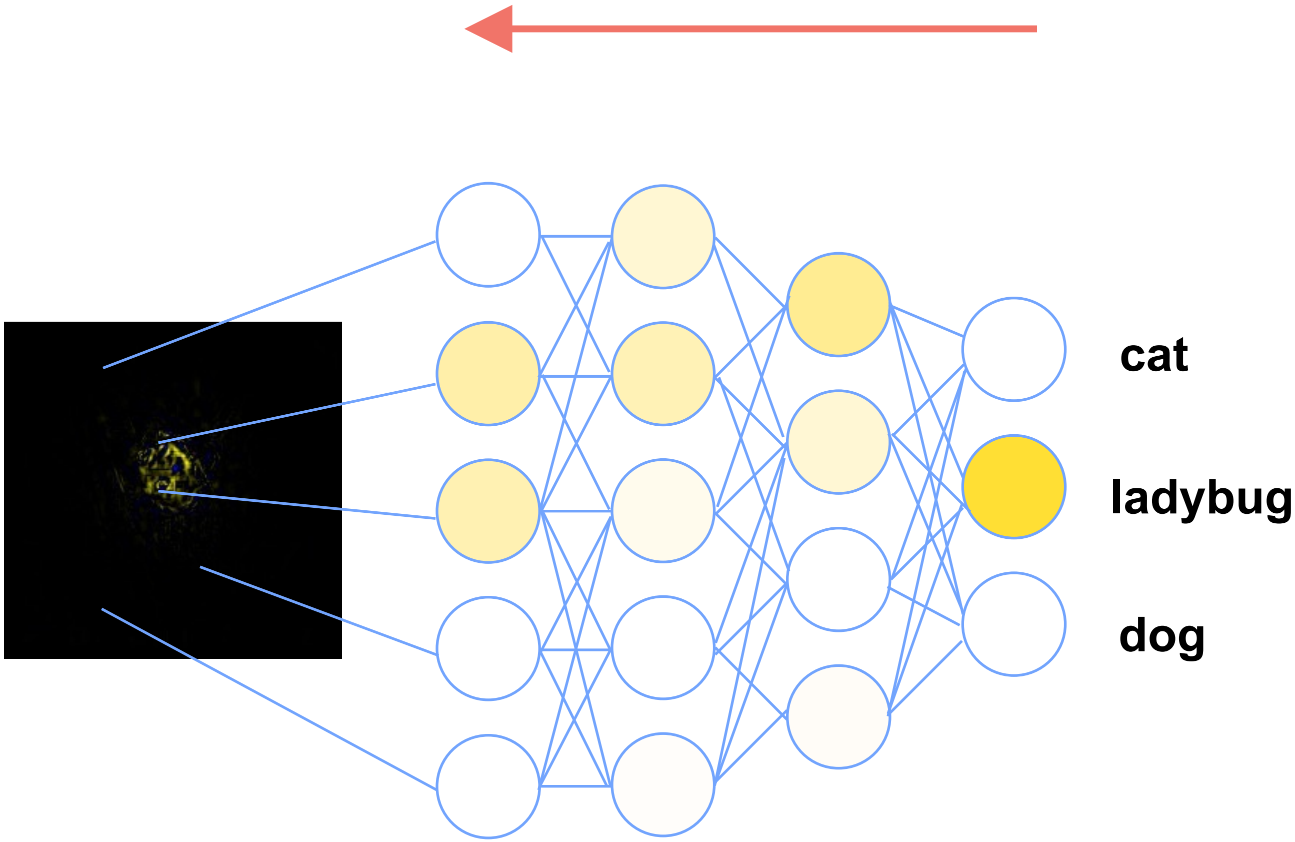
Theoretical interpretation

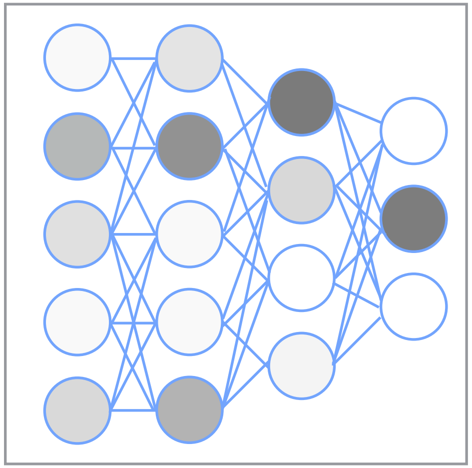
(Deep) Taylor Decomposition

*(Montavon et al., **arXiv** 2015)*

Relevance of upper layers is redistributed to lower layers <u>proportionally</u> (depending on activations & weights).

Relevance Conservation Property



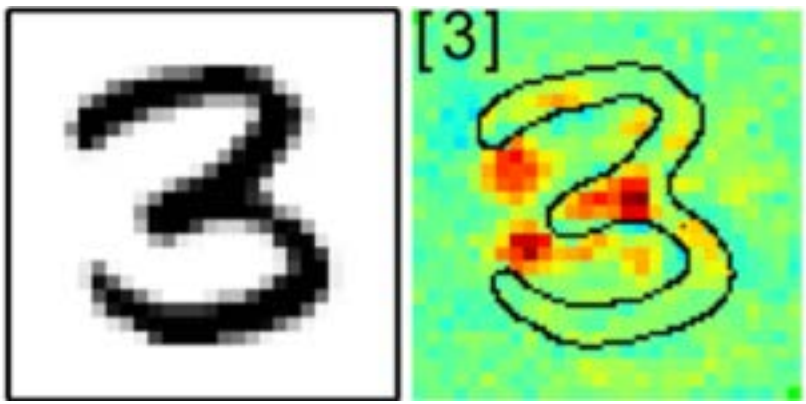


Relevance Conservation Property

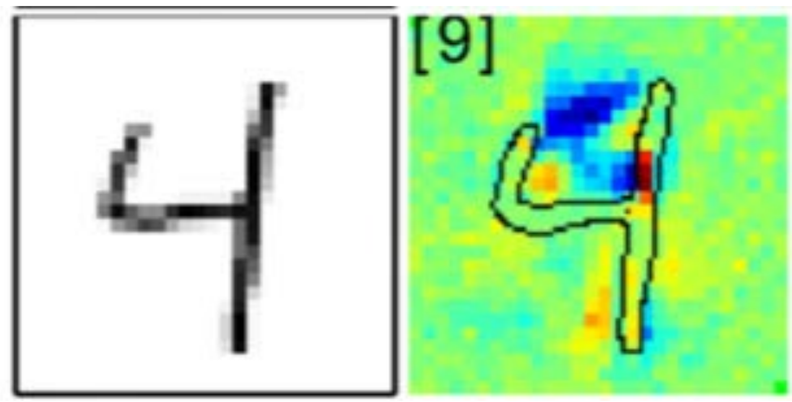
$$\sum_p r_p = \ldots = \sum_i r_i = \sum_j r_j = \ldots = f(x)$$

ML Decomposition Examples

what speaks for / against classification as “3”



what speaks for / against classification as “9”



[*number*]: explanation target class

red color: evidence for prediction

blue color: evidence against prediction

(*Bach et al.*, ***PLOS ONE*** *2015*)

ML Decomposition distinguishes between positive and negative evidence

Summary: Machine Learning and Communication are converging

- **Video Coding Standards and Machine Learning:**
  - H.264 → H.265 → H.266
  - Improve Video Encoding using ML
- **Data Communication and Machine Learning:**
  - Next Generation 5G: High bitrates, low latencies (Tactile Internet), Sensors
  - Machine Learning necessary for efficient
- **Interpretable Machine Learning:**
  - Decomposition explains classification results
  - Explanation required for Decision Making!

Acknowledgement & Support

ITU-T VCEG & ISO/IEC MPEG colleagues

HHI/TUB members and research associates

- H. Schwarz, D. Marpe, T. Hinz, P. Helle
- T. Schierl, C. Hellge, R. Skupin, Y. Sanchez
- S. Bosse, B. Blankertz, A. Norcia, G. Curio
- K.-R. Müller, W. Samek
- …

Supported by Federal Ministry of Economics and Technology

SPONSORED BY THE Federal Ministry of Education and Research

Deutsche Forschungsgemeinschaft DFG