Convergence of Communication and Machine Learning

Thomas Wiegand

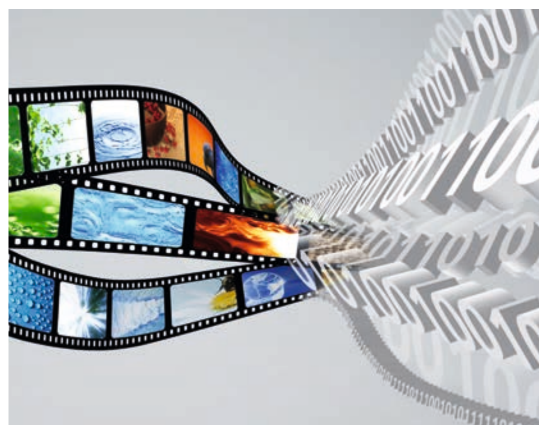
TU berlin

Fraunhofer
Heinrich Hertz Institute

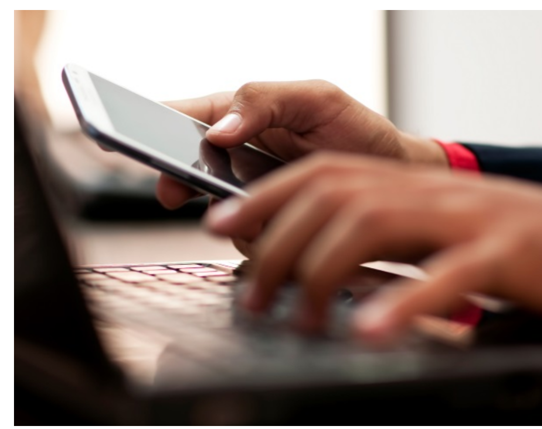
Fraunhofer Heinrich Hertz Institute

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

10^0 – 10^2 – 10^4 Gbps



H.264 – H.265 – H.266



3G – 4G – 5G

Outline

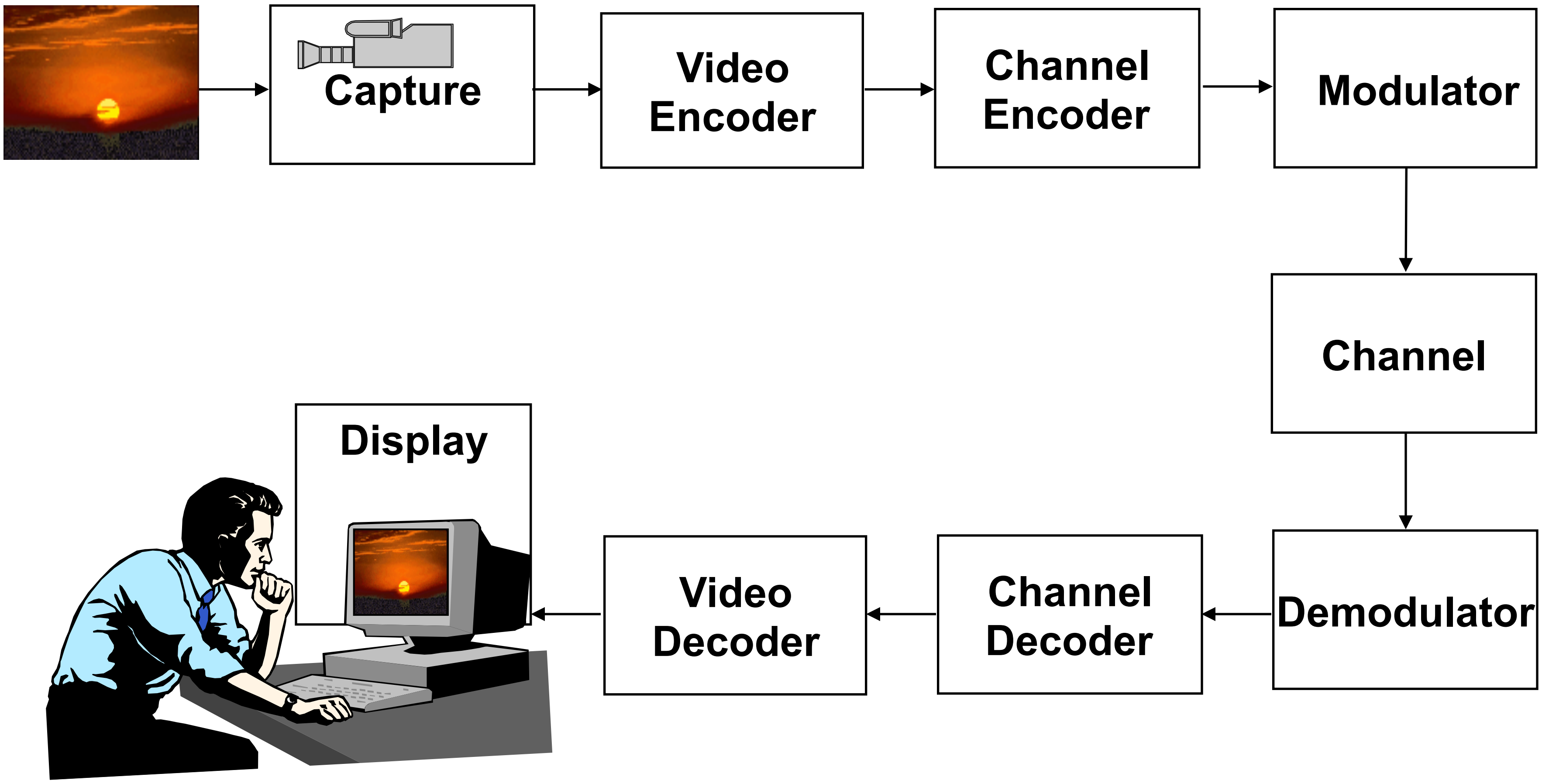
Machine Learning and

- Video Coding Standards
- Data Communication
- Decision Making Explained

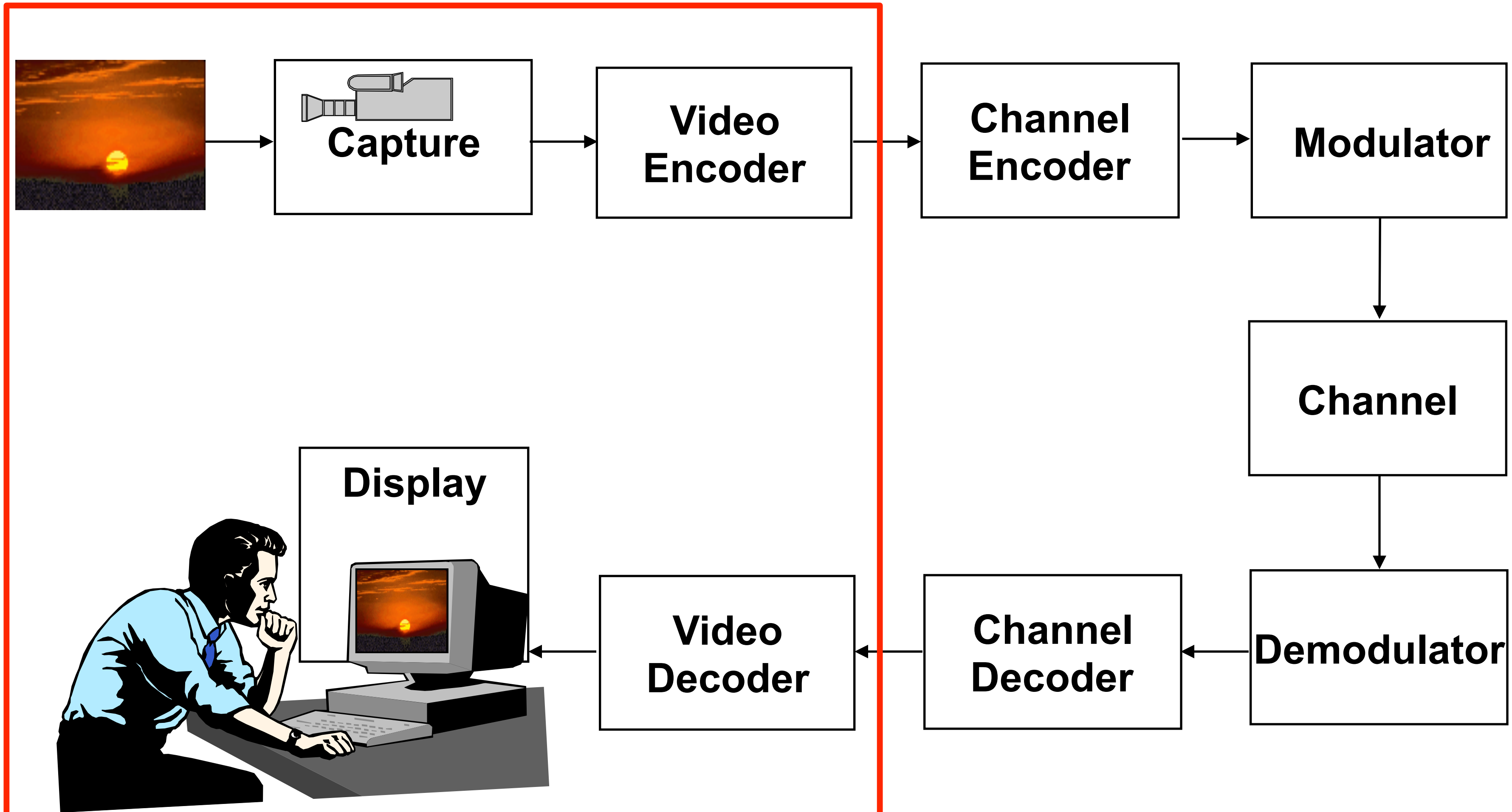
Machine Learning and Video Coding Standards

Fraunhofer
Heinrich Hertz Institute

Visual Communication Systems



Visual Communication Systems



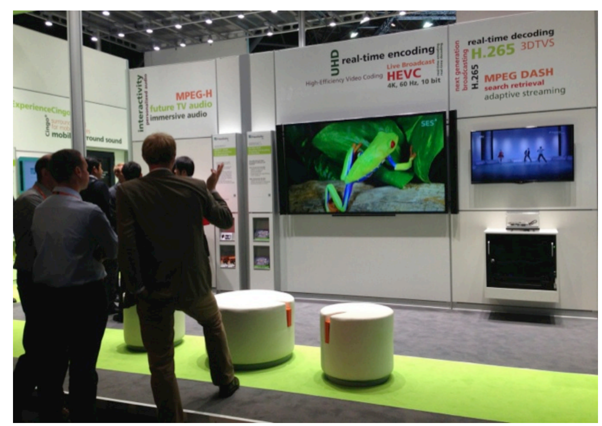
Video Coding Standards

International standardization of video coding:

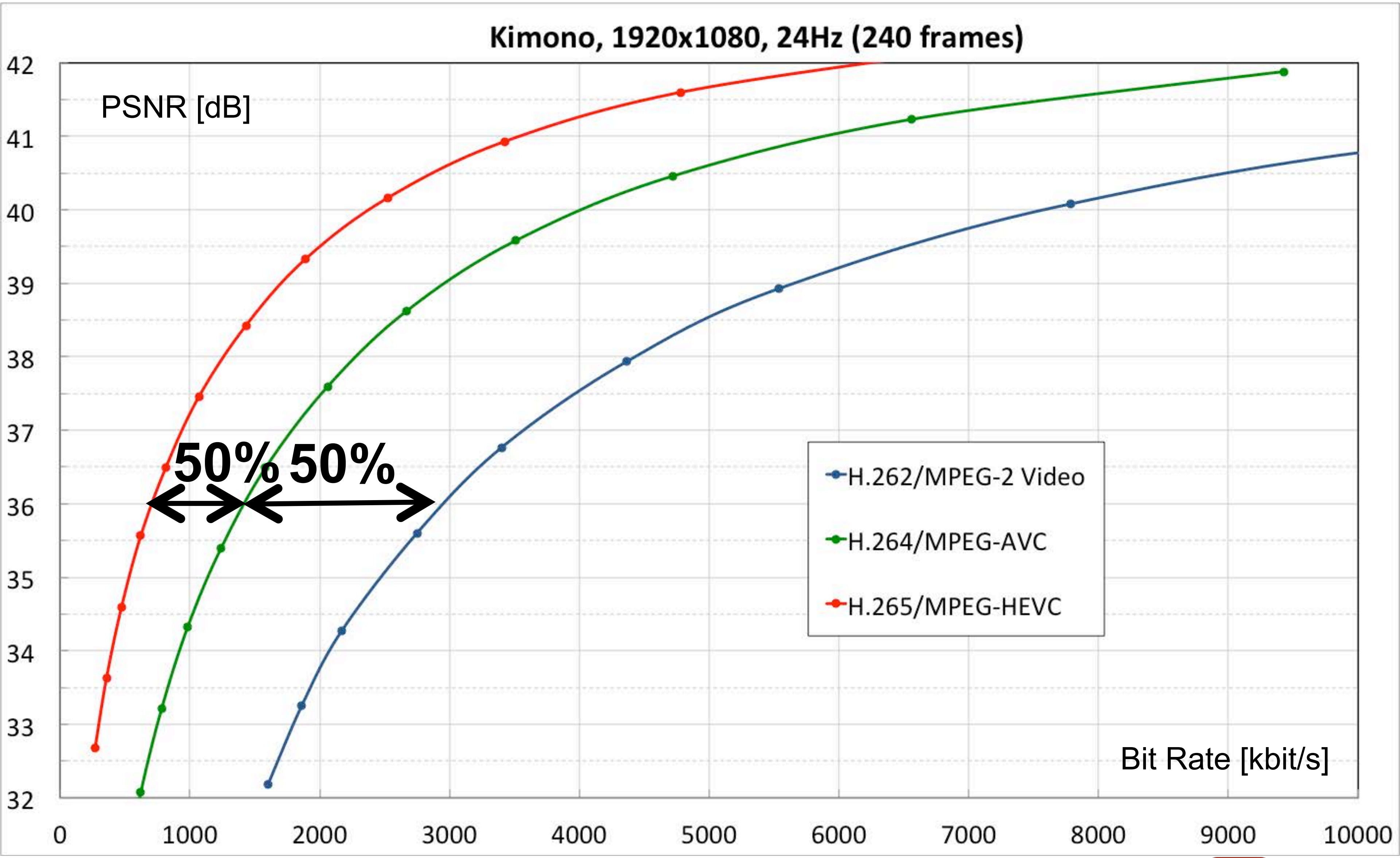
- Every 2nd bit on the Internet is H.264
- H.265 is starting to become relevant (12/2016: about 1 Billion devices)
- H.266 is in future planning stage

Implementations of video coding standards:

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



Performance of Video Standards



Machine Learning

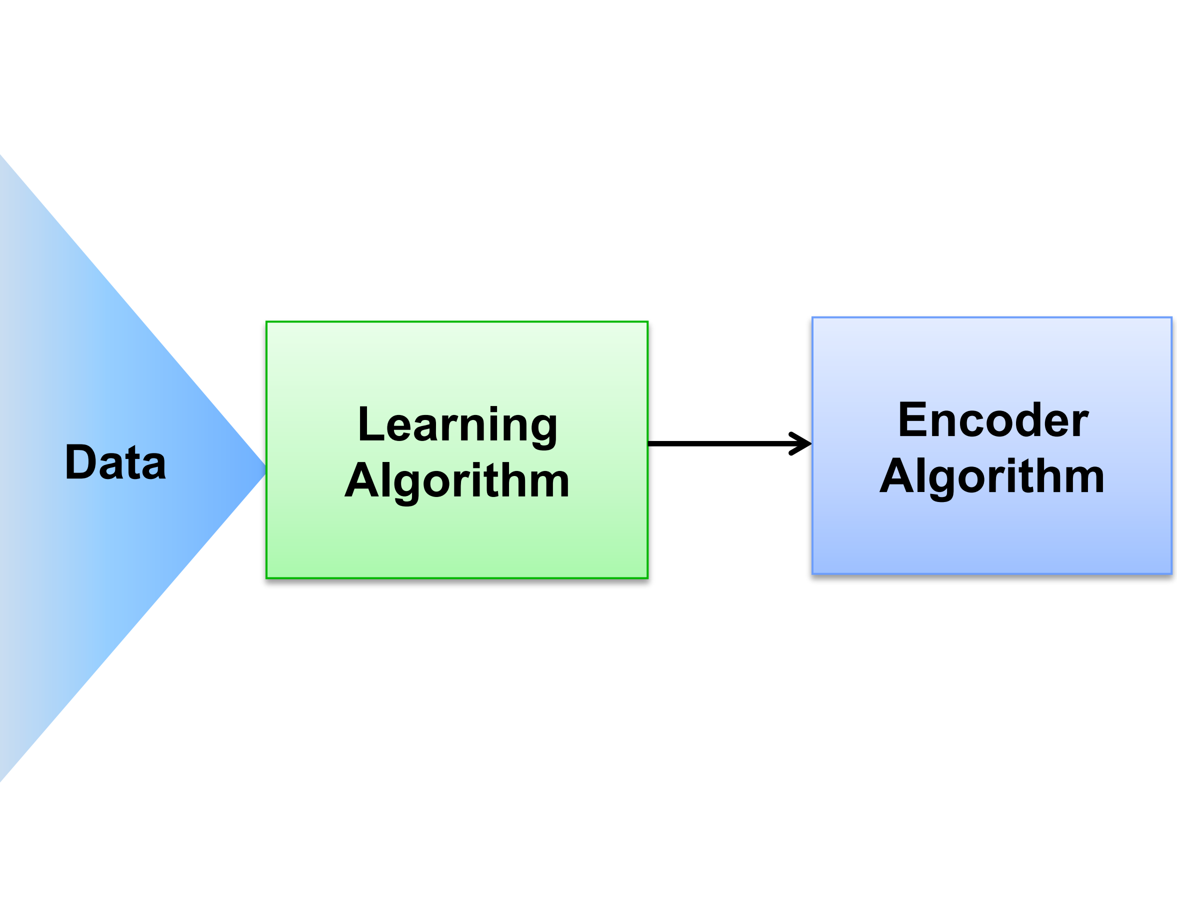
- Natural video

- H.265/MPEG-HEVC

- Boundary conditions

  Rate <= R,

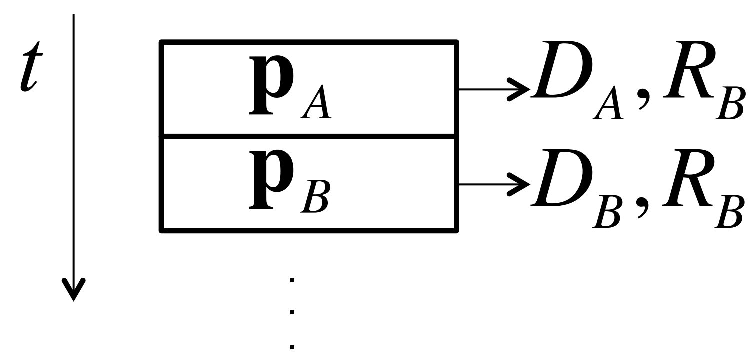
  Time <= T, ...



Learn to Encoder Program

- Video encoder needs to find a good parameter vector ***p*** fast (e.g. real time encoding)

Encoder program:



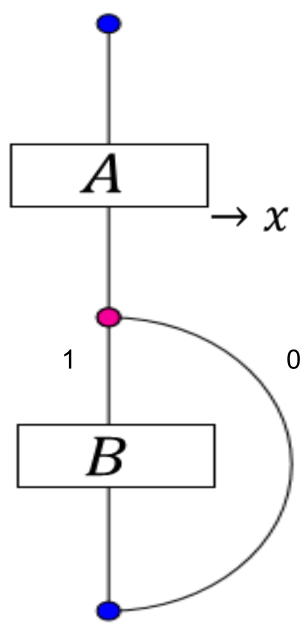
- Calculating D,R values takes time
- Trade-off between rate, distortion and computational complexity

Construct a learning problem

- Continue or terminate the search for a better R,D?
- As cost, use RD-cost $J = D + \lambda R$ and time T
- Base decision on known information x (features)

$J_0 = J_A$
$T_0 = 0$

$J_1 = \min\{J_A, J_B$
$T_1 = T_B$



- $C_k = J_k + \mu * T_k$ Cost for decision k
- $y = I\{C_1 < C_0\}$ Optimal decision (I is indicator func.)

- Collect data $(x, J_A, J_B, T_B)_{m=0..M-1}$ from the encoder

- construct a **supervised learning problem**, i.e. find a function predicting y from x.

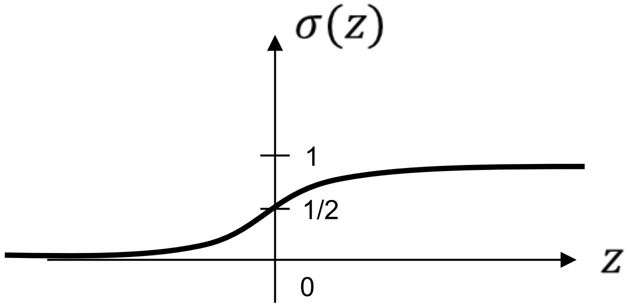
$$\hat{y} = f(x)$$

Binary classification

- The target is $y \in \{0,1\}$, we have a **binary classification** problem
- Use logistic regression to find f.
- As hypothesis, logistic regression uses a linear combination of features $\theta^T x$, surrounded by the non-linear **logistic function** σ:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

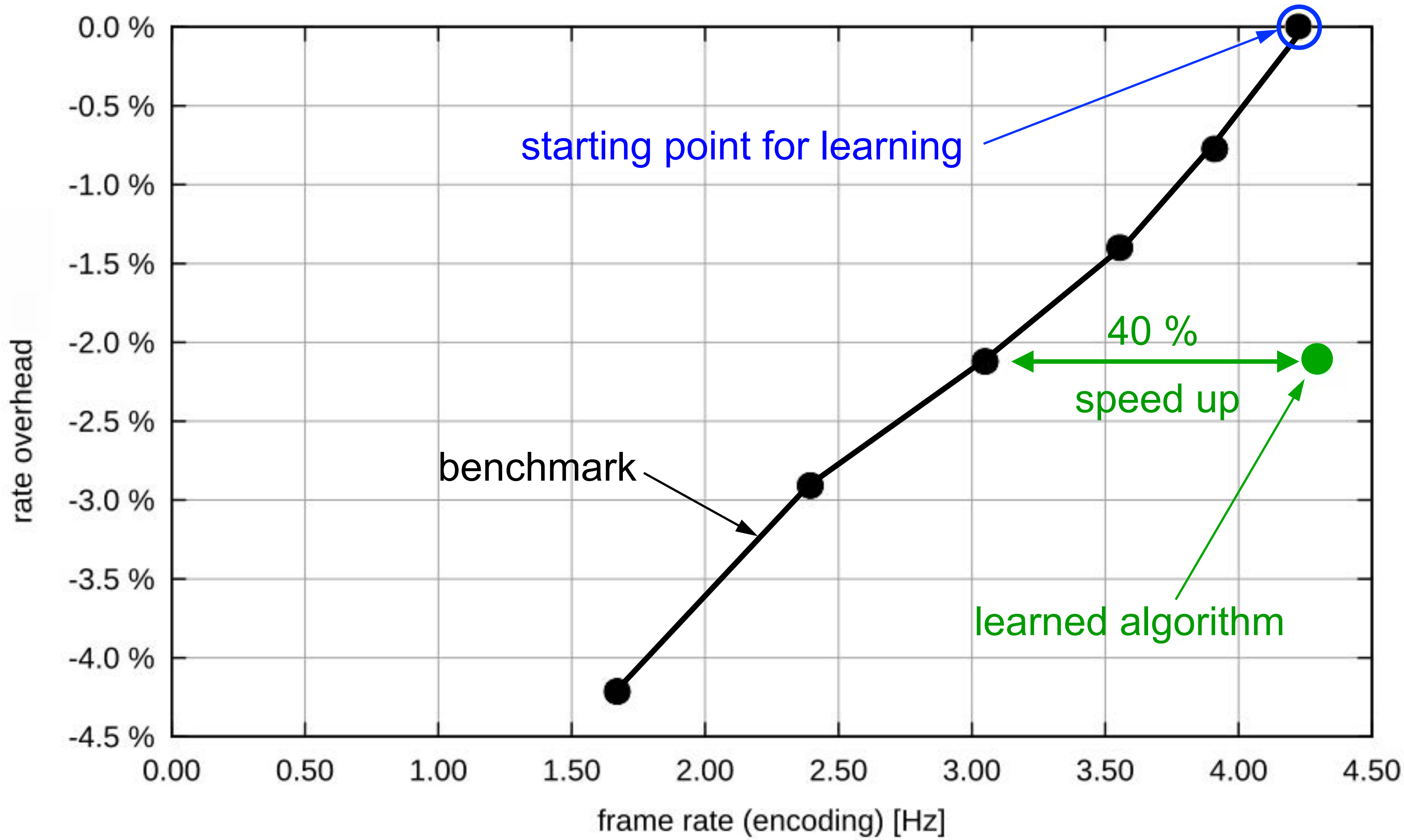
$$f_\theta(x) = \sigma(\theta^T x)$$



- The hypothesis is continuous: $f_\theta(x) \in [0,1]$
- **Interpretation**: $f_\theta(x)$ is an estimate of the probability that $y = 1$.

$$f_\theta(x) = P(y = 1 | x; \theta)$$

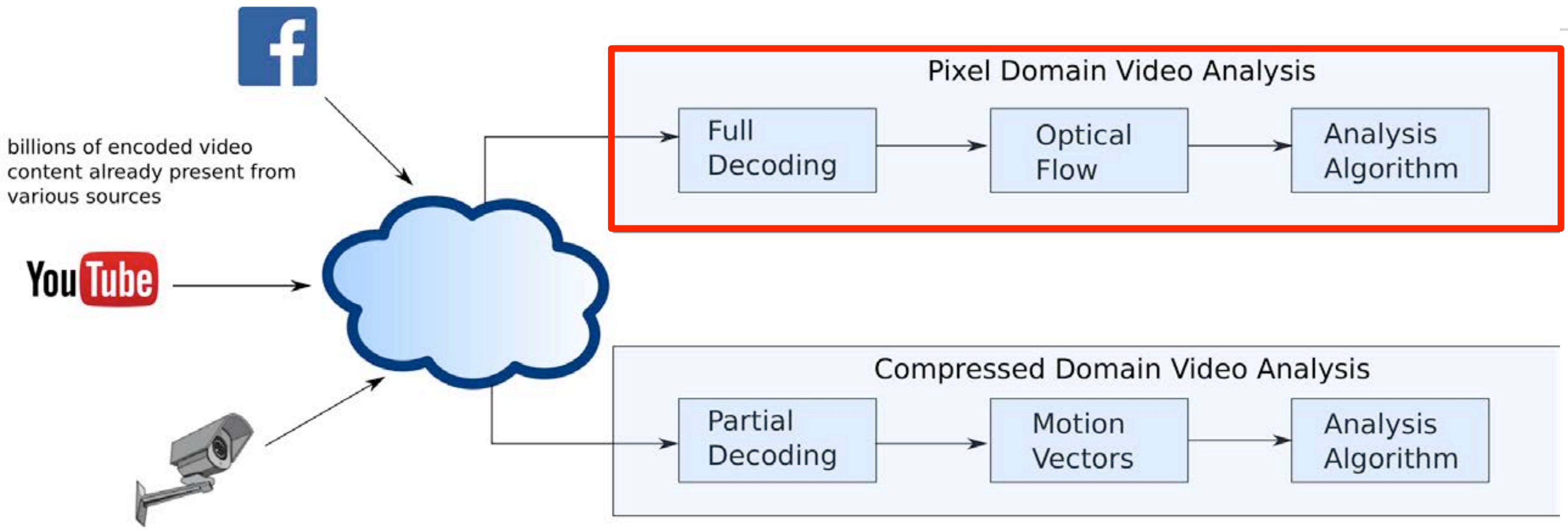
First Results: Fraunhofer HHI H.265 Encoder



Compressed-Domain Video Analysis

Fraunhofer
Heinrich Hertz Institute

Compressed Domain Video Analysis



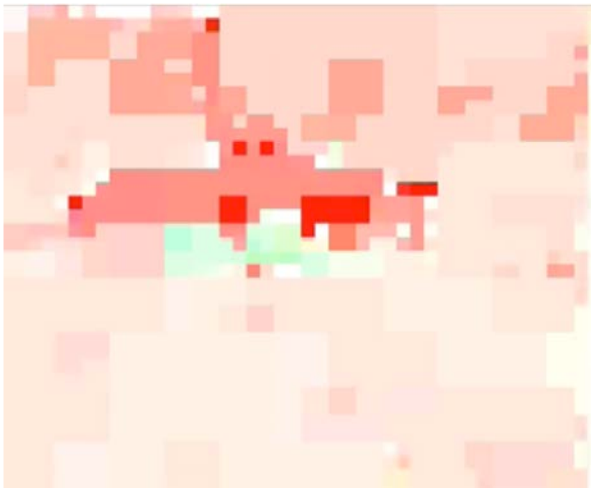
Conventional video analysis in **pixel-domain:**
Full decoding + processing on pixel levels

High complexity and storage requirements: a bottleneck for real-time analysis of multiple video streams

Billions of videos already stored in **compressed** form !

Compressed Domain Object Tracking

- Spatio-temporal Markov Random Field (ST-MRF) model the evolution of the MV field [Khatoonabadi14]
- In compressed domain – Motion Vectors available



→Motion vectors may be ambiguous.
→Use hybrid approach with inclusion of I Frames

Compressed Domain Object Tracking

Hall Monitor

Motion vectors (HEVC)

Optical flow [Brox04]

Coastguard

Motion vectors (HEVC)

Optical flow [Brox04]

Tracking accuracy (%):

		MV	MV+I	OF
Coastguard	Precision	55.9	63.2	61.8
	Recall	90.9	89.6	94.1
	F-Measure	68.6	73.3	73.3
Hall Monitor	Precision	69.6	77.9	79.1
	Recall	79.4	72.6	85.6
	F-Measure	74.0	74.9	81.2

- Higher tracking performance with OF input
- MVs only show performance degradation
- MVs + I comparable performance

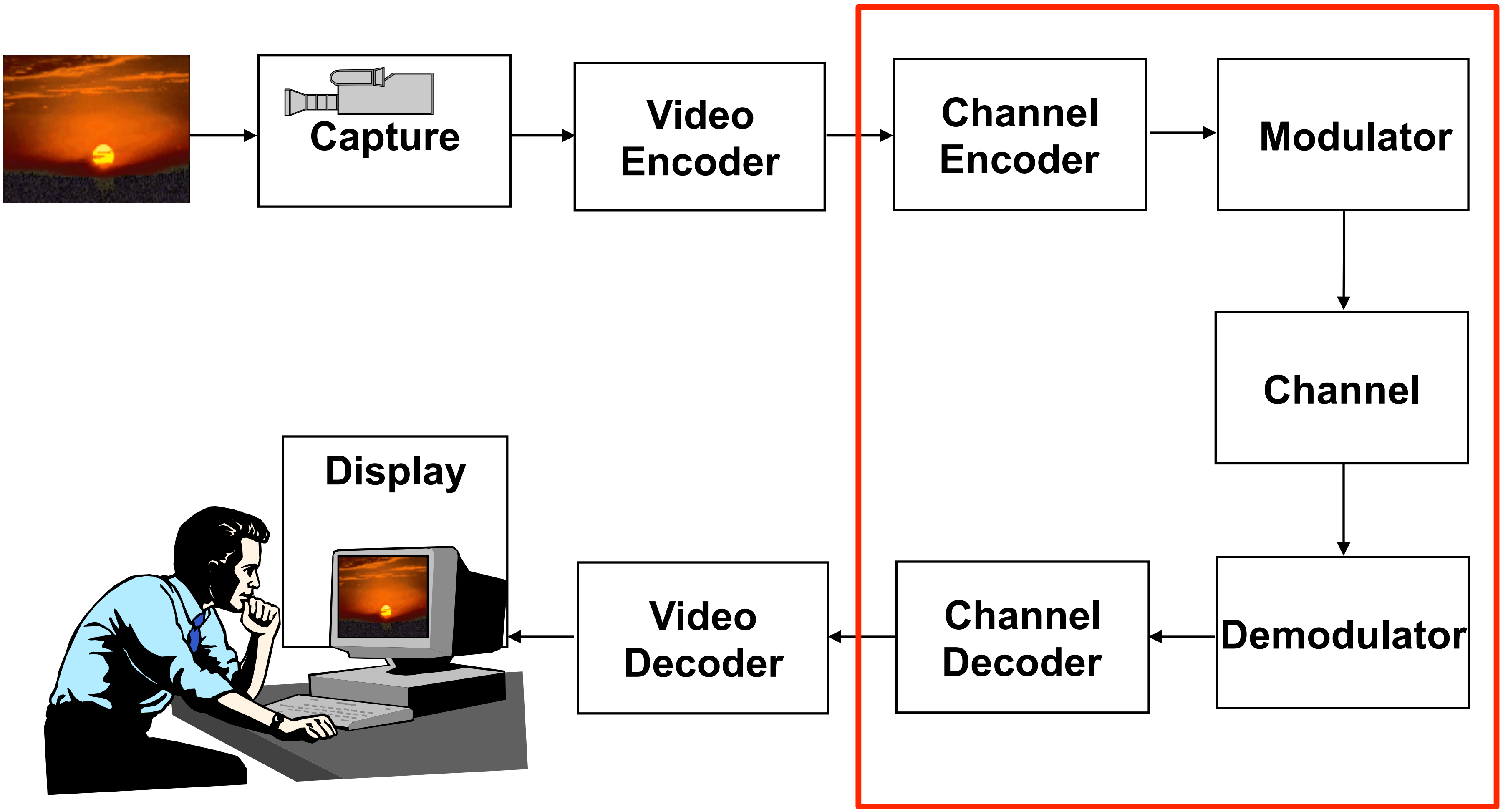
(TP) true positives
(FP) false positives
(FN) false negatives

Machine Learning and Data Communication

TU berlin

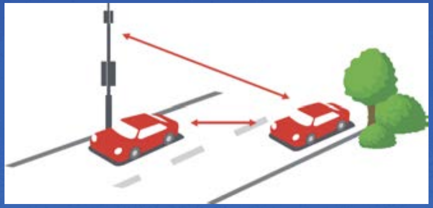
Fraunhofer
Heinrich Hertz Institute

Visual Communication Systems

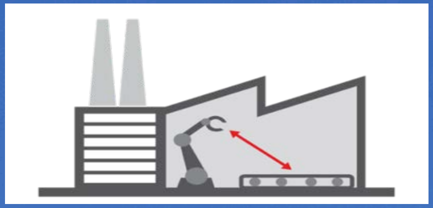


The Next Generation: 5G Network

Mobile High Speed Internet

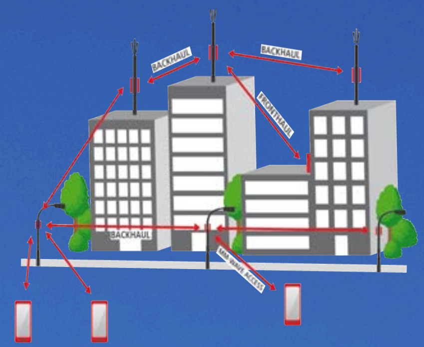


Car2Car & Car2X Communications



Industrial Wireless

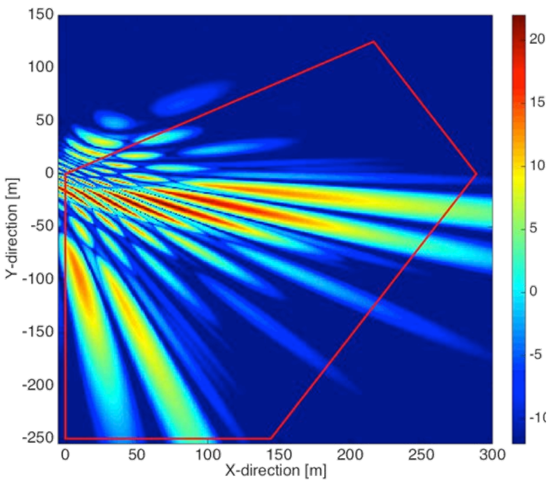
Requirements
- 1000 x throughput
- 100 x devices
- 10 x battery life
- 1 ms latency



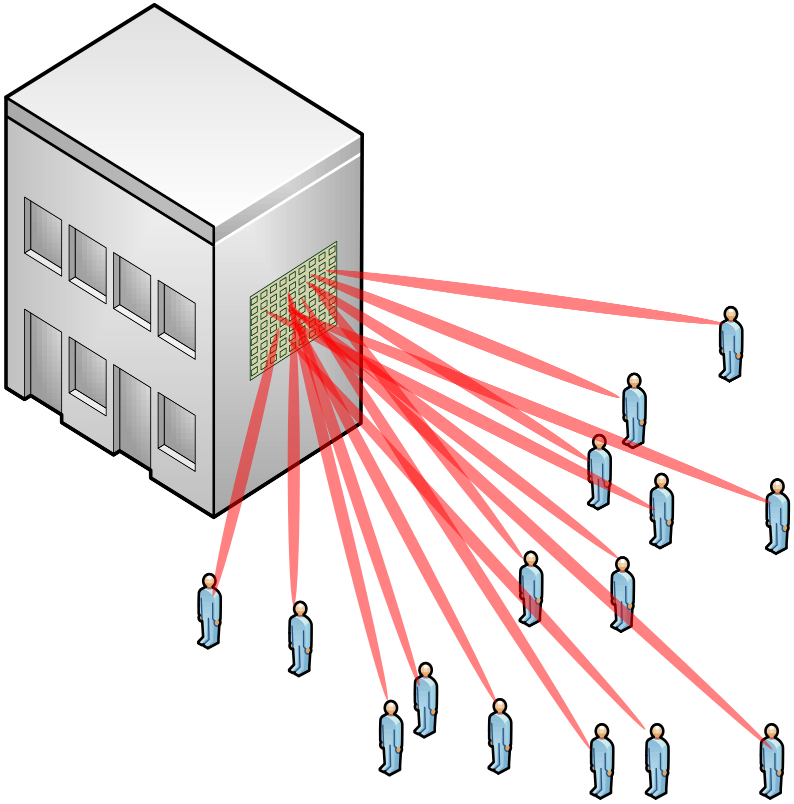
Technology
- DSL boxes and street lights become senders
- Optical fiber

Wireless Fiber and Location Sensing

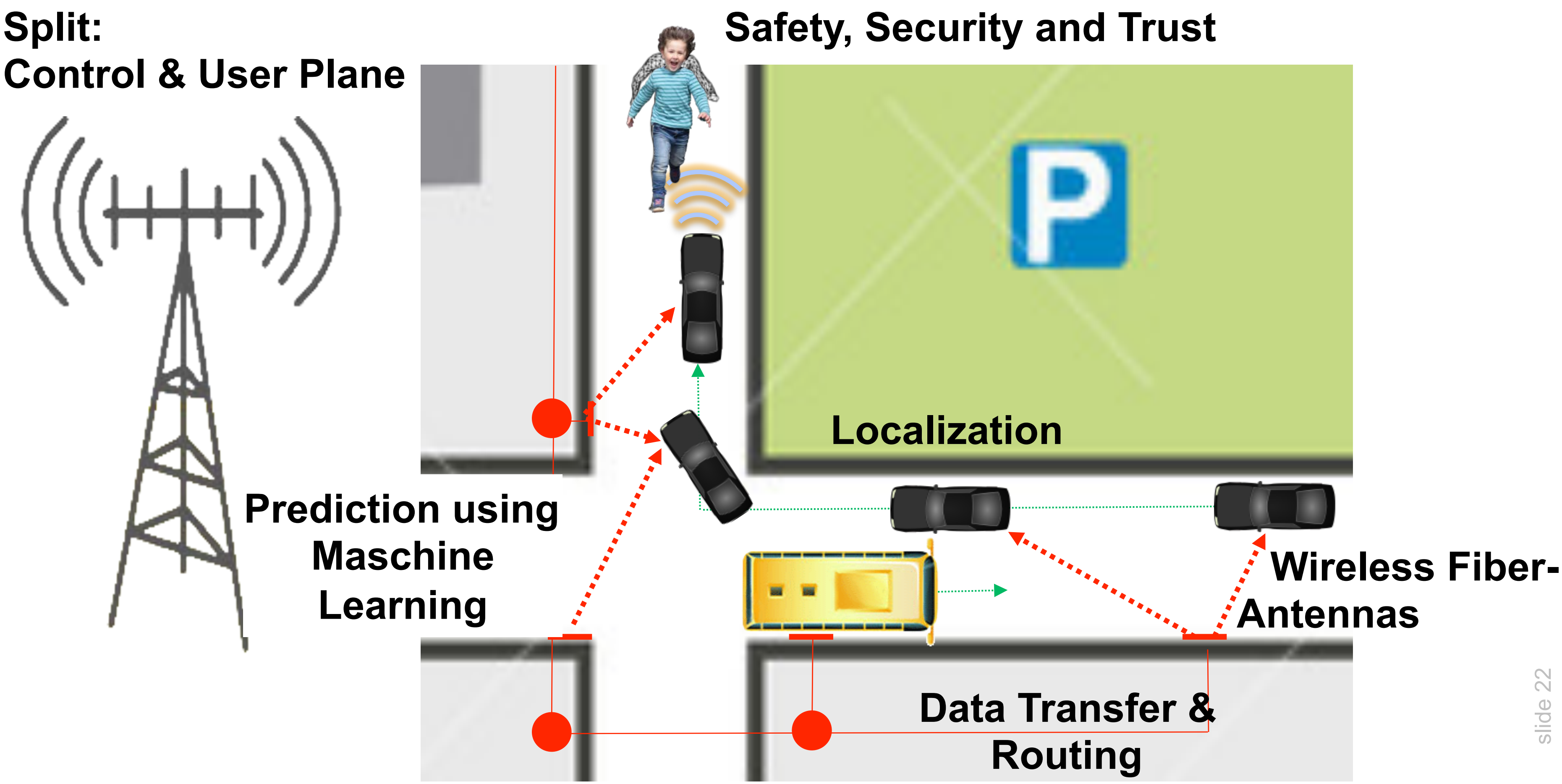
- 3D beamforming with MIMO Antennas



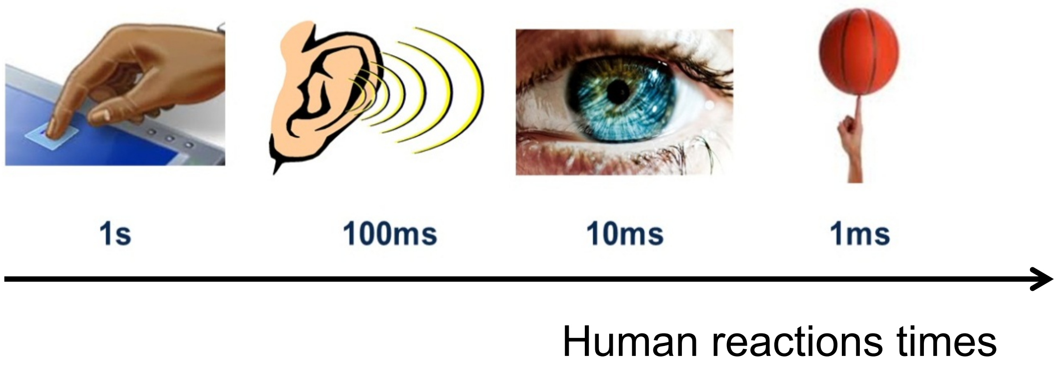
- Location of users via sensors



Future Mobile Digital Infrastructure Example: Networked Autonomous Driving



The Tactile Internet



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

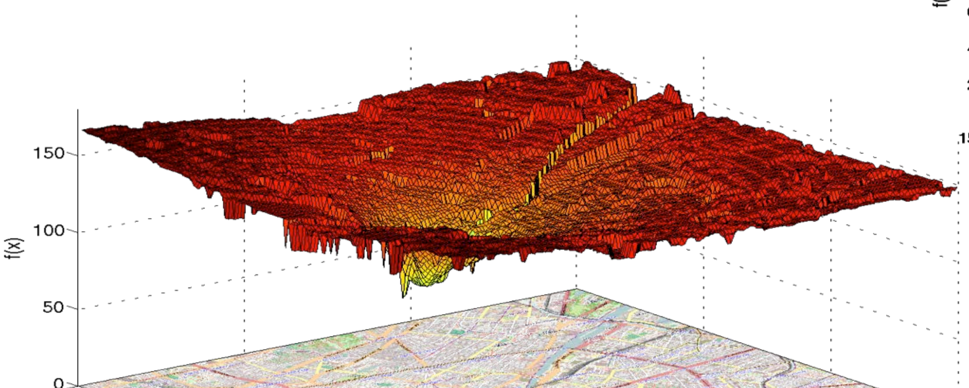
Driver assistance with AR of potentially dangerous objects and situations

Cognitive Network Management

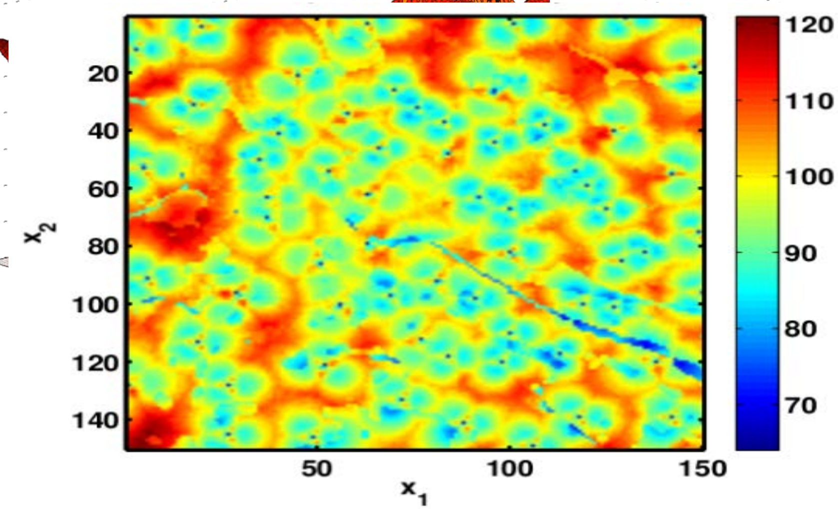
- Develop awareness at the node level (e.g. nodal knowledge about network state) through cognition, real-time (machine) learning and stochastic control amidst network uncertainties
- Bring the awareness into the self-management loop to enable autonomic network operation via distributed adaptive (multi-objective) optimization and in-network processing
- Enhance network reliability and robustness by coping with resource and objective conflicts
- Counterfeit malicious and abnormal behavior through distributed fault diagnosis and network response mechanisms towards nullifying the malignant effects in the network

Learning of Radio Maps

- Radio map: **unknown** function f(x) that relates a geographic location x to a radio system parameter (e.g. path-loss)
- Path-loss map for one base station



- Path-loss map where each location is related only to the base station with lowest path-loss
- 2D view:

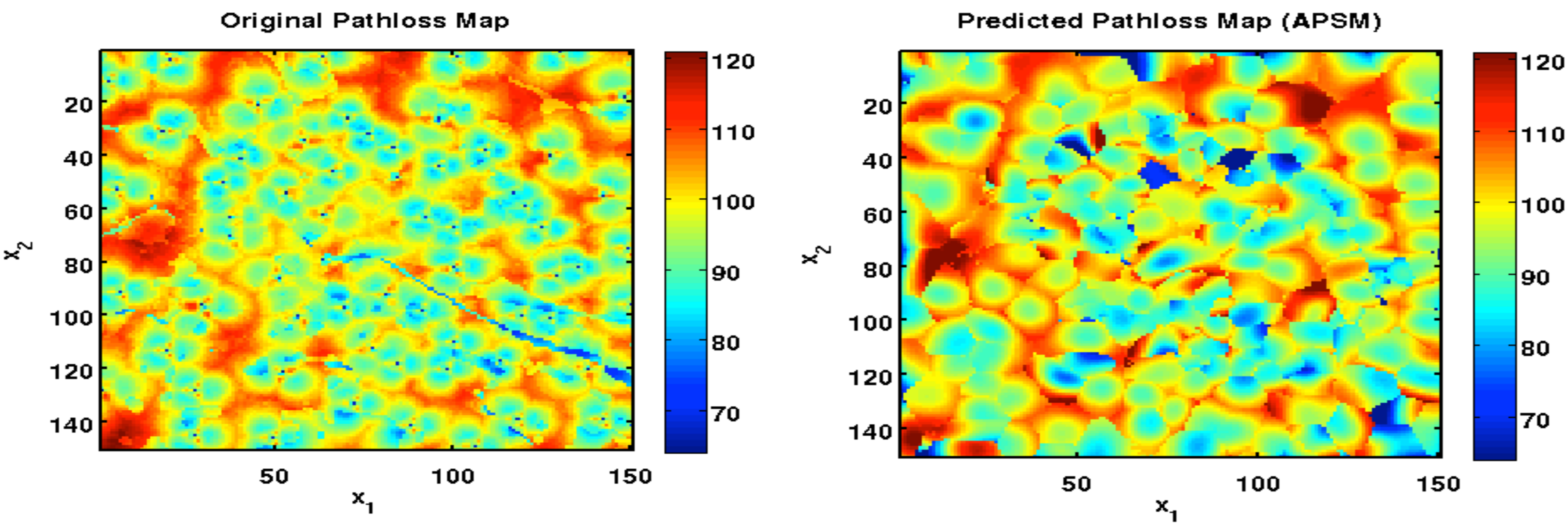


Goal: Online reconstruction and prediction of radio maps from user measurements

Example: Path-loss Map Reconstruction

Berlin path-loss data (real measurement data):

- Size of area: 150x150 pixels, each pixel is an area of size 50x50 meters
- 187 base stations (BS)
- For each BS, there is path-loss data from the BS to each pixel
- Cells are defined by assigning each pixel to a BS with lowest path-loss



M. Kasparick et.al., "Kernel-Based Adaptive Online Reconstruction of Coverage Maps With Side Information," in *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5461-5473, July 2016

Interpretable Machine Learning

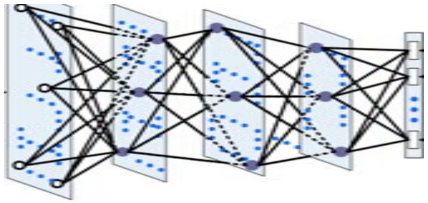
Classification using Machine Learning

Big Data

14.2 Million images, 22.000 classes

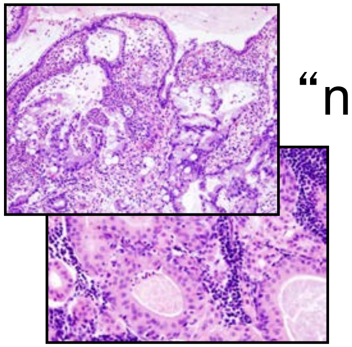
+

Machine Learning



=

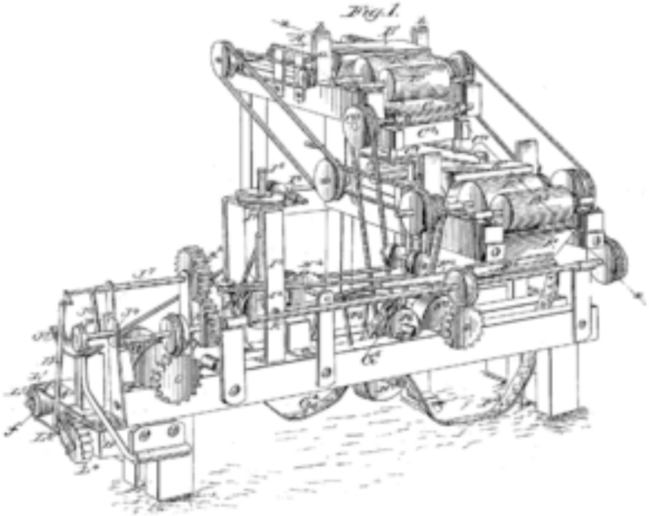
Automatic Annotation



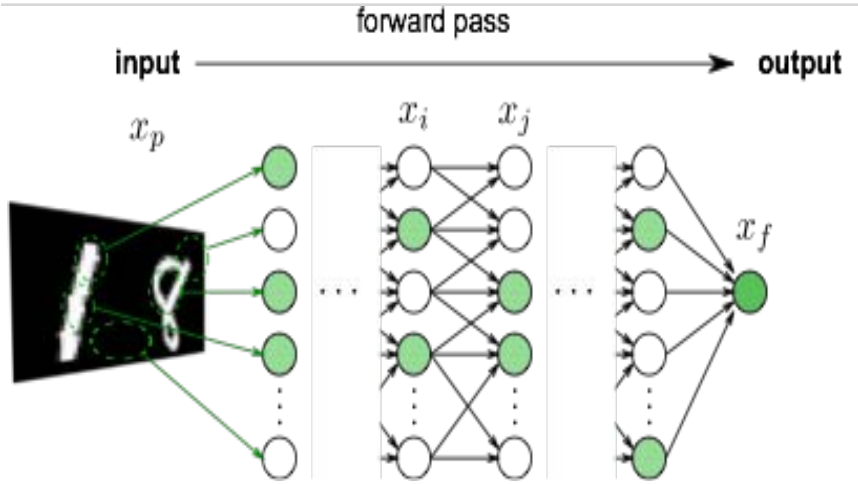
"no cancer"

"cancer"

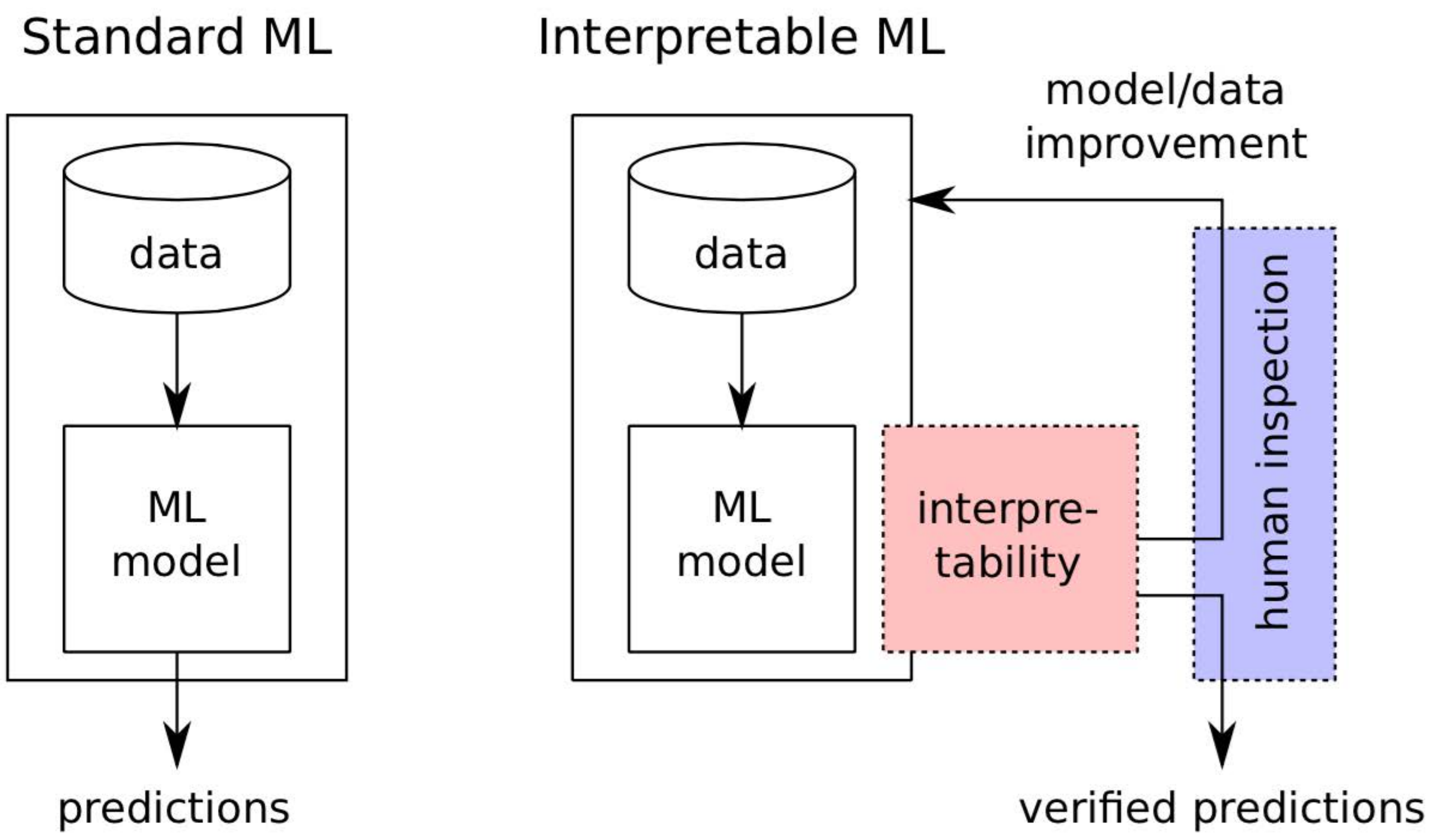
Do we trust the machine ???



Revert the Deep Neural Network



Interpretability of Machine Learning

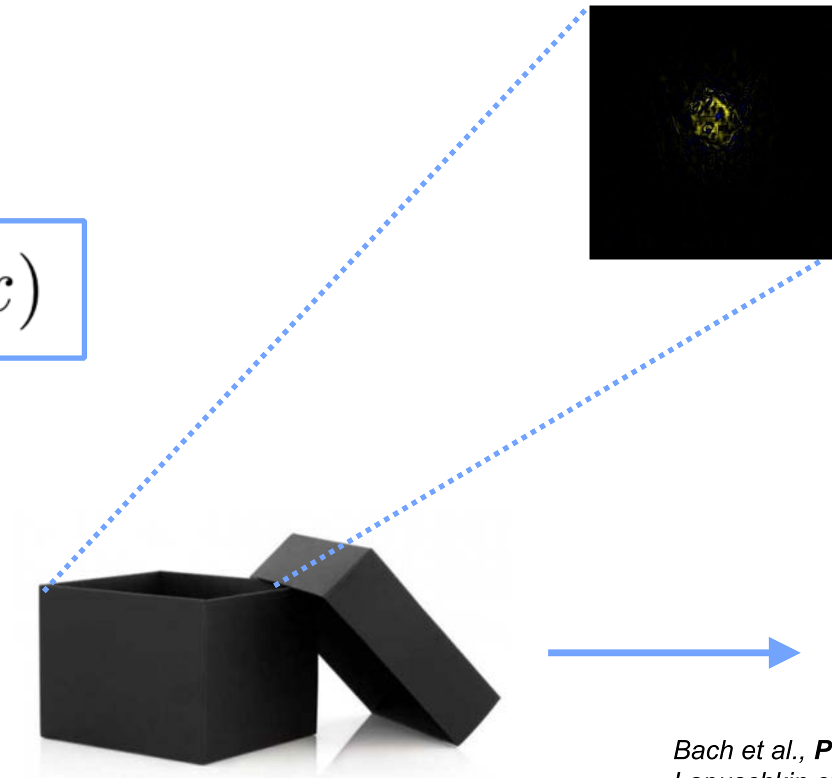


Interpretability is first step towards making sure (i.e. verifying) that ML algorithms do the right thing !

Idea for Interpretable Machine Learning

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

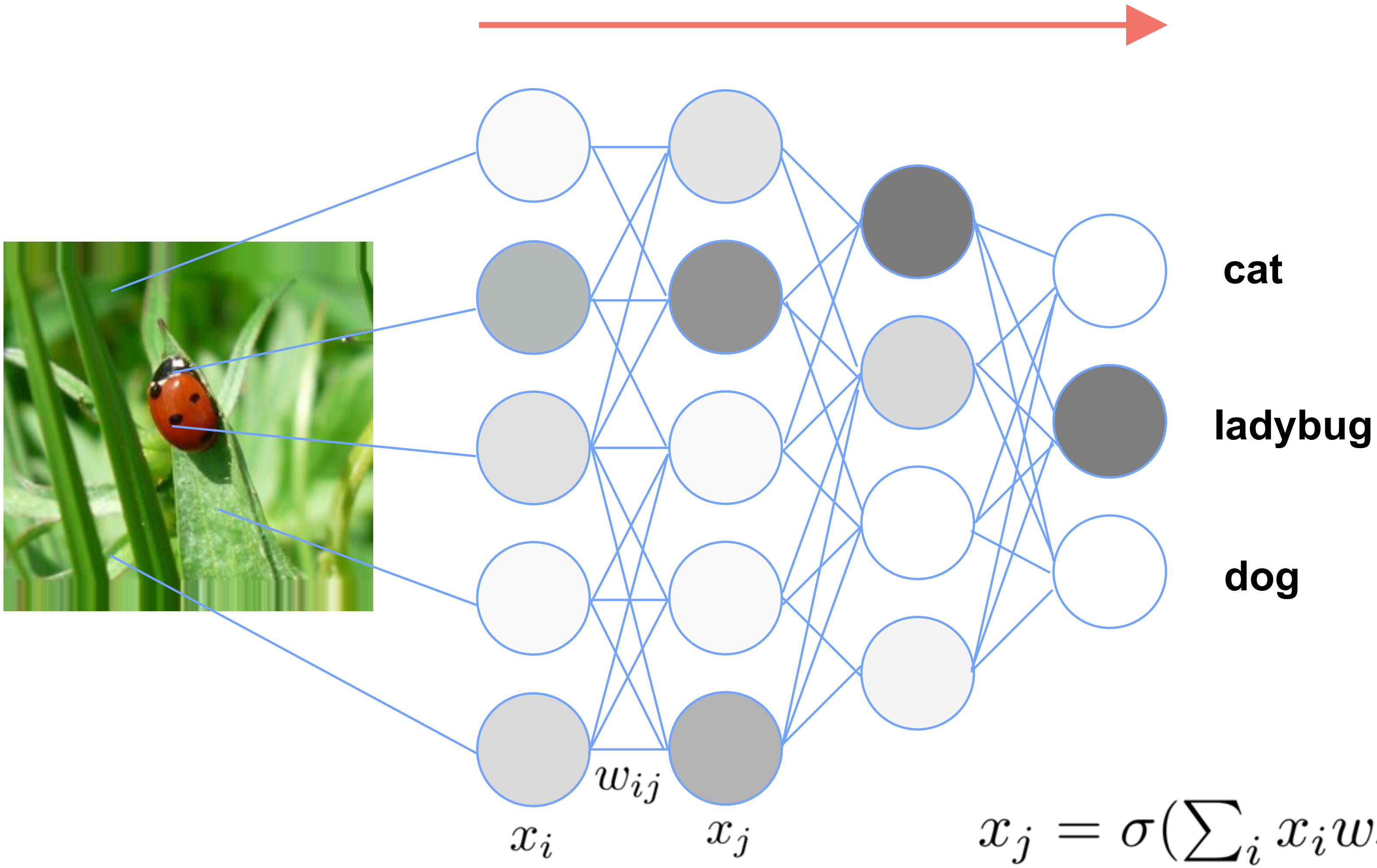
Main idea: $\sum_p r_p = f(x)$



"ladybug"

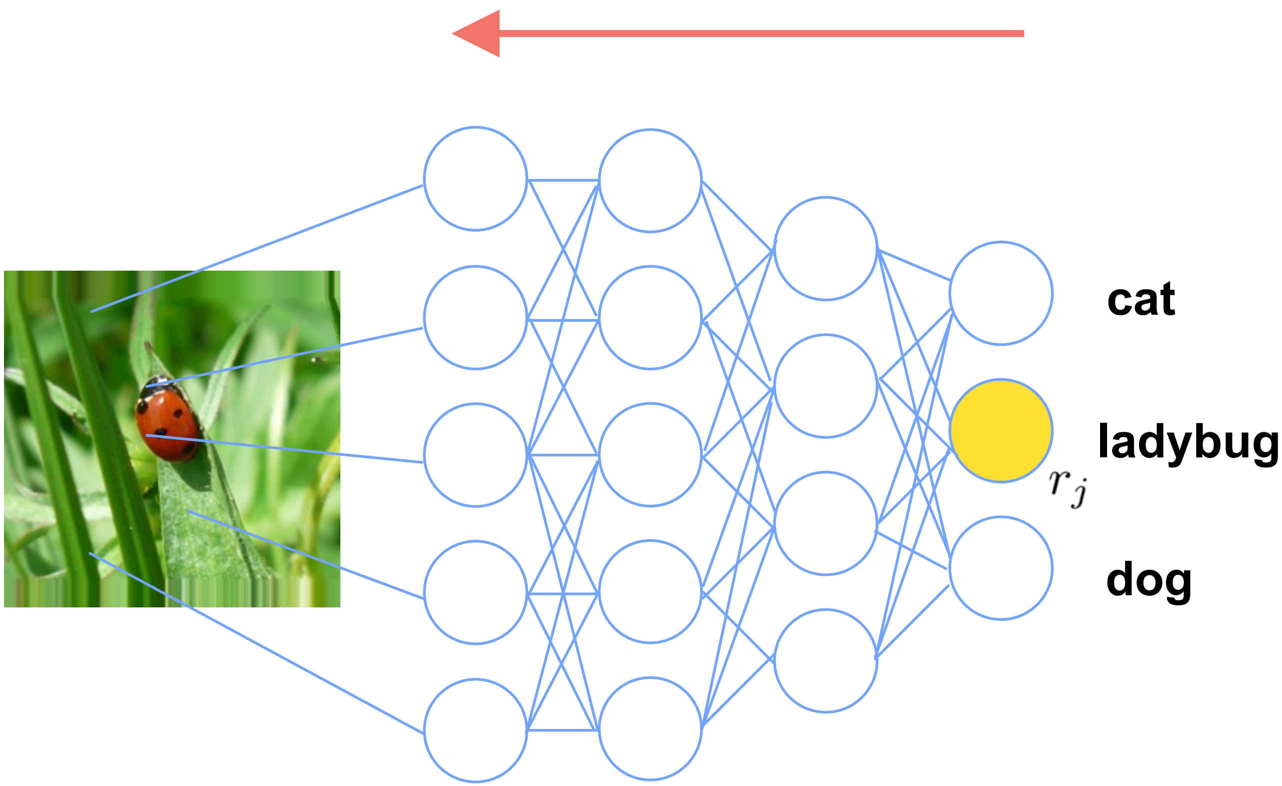
Bach et al., ***PLOS ONE****, 2015*
Lapuschkin et al., ***CVPR****, 2016*
Samek et al., ***TNNLS****, 2016*
….

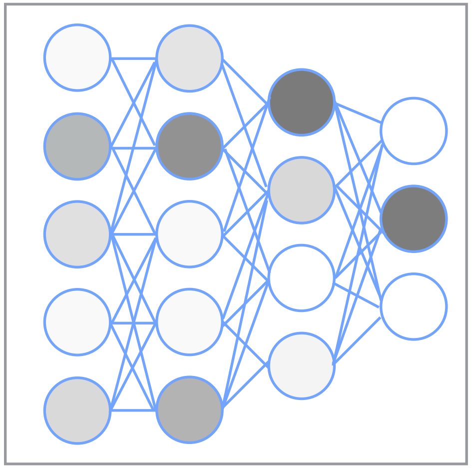
Classification



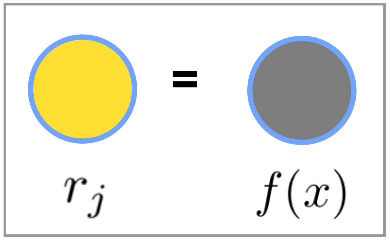
$$x_j = \sigma(\textstyle\sum_i x_i w_{ij} + b_j)$$

Explanation

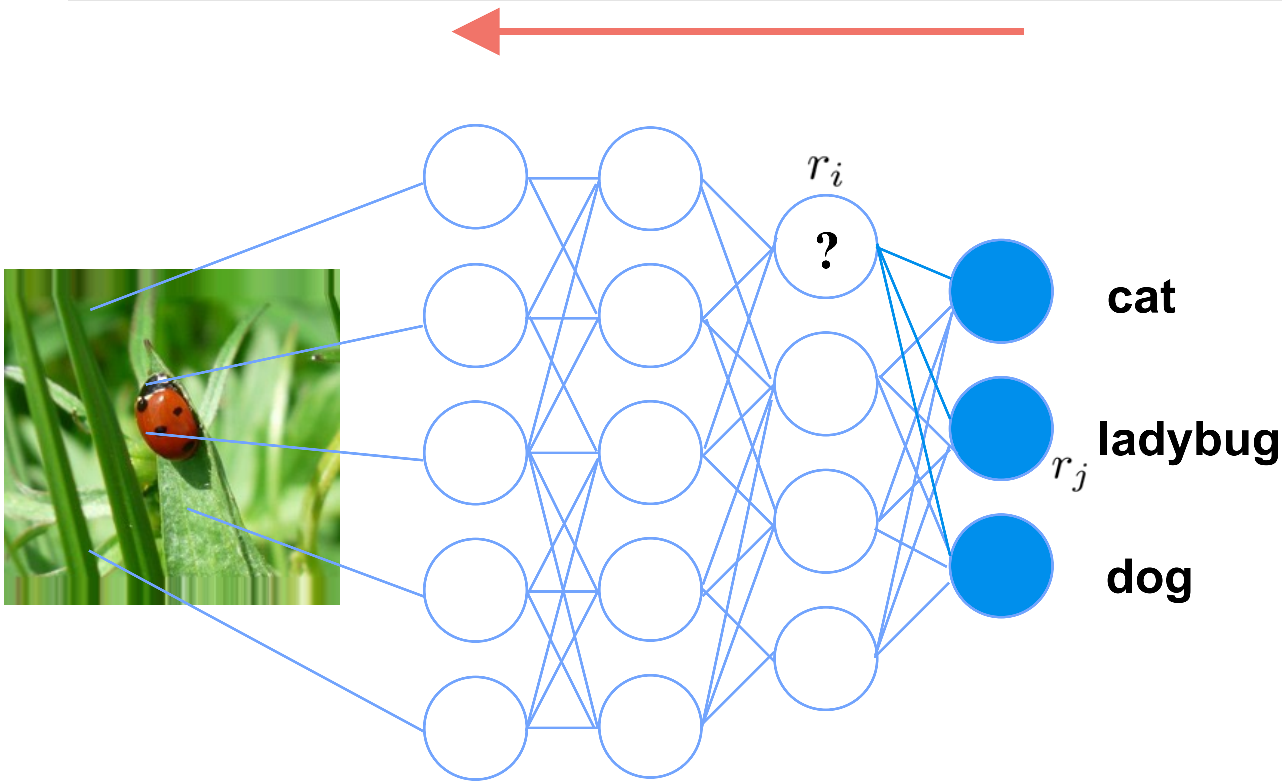


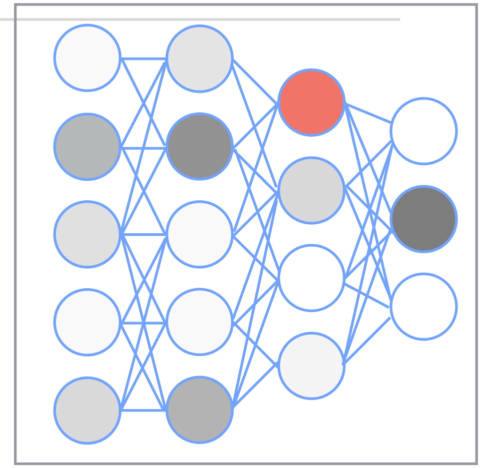


Initialization



Relevance Propagation





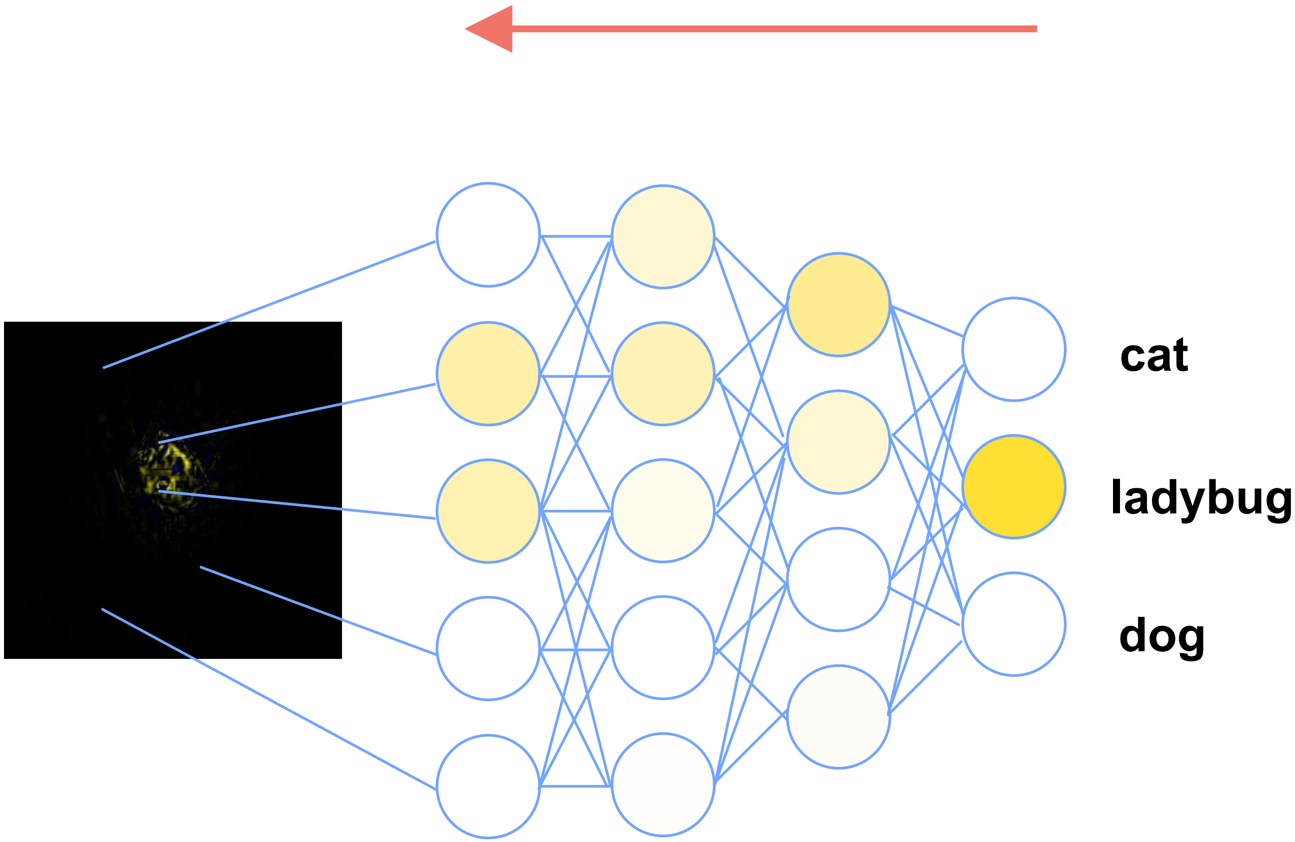
$$r_i = x_i \sum_j \frac{w_{ij} r_j}{\sum_i x_i w_{ij}} = x_i c_i$$

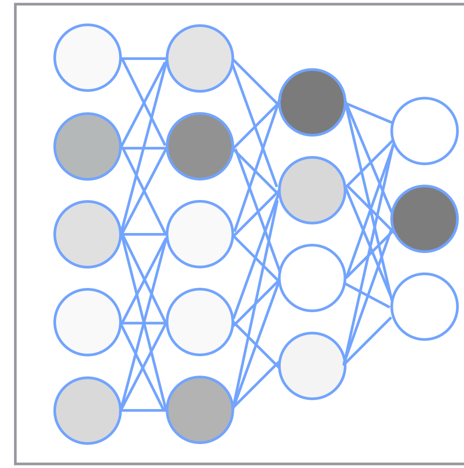
Theoretical interpretation
(Deep) Taylor Decomposition

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

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

Relevance Conservation Property





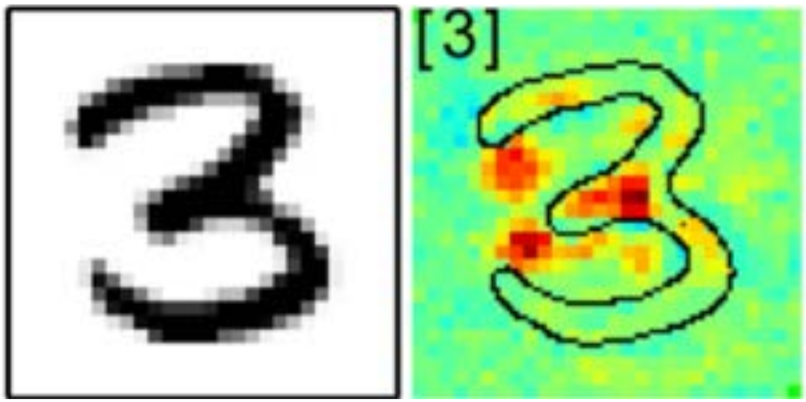
Relevance Conservation Property

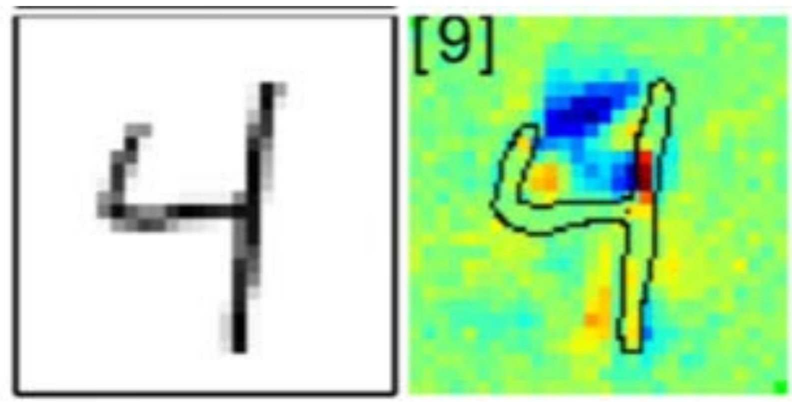
$$\sum_p r_p = \ldots = \sum_i r_i = \sum_j r_j = \ldots = f(x)$$

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 communication
- **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
- S. Stanczak, T. Haustein, M. Kasparick
- …

Supported by Federal Ministry of Economics and Technology

SPONSORED BY THE Federal Ministry of Education and Research

Deutsche Forschungsgemeinschaft DFG