Zürcher Hochschule
für Angewandte Wissenschaften

A Winterthur Perspective on Data Science and Machine Learning

Thilo Stadelmann, 29th of January, 2016

www.zhaw.ch/datalab

Zürcher Fachhochschule

(explained later upon request)

Zürcher Hochschule
für Angewandte Wissenschaften

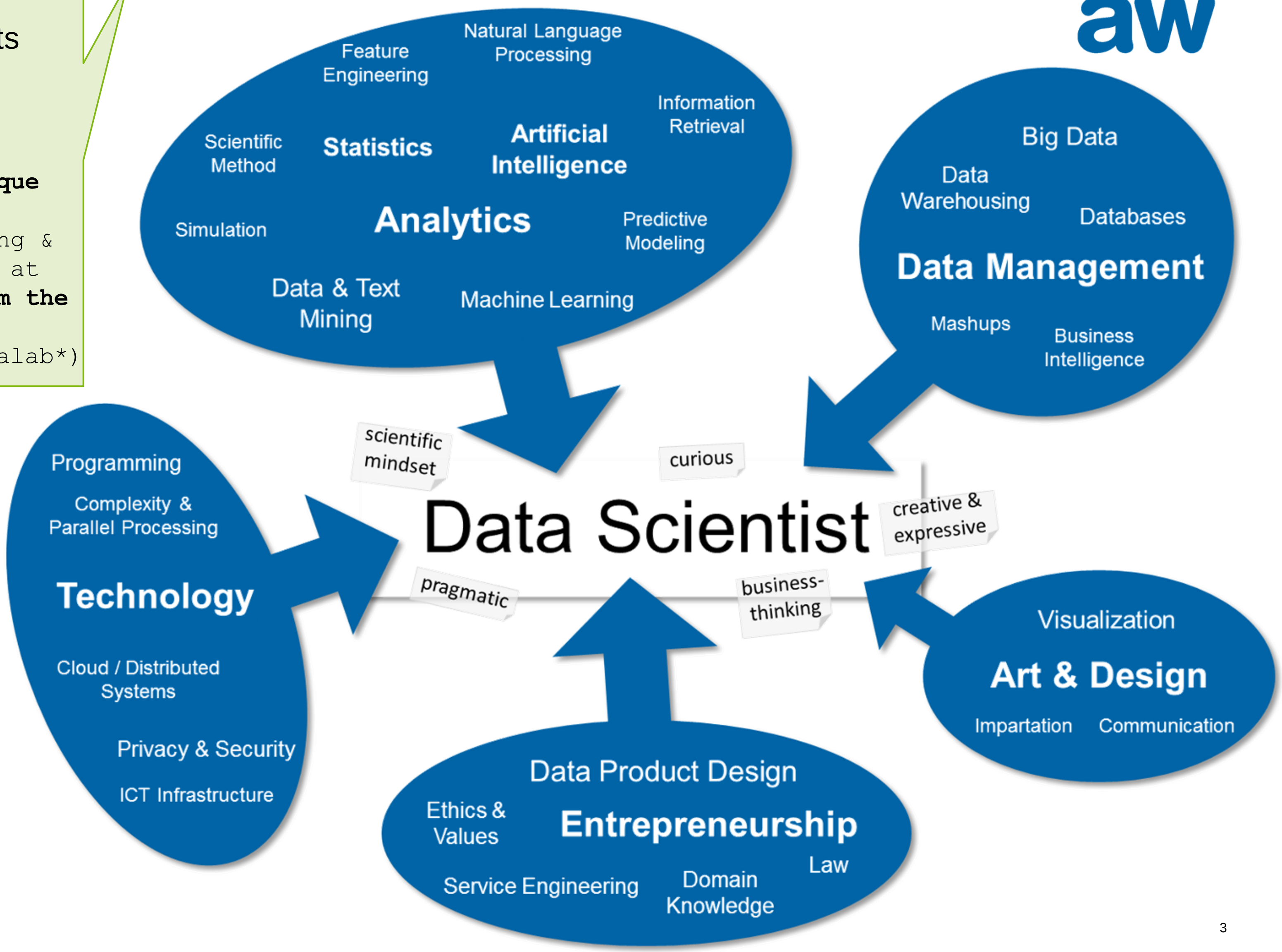
1

The ZHAW Data Science Laboratory

What is Data Science?

Enables Data Products
➔ Applied Science
➔ Interdisciplinary

Data Science := "**Unique blend of skills** from analytics, engineering & communication aiming at generating **value from the data** itself […]"
(ZHAW Datalab*)



*) Stadelmann, Stockinger, Braschler, Cieliebak, Baudinot, Dürr and Ruckstuhl (2013). *Applied Data Science in Europe* . ECSS 2013, Amsterdam.

A Personal Story

- Fascinated by AI
- Studied computer science
- Researched ML & IR during Ph.D.
- Used DWH & DM professionally

- Difficult to briefly explain professional interests

➔ Excited about term «Data Scientist»

ZHAW Datalab: Est. 2013

Zürcher Hochschule für Angewandte Wissenschaften

Forerunner

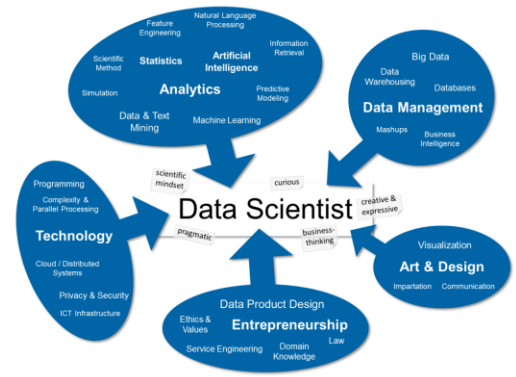
- **One of the first** interdisciplinary data science initiatives in Europe
- One of the first interdisciplinary centers at ZHAW

Foundation

- **People**: ca. **60 researchers** from **5 institutes** / **3 departments** opted in
- Vision: **Nationally leading** and **internationally recognized** center of excellence
- Mission: **Generate projects** through critical mass and mutual relationships
- Competency: **Data Product design** with structured and unstructured data

Success factors

- **Lean** organization and operation → geared towards projects
- Years of successful **pre-Datalab collaboration**



Zürcher Hochschule
für Angewandte Wissenschaften

Education

Undergraduate

- Involved in numerous courses of B.Sc. Programs: e.g., «Scripting» for industrial engineers

Graduate and post-graduate

- Several **M. Eng. modules**: Machine Learning and Predictive Modeling in Swiss-wide program
- Actively seeking collaborations: Ambitions for Data Science **M.Sc. and Ph.D. programs**

Professional education

- Diploma of Advanced Studies (**DAS**) **in Data Science**
- Planned **Master of Advanced Studies (MAS)** in Data Science
- **Sole technical** oriented data science program in Switzerland!
- **Completely booked** for fall 2016, few free seats for fall 2017

Community Outreach

data service
Swiss Alliance for Data-Intensive Services

Generating impact

- Leader of **National Initiative: Swiss Alliance for Data-Intensive Services**
- Workshop organization: e.g. SwissText 2016 (Swiss conference on text understanding)
- Keynotes: e.g. IBM Business Connect 2013, SwissICT 2014, **SAS Forum 2016**
- Overview publications: e.g. **book on applied data science** (to appear with Springer)

SDS – Swiss Conference on Data Science

- SDS|2014: ca. 120 participants (planned 60)
- SDS|2015: ca. **190 participants**, ca. 45’000 CHF budget
- SDS|2016: planned, ca. **79’000 CHF budget**, several international keynote speakers invited

R&D

Volume

- \> 450’000 CHF in first half year
- \> 3.3 Mio. CHF since foundation
- **Overall turnover** of Datalab projects:
  > 8 Mio. CHF in **< 3 years**

Spin-offs

- Prognosix – a ZHAW IAS spin-off
- SPINNINGBYTES – a joint spin-off from ZHAW and ETH Zurich

Topics

- E-Health (e.g. CTI application «SenSkin»)
- Industry 4.0 (e.g., CTI project «DaCoMo»)
- FinTech (e.g., CTI application «DatFisMo»)
- Mobility (e.g., project «Placebook»)
- Sustainability (e.g., CTI project «EAT-IT CO_2»)
- Technology (e.g., CTI project «Zurich NoSQL»)
- …

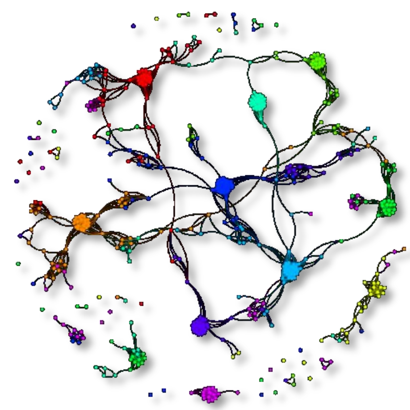


Figure: Visualizing the relationships of all Swiss foundations, based on the similarity of goals as expressed in their statutes. A proud collaboration of InIT and IDP within CTI project «Stiftungsregister SR 2.0»

Swiss Confederation
Commission for Technology and Innovation CTI

HASLER STIFTUNG

FNSNF

3 Years of Datalab: Lessons Learned

PRO

- We **backed the right horse**: Buzzwords come and go, disciplines stay
- **Lean** is beautiful: create an opportunity space, and that’s it
- A «**coalition of the willing**» helps settle power issues in interdisciplinary, matrix-like structures

CON

- Interdisciplinary networks **stretch official processes** and structures
- Challenging to **engage a larger number** of coworkers

Zürcher Hochschule
für Angewandte Wissenschaften

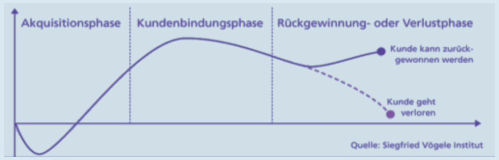
2

Project Examples – A Deep Learning Story

Zürcher Hochschule
für Angewandte Wissenschaften

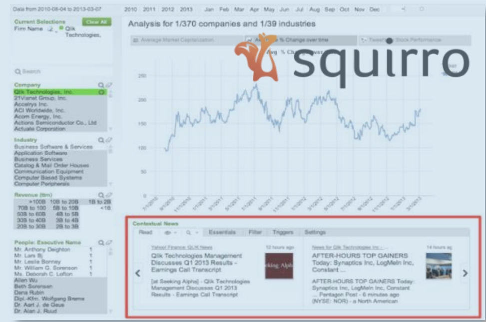
Some Datalab projects

Predicting churn (marketing)

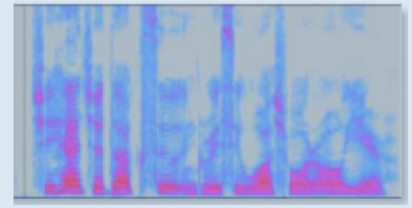


swisscom

Enterprise Knowledge Curation



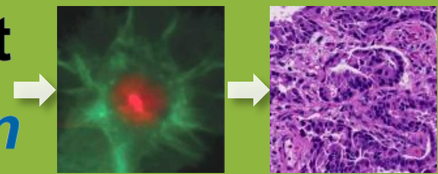
Talkalyzer (voice recognition)



Deep Leaning

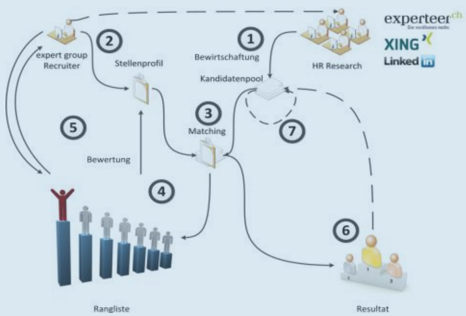
Project

πVision

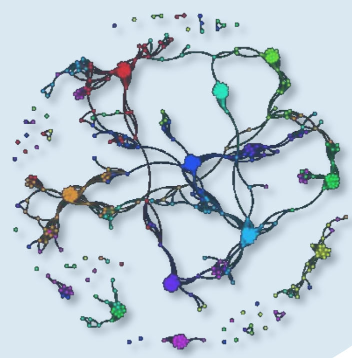


…and several created spin-offs

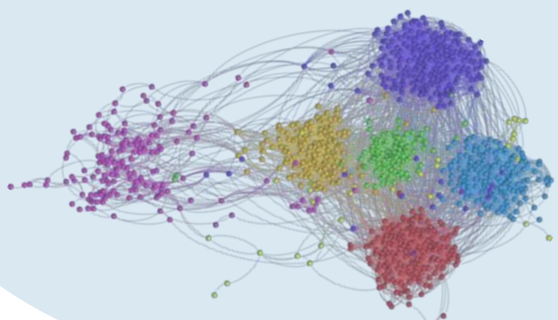
Expert Match (talent search)



Foundation Register 2.0

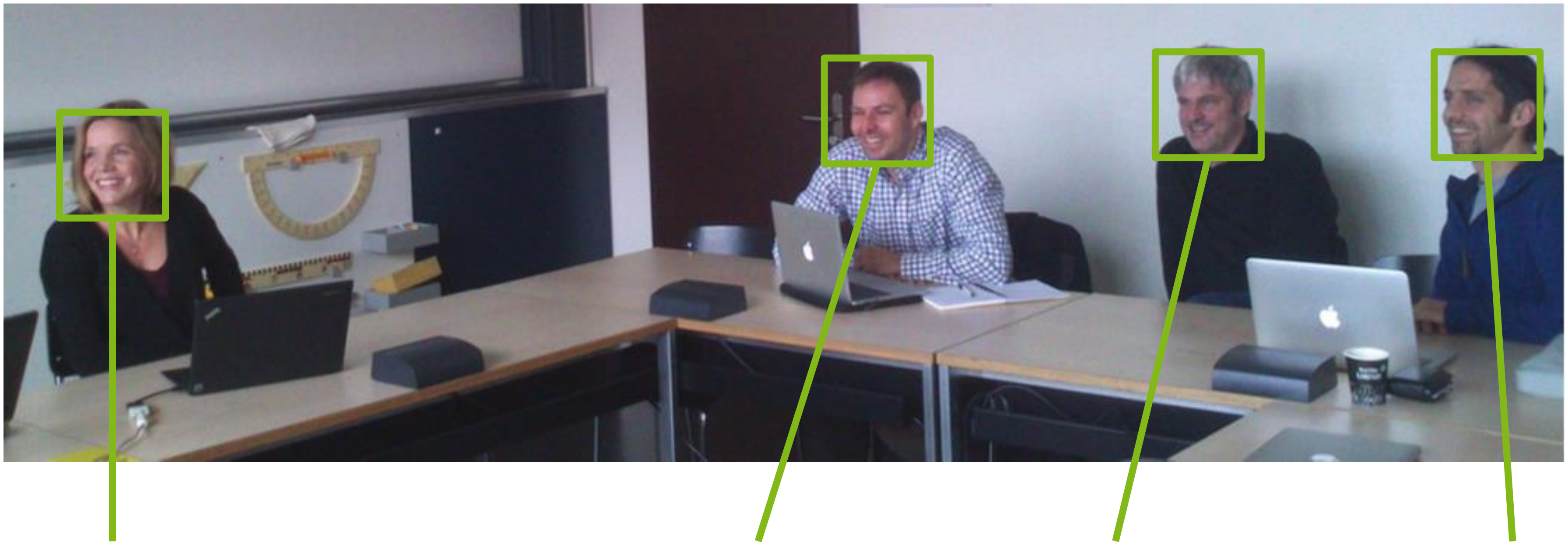


Influencer Detection in Social Media



Zürcher Hochschule
für Angewandte Wissenschaften

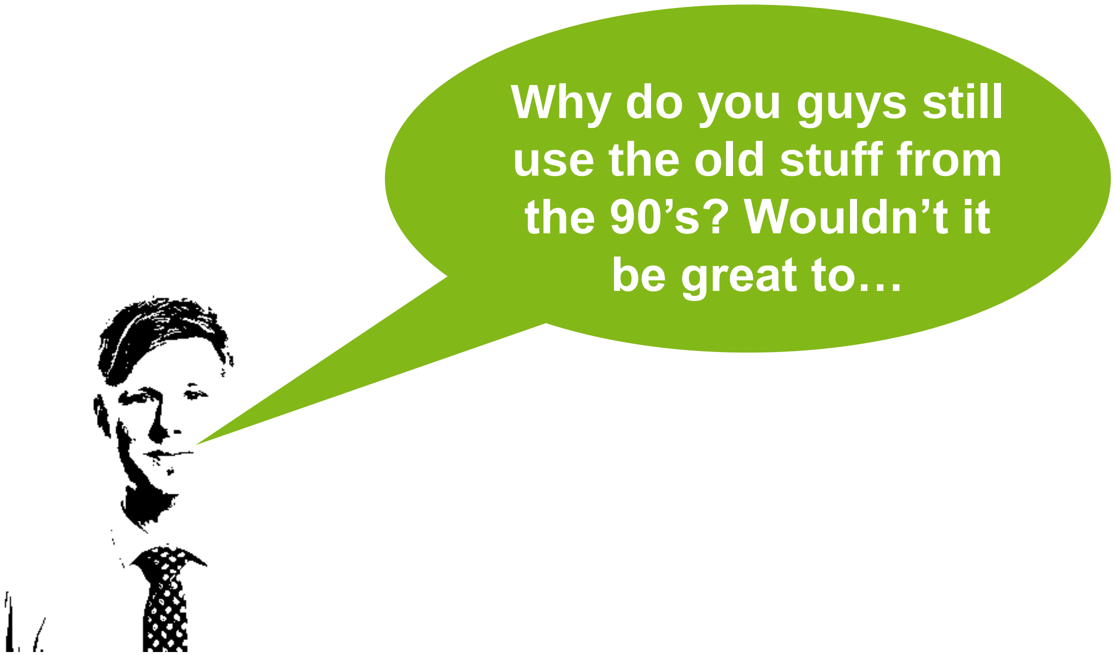
Beginning: The π-Vision Project
(internal seed funding)



Idea: Bring face recognition to a Raspberry Pi, foster exchange and knowledge.

Introducing Deep Learning

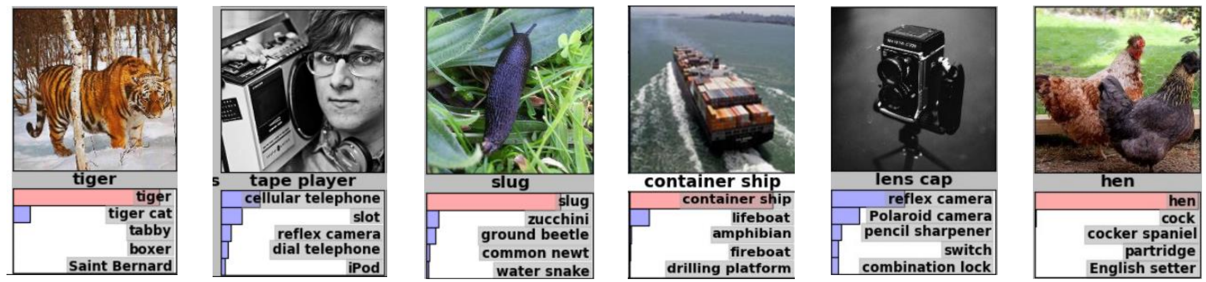
Half way through the project, during a cup of tea in the cafeteria…



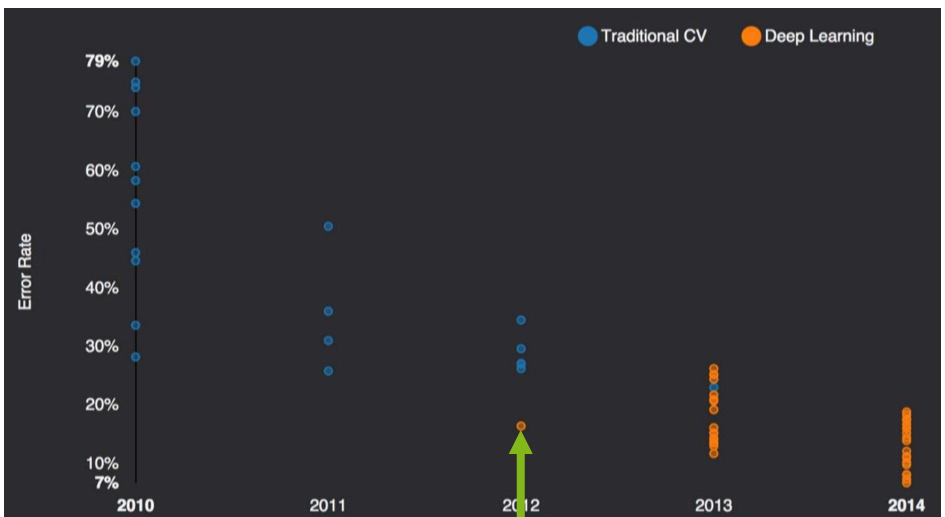
Background I: …what is he talking about?

The ImageNet competition

1000 Classes
1 Mio. samples



…



A. Krizhevsky uses first CNN in 2012.
Trained on Gaming Graphic Cards

2015: It gets tougher

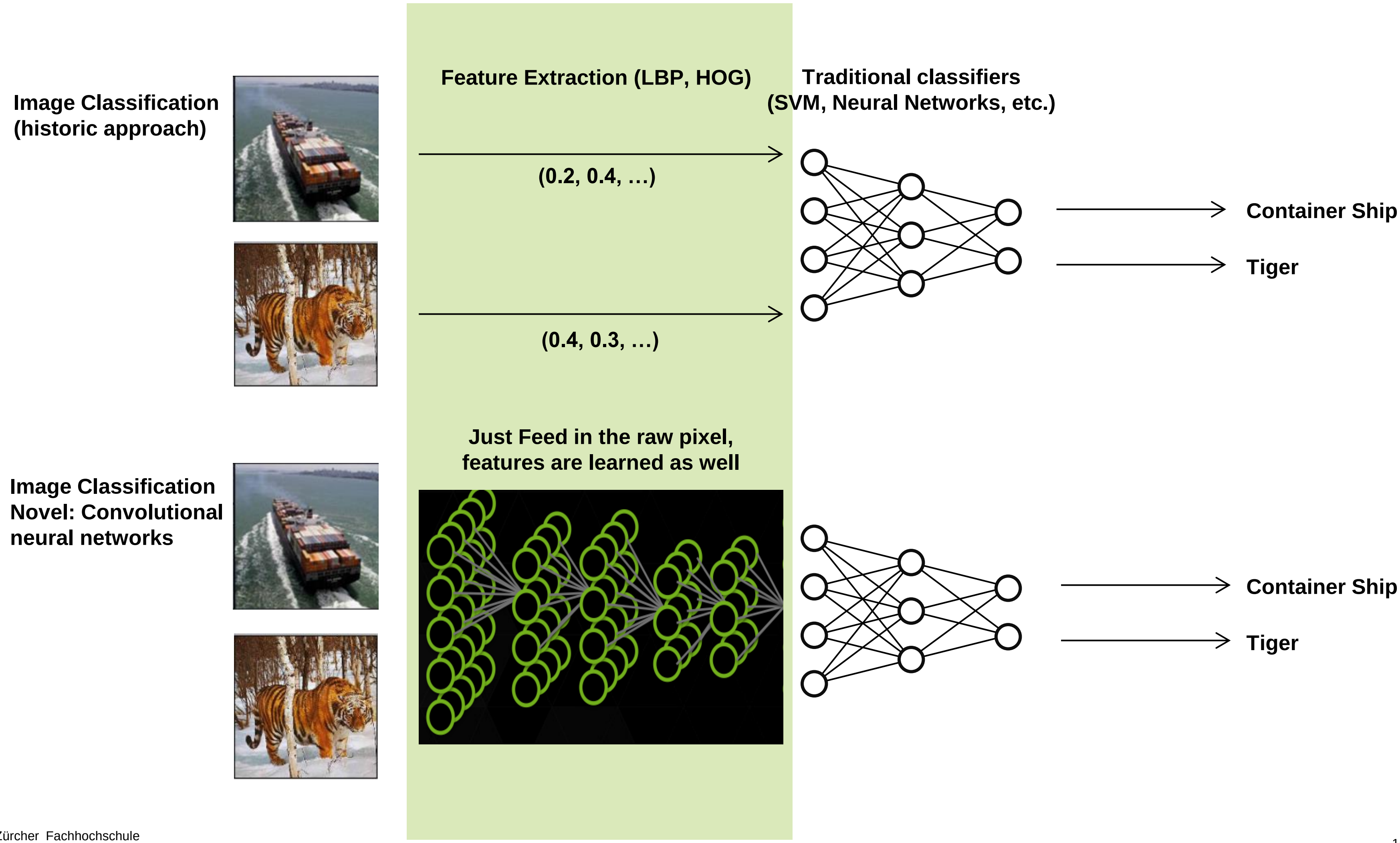
4.95% Microsoft (Feb 6)
→ surpassing human performance of 5.1%

4.8% Google (Feb 11)

4.58% Baidu (May 11)

→ **Computers learn** to identify objects

Background II: Key idea “feature learning”

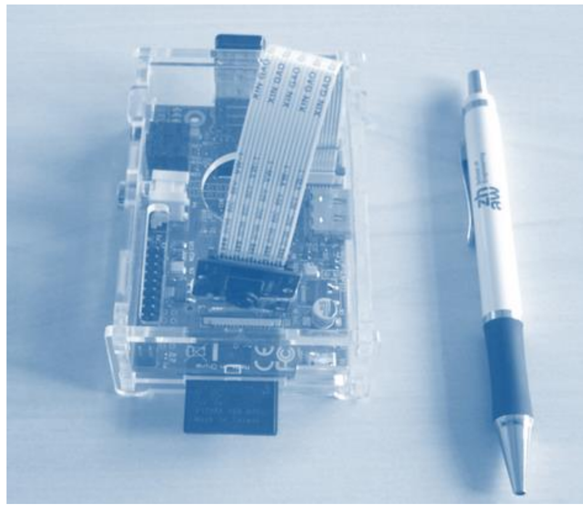


Adapting feature learning for Raspberry Pi
Approach

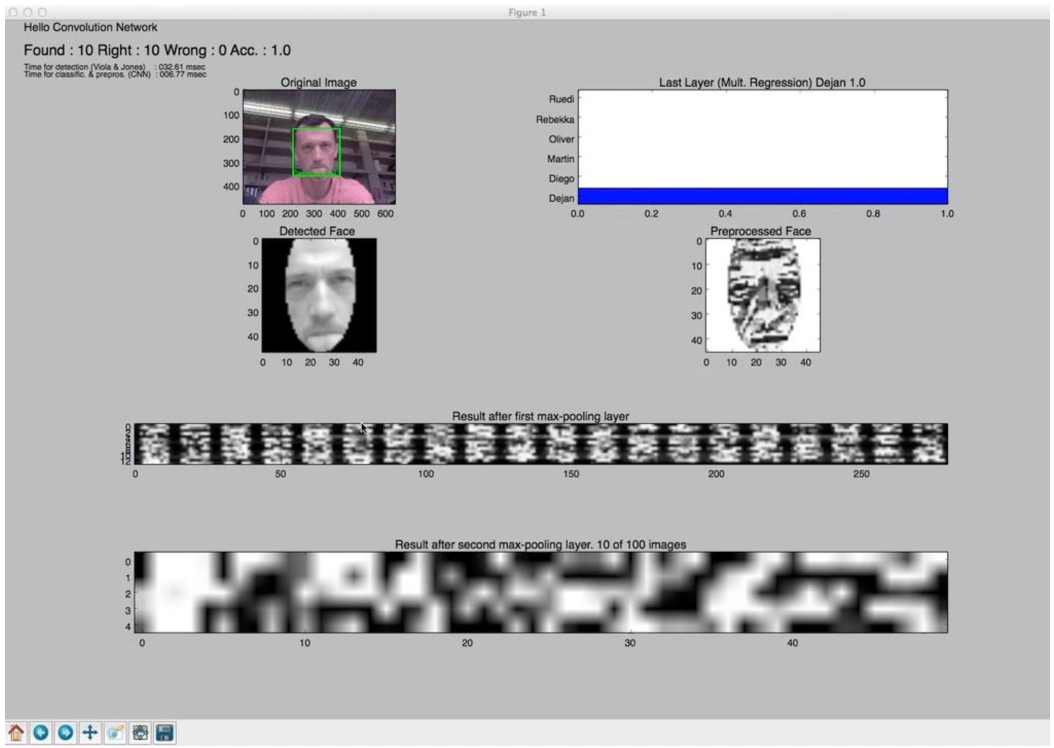
- Changed architecture from traditional to novel approach
- Caring for a small (embedded) target system
  - Train on GPU (Datalab Servers with Gaming Graphic Cards)
  - Run on Raspberry Pi

Learn on GPU

Run on Raspberry Pi

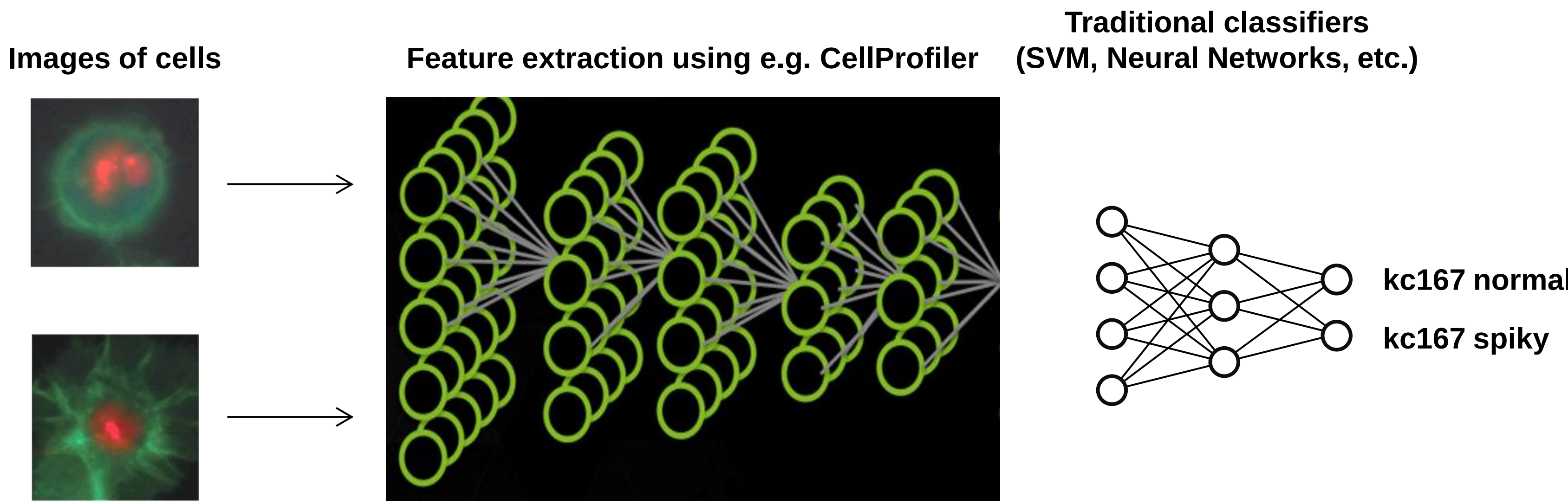


Adapting feature learning for Raspberry Pi
Results



Dürr, Pauchard, Browarnik, Axthelm, Loeser (2015): *Deep Learning on a Rasperry Pi for Real Time Face Recognition*. EG 2015 – Posters 11-12.

Further applications I
Talking to a biologist



Dürr, Sick (2015): *Deep learning: A novel approach to classify phenotypes in high content screening*, Swiss Image Based Screening Conference.

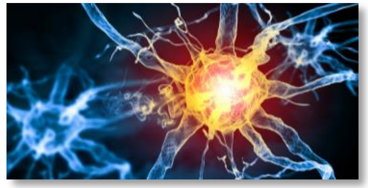
Dürr, Sick (2016): *Single cell phenotype classification using deep convolutional neural networks*, J. of Biomolecular Screening.

Further applications II
Deep Learning picked up in teaching

Zürcher Hochschule
für Angewandte Wissenschaften

CAS

- Data Science Applications, module «Machine Learning»



M.Eng.

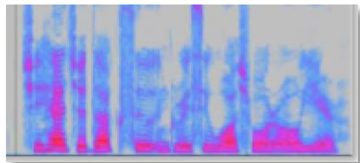
- TSM_MachLe «Machine Learning»

Thesis

- E. Murnia, M. Hirt: **Top 10%** in intl. Diabetes retinopathy **Kaggle® competition** ($ 100k)
- G. Eyyi: Speaker identification

Thesis

- Y. Lukic, C. Vogt: **Speaker clustering**



Further applications III
Data-driven Condition Monitoring Project

Situation: Maintaining big (rotating) machinery is expensive, defect is more expensive

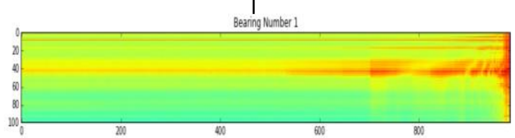
Goal: Schedule maintenance shortly before defect is expected, not merely regularly

Challenge: Develop an approach that adapts to each new machine automatically

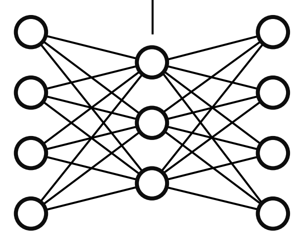
Solution: Use machine learning approaches for anomaly detection to learn the normal state of each machine and deviations of it purely from observed sensor signals; the approach combines classic and industry-proven features with e.g. deep learning auto-encoders

vibration sensors

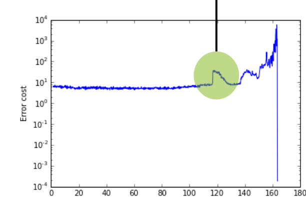
feature extraction



e.g., RNN autoencoder



early detection of fault



PANOPTES – Automated Article Segmentation of Newspaper Pages for "Real Time Print Media Monitoring"

M. Arnold, M. Cieliebak, T. Stadelmann, J. Stampfli, and F. Uzdilli

Further applications IV

Automated Article Segmentation in Newspapers

Overview

Partners
Who are we

ARGUS der Presse AG
- Switzerland's leading media monitoring and information provider
- Experience of more than 100 years

ZHAW Datalab
- Interdisciplinary research group at Zurich University of Applied Sciences
- Combining the knowledge of different fields related to machine learning

The Project
What do we do

Goal
- Real Time Print Media Monitoring
  - Extraction of relevant articles from newspaper pages
  - Delivering articles to customers

Problem
- Fully automated article segmentation
- Identification of article elements (e.g. title, subtitle, etc.)

Approach

Rule based
Segmentation based on hardcoded rules

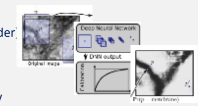
Rule examples
- Each article must contain a title
- Titles define article's width
- Articles are graphically separated by e.g. lines
- etc.

Pros
- Performance increases the more time is spent for finding rules
- Adding new rules is simple

Cons
- Not every case can be covered
- Adaptation to new layouts is costly manual work

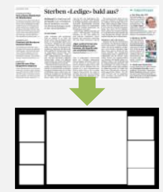
Image based
Segmentation based on visual features and deep learning

Approach
- Pixel classification (article/border) based on [1]



Pros
- Rules can be learned implicitly
- New layouts can be adapted automatically

Cons
- Success factors on new data and problems are unknown
- Training requires a huge amount of data

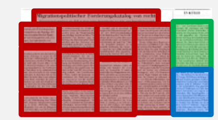


Text based
Segmentation based on textual features and neural nets

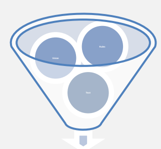
Approach
- Text block clustering (semantic distance) based on [2]

Pros
- Rules can be learned implicitly
- Not layout dependent

Cons
- Only text can be processed



Combination
Combination of rules, visual and textual features



Final segmentation

Result

References

[1] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. *Deep neural networks segment neuronal membranes in electron microscopy images.* In *NIPS*, pages 2852–2860, 2012.
[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean. *Efficient Estimation of Word Representations in Vector Space.* In Proceedings of Workshop at *ICLR*, 2013.

This project was funded by CTI under project number 17719.1 PFES-ES

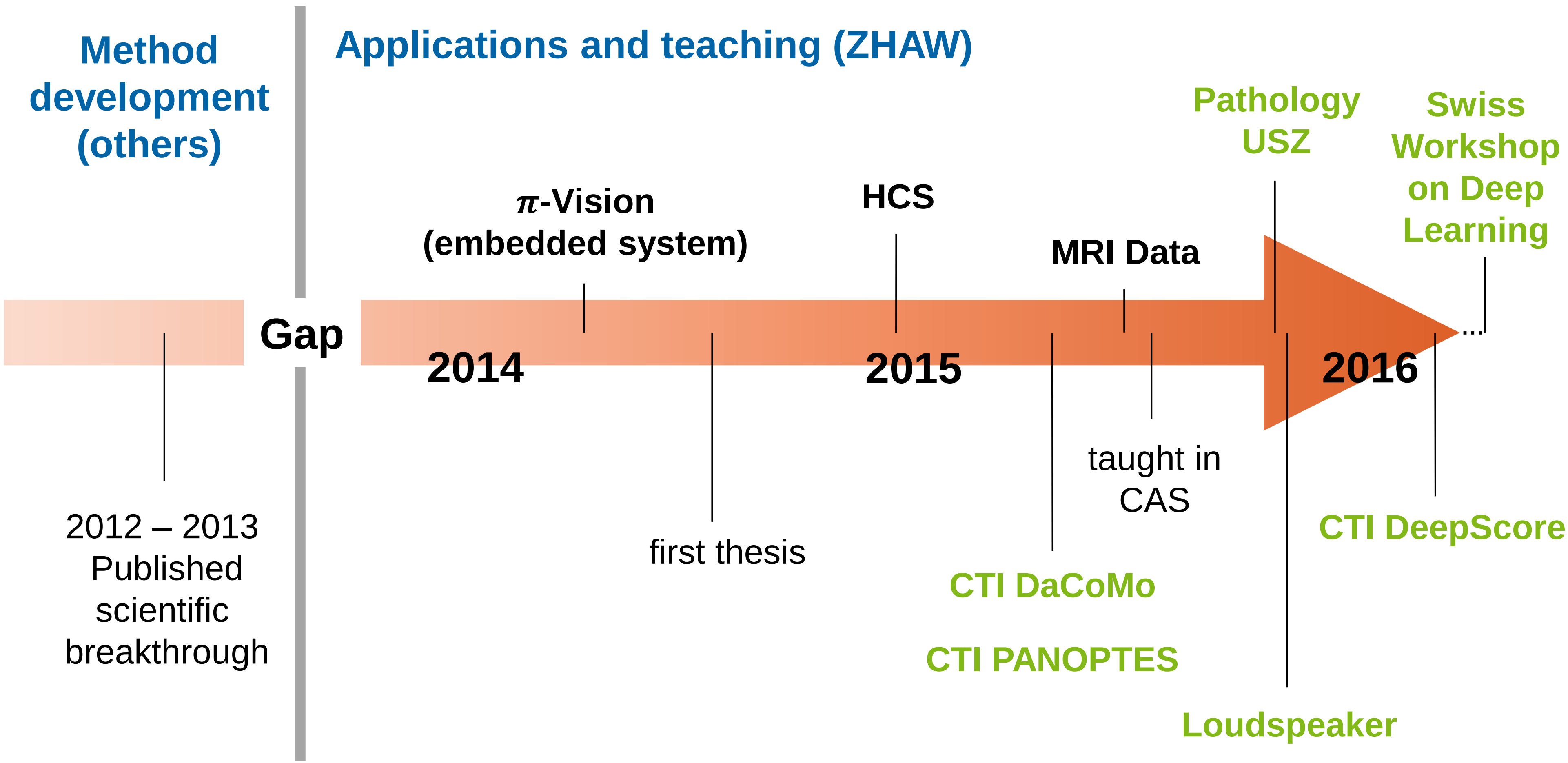
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
Commission for Technology and Innovation CTI
Innovation Promotion Agency

swiss group for artificial intelligence and cognitive science

Further applications V

Timeline of Deep Learning @ Datalab



Zürcher Hochschule
für Angewandte Wissenschaften

3

Open Machine Learning Questions: Sequential Learning & Inductive Biases

Open Question #1
Overcoming the Bag of Features Approach

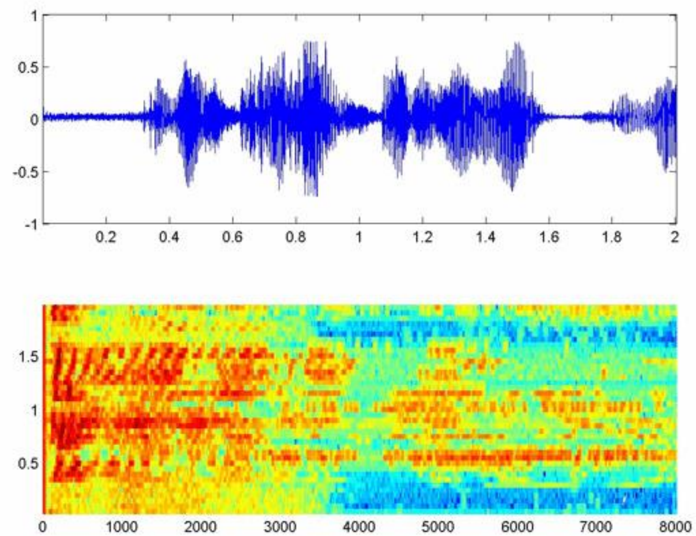
Supervised Learning

Incanter Dataset

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa
5.4	3.4	1.7	0.2	setosa
5.1	3.7	1.5	0.4	setosa

feature vectors labels

Sequential Supervised Learning



Typical assumption on data:

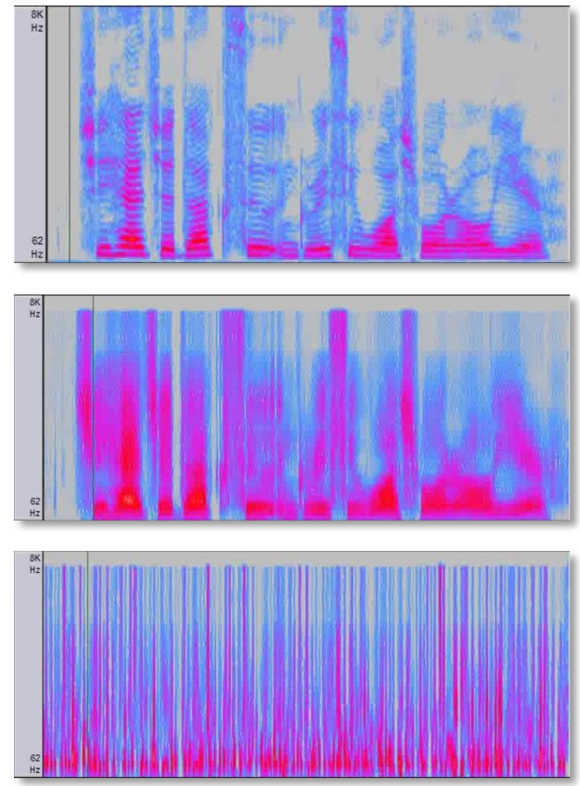
- i.i.d.

Typical finding:

- **Sequence** information **matters**

Example: Voice Recognition
Question #1

Temporal context matters
- Typical approaches are based on modeling the statistical distribution (GMM) of bag of features (MFCC)
- But: Voice emerges in chunks of ca. 120 ms length
- Literature promises **one order of magnitude improvement** on correct dealing with this issue*



Problem description
- Need to go **from** “feature **vector**” **to** “feature **matrix**”

$$(x_1, x_2, \cdots, x_d) \rightarrow \begin{pmatrix} x_{1,1} & \cdots & x_{1,d} \\ \vdots & \ddots & \vdots \\ x_{\Delta t,1} & \cdots & x_{\Delta t,d} \end{pmatrix}$$

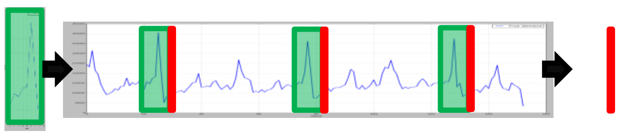
- I.e., **local structure** of features has to be **considered**

*) Stadelmann, Freisleben, „Unfolding Speaker Clustering Potential – A Biomimetic Approach“, 2009

Why is this a fundamental, open question?
Question #1

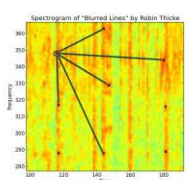
MULTIPLE USE CASES

- Food consumption forecasting for grocers



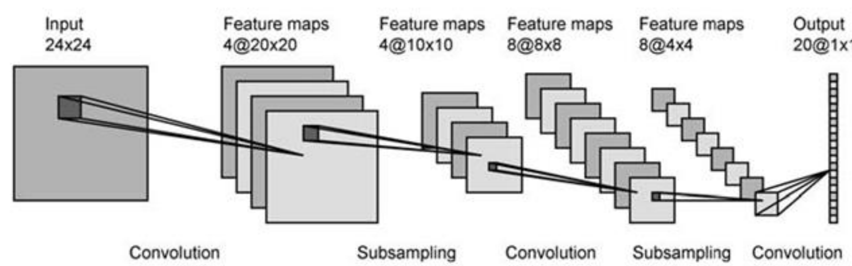
- Music-OCR

- Content-based music similarity



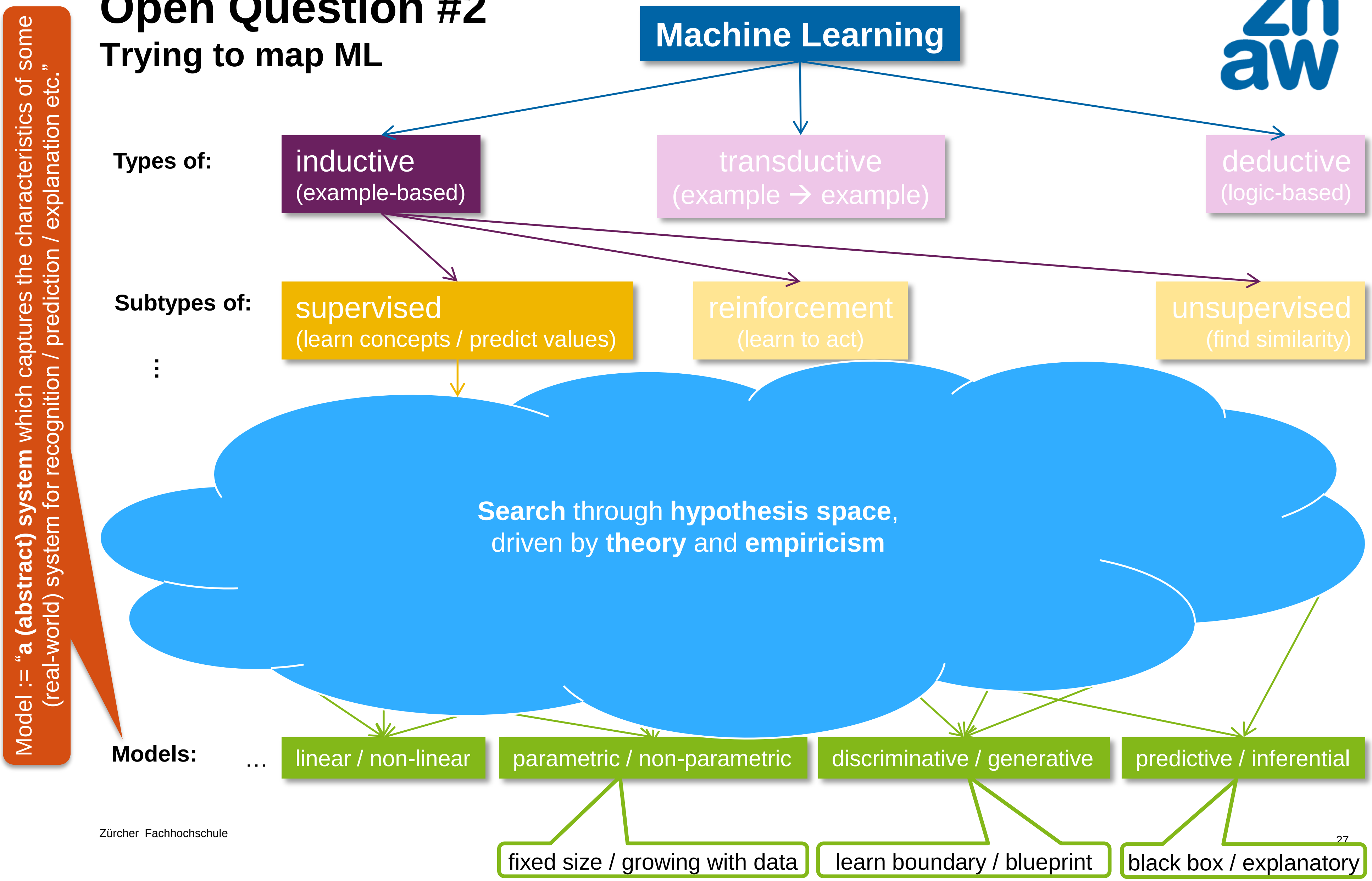
INSUFFICIENT SOLUTIONS

- δ / $\delta\delta$ features, other "**super vectors**"
  → **yet another** form of **bag**s of features
- **Clustering**s of multi-dim patterns
  → problem of pattern **representation** remains
- **CNN** and other multi-dimensional models
  →are **features** like locality **desirable**?
- **Sequence models** like HMM/RNN
  →**difficult** to train, enough **data**?



Open Question #2

Trying to map ML



What makes Learning Algorithms learn?
Question #2

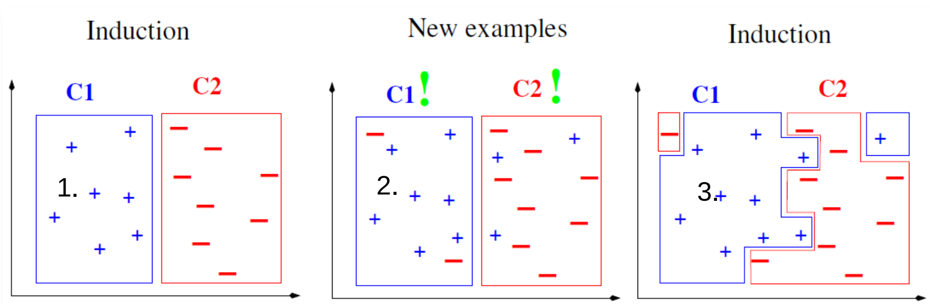
Background: The goal of Machine Learning
- Discover **general** concepts **from** a **limited** set of **examples** (experience)

Methods are based on **inductive reasoning**
- Obtain **general** knowledge (a **model**) **from specific** information
- This is **heuristic** in nature (i.e., no well-founded theory)
- Most of the **human learning is inductive**

➔ **Assumption**: A model fitted to sufficiently large example set will **generalize** to unseen data

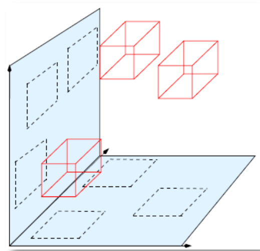
called the «inductive learning hypothesis»

What is sufficient? Another Basic ML research question!

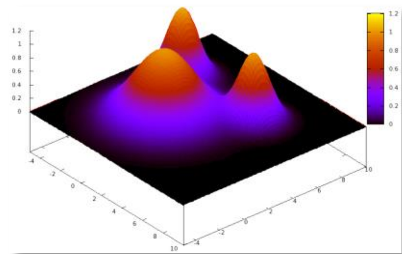


Background I: Learning as search through a hypothesis space $\mathcal{H}$
Question #2

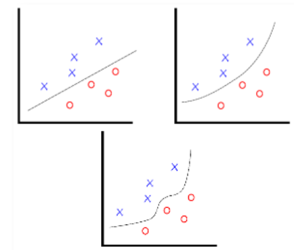
Hypothesis spaces



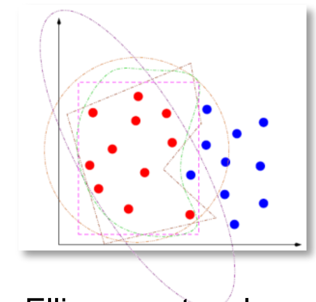
Logic formulas in DNF



Probabilistic models (PDF)



Linear/non-linear functions



Ellipse, rectangle, …

- $\mathcal{H}$ contains all possible hypothesis that can be built with the chosen representation

What is best? Driven by the **inductive bias**.

Formal goal

- **Find** the **hypothesis** $h^*(x,\theta) = \hat{y}$ **that *best fits*** *the training* ***data***…
  - …according to a **loss function** $L(h(x,\theta),y)$…
  - …by searching the hypothesis space $\mathcal{H} = \{h(x,\theta)|\theta \in P\}$ (P is the set of all possible parameters)
  - That is: find $h^* = \arg\min_{h\in\mathcal{H}} E_{emp}(h)$…
  - …by minimizing the **empirical error** $E_{emp}(h) = \frac{1}{N}\sum_{i=1}^{N} L(h(x_i,\theta),y_i)$, with e.g. $L(\hat{y},y) = \begin{cases} 0 \; if \; y = \hat{y} \\ 1 \; else \end{cases}$

➔ Very useful decomposition of inductive bias for practice:
language bias | search bias | overfitting-avoidance bias

Zurich University of Applied Sciences

Background II: The Inductive bias guides the search through $\mathcal{H}$
Question #2

*«A learner that makes **no a priori assumptions** regarding the identity of the target concept has **no rational basis for classifying** any unseen instances»*
[Mitchell, 1997, Ch. 2.7.3]

Inductive bias of a learning algorithm $\mathcal{L}$ for instances in X

- Any **minimal set of assertions** B that, together with $\mathcal{L}$ and the training set $D = \{(x_i, y_i)\}, i = 1..N$, **allows for deductively inferring** the y' for a new $x' \in X$
- That is: Make all assumptions **explicit** in B such that $\forall x' \in X$: $(B, \mathcal{L}, D, x') \Rightarrow y'$ is provable

i.e.: based on a priori knowledge

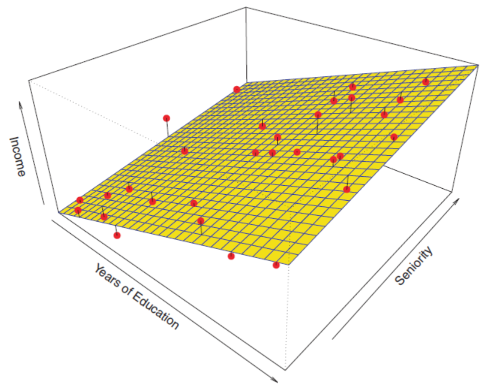
➔ Ultimately, ML depends on intelligent choice of the class of $\mathcal{H}$; $\mathcal{L}$ then optimizes the details
➔ We can characterize ML algorithms by (the strength of) their inductive bias

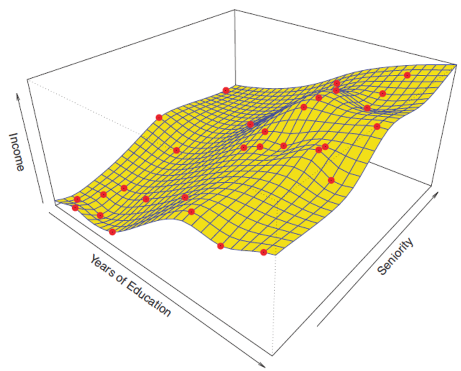
Is there a universal guide to learning?
Question #2

application level

fundamental level

there's a linear relationship between inputs & outputs

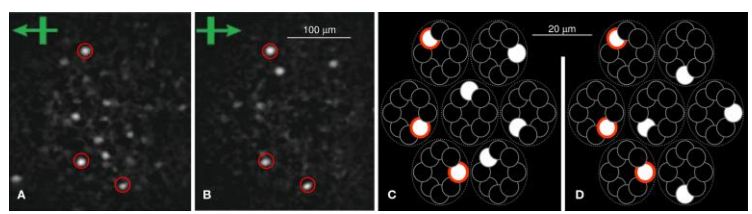




the hypothesis space is **smooth**

- ➔ Can the bias guide algorithm choice?
- ➔ Can general prior knowledge facilitate AI?

learn **sparse**, **distributed** representations



Summary

This talk has been about introduction:

- To our organization and achievements ➔ to inspire organizational out-of-the-box thinking
- To some of our work (R&D, education, community) ➔ to inspire research
- To our research questions ➔ to gain collaborators

Looking forward to getting in touch!

datalab
www.zhaw.ch/datalab

data+service
Swiss Alliance for Data-Intensive Services

More about me:

- Head of ZHAW Datalab
- thilo.stadelmann@zhaw.ch
- +41 58 934 72 08
- www.zhaw.ch/~stdm

More about Data Science?

- Education www.zhaw.ch/datalab
- Conference: www.zhaw.ch/datalab/sds2016
- Projects: datalab@zhaw.ch
- Association data+service: Being founded, open for members

➔ Please get in touch.