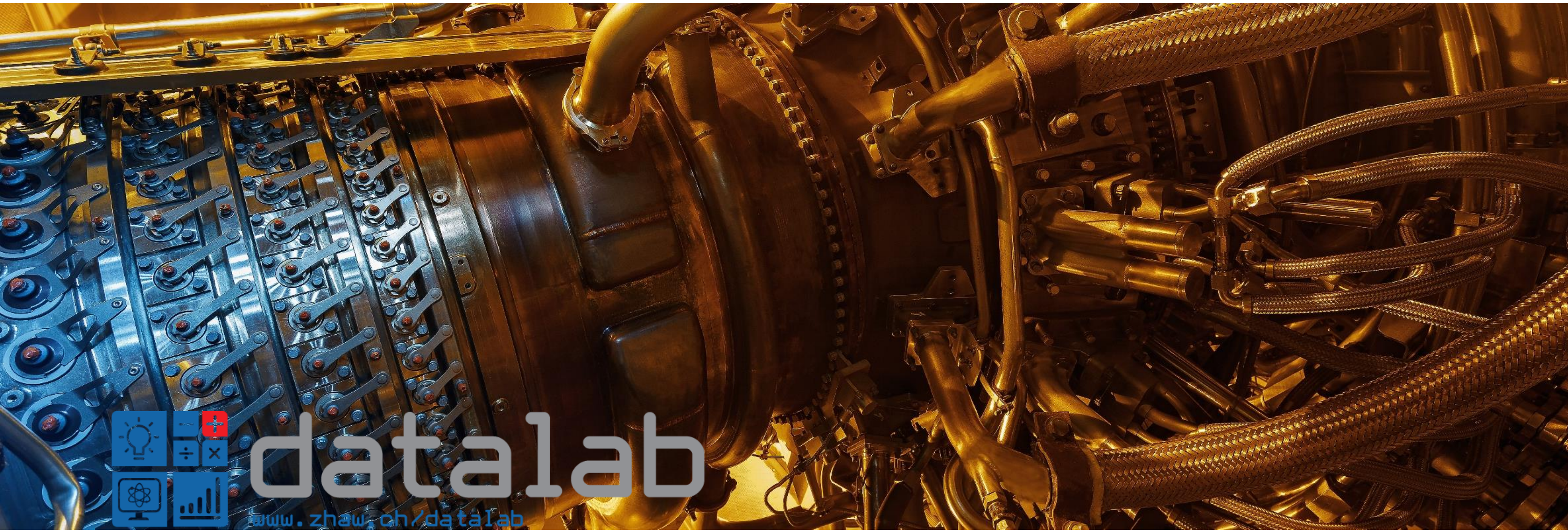
Zürcher Hochschule
für Angewandte Wissenschaften

Deep Learning in an industrial context: predictive maintenance, inspection and beyond

Data+Service Expert Group Predictive Maintenance, May 10, 2019

Thilo Stadelmann



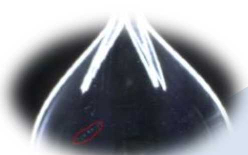
data service
Swiss Alliance for
Data-Intensive Services

Zürcher Hochschule
für Angewandte Wissenschaften

Agenda

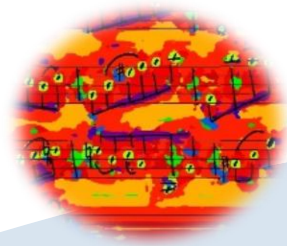
1. The group

2. Predictive maintenance

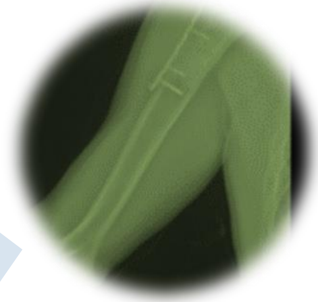


3. Industrial quality control (inspection)

4. Face recognition



5. Optical music recognition



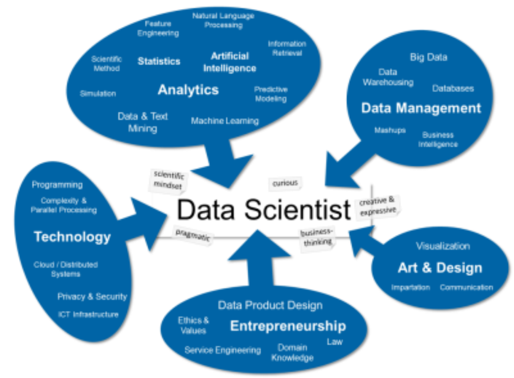
6. Lessons learned

1. ZHAW Datalab: Est. 2013

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

Foundation
- **People**: ca. **90 researchers** from **7 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

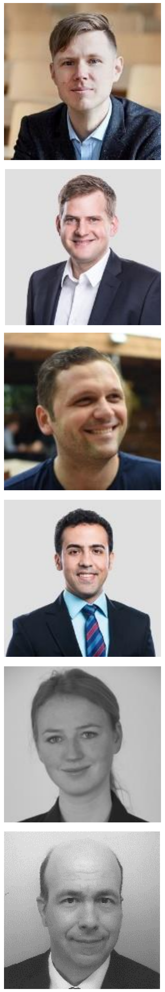
Success factors
- **Lean** organization and operation → geared towards projects
- Years of successful **pre-Datalab collaboration**



1. ML @ Information Engineering Group

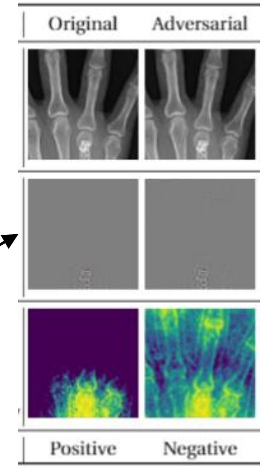
Institute of Applied Information Technology, School of Engineering

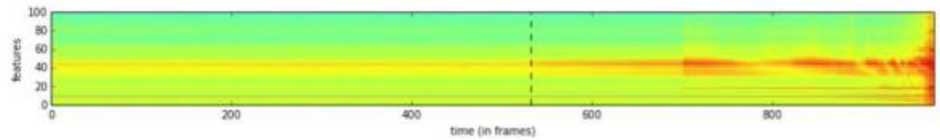
Zürcher Hochschule
für Angewandte Wissenschaften



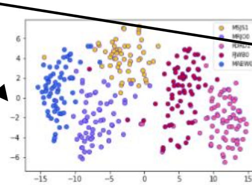
Machine learning-based Pattern Recognition

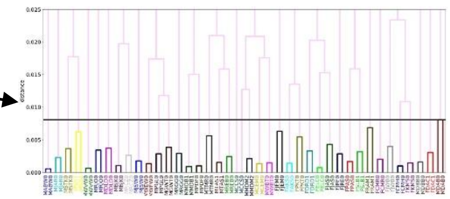
Robust Deep Learning





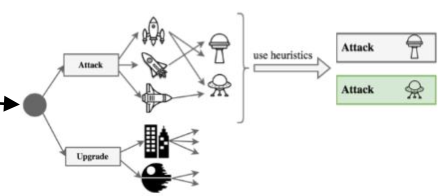
Voice Recognition

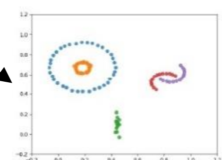


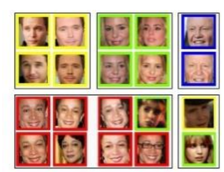


Document Analysis

Learning to Learn & Control







Zürcher Hochschule
für Angewandte Wissenschaften

2. Data-driven Condition Monitoring

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

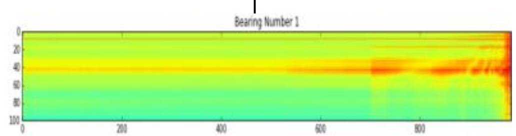
Goal: Schedule maintenance shortly before defect is expected, not merely regularly

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

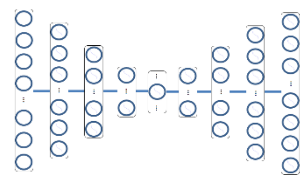
Solution: Use machine learning approaches for **anomaly detection** to learn the normal state of each machine and deviations of it purely from observed sensor signals; the approach combines classic and industry-proven features with e.g. deep learning auto-encoders

vibration sensors

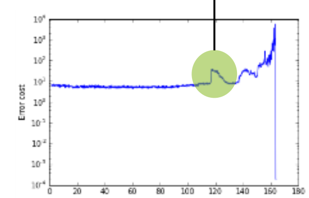
feature extraction



e.g., RNN autoencoder



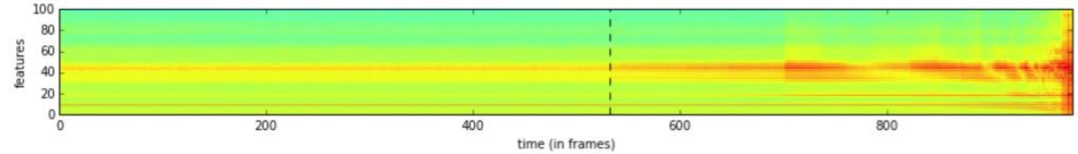
early detection of fault



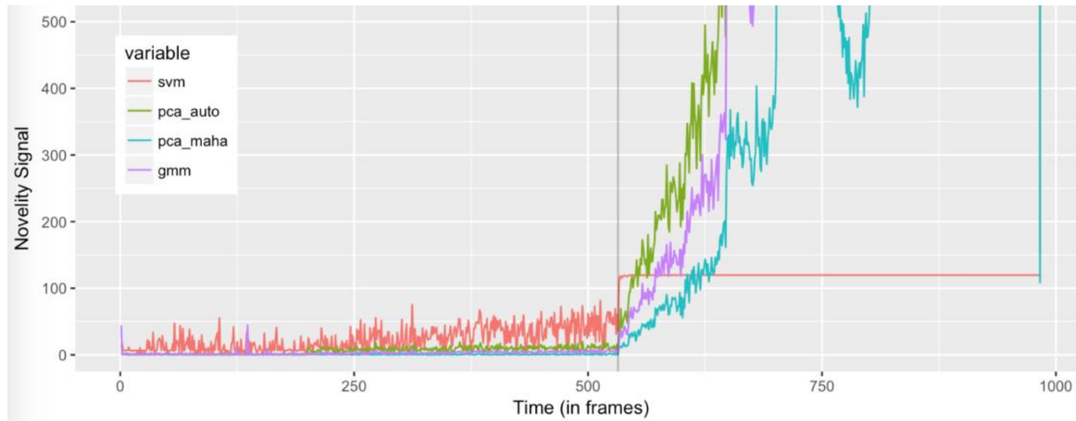
Stadelmann, Tolkachev, Sick, Stampfli & Dürr (2019): *«Beyond ImageNet–Deep Learning in Industrial Practice»*. In: Braschler et al. (Ed.), *«Appl. Dat. Sci.», Springer.*

2. Data-driven Condition Monitoring: Results

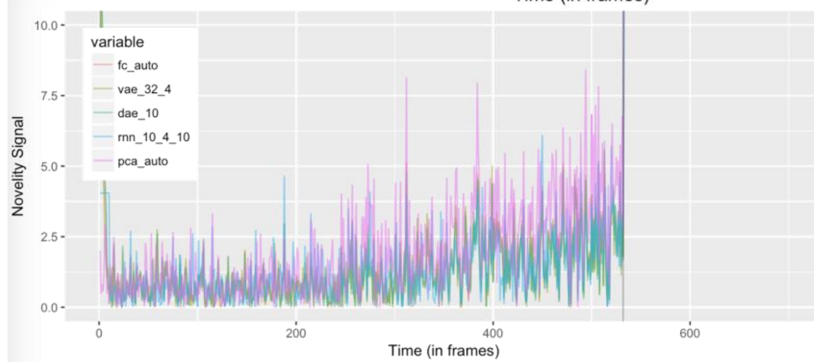
Signal:



Shallow learning methods:

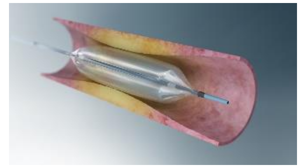


Deep learning methods:



➔ DL and standard methods detect the defect time; DL show **less novelty** where there is **still no defect**

3. Industrial quality control



Zürcher Hochschule
für Angewandte Wissenschaften

zh
aw

BW-TEC

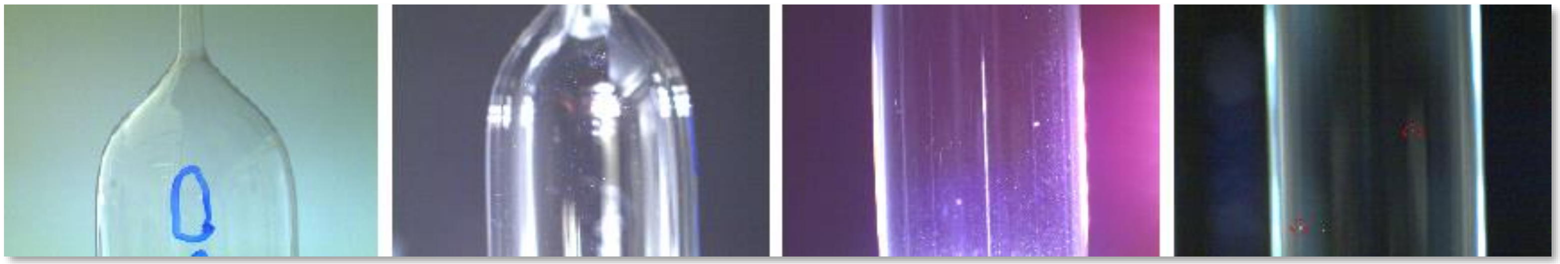
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

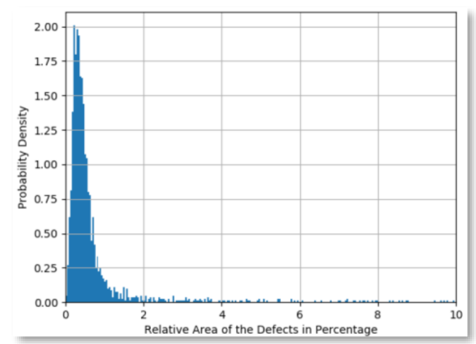
Task

- Reliably **sort out faulty balloon catheters** in image-based production quality control



Challenges

- **Non-natural** image source, class **imbalance**, **optic**al conditions, **variation** in defect size & shape



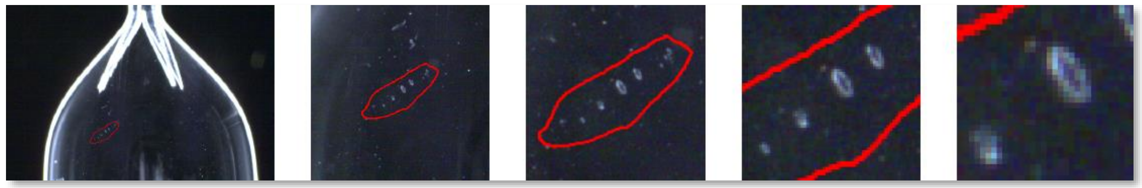
Stadelmann, Amirian, Arabaci, Arnold, Duivesteijn, Elezi, Geiger, Lörwald, Meier, Rombach & Tuggener (2018). *«Deep Learning in the Wild»*. ANNPR'2018.

Zürcher Hochschule für Angewandte Wissenschaften

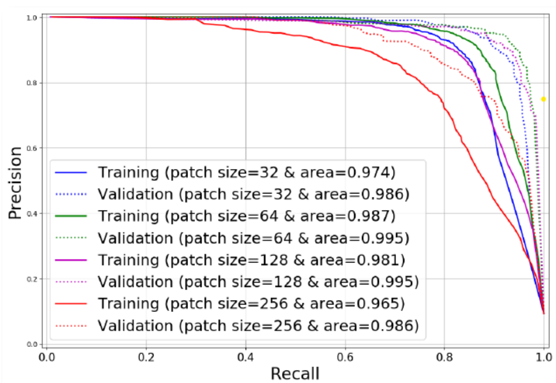
3. Industrial quality control – baseline results

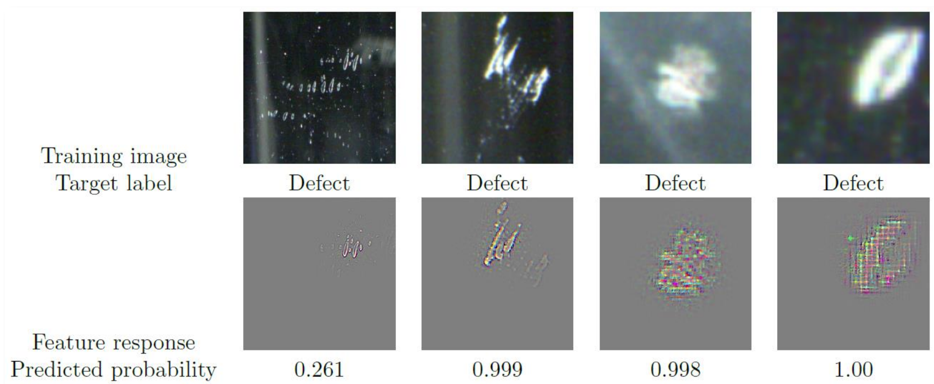
Ingredients

- Weighted loss
- Defect cropping
- Careful customization



Interim results





3. Industrial quality control – recent results

- Human performance isn't flawless
- Tailoring pays off
- Data shortage may be outsmarted

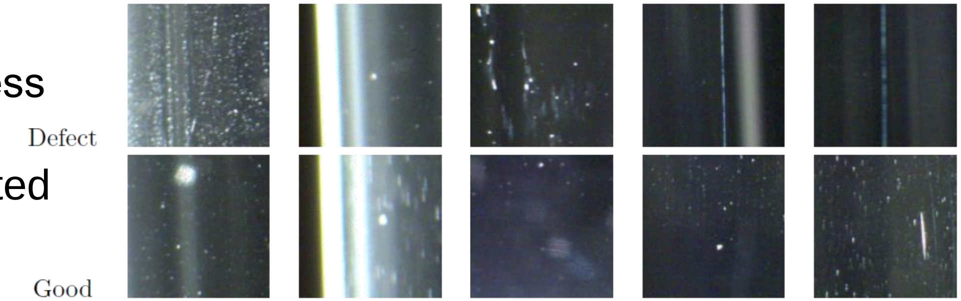
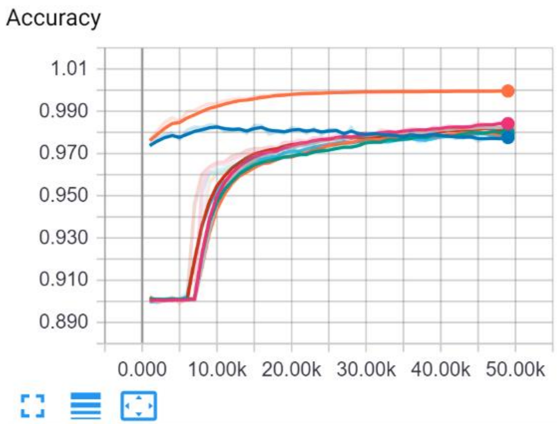
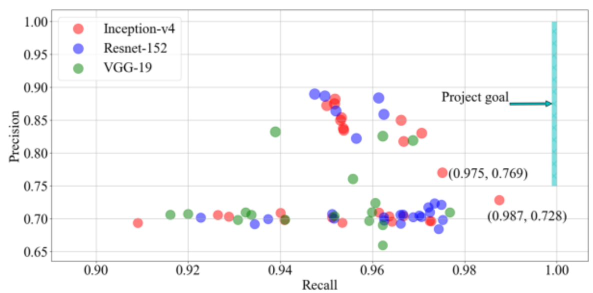
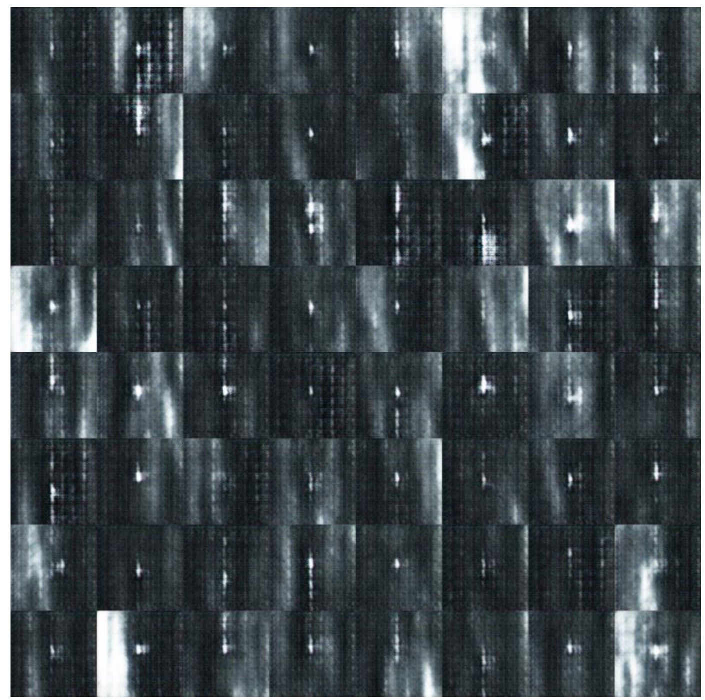


Figure 2: Samples of failure cases in classification. The shown *defect* samples in the table are not recognized as a defects, and the *good* images are misclassified as defects.





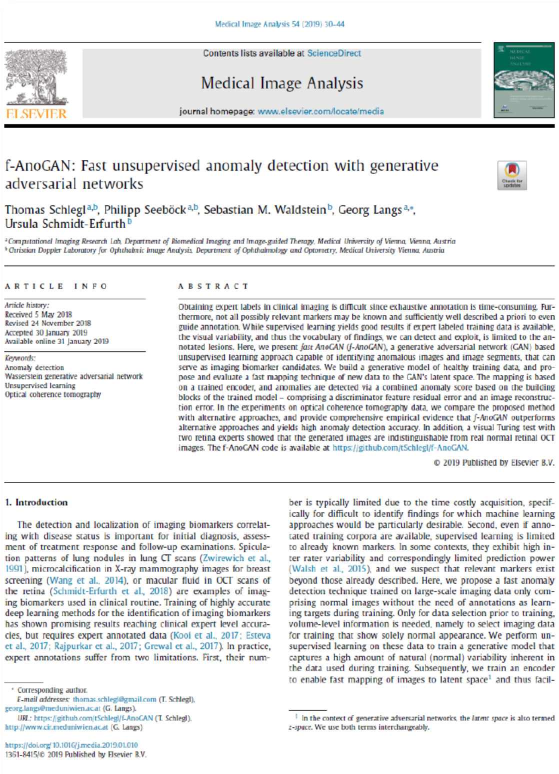
Name	Smoothed	Value	Step	Time	Relative
QualitAI_VGG19_Full_Pretrained\train	0.9996	0.9996	49.00k	Tue Jan 22, 02:32:13	8h 30m 56s
QualitAI_VGG19_Full_Pretrained\validation	0.9776	0.9783	49.00k	Tue Jan 22, 02:32:24	8h 30m 56s
QualitAI_VGG19_Full_Random\train	0.9841	0.9841	49.00k	Thu Jan 24, 19:28:02	10h 29m 2s
QualitAI_VGG19_Full_Random\validation	0.9798	0.9798	49.00k	Thu Jan 24, 19:28:14	10h 29m 2s
QualitAI_VGG19_Half\train	0.9827	0.9835	49.00k	Thu Jan 24, 13:01:47	4h 9m 12s
QualitAI_VGG19_Half\validation	0.9792	0.9798	49.00k	Thu Jan 24, 13:01:54	4h 9m 11s
QualitAI_VGG19_Quarter\train	0.9817	0.9823	49.00k	Thu Jan 24, 10:53:52	2h 17m 21s
QualitAI_VGG19_Quarter\validation	0.9791	0.9806	49.00k	Thu Jan 24, 10:53:56	2h 17m 21s



Zürcher Hochschule für Angewandte Wissenschaften
zh aw
BW-TEC
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Innosuisse – Swiss Innovation Agency

3. Industrial quality control – future work

Trying to overcome class imbalance and small training set sizes



Medical Image Analysis 54 (2019) 30–44

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks

Thomas Schlegl[a,b], Philipp Seeböck[a,b], Sebastian M. Waldstein[b], Georg Langs[a,*], Ursula Schmidt-Erfurth[b]

[a] Computational Imaging Research Lab, Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Vienna, Austria
[b] Christian Doppler Laboratory for Ophthalmic Image Analysis, Department of Ophthalmology and Optometry, Medical University Vienna, Austria

ARTICLE INFO

Article history:
Received 5 May 2018
Revised 24 November 2018
Accepted 30 January 2019
Available online 31 January 2019

Keywords:
Anomaly detection
Wasserstein generative adversarial network
Unsupervised learning
Optical coherence tomography

ABSTRACT

Obtaining expert labels in clinical imaging is difficult since exhaustive annotation is time-consuming. Furthermore, not all possibly relevant markers may be known and sufficiently well described a priori to even guide annotation. While supervised learning yields good results if expert labeled training data is available, the visual variability, and thus the vocabulary of findings, we can detect and exploit, is limited to the annotated lesions. Here, we present *fast AnoGAN (f-AnoGAN)*, a generative adversarial network (GAN) based unsupervised learning approach capable of identifying anomalous images and image segments, that can serve as imaging biomarker candidates. We build a generative model of healthy training data, and propose and evaluate a fast mapping technique of new data to the GAN's latent space. The mapping is based on a trained encoder, and anomalies are detected via a combined anomaly score based on the building blocks of the trained model – comprising a discriminator feature residual error and an image reconstruction error. In the experiments on optical coherence tomography data, we compare the proposed method with alternative approaches, and provide comprehensive empirical evidence that *f-AnoGAN* outperforms alternative approaches and yields high anomaly detection accuracy. In addition, a visual Turing test with two retina experts showed that the generated images are indistinguishable from real normal retinal OCT images. The f-AnoGAN code is available at https://github.com/tSchlegl/f-AnoGAN.

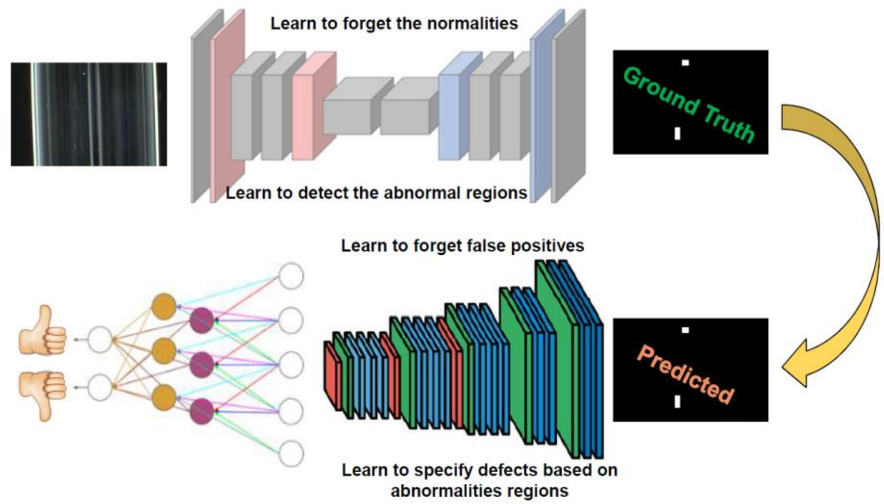
1. Introduction

The detection and localization of imaging biomarkers correlating with disease status is important for initial diagnosis, assessment of treatment response and follow-up examinations. Spiculation patterns of lung nodules in lung CT scans (Zwirewich et al., 1991), microcalcification in X-ray mammography images for breast screening (Wang et al., 2014), or macular fluid in OCT scans of the retina (Schmidt-Erfurth et al., 2018) are examples of imaging biomarkers used in clinical routine. Training of highly accurate deep learning methods for the identification of imaging biomarkers has shown promising results reaching clinical expert level accuracies, but requires expert annotated data (Kooi et al., 2017; Esteva et al., 2017; Rajpurkar et al., 2017; Grewal et al., 2017). In practice, expert annotations suffer from two limitations. First, their number is typically limited due to the time costly acquisition, specifically for difficult to identify findings for which machine learning approaches would be particularly desirable. Second, even if annotated training corpora are available, supervised learning is limited to already known markers. In some contexts, they exhibit high inter rater variability and correspondingly limited prediction power (Walsh et al., 2015), and we suspect that relevant markers exist beyond those already described. Here, we propose a fast anomaly detection technique trained on large-scale imaging data only comprising normal images without the need of annotations as learning targets during training. Only for data selection prior to training, volume-level information is needed, namely to select imaging data for training that show solely normal appearance. We perform unsupervised learning on these data to train a generative model that captures a high amount of natural (normal) variability inherent in the data used during training. Subsequently, we train an encoder to enable fast mapping of images to latent space[1] and thus facil-

* Corresponding author.
E-mail addresses: thomas.schlegl@gmail.com (T. Schlegl), georg.langs@meduniwien.ac.at (G. Langs).
URL: https://github.com/tSchlegl/f-AnoGAN (T. Schlegl), http://www.cir.meduniwien.ac.at (G. Langs)

[1] In the context of generative adversarial networks, the *latent space* is also termed *z-space*. We use both terms interchangeably.

https://doi.org/10.1016/j.media.2019.01.010
1361-8415/© 2019 Published by Elsevier B.V.



3. Face matching

Zürcher Hochschule
für Angewandte Wissenschaften

3. Face matching – challenges & solutions

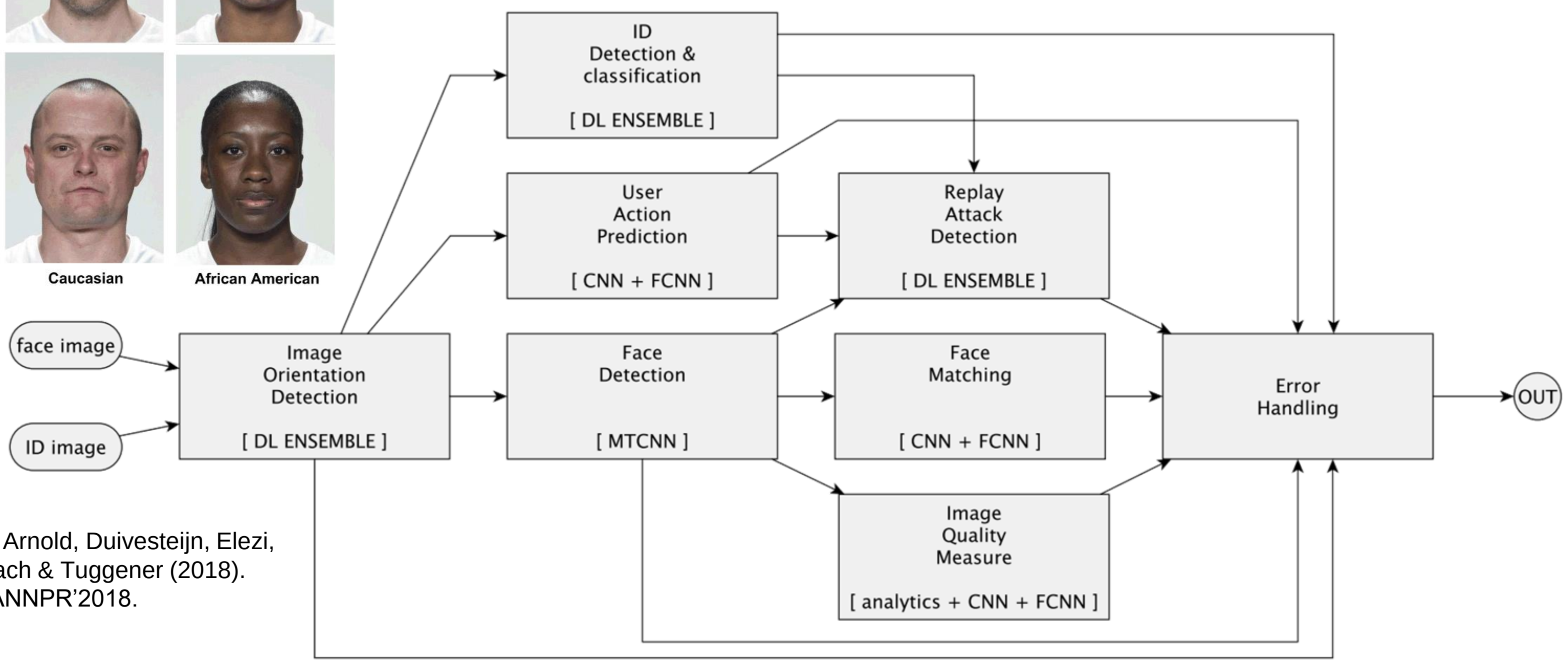
[!] DEEPIMPACT

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

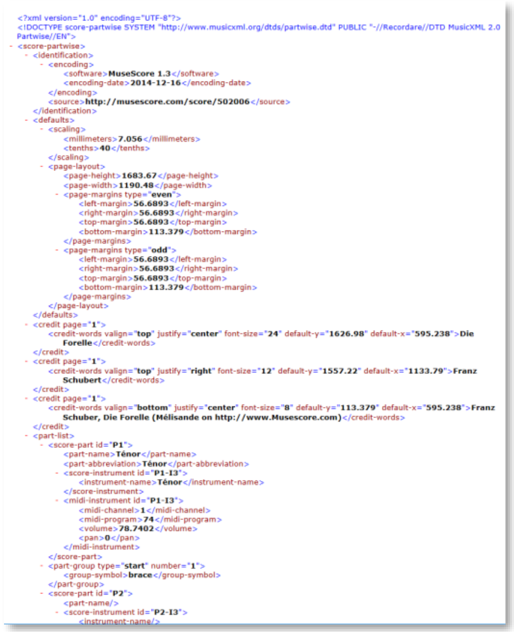
Asian Indian | East Asian | Caucasian | African American



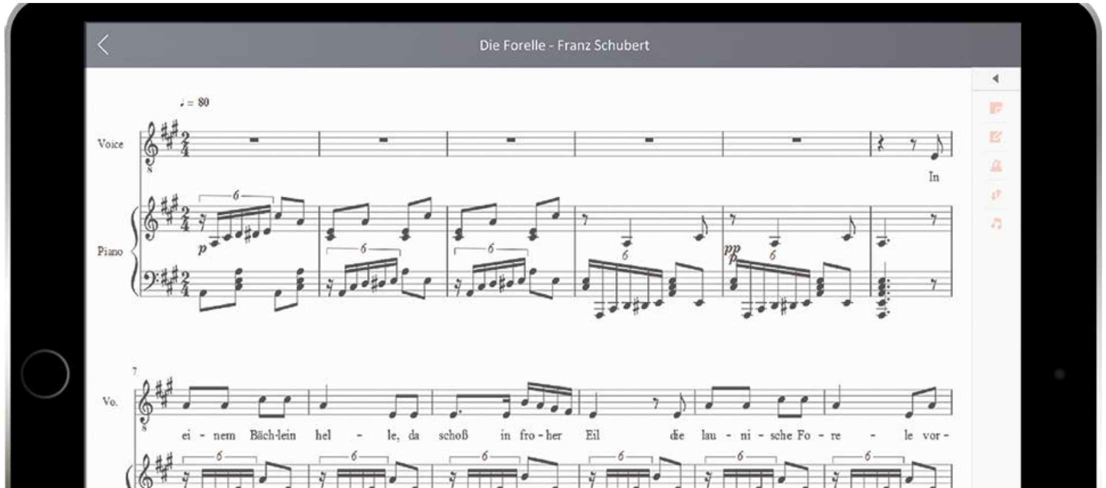
Stadelmann, Amirian, Arabaci, Arnold, Duivesteijn, Elezi, Geiger, Lörwald, Meier, Rombach & Tuggener (2018). *«Deep Learning in the Wild»*. ANNPR'2018.

Zürcher Hochschule
für Angewandte Wissenschaften

4. Music scanning



```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE score-partwise SYSTEM "http://www.musicxml.org/dtds/partwise.dtd" PUBLIC "-//Recordare//DTD MusicXML 2.0 Partwise//EN">
- <score-partwise>
  - <identification>
    - <encoding>
        <software>MuseScore 1.3</software>
        <encoding-date>2014-12-16</encoding-date>
      </encoding>
      <source>http://musescore.com/score/502006</source>
    </identification>
  - <defaults>
    - <scaling>
        <millimeters>7.056</millimeters>
        <tenths>40</tenths>
      </scaling>
    - <page-layout>
        <page-height>1683.67</page-height>
        <page-width>1190.48</page-width>
      - <page-margins type="even">
          <left-margin>56.6893</left-margin>
          <right-margin>56.6893</right-margin>
          <top-margin>56.6893</top-margin>
          <bottom-margin>113.379</bottom-margin>
        </page-margins>
      - <page-margins type="odd">
          <left-margin>56.6893</left-margin>
          <right-margin>56.6893</right-margin>
          <top-margin>56.6893</top-margin>
          <bottom-margin>113.379</bottom-margin>
        </page-margins>
      </page-layout>
    </defaults>
  - <credit page="1">
      <credit-words valign="top" justify="center" font-size="24" default-y="1626.98" default-x="595.238">Die Forelle</credit-words>
    </credit>
  - <credit page="1">
      <credit-words valign="top" justify="right" font-size="12" default-y="1557.22" default-x="1133.79">Franz Schubert</credit-words>
    </credit>
  - <credit page="1">
      <credit-words valign="bottom" justify="center" font-size="8" default-y="113.379" default-x="595.238">Franz Schuber, Die Forelle (Mélisande on http://www.Musescore.com)</credit-words>
    </credit>
  - <part-list>
    - <score-part id="P1">
        <part-name>Ténor</part-name>
        <part-abbreviation>Ténor</part-abbreviation>
      - <score-instrument id="P1-I3">
          <instrument-name>Ténor</instrument-name>
        </score-instrument>
      - <midi-instrument id="P1-I3">
          <midi-channel>1</midi-channel>
          <midi-program>74</midi-program>
          <volume>78.7402</volume>
          <pan>0</pan>
        </midi-instrument>
      </score-part>
    - <part-group type="start" number="1">
        <group-symbol>brace</group-symbol>
      </part-group>
    - <score-part id="P2">
        <part-name/>
      - <score-instrument id="P2-I3">
          <instrument-name/>
```

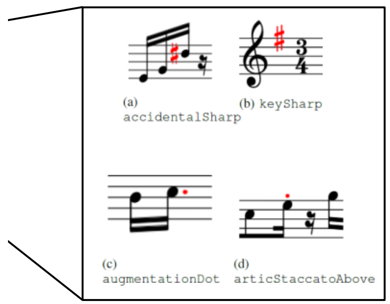



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

4. Music scanning – challenges & solutions



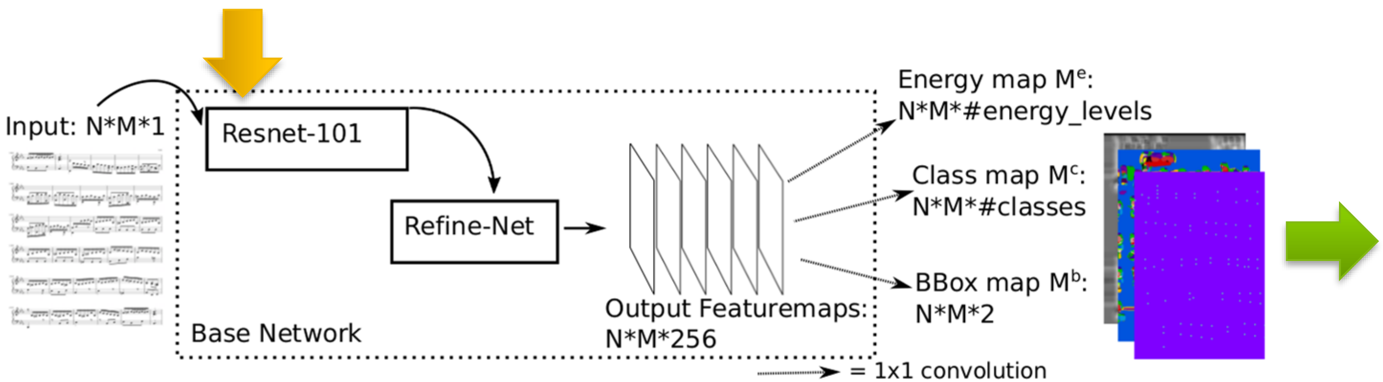
,

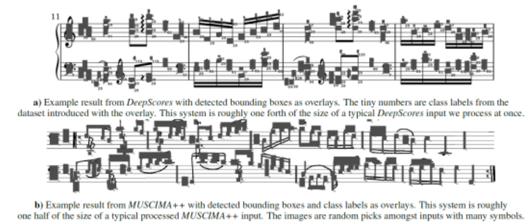
SCOREPAD

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

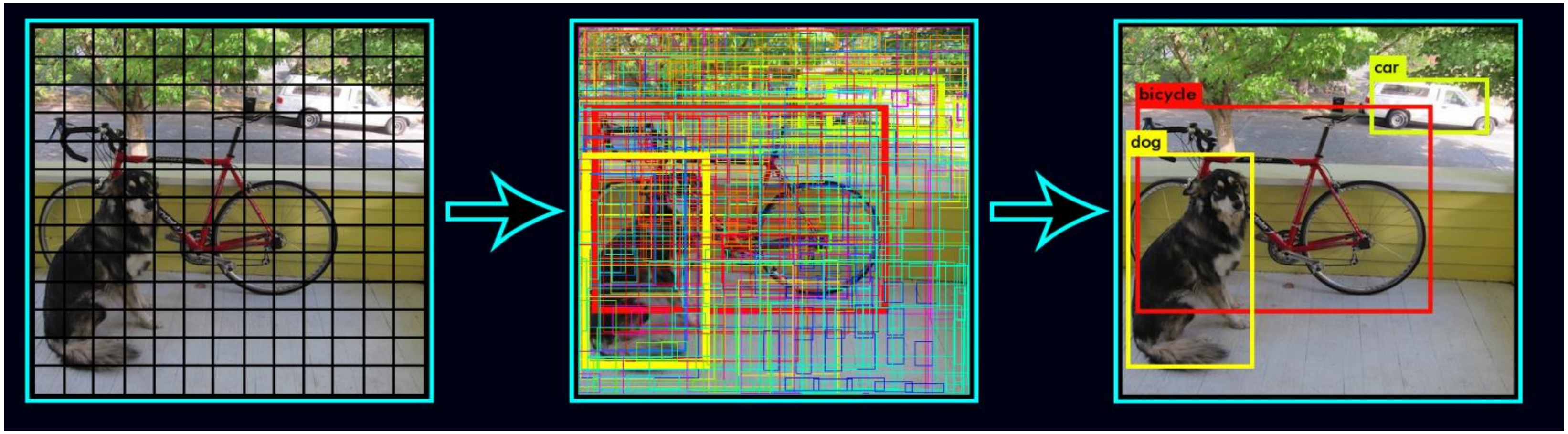




Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR’2018.
Tuggener, Elezi, Schmidhuber & Stadelmann (2018). *«Deep Watershed Detector for Music Object Recognition»*. ISMIR’2018.

4. Music scanning – methodology
OMR vs state of the art object detectors

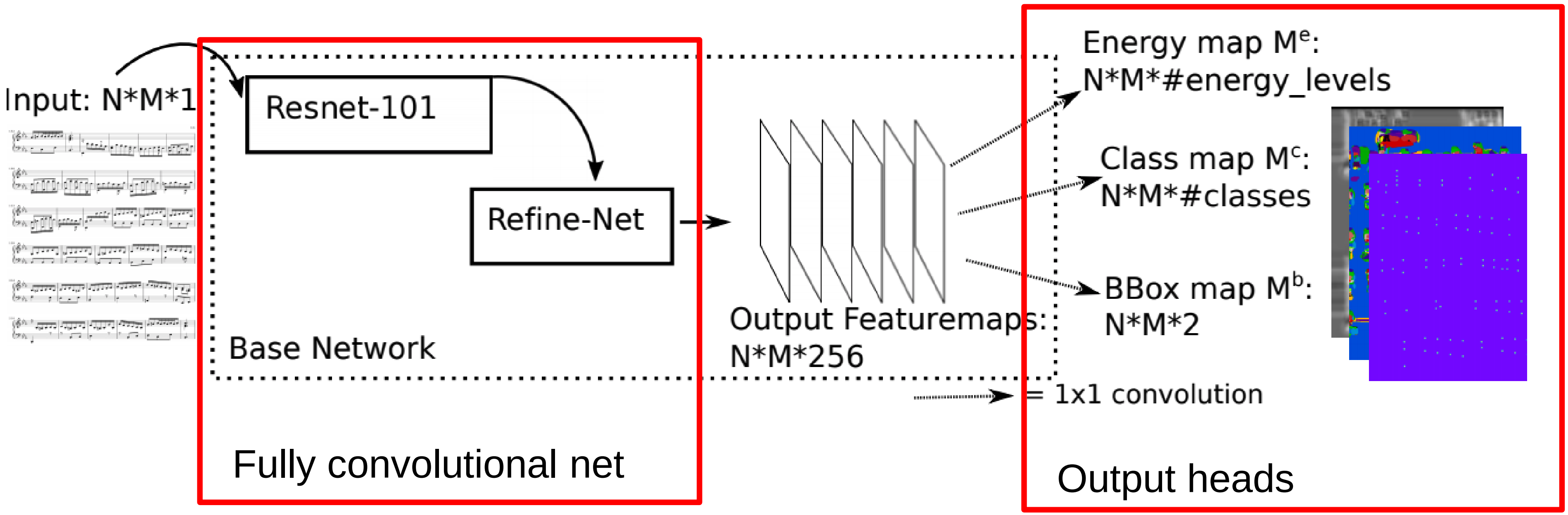
YOLO/SSD-type detectors



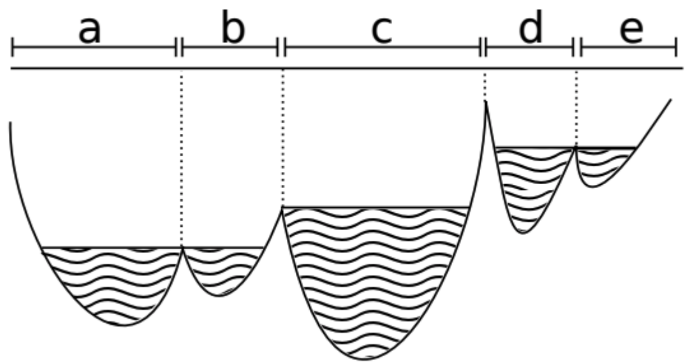
Source: https://pjreddie.com/darknet/yolov2/ (11.09.2018)

R-CNN
- Two-step proposal and refinement scheme
- Very large amount of proposals at high resolution needed

4. Music scanning – methodology (contd.)
The deep watershed detector

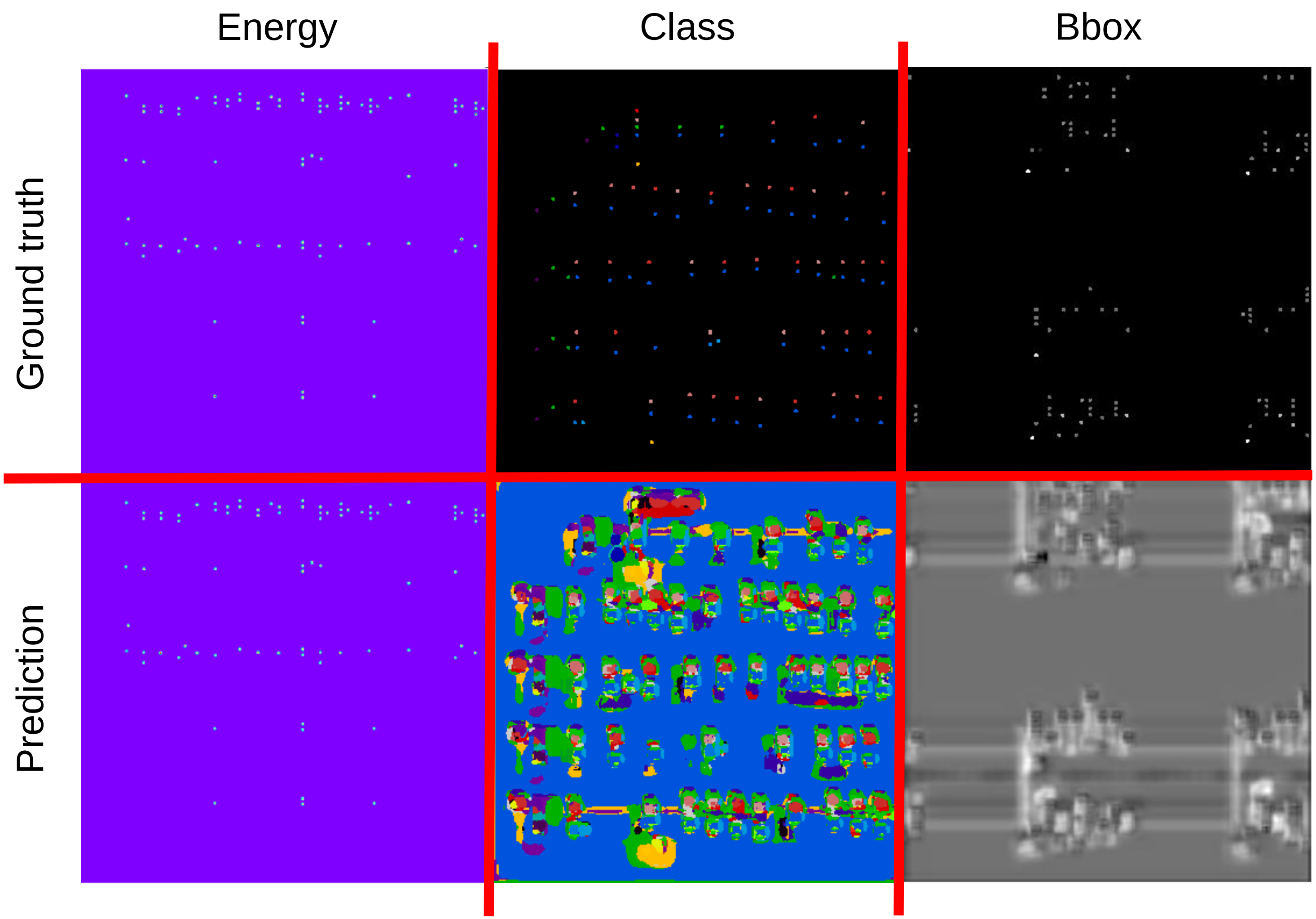


4. Music scanning – methodology (contd.)
The (deep) watershed transform



4. Music scanning – methodology (contd.)

Output heads of the deep watershed detector



4. Music scanning – industrialization

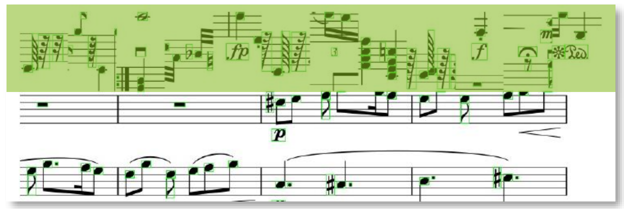
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Recent results on class imbalance and robustness challenges

1. Added sophisticated **data augmentation** in every page's margins



2. Put additional effort (and compute) into hyperparameter **tuning** and **longer training**
3. Trained also on scanned (more **real-worldish**) scores

➔ **Improved** our **mAP** from 16% (on purely synthetic data) **to 73%** on more challenging real-world data set (additionally, using Pacha et al.'s evaluation method as a 2nd benchmark: from 24.8% to 47.5%)

Elezi, Tuggener, Pelillo & Stadelmann (2018). *«DeepScores and Deep Watershed Detection: current state and open issues»*. WoRMS @ ISMIR'2018.
Pacha, Hajic, Calvo-Zaragoza (2018). *«A Baseline for General Music Object Detection with Deep Learning»*. Appl. Sci. 2018, 8, 1488, MDPI.

5. Lessons learned

Data is key.

- Many real-world projects miss the required **quantity & quality** of data
  → even though «big data» is not needed
- **Class imbalance** needs careful dealing
  → special loss, resampling (also in unorthodox ways), exploitation of every possible learning signal
- **Unsupervised** methods need to be used creatively
- Users & label providers need to be **trained**

Robustness is important.

- **Training processes** can be tricky
  → give hints via a unique loss, proper preprocessing and pretraining




5. Lessons learned – model interpretability

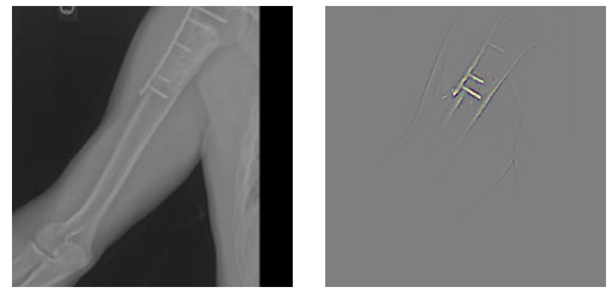
Zürcher Hochschule
für Angewandte Wissenschaften

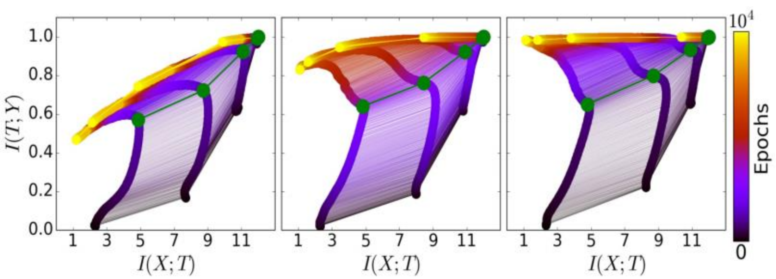
Interpretability is required.

- Helps the developer in «debugging», needed by the user to trust
  → visualizations of learned features, training process, learning curves etc. should be «always on»

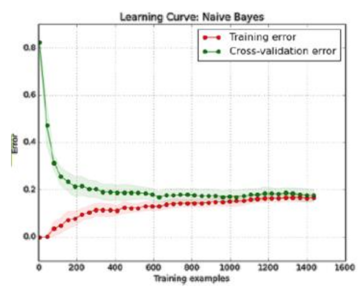
negative X-ray

positive X-ray

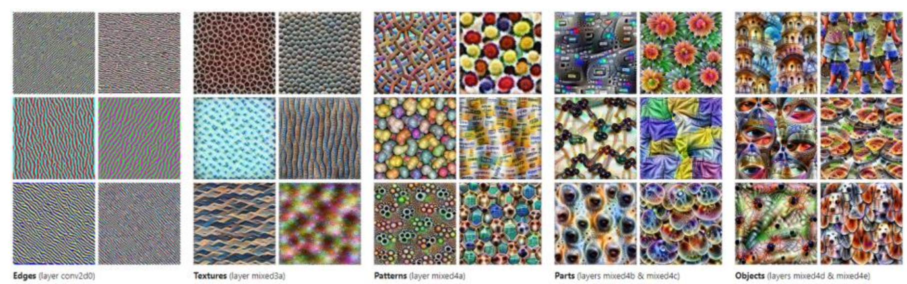




DNN training on the Information Plane



a learning curve



feature visualization

Stadelmann, Amirian, Arabaci, Arnold, Duivesteijn, Elezi, Geiger, Lörwald, Meier, Rombach & Tuggener (2018). *«Deep Learning in the Wild»*. ANNPR'2018.
Schwartz-Ziv & Tishby (2017). *«Opening the Black Box of Deep Neural Networks via Information»*.
https://distill.pub/2017/feature-visualization/, https://stanfordmlgroup.github.io/competitions/mura/

Zürcher Hochschule
für Angewandte Wissenschaften

Conclusions

- Important for DL in practice, and hence target of applied research:
  sample efficiency, **robustness**, **interpretability**
- Future work will include:
  Unsupervised and semi-supervised learning approaches
  Novel **object detection** approaches **for many tiny objects**
  Work on **explainable DL**

Swiss Alliance for
Data-Intensive Services

datalab
www.zhaw.ch/datalab

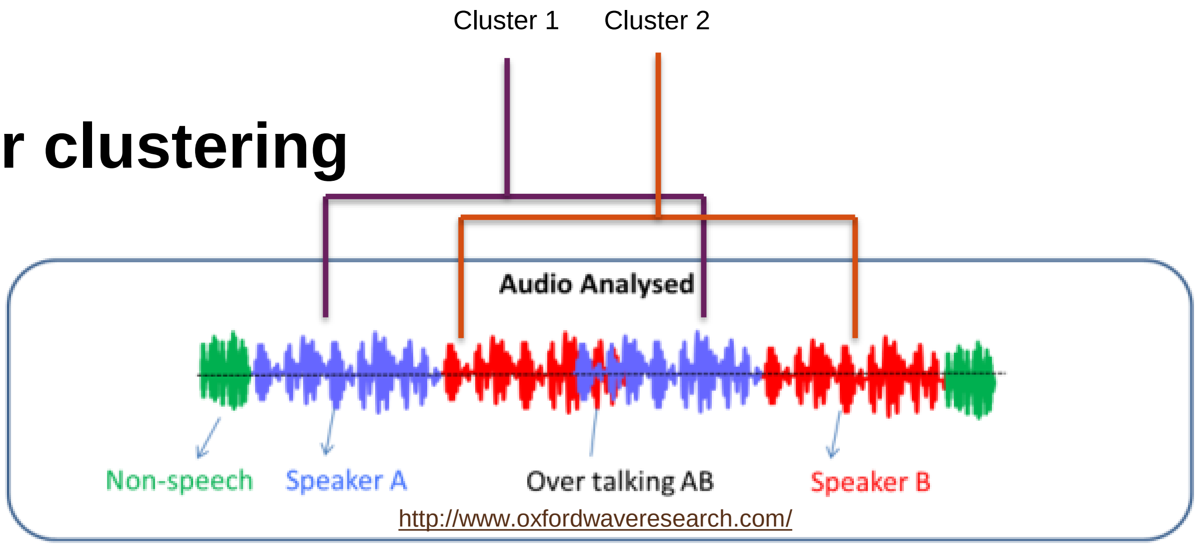
On me:
- Prof. AI/ML, scientific director ZHAW digital, board Data+Service
- thilo.stadelmann@zhaw.ch
- 058 934 72 08
- https://stdm.github.io/

Further contacts:
- Data+Service Alliance: www.data-service-alliance.ch
- Collaboration: datalab@zhaw.ch

➔ Happy to answer questions & requests.

APPENDIX

6. Speaker clustering



For the 630 training utterances, GMMs with 32 mixtures are built a priori, then an identification experiment is run for the 630 test utterances. It yields a satisfactory 0.5% closed set identification error.

[34]. Evaluations typically concentrate on data sets built from broadcast news/shows and meeting recordings, where diarization error rates ranging from 8% to 24% are reported [28][34][45]. These results are confirmed by more recent

The hypothesis of this paper is: the techniques originally developed for speaker verification and identification are not suitable for speaker clustering, taking into account the escalated difficulty of the latter task. However, the processing chain for speaker clustering is quite large – there are many potential areas for improvement. The question is: *where should improvements be made to improve the final result?*

The interpretation of our results has shown that it is the stage of modeling that bears the highest potential: the inclusion of temporal context information among feature vectors is what is crucially missing there. Furthermore, the inclusion

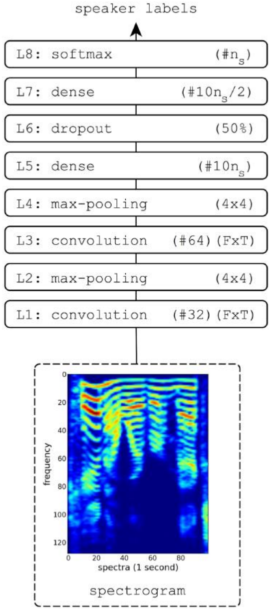
context vector. This corresponds to a syllable length of 130 ms and is found to best capture speaker specific sounds in informal listening experiments over a range of 32–496 ms (in intervals of 16 ms). Our context vector step is one orig-

Stadelmann & Freisleben (2009). *«Unfolding Speaker Clustering Potential: A Biomimetic Approach»*. ACMMM'2009.

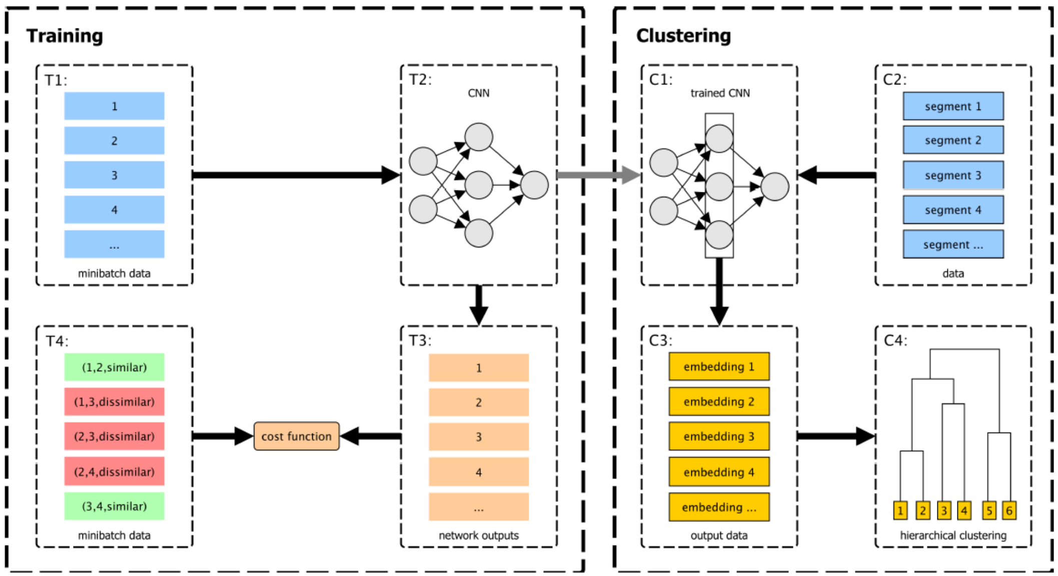
Zürcher Hochschule
für Angewandte Wissenschaften

6. Speaker clustering – exploiting time information

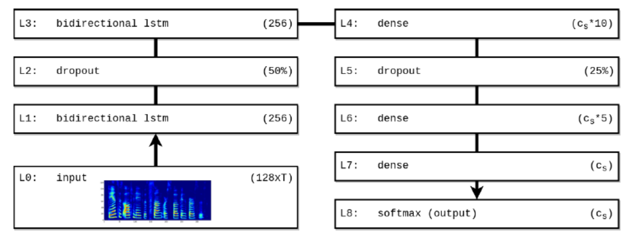
CNN (MLSP'16)



CNN & clustering-loss (MLSP'17)



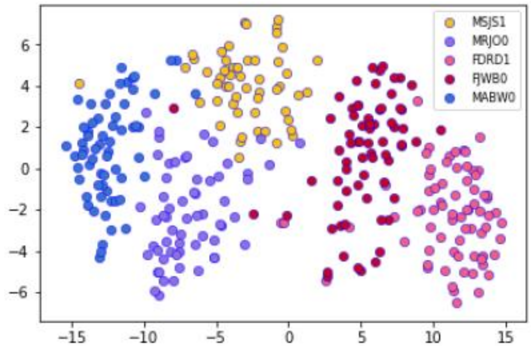
RNN & clustering-loss (ANNPR'18)

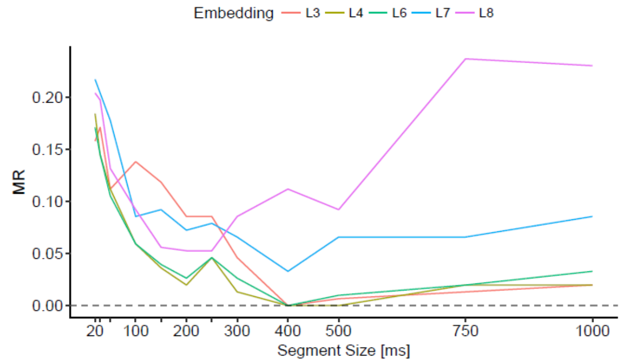


Method	MR	MR (legacy)
RNN /w PKLD	**2.19%** ($\frac{1.25\%+2.5\%+1.25\%+3.75\%}{4}$)	**4.38%** (average of 4 runs)
CNN /w PKLD [24]	-	**5%**
CNN /w cross entropy [23]	-	**5%**
ν-SVM [40]	6.25%	-
GMM/MFCC [40]	12.5%	-

Lukic, Vogt, Dürr & Stadelmann (2016). «Speaker Identification and Clustering using Convolutional Neural Networks». MLSP'2016.
Lukic, Vogt, Dürr & Stadelmann (2017). «Learning Embeddings for Speaker Clustering based on Voice Equality». MLSP'2017.
Stadelmann, Glinski-Haefeli, Gerber & Dürr (2018). «Capturing Suprasegmental Features of a Voice with RNNs for Improved Speaker Clustering». ANNPR'2018.

6. Speaker clustering – learnings & future work





«Pure» voice modeling seems largely solved

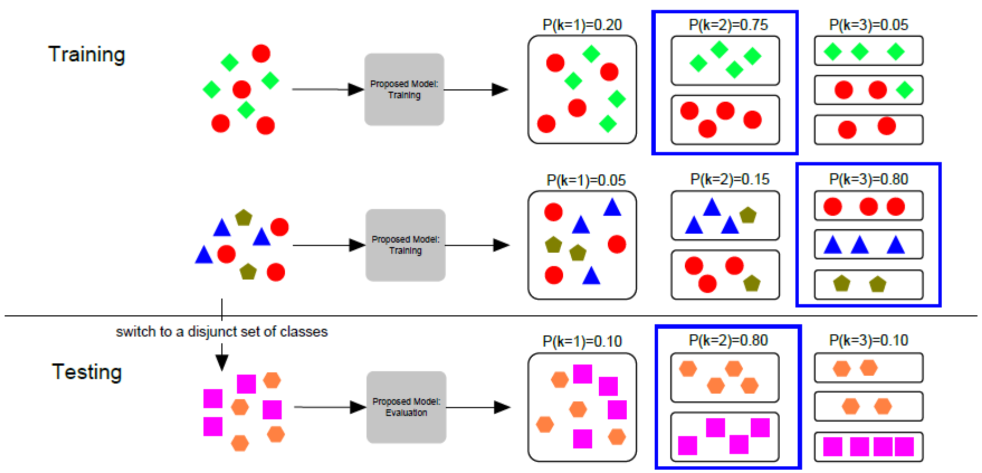
- RNN **embeddings work well** (see t-SNE plot of single segments)
- RNN model robustly exhibits *the predicted* **«sweet spot» for** the used **time information**
- Speaker clustering on clean & reasonably long input works **an order of magnitude better** (*as predicted*)
- Additionally, using a smarter clustering algorithm on top of embeddings makes **clustering on TIMIT as good as identification** (see ICPR'18 paper on dominant sets)

Future work

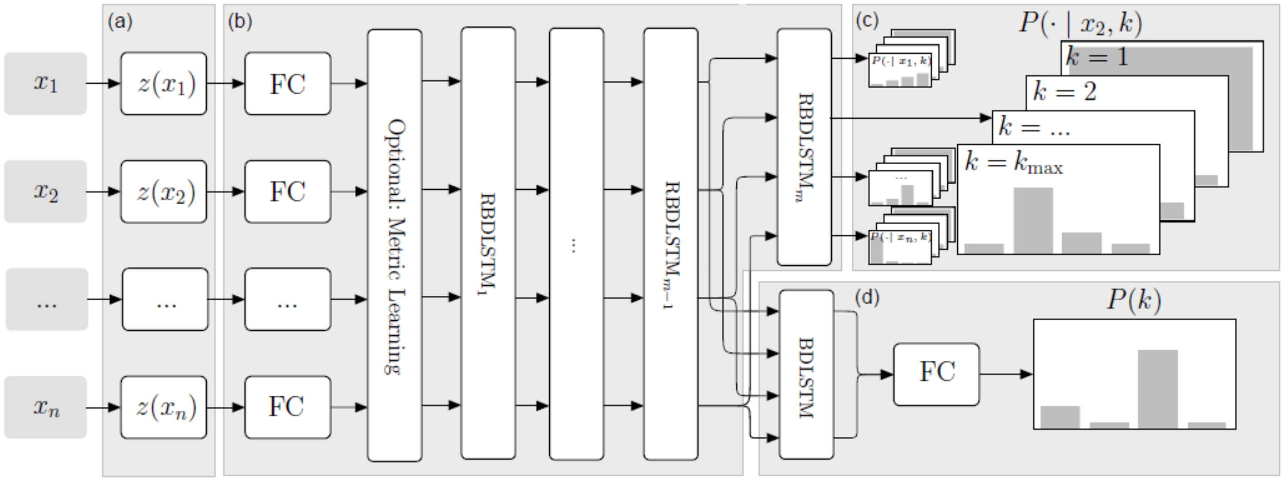
- Make models robust on **real-worldish data** (noise and more speakers/segments)
- Exploit findings for robust reliable **speaker diarization**
- **Learn** embeddings and the clustering algorithm **end to end**

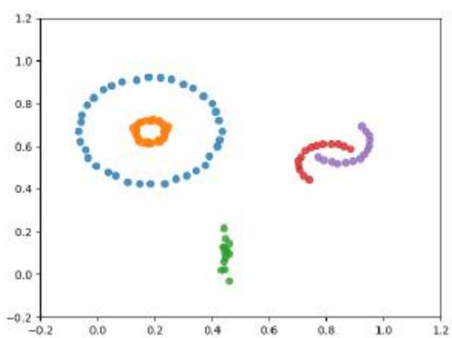
Hibraj, Vascon, Stadelmann & Pelillo (2018). «Speaker Clustering Using Dominant Sets». ICPR'2018.
Meier, Elezi, Amirian, Dürr & Stadelmann (2018). «Learning Neural Models for End-to-End Clustering». ANNPR'2018.

7. Learning to cluster



7. Learning to cluster – architecture & examples





Meier, Elezi, Amirian, Dürr & Stadelmann (2018). *«Learning Neural Models for End-to-End Clustering»*. ANNPR'2018.

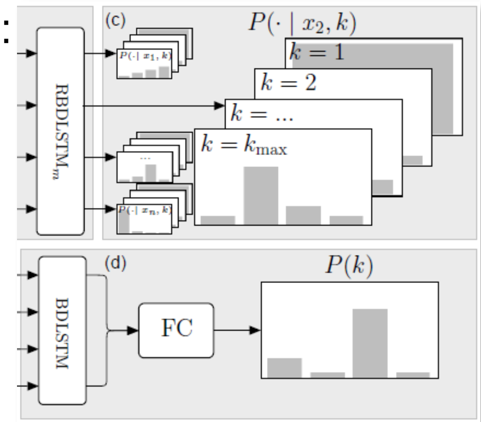
7. Learning to cluster – loss

Probability of two instances ***i*, *j*** being **in the same cluster** ℓ (of k clusters):

$$P_{ij}(k) = \sum_{\ell=1}^{k} P(\ell \mid x_i, k) P(\ell \mid x_j, k).$$

Probability of two instances i, j being in the same cluster ℓ **in general**:

$$P_{ij} = \sum_{k=1}^{k_{\max}} P(k) \sum_{\ell=1}^{k} P(\ell \mid x_i, k) P(\ell \mid x_j, k).$$



Cluster assignment loss (with $y_{ij} = 1$ *iif* the two instances are from the same cluster, 0 otherwise):
Weighted binary cross entropy (weights account for imbalance due to more dissimilar pairs)

$$L_{\text{ca}} = \frac{-2}{n(n-1)} \sum_{i<j} \left(\varphi_1 y_{ij} \log(P_{ij}) + \varphi_2 (1 - y_{ij}) \log(1 - P_{ij})\right)$$

Number of cluster loss:
Categorical cross entropy

$$L_{\text{cc}} = -\log(P(k))$$

Total loss:

$$L_{\text{tot}} = L_{\text{cc}} + \lambda L_{\text{ca}}$$

Swiss Alliance for Data-Intensive Services

Zürcher Hochschule für Angewandte Wissenschaften

The Swiss Alliance for Data-Intensive Services provides a significant contribution to **make Switzerland an internationally recognized hub for data-driven value creation**.

In doing so, we rely on **cooperation in an interdisciplinary expert network** of innovative **companies** and **universities** to combine knowledge from different fields into marketable products and services.

Industrial Members

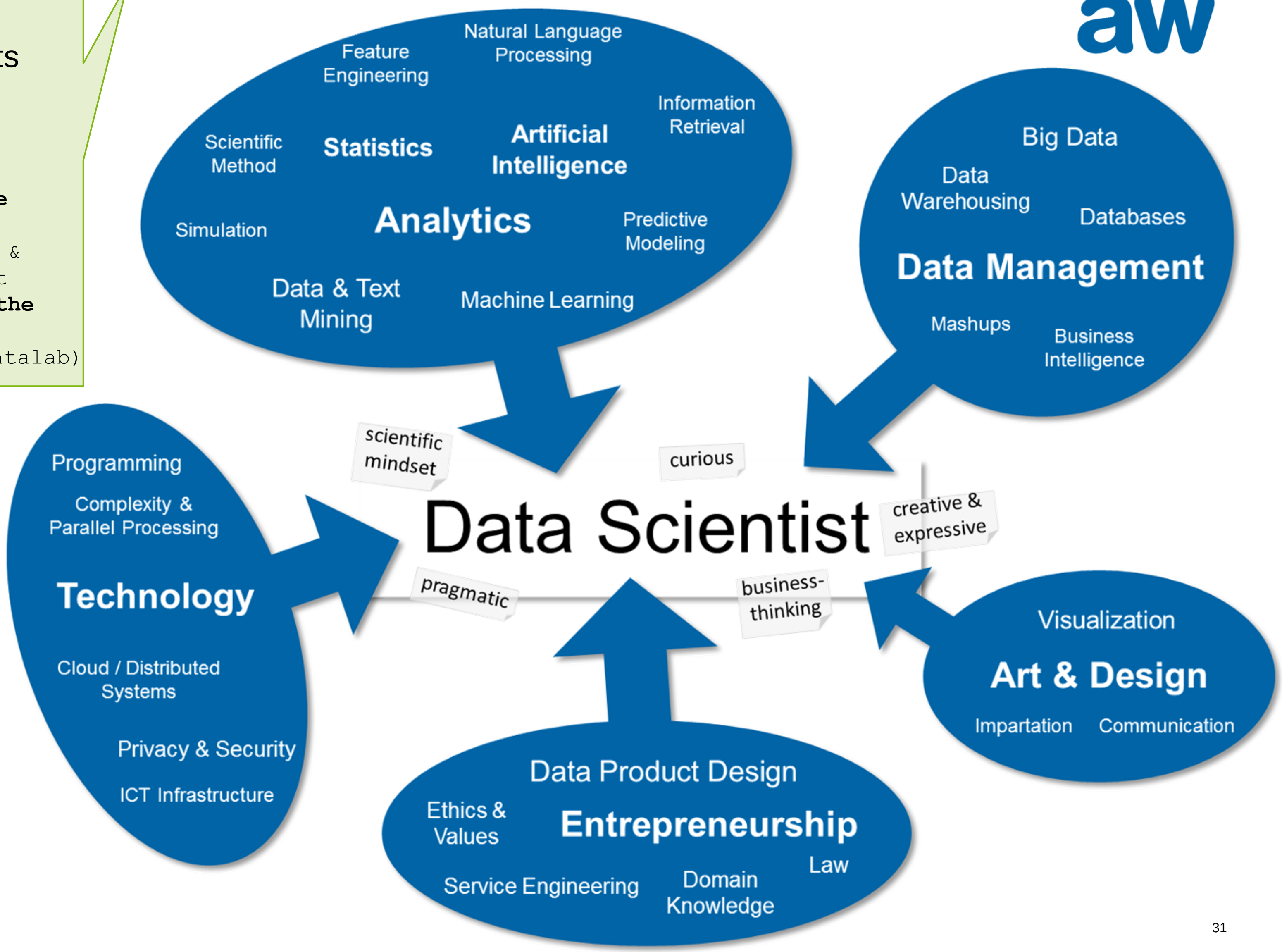
Academic Members

National & International Partners

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.

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

M. Cieliebak & T. Stadelmann, ZHAW

ARGUS MEDIA BASED INTELLIGENCE

zh aw School of Engineering

datalab www.zhaw.ch/datalab

Overview

Partners
Who are we

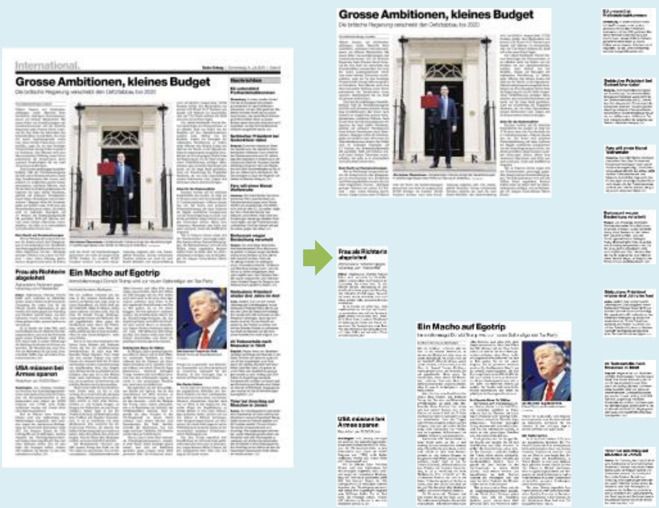
ARGUS der Presse AG
- Switzerland's leading media monitoring and information provider
- Experience of more than 100 years

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

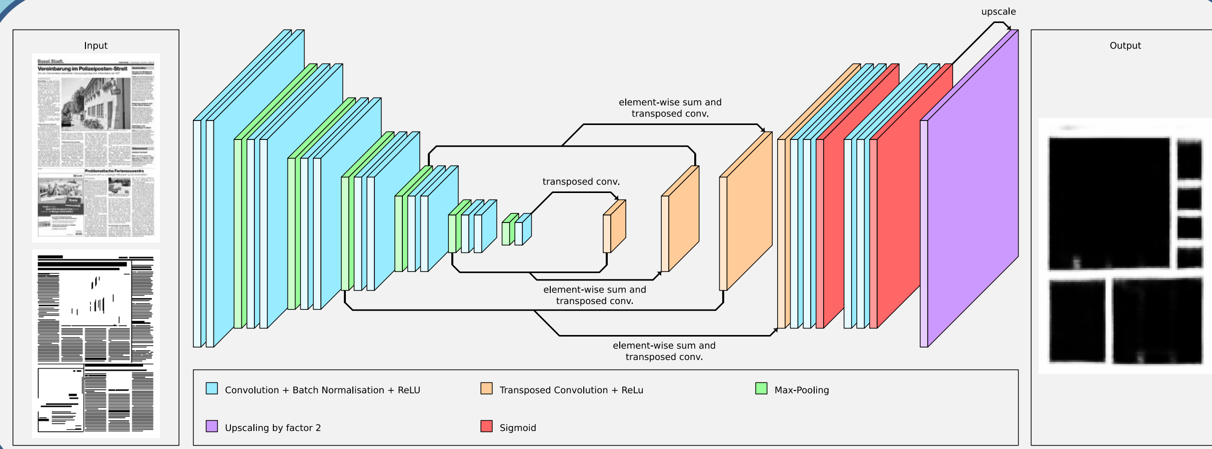
The Project
What do we do

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

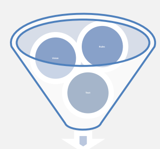
Problem
- Fully automated article segmentation
- Identification of article elements (e.g. title, subtitle, etc.)



Most Successful Approach [3]



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.
[3] B. Meyer, T. Stadelmann, J. Stampfli, M. Arnold, M. Cieliebak. *Fully Convolutional Neural Networks for Newspaper Article Segmentation.* In Proceedings of ICDAR, Kyoto, Japan, 2018.

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