

4 Years of Deep Learning Research @ ZHAW: an Information Engineering Perspective

InIT Meeting, Oct 17, 2018

*T. Stadelmann, M. Amirian, I. Arabaci, M. Arnold, G. F. Duivesteijn, I. Elezi,
M. Geiger, S. Lörwald, B. B. Meier, K. Rombach & L. Tuggener*



How time flies...

First Inspiration @ Zurich ML Meetup #1, Feb 25, 2014

First presentation of DL activities @ SP IE, Nov 14, 2014



How time flies...

First Inspiration @ Zurich ML Meetup #1, Feb 25, 2014

First presentation of DL activities @ SP IE, Nov 14, 2014



Thema «Deep Learning»

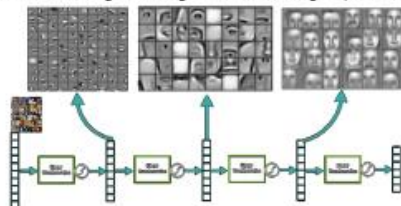


Aktuell «heissestes Thema» im Machine Learning

- (Aktueller) Kick-Off: Bengio et al., «A fast learning algorithm for deep belief nets», 2006
- Seitdem: Viele Pattern Recognition Benchmarks (teils um Grössenordnungen) durch DL-Ansätze verbessert (z.B. Schmidhuber, «Deep Learning in Neural Networks: An Overview», 2014)

Technisches

- Neuronales Netz mit vielen Schichten («deep»), einzeln vortrainiert z.B. als Autoencoder, ganzheitlich feinstabgestimmt durch Backpropagation
- Spezialität: «Unsupervised Feature Learning» → Verfahren lernt selbständig Hierarchien «guter» (d.h. ähnlich wie der Mensch) Repräsentationen der Daten
- Anwendung: z.Zt. Pattern Recognition auf grossen Datenmengen (z.B. Bilder, Text, Audio)



Impact? Markt? Research?



Strategische Bedeutung

- Vor 2 Jahren etwa <10 Forschungsgruppen an Top-Unis (u.a. IDSIA in der Schweiz)
- Januar 2014: Google kauft Fa. DeepMind für 500 Mio. \$ (Gründung u.a. von Postdocs IDSIA)
- Aktuell:
 - Wissenschaftlich an der Grenze zwischen Forschung und Anwendung → stark abhängig von Anwendungsfall
 - Bücher, Libraries, Firmen entstehen erst
 - Prognose: in 2-4 Jahren ein Toolam Markt wie «SVM»



Deep Learning an der ZHAW

- Angeschaffte Hardware: 2 Power-Workstations mit leistungsstarken GPUs «ready» (InIT & IDP)
- Deep Learning Journals Club: 12 Personen lesen gemeinsam UFLDL Tutorial aus Stanford
- Mailingliste: deeplearning@downbim.zhaw.ch (für den Journals Club)
- DL Literatur-Archiv: <https://www.dropbox.com/sh/duw20to4uqec73t/AADCPCB9xoHcMIUmK1v8tueGBa7d=0>
- Verschiedene Ideen für interne Anträge (duco am IDP; stdm mit baud; ciel mit uzdi)
- Erste KTI Anträge entstehen (acke mit stdm)

→ Für stdm/musy/stmf, den Schwerpunkt und das Datalab ein Fokusthema für die nächsten Jahre

How time flies...

First Inspiration @ Zurich ML Meetup #1, Feb 25, 2014

First presentation of DL activities @ SP IE, Nov 14, 2014



Thema «Deep Learning»

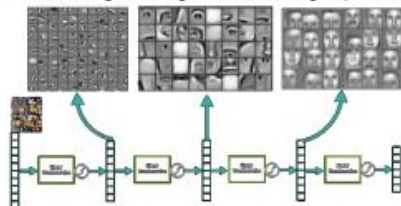


Aktuell «heissestes Thema» im Machine Learning

- (Aktueller) Kick-Off: Bengio et al., «A fast learning algorithm for deep belief nets», 2006
- Seitdem: Viele Pattern Recognition Benchmarks (teils um Grössenordnungen) durch DL-Ansätze verbessert (z.B. Schmidhuber, «Deep Learning in Neural Networks: An Overview», 2014)

Technisches

- Neuronales Netz mit vielen Schichten («deep»), einzeln vortrainiert z.B. als Autoencoder, ganzheitlich feinstabgestimmt durch Backpropagation
- Spezialität: «Unsupervised Feature Learning» → Verfahren lernt selbständig Hierarchien «guter» (d.h. ähnlich wie der Mensch) Repräsentationen der Daten
- Anwendung: z.Zt. Pattern Recognition auf grossen Datenmengen (z.B. Bilder, Text, Audio)



Impact? Markt? Research?



Strategische Bedeutung

- Vor 2 Jahren etwa <10 Forschungsgruppen an Top-Unis (u.a. IDSIA in der Schweiz)
- Januar 2014: Google kauft Fa. DeepMind für 500 Mio. \$ (Gründung u.a. von Postdocs IDSIA)
- Aktuell:
 - Wissenschaftlich an der Grenze zwischen Forschung und Anwendung → stark abhängig von Anwendungsfall
 - Bücher, Libraries, Firmen entstehen erst
 - Prognose: in 2-4 Jahren ein Tool am Markt wie «SVM»



Deep Learning an der ZHAW

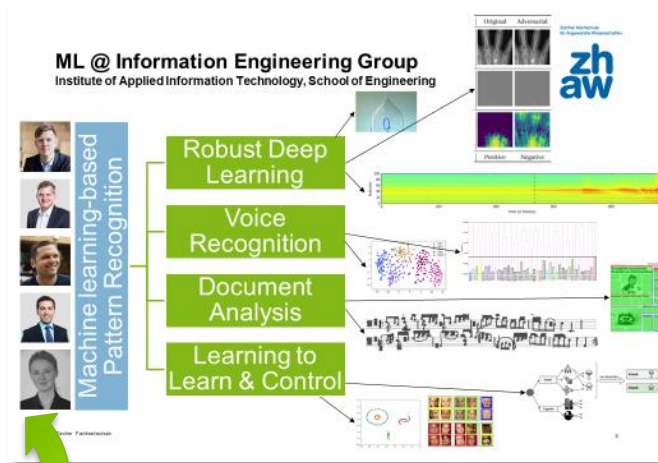
- Angeschaffte Hardware: 2 Power-Workstations mit leistungsstarken GPUs «ready» (InIT & IDP)
- Deep Learning Journals Club: 12 Personen lesen gemeinsam UFLDL Tutorial aus Stanford
- Mailingliste: deeplearning@downbim.zhaw.ch (für den Journals Club)
- DL Literatur-Archiv: <https://www.dropbox.com/sh/duw20to4uqec73t/AADCPCB9xoHcMIUjK1v8ueGBa7d=0>
- Verschiedene Ideen für interne Anträge (duco am IDP; stdm mit baud; ciel mit uzdi)
- Erste KTI Anträge entstehen (acke mit stdm)

→ Für stdm/musy/stmf, den Schwerpunkt und das Datalab ein Fokusthema für die nächsten Jahre

How time flies...

First Inspiration @ Zurich ML Meetup #1, Feb 25, 2014

First presentation of DL activities @ SP IE, Nov 14, 2014



Thema «Deep Learning»

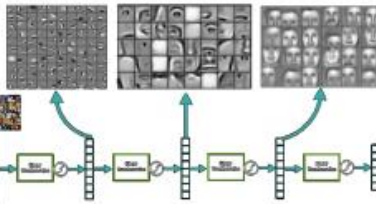


Aktuell «heissestes Thema» im Machine Learning

- (Aktueller) Kick-Off: Bengio et al., «A fast learning algorithm for deep belief nets», 2006
- Seitdem: Viele Pattern Recognition Benchmarks (teils um Grössenordnungen) durch DL-Ansätze verbessert (z.B. Schmidhuber, «Deep Learning in Neural Networks: An Overview», 2014)

Technisches

- Neuronales Netz mit vielen Schichten («deep»), einzeln vortrainiert z.B. als Autoencoder, ganzheitlich feinstabgestimmt durch Backpropagation
- Spezialität: «Unsupervised Feature Learning» → Verfahren lernt selbständig Hierarchien «guter» (d.h. ähnlich wie der Mensch) Repräsentationen der Daten
- Anwendung: z.Zt. Pattern Recognition auf grossen Datenmengen (z.B. Bilder, Text, Audio)



Impact? Markt? Research?



Strategische Bedeutung

- Vor 2 Jahren etwa <10 Forschungsgruppen an Top-Unis (u.a. IDSIA in der Schweiz)
- Januar 2014: Google kauft Fa. DeepMind für 500 Mio. \$ (Gründung u.a. von Postdocs IDSIA)
- Aktuell:
 - Wissenschaftlich an der Grenze zwischen Forschung und Anwendung → stark abhängig von Anwendungsfall
 - Bücher, Libraries, Firmen entstehen erst
 - Prognose: in 2-4 Jahren ein Tool am Markt wie «SVM»



Deep Learning an der ZHAW

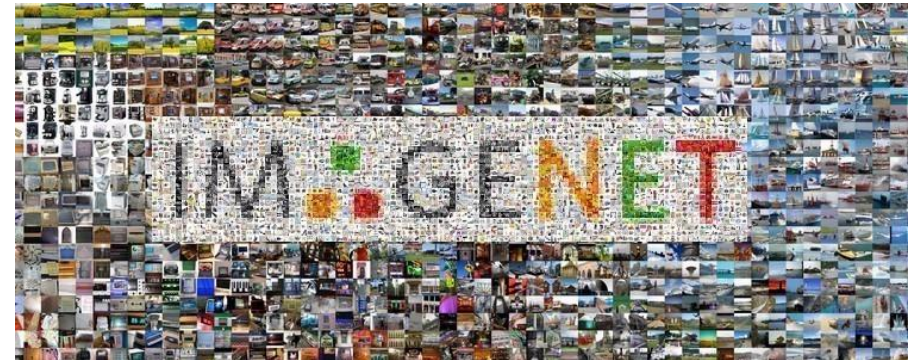
- Angeschaffte Hardware: 2 Power-Workstations mit leistungsstarken GPUs «ready» (InIT & IDP)
- Deep Learning Journals Club: 12 Personen lesen gemeinsam UFLDL Tutorial aus Stanford
- Mailingliste: deeplearning@downbird.zhaw.ch (für den Journals Club)
- DL Literatur-Archiv: <https://www.dropbox.com/sh/3duw20to4uqec73t/AADCPCB9xoHcMIJmK1v8tueGBa7d=0>
- Verschiedene Ideen für interne Anträge (duo am IDP; stdm mit baud; ciel mit uzdi)
- Erste KTI Anträge entstehen (acke mit stdm)

→ Für stdm/musy/stmf, den Schwerpunkt und das Datalab ein Fokusthema für die nächsten Jahre

Research in the wild and in the lab



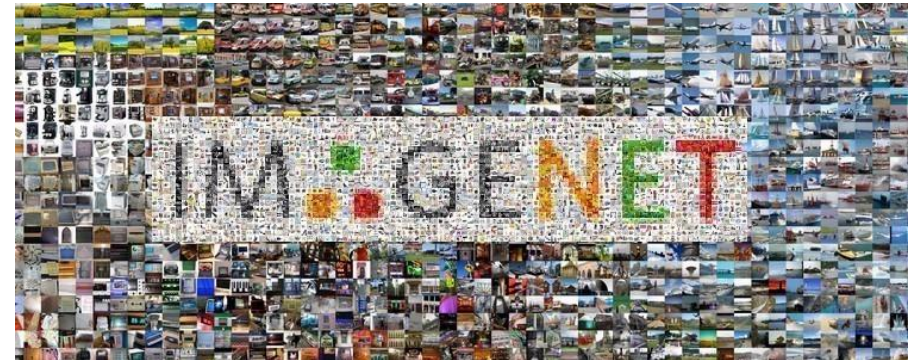
Research in the wild and in the lab



Motivated by general progress

- Given *known environment* (learning target, data, evaluation metric)
- Goal: *fundamental advance* in method

Research in the wild and in the lab



Motivated by application

- Facing *unclear/unprecedented learning target & data quality / quantity* issues

→ Goal: *new product & advance in method*

Motivated by general progress

- Given *known environment* (learning target, data, evaluation metric)

→ Goal: *fundamental advance in method*

Research in the wild and in the lab



Motivated by application

- Facing *unclear/unprecedented learning target & data quality / quantity* issues

→ Goal: *new product & advance in method*

LETTER

Human-level control through deep reinforcement learning

Vedantoyee Khadka¹, Kumar Karanikaugath², David Silver³, Andrew A. Senior⁴, Joel Veness⁵, Marc G. Bellemare⁶, Alex Graves⁷, Marcio de Almeida⁸, Andrew B. Dalrymple⁹, George Gordon¹⁰, Greg De Raedt¹¹, Charles Bessière¹², Armin Burchard¹³, Saurabh Amankar¹⁴, Hideo King¹⁵, Hitarshan Karmakar¹⁶, Dhan Winterer¹⁷, Shantnu Laha¹⁸, R. Dhanraj Hanmani¹⁹

The theory of reinforcement learning provides a normative account, deeply rooted in psychological and neuroscience perspectives on animal behaviour, of how agents may optimise their control of an environment. To use reinforcement learning successfully in situations approaching real-world complexity, however, agents are confronted with a difficult task: they must discover efficient representations of the environment that allow them to generalise past experience to new situations. Remarkably, humans and other animals seem to solve this problem through a harmonious combination of reinforcement learning and hierarchical sensory processing systems¹, the former exhibiting a wealth of internal state and the latter providing a means for abstracting the most salient features of the environment to a limited set of actions². Their applicability has particularly been limited to domains in which useful features can be hand-coded, or to domains with fully observed, low-dimensional state spaces. Here we use recent advances in training deep neural networks^{3,4}, to develop a novel artificial agent, termed a deep Q-network, that can learn successful policies across diverse high-dimensional sensory inputs using only raw and unlabelled training data. We tested this agent on the challenging domain of classic Atari 2600 games⁵. We demonstrate that the deep Q-network agent, receiving only the pixels and the game score as inputs, was able to surpass the performance of all previous algorithms and achieve a level comparable to that of professional human games players across a set of 49 games, using the same algorithm, network architecture and hyperparameters. This work bridges the divide between high-dimensional sensory inputs and actions, resulting in the first artificial agent that is capable of learning to excel at a diverse array of challenging tasks.

We used to create a single algorithm that needed to be able to develop a wide range of competencies on a varied range of challenging tasks—a central goal of general artificial intelligence⁶ that has eluded previous efforts⁷. In addition, this domain-based agent, termed a deep Q-network (DQN), which is able to combine reinforcement learning with a class of artificial neural networks⁸ known as deep neural networks, has led to recent advances in deep neural networks⁹, in which several layers of nodes are used to build up progressively more abstract representations of the data. Here we made it possible for artificial neural networks to learn concepts such as object categories directly from raw sensory data. We use one particularly successful architecture, the deep convolutional neural network¹⁰, which uses hierarchical layers of local convolutional filters to mimic the effects of receptive fields—inspired by Hubel and Wiesel's seminal work on feature processing in early visual cortex¹¹—and by exploiting the local spatial correlations present in images. The resulting networks are robust to natural transformations such as change of scale.

We consider tasks in which the agent interacts with an environment through a sequence of observations, actions and rewards. The goal of the

agent is to select actions in a fashion that maximises cumulative future reward. Here formally we use a deep convolutional neural network to approximate the optimal action value function

$$Q^*(s, a) = \max_{a'} \mathbb{E} [\sum_{t=0}^{\infty} \gamma^t r_{t+1} | s, a]$$

which is the maximum over of rewards r , discounted by γ at each time step. Additionally, we use a neural network to approximate the optimal action value function $Q^*(s, a)$ and taking an action a (see Methods).¹² Reinforcement learning is typically not even able to develop when a nonlinear function approximator such as a neural network is used to represent the action value (see below as Q-learning¹³). This instability has several causes: the correlation present in the sequence of observations, the fact that small updates to Q may significantly change the policy and therefore change the data distribution, and the correlation between the action values Q and the target values $r + \gamma \max_{a'} Q(s', a')$. We address these conditions with a novel variation of Q-learning, which uses two key ideas. First, we used a biologically inspired mechanism termed experience replay¹⁴, that randomises over the data, thereby removing correlations in the observations and associated state changes in the data distribution (see below for details). Second, we used an iterative update that adjusts the action values Q towards target values that are only periodically updated, thereby reducing correlations with the target.

While other stable methods exist for training neural networks in the reinforcement learning setting, this is the second best Q-learning¹⁵. These methods involve the repeated training of networks to zero on hundreds of iterations. Consequently, these methods, unlike our algorithm, are too inefficient to be used successfully with large neural networks. We parameterise an approximate value function $Q(s, a)$ using the deep convolutional neural network shown in Fig. 1, in which θ are the parameters that Q depends on. We store the agent's experiences $s_t = (s_t, a_t, r_t, s_{t+1})$ at each time-step t by a database $\mathcal{D} = \{s_t\}_{t=0}^{\infty}$. During learning, we apply Q-learning updates, on samples for minibatches of experience $(s_t, a_t, r_t, s_{t+1}) \in \mathcal{D}$, drawn uniformly at random from the pool of stored samples. The Q-learning update at iteration i uses the following loss function

$$L_i(\theta) = \mathbb{E}_{(s, a, r, s') \sim \mathcal{D}} [(r + \gamma \max_{a'} Q(s', a') - Q(s, a; \theta))^2]$$

in which θ is the discrete factor denoting the agent's behaviour. θ are parameters of the Q-network at iteration i and θ' are the network used to compute the target at iteration i . The target network $Q(s, a; \theta')$ are only updated with the Q-network parameters θ and are held fixed between individual updates (see Methods).

We consider tasks in which the agent interacts with an environment through a sequence of observations, actions and rewards. The goal of the

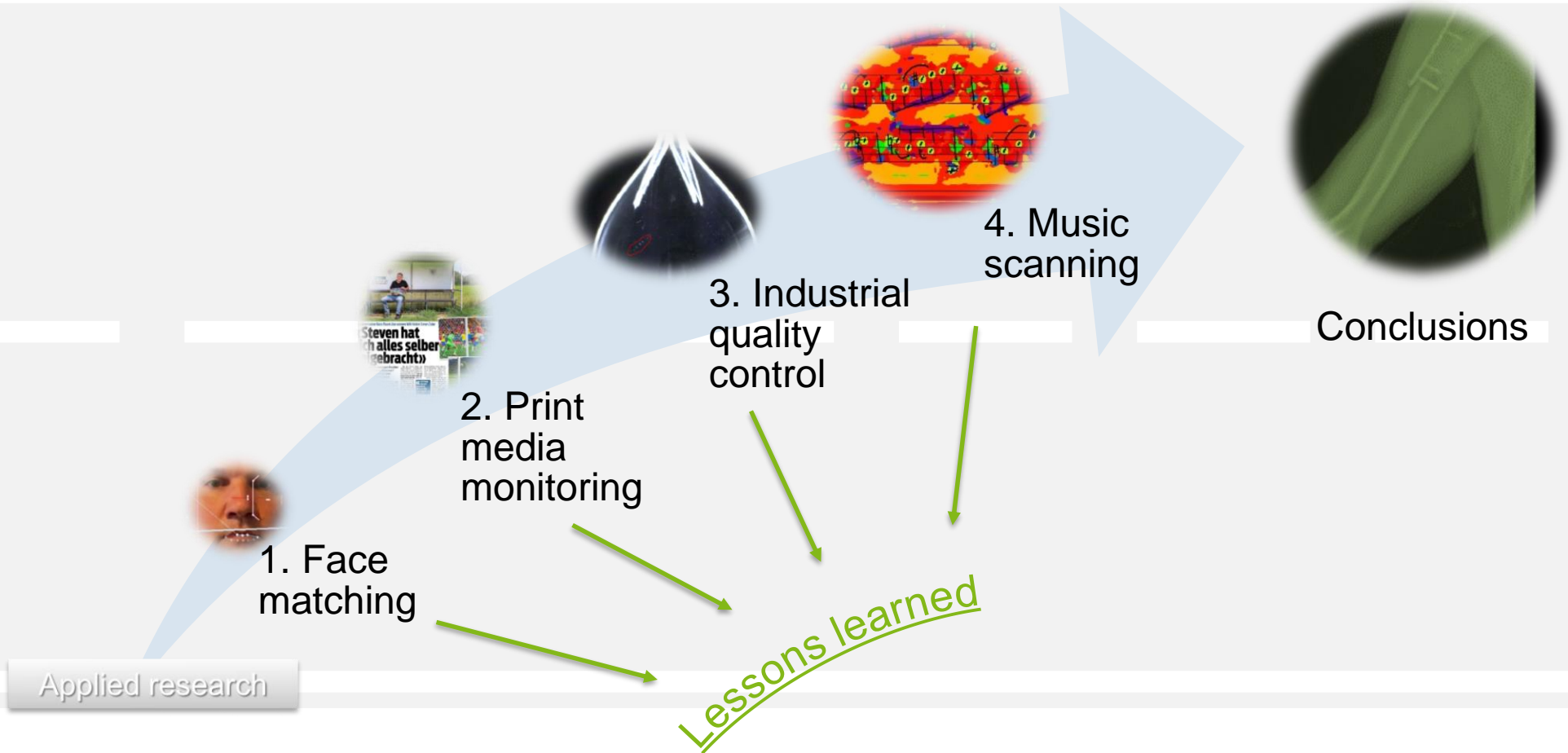
Motivated by general progress

- Given *known environment (learning target, data, evaluation metric)*

→ Goal: *fundamental advance in method*

e.g.


Roadmap



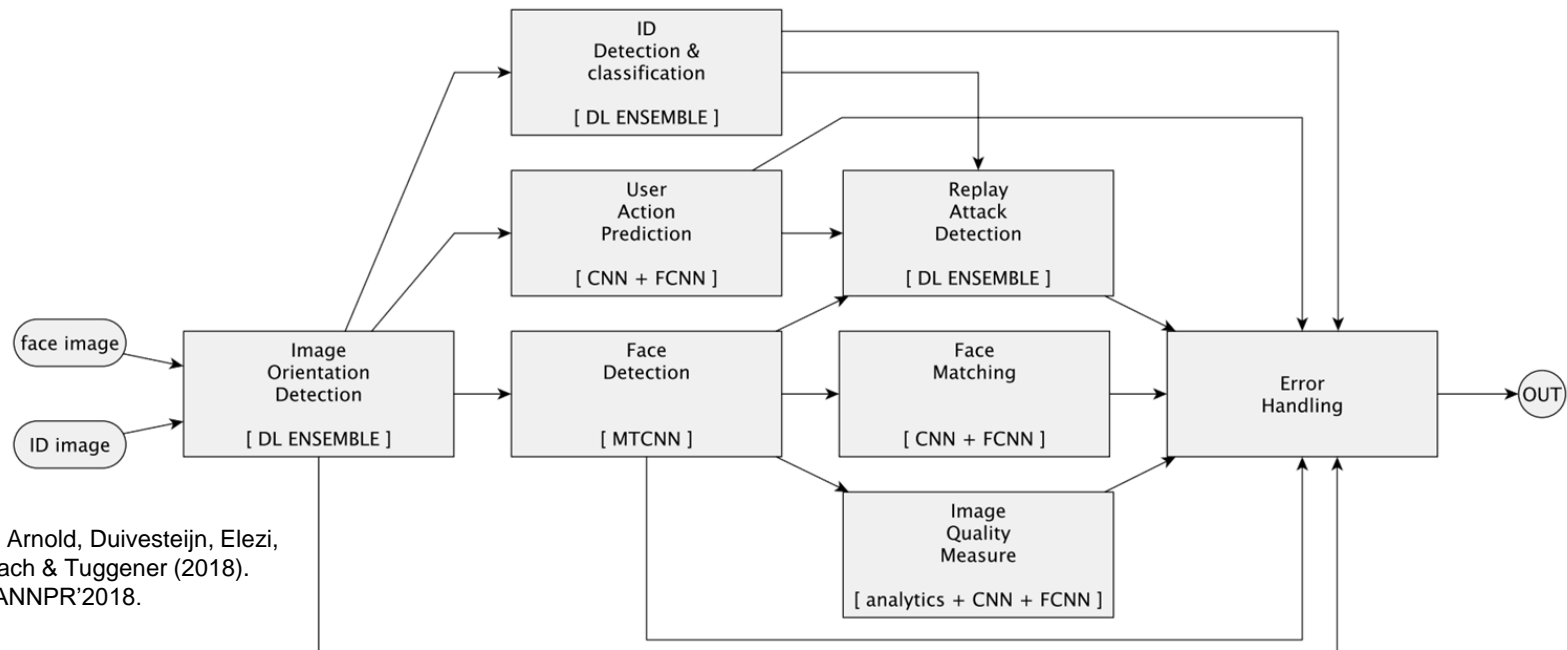
1. Face matching



 **DEEPIIMPACT**

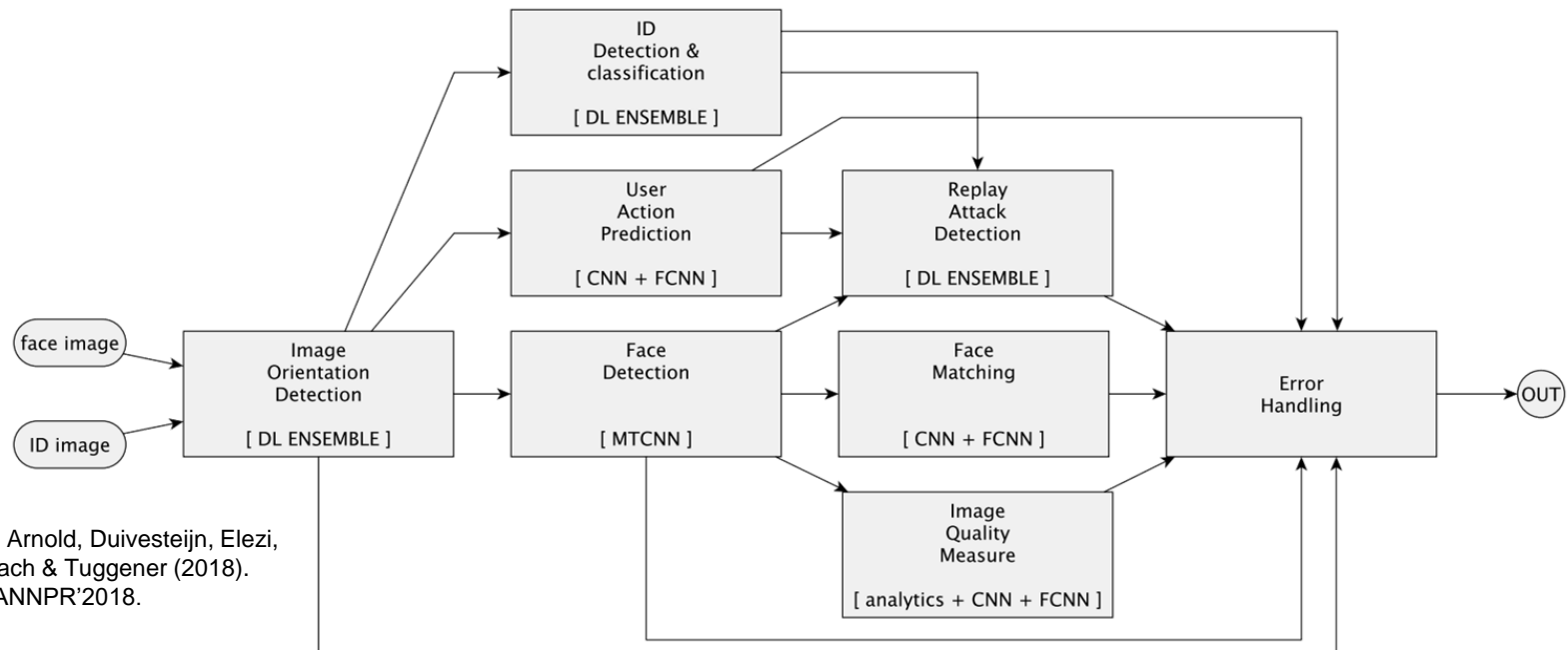
 Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Innosuisse – Swiss Innovation Agency

1. Face matching – challenges & solutions



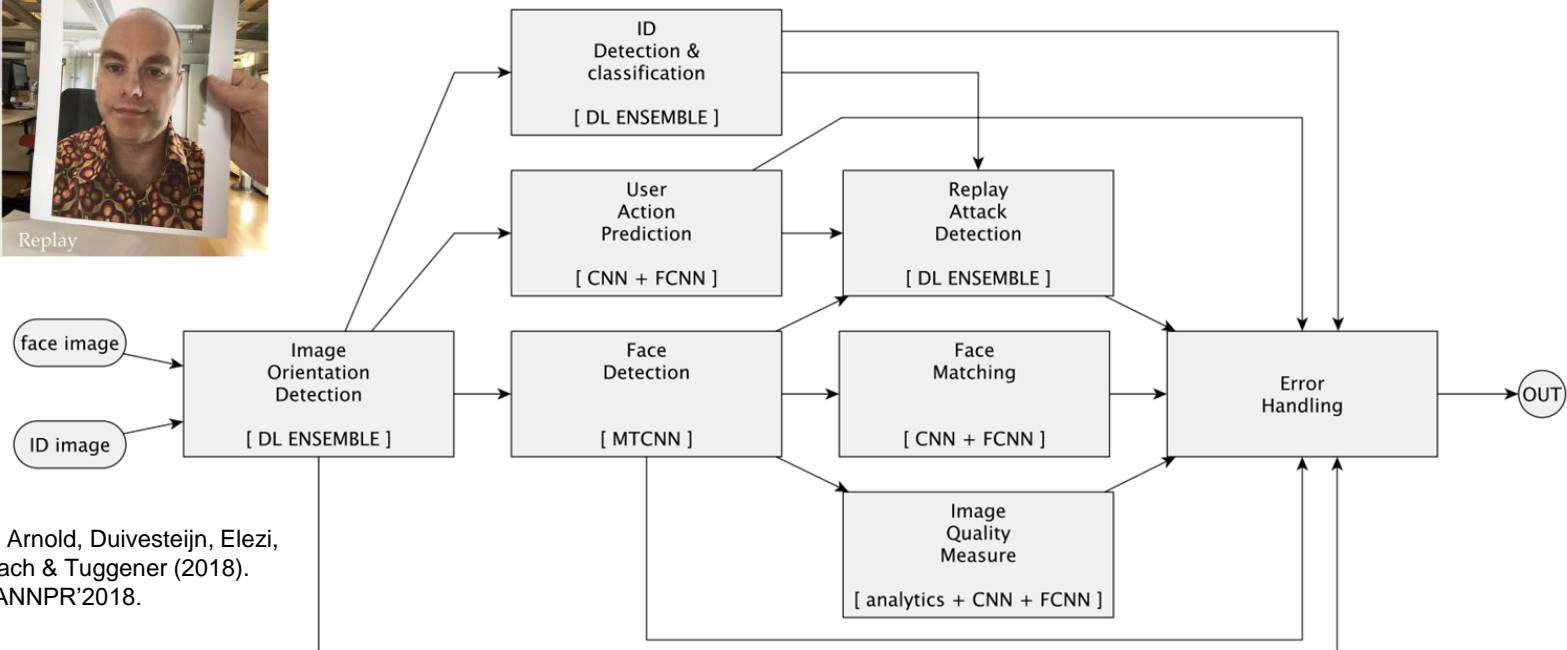
Stadelmann, Amirian, Arabaci, Arnold, Duivesteijn, Elezi, Geiger, Lörwald, Meier, Rombach & Tuggener (2018). «Deep Learning in the Wild». ANNPR'2018.

1. Face matching – challenges & solutions



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

1. Face matching – challenges & solutions



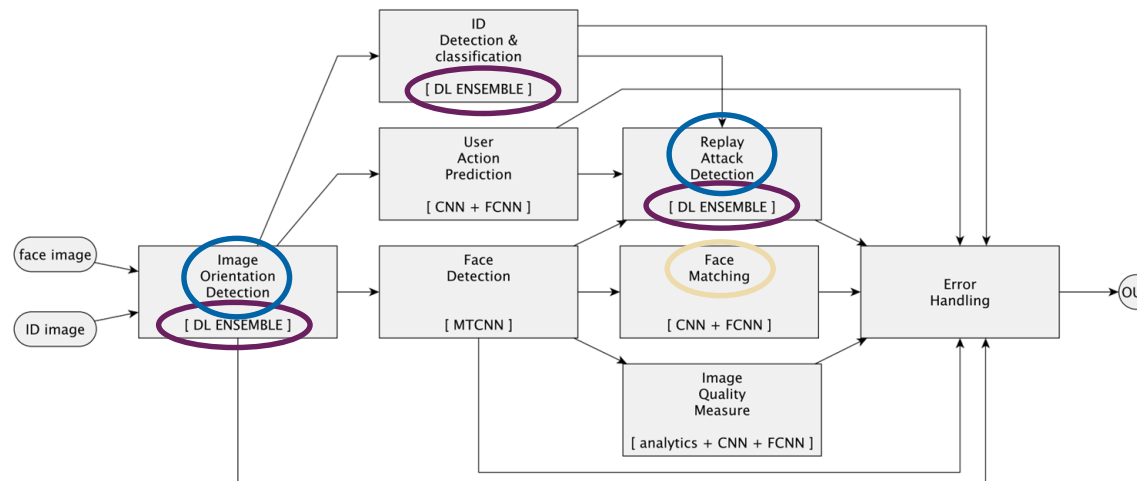
Stadelmann, Amirian, Arabaci, Arnold, Duivesteijn, Elezi, Geiger, Lörwald, Meier, Rombach & Tuggener (2018). «Deep Learning in the Wild». ANNPR'2018.

Lessons learned 1/4



Deployment

- Might involve the buildup of up to dozens of **other machine learning** models to flank the **original core part**.



- **Specialized models** for identifiable sub-problems increase the accuracy in production systems over all-in-one solutions, and **ensembles** of experts help where no single method reaches adequate performance.



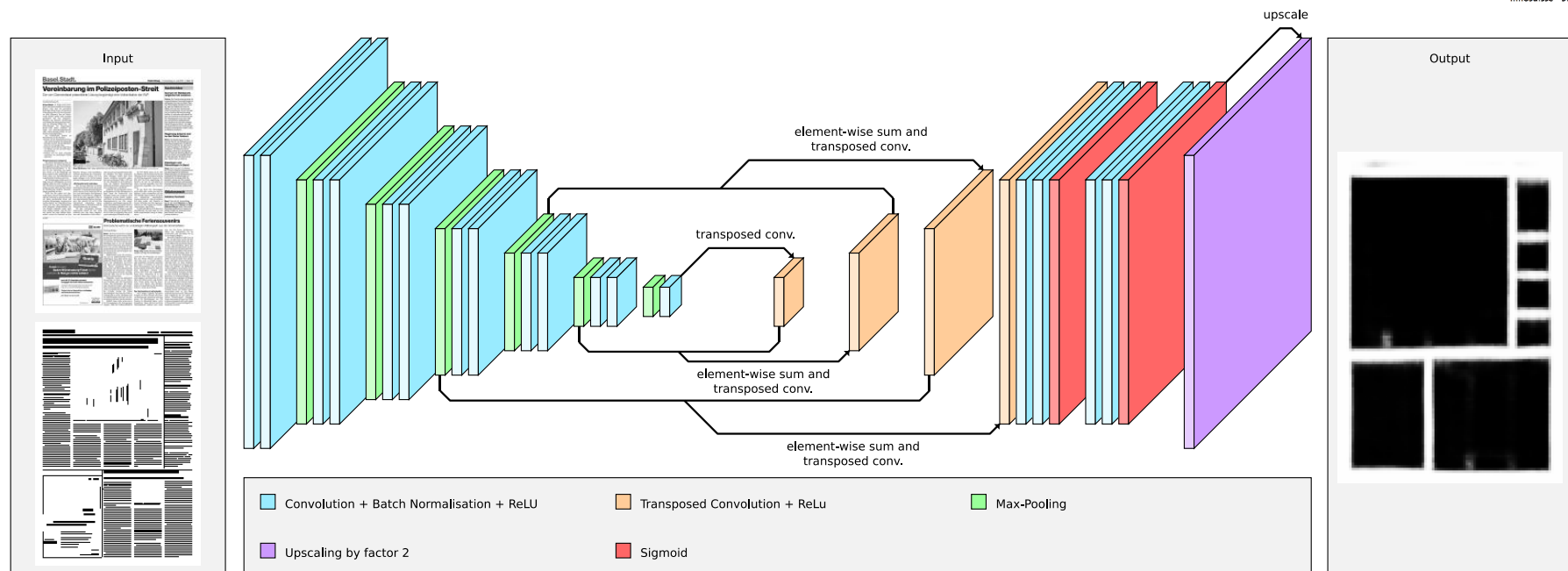
2. Print media monitoring

Task

Challenge

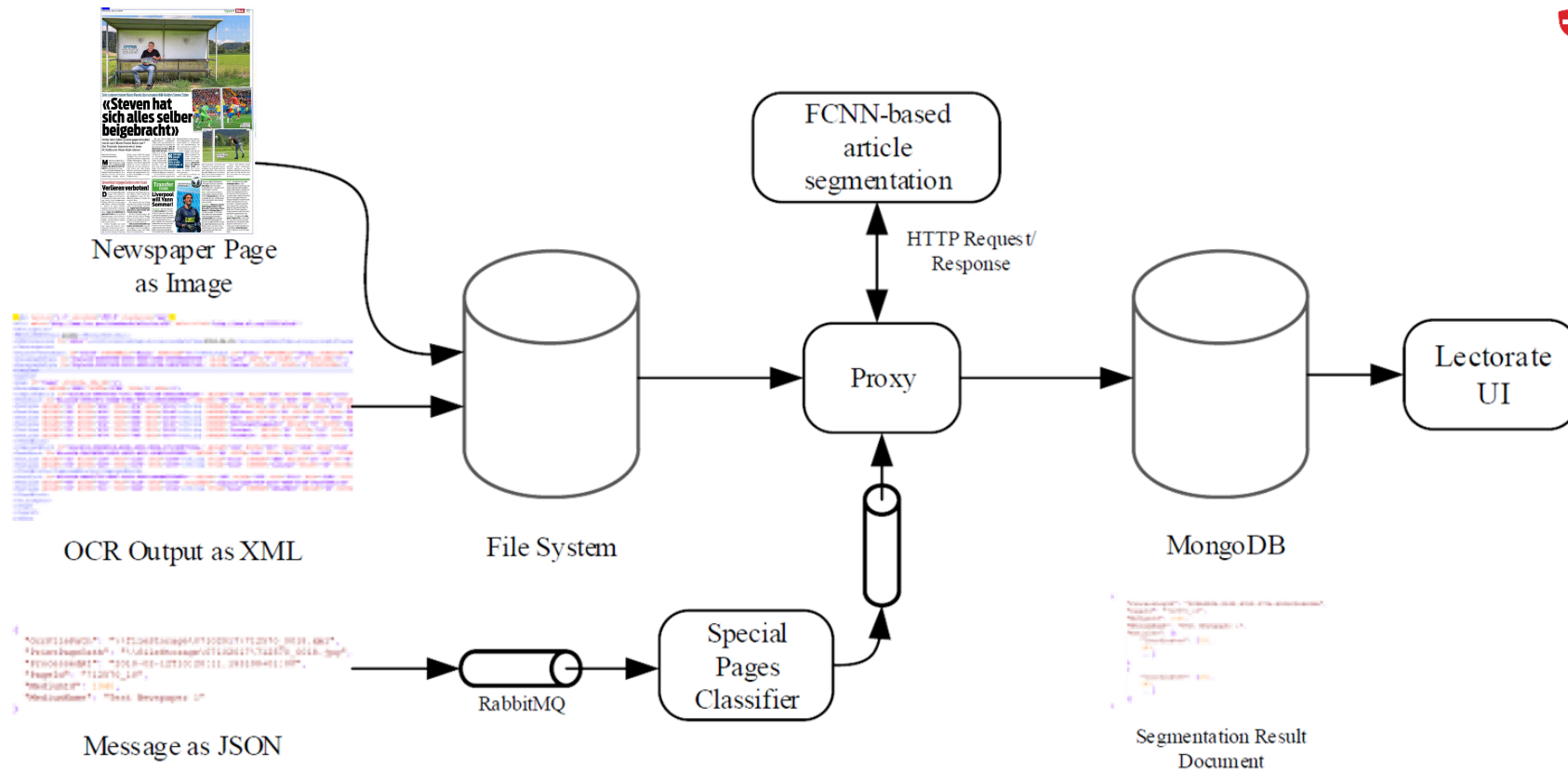
Nuisance

2. Print media monitoring – ML solution



Meier, Stadelmann, Stampfli, Arnold & Cieliebak (2017). «Fully Convolutional Neural Networks for Newspaper Article Segmentation». ICDAR'2017.
 Stadelmann, Tolkachev, Sick, Stampfli & Dürri (2018). «Beyond ImageNet - Deep Learning in Industrial Practice». In: Braschler et al., «Applied Data Science», Springer.

2. Print media monitoring – deployment



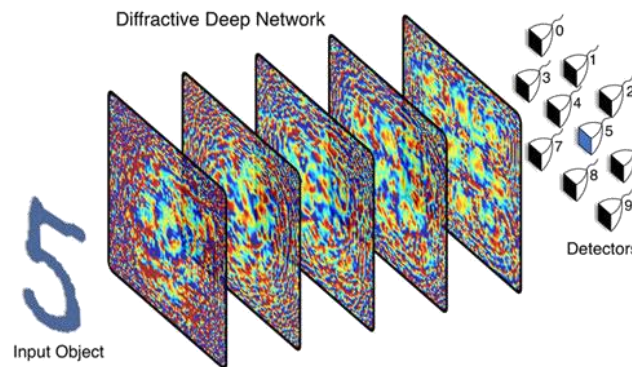
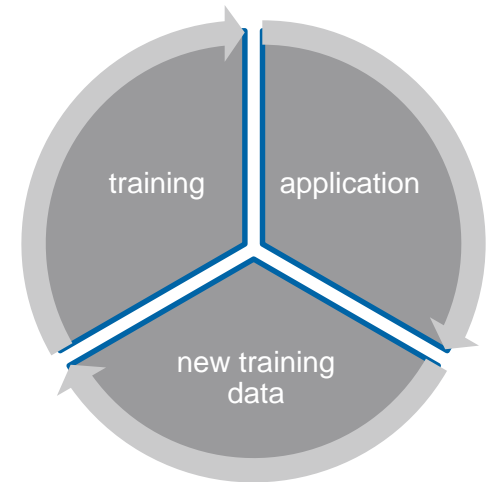
Stadelmann, Amirian, Arabaci, Arnold, Duivesteyn, Elezi, Geiger, Lörwald, Meier, Rombach & Tuggener (2018). «Deep Learning in the Wild». ANNPR'2018.

Lessons learned 2/4



Deployment

- Should include **continuous learning**
- Needs to take care of **processing speed / efficiency**



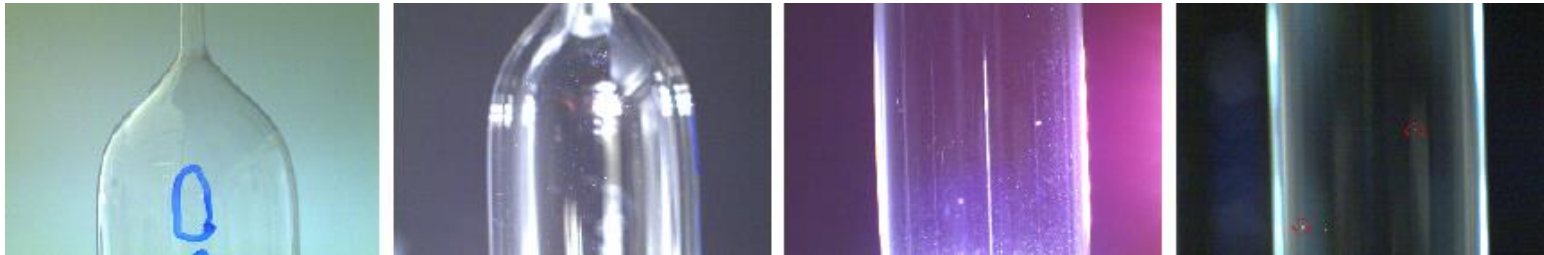
Symbolic image: a CNN in (optical) hardware (Lin et al., 2018).

Lin, Rivenson, Yardimci, Veli, Luo, Jarrahi & Oczan (2018). «All-optical machine learning using diffractive deep neural networks». Science, 26. Jul 2018.

3. Industrial quality control

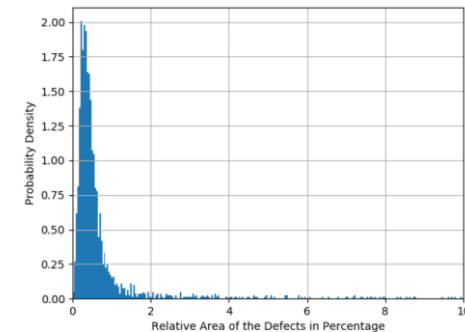
Task

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



Challenges

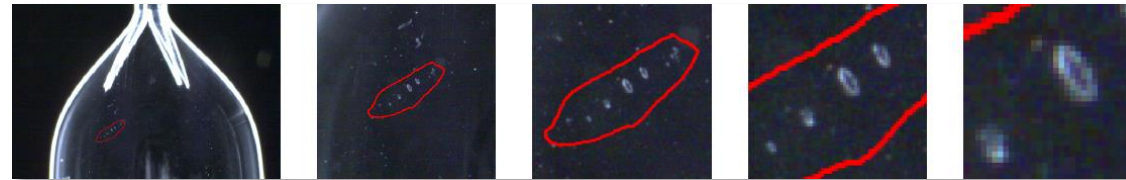
- Non-natural** image source, class **imbalance**, **optical** conditions, **variation** in defect size & shape



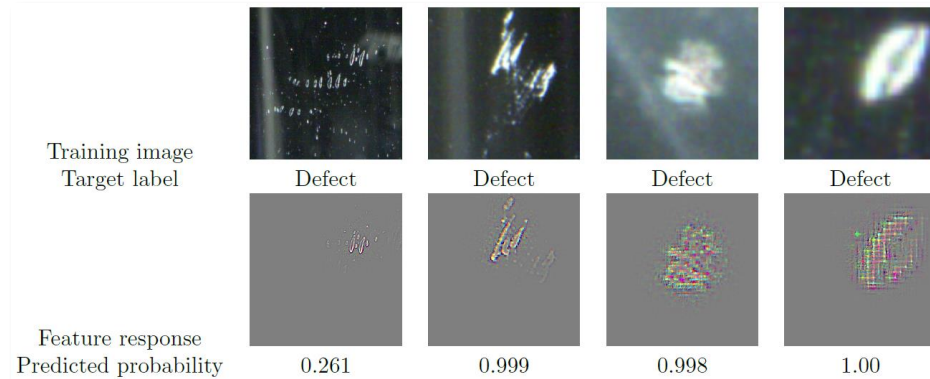
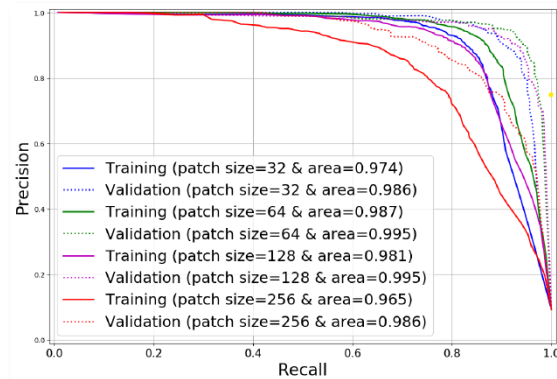
3. Industrial quality control – solutions (Work in progress)

Ingredients

- Weighted loss
- Defect cropping
- Secret sauce



Preliminary results

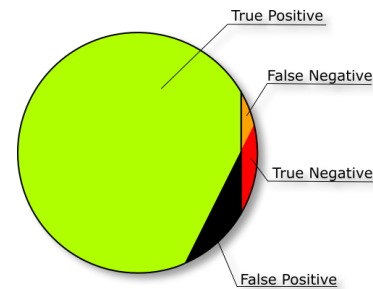


Lessons learned 3/4



Data

- Acquisition usually **needs much more time** than expected, yet is the basis for all subsequent success
- Class **imbalance** & **covariate shift** are usual

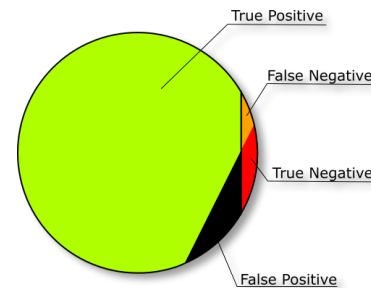


Lessons learned 3/4



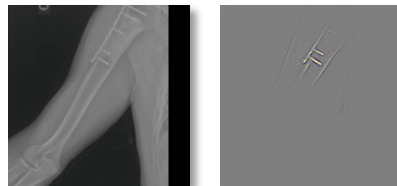
Data

- Acquisition usually **needs much more time** than expected, yet is the basis for all subsequent success
- Class **imbalance** & **covariate shift** are usual



Understanding

- **What has been learned** and **how decisions emerge** help both the user and the developer of neural networks to build trust and improve quality



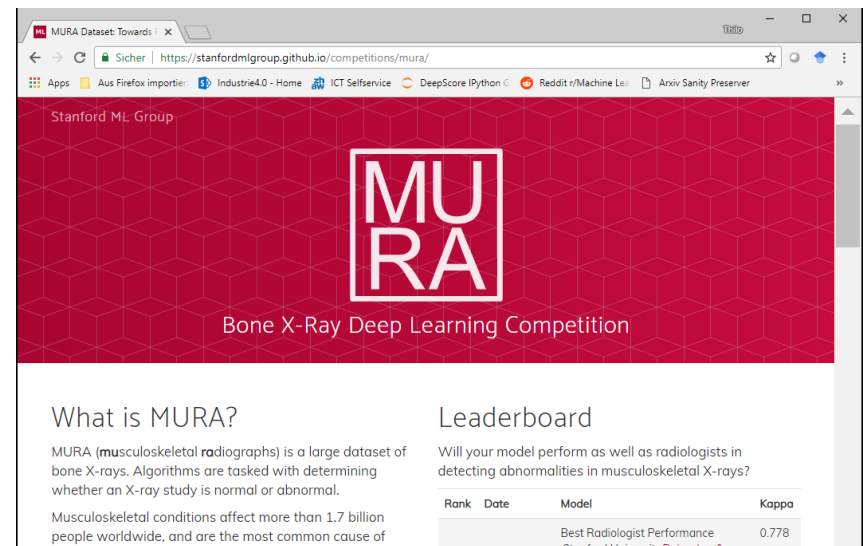
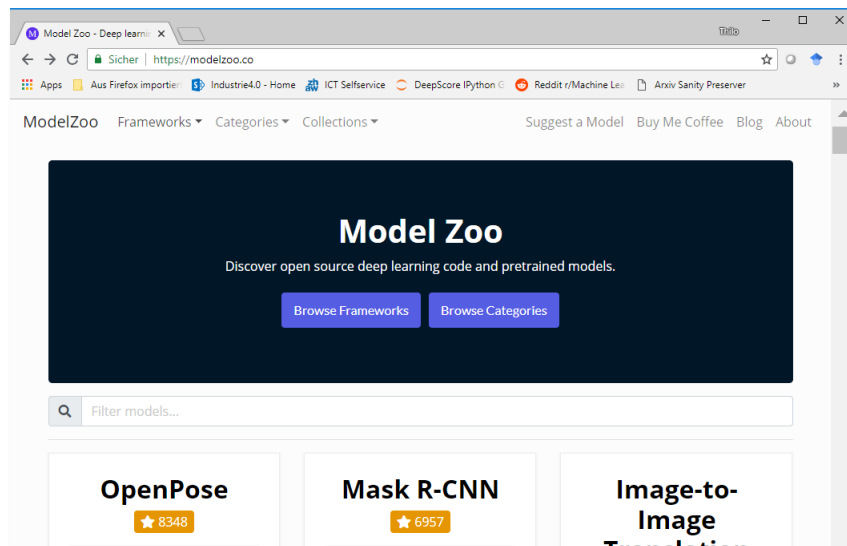
- **Operators and business owners** need a basic understanding of used methods to produce usable ground truth and provide relevant subject matter expertise

Lessons learned 3/4 (contd.)



Simple baselines

- Do a good job in **determining the feasibility** as well as the **potential** of the task at hand when final datasets or novel methods are not yet seen



4. Music scanning

Die Forelle.
Op. 52, No. 14, D. 580.
Für eine Singstimme mit Begleitung des Pianoforte.
Schubert's Werke. Franz Schubert.
Erste Fassung. N. 201

Melodie:
Singsstimme: In mir war Büchlein hel- le, da schoß in fro-her Eil die lau-ni-sche Fo-re-le vor-
Pianoforte: Pi-was mit der Be- le walt an dem U-fer stand und
sah die Forelle, die so schön war.
Ich sah sie nicht, die Forelle, die so schön war.
Ich sah sie nicht, die Forelle, die so schön war.
Ich sah sie nicht, die Forelle, die so schön war.
Ich sah sie nicht, die Forelle, die so schön war.



```
<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE score-partwise SYSTEM "http://www.musicxml.org/@/partwise.dtd" PUBLIC "-//Recordare/DTO 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="1552.22" 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  
Schubert, Die Forelle (Mollisande 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="P1-13">  
        <instrument-name>Ténor</instrument-name>  
      </score-instrument>  
      <midi-instrument id="P1-13">  
        <midi-channel>1</midi-channel>  
        <midi-program>74</midi-program>  
        <volume>78.7402</volume>  
      </midi-instrument>  
    </score-part>  
    <score-part id="P2">  
      <part-name>Piano</part-name>  
      <score-instrument id="P2-13">  
        <instrument-name>Piano</instrument-name>  
      </score-instrument>  
    </score-part>  
  </part-list>  
</score-partwise>
```



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Innosuisse – Swiss Innovation Agency



Die Forelle - Franz Schubert

$\text{♩} = 80$

Voice: In

Piano

7

Vo. ei - nem Büch-lein hel - le, da schoß in fro-her Eil die lau-ni-sche Fo-re-le vor-

4. Music scanning – challenges & solutions



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.

4. Music scanning – challenges & solutions



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.

4. Music scanning – challenges & solutions

The image shows a musical score with various annotations. A callout box highlights four specific annotations:

- (a) accidentalSharp
- (b) keySharp
- (c) augmentationDot
- (d) articStaccatoAbove

Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). «DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects». ICPR'2018.

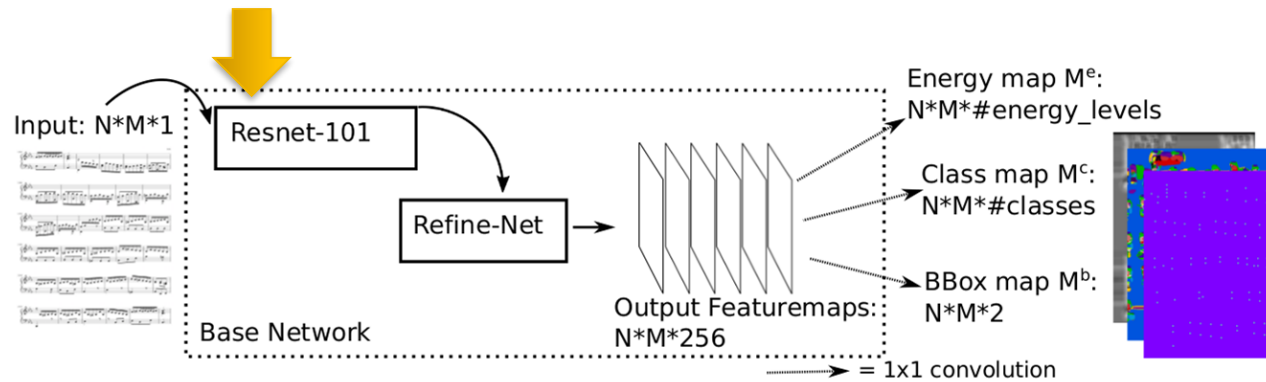
4. Music scanning – challenges & solutions

(a) accidentalSharp (b) keySharp

(c) augmentationDot (d) articStaccatoAbove

Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). «DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects». ICPR'2018.

4. Music scanning – challenges & solutions



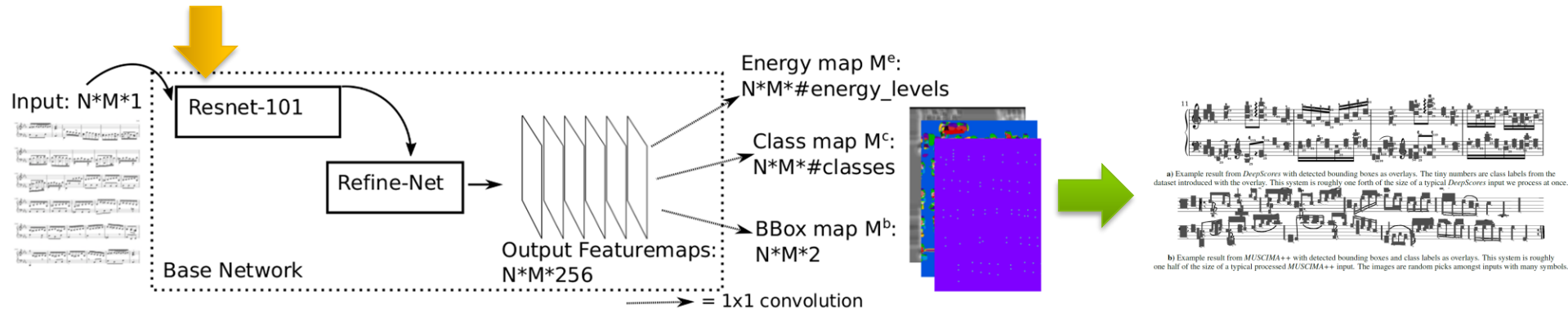
Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). «DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects». ICPR'2018.
 Tuggener, Elezi, Schmidhuber & Stadelmann (2018). «Deep Watershed Detector for Music Object Recognition». ISMIR'2018.

4. Music scanning – challenges & solutions

(a) accidentalSharp (b) keySharp

(c) augmentationDot (d) articStaccatoAbove

SCOREPAD
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Innosuisse – Swiss Innovation Agency

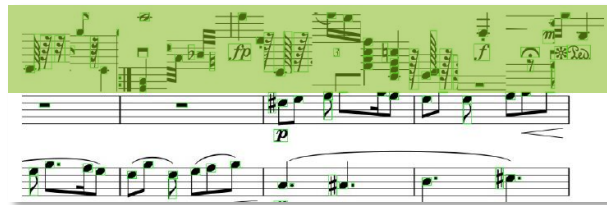


Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). «DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects». ICPR'2018.
Tuggener, Elezi, Schmidhuber & Stadelmann (2018). «Deep Watershed Detector for Music Object Recognition». ISMIR'2018.

4. Music scanning – industrialization (Work in progress)

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.

Lessons learned 4/4



Loss shaping

- Usually necessary to **enable learning** of very complex target functions

*“Initially, the training was **unstable** [...] if directly trained on the **combined weighted loss**. Therefore, we now **train** [...] on each of the **three tasks separately**.*

*We further observed that while the network gets trained on the bounding box prediction and classification, the energy level predictions get worse. To avoid this, the network is **fine-tuned only for the energy level loss** [...]. Finally, the network is retrained on the combined task [...] for a few thousand iterations [...].”*

- This includes **encoding expert knowledge** manually into the model architecture or training setup

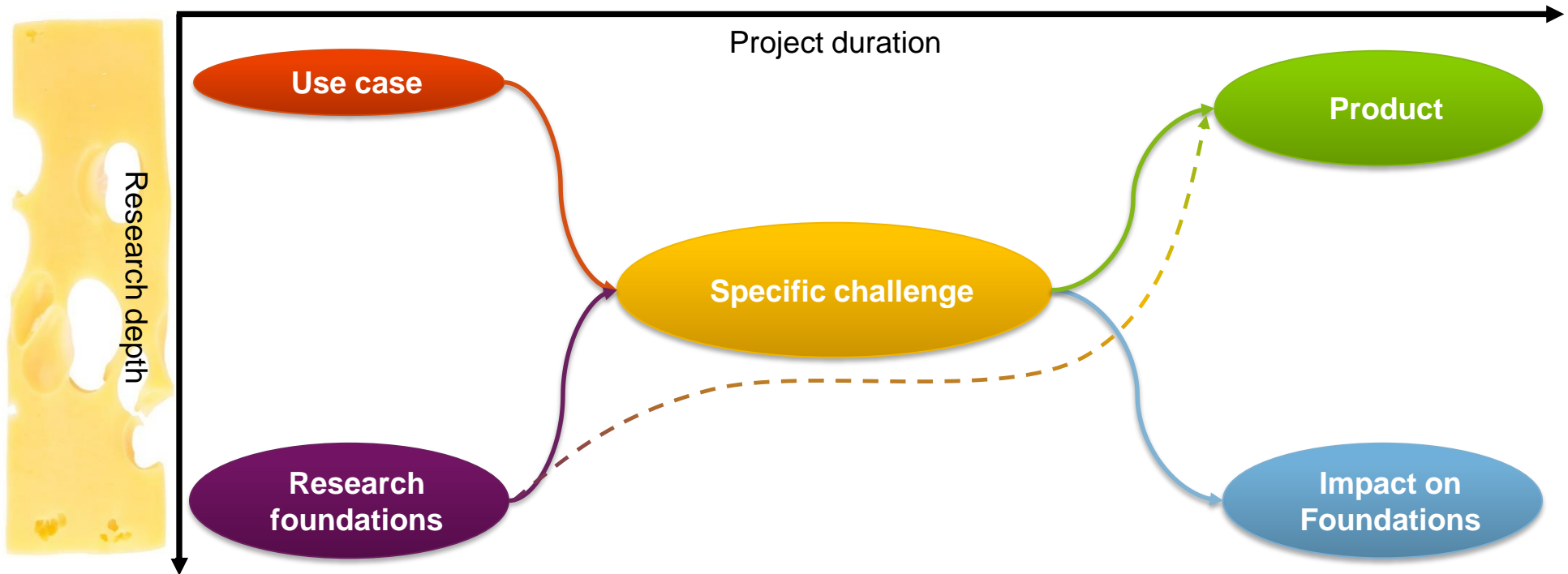
*“The **size of the anomaly** in classifying balloon catheters as good or bad is **quite decisive**. Thus, rescaling the training images is not allowed, and we used a fixed size window around the center of each defect to extract the training images.”*

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

Hypothesis: basic & applied research converge

Speed of “digital” innovation makes complementary skills necessary *at the same time*:

- *Rigor* to come up with completely new methodical approaches
- *Creativity* to solve completely new scenario, thereby “filling wholes”



Conclusions

- **Latest research** is **applied** and deployed in «normal» businesses (non-AI, SME)
- It does not need big-, but some **data (effort usually underestimated)**
- DL/RL **training** for new use cases **can be tricky** (→ needs thorough experimentation)
- The **simultaneity** of research **types A^{pp}lied and B^{asic}** speaks out loud for **collaboration**



On me:

- Prof. AI/ML, head ZHAW Datalab, board SGAICO & Data+Service
- thilo.stadelmann@zhaw.ch
- +41 58 934 72 08
- <https://stdm.github.io/>

On the topics:

- AI: <https://sgaico.swissinformatics.org/>
- Data+Service Alliance: www.data-service-alliance.ch
- Collaboration: datalab@zhaw.ch

→ Happy to answer questions & requests.





APPENDIX

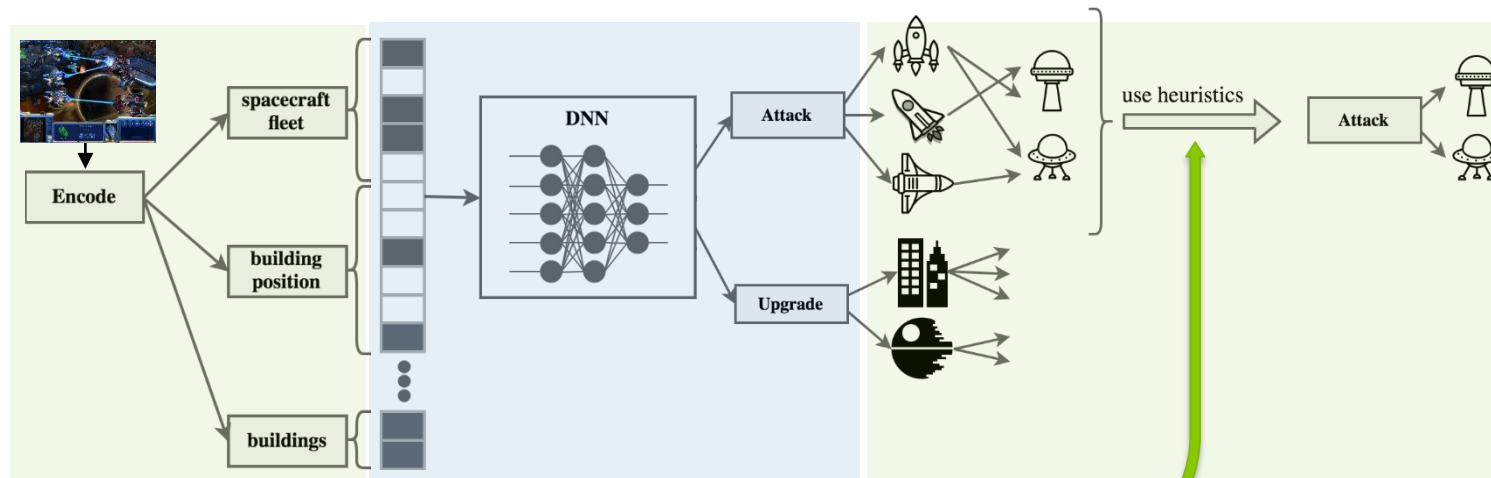
5. Game playing (work in progress)



(symbolic figure)



5. Game playing – challenges & solutions (work in progress)



Reinforcement learning: deep Q network

Large discrete action space → use heuristic

- makes exploration difficult
- elongates training time

Delayed and sparse reward → do reward shaping

- sequence of actions crucial to get a reward



Distance encoding → use reference points

Transfer Learning → difficult: more complex environment needs other action sequence

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

6. Automated machine learning (work in progress)

The project

- Target: in-house solution of industrial partner to improve turnover in standard analytics projects
- Challenge: optimize hyperparameters smarter than with well initialized random perturbations
- Idea: use reinforcement learning to meta-learn from past analytics projects

Initial experiments

Dataset	Task	Metric	Auto-Sklearn		TPOT		DSM	
			Validation	Test	Validation	Test	Validation	Test
Cadata	Regression	Coefficient Of Determination	0.7913	0.7801	0.8245	0.8017	0.7078	0.7119
Christine	Binary Classification	Balanced Accuracy Score	0.7380	0.7405	0.7435	0.7454	0.7362	0.7146
Digits	Multiclass Classification	Balanced Accuracy Score	0.9560	0.9556	0.9500	0.9458	0.8900	0.8751
Fabert	Multiclass Classification	Accuracy Score	0.7245	0.7193	0.7172	0.7006	0.7112	0.6942
Helena	Multiclass Classification	Balanced Accuracy Score	0.3404	0.3434	0.2654	0.2667	0.2085	0.2103
Jasmine	Binary Classification	Balanced Accuracy Score	0.7987	0.8348	0.8188	0.8281	0.8020	0.8371
Madeline	Binary Classification	Balanced Accuracy Score	0.8917	0.8769	0.8885	0.8620	0.7707	0.7686
Philippine	Binary Classification	Balanced Accuracy Score	0.7787	0.7486	0.7839	0.7646	0.7581	0.7406
Sylvine	Binary Classification	Balanced Accuracy Score	0.9414	0.9454	0.9512	0.9493	0.9414	0.9233
Volkert	Multiclass Classification	Accuracy Score	0.7174	0.7101	0.6429	0.6327	0.5220	0.5153
Average Performance			0.7678	0.7654	0.7586	0.7497	0.7048	0.6991

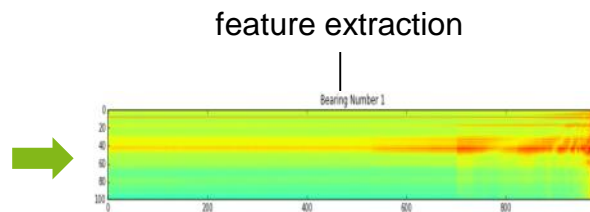
Stadelmann, Amirian, Arabaci, Arnold, Duivesteyn, Elezi, Geiger, Lörwald, Meier, Rombach & Tuggener (2018). «Deep Learning in the Wild». ANNPR'2018.

7. Condition monitoring

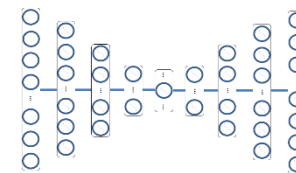
Maintaining machines on predicted failure only

We 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.

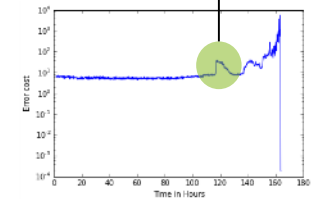
vibration sensors



e.g., RNN autoencoder









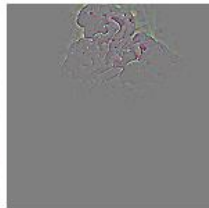
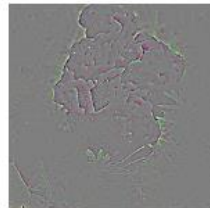
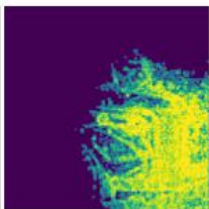
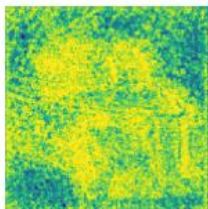
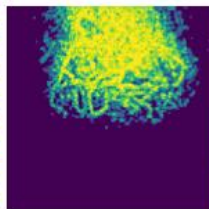
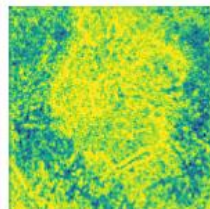
early detection of fault



Stadelmann, Tolkachev, Sick, Stampfli & Dürr (2018). «Beyond ImageNet - Deep Learning in Industrial Practice». In: Braschler et al., «Applied Data Science», Springer.

8. Trace & detect adversarial attacks

...using average local spatial entropy of feature response maps

	Original	Adversarial	Original	Adversarial
Image:				
Feature response:				
Local spatial entropy:				

Amirian, Schwenker & Stadelmann (2018). «Trace and Detect Adversarial Attacks on CNNs using Feature Response Maps». ANNPR'2018.

ML @ Information Engineering Group

Institute of Applied Information Technology, ZHAW School of Engineering

Machine learning-based Pattern Recognition



Robust Deep Learning

Voice Recognition

Document Analysis

Learning to Learn & Control

