Zürcher Hochschule
für Angewandte Wissenschaften

Deep Learning für Dokumentenanalyse

Kolloquium der ETH Bibliothek, online, 25. März 2021

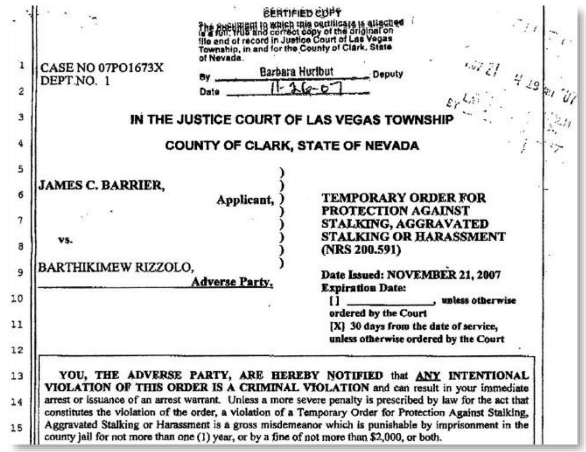
Thilo Stadelmann

zhaw.ch/datalab

Zürcher Fachhochschule

Document analysis?


Das Steirerblatt
Amerika beantragt UN-Kontrolle der Donau
Vor neuer Vorsprache bei Stalin
Grauer Zucker wird wieder schwarz


CERTIFIED COPY
The document to which this certificate is attached is a full, true and correct copy of the original on file and of record in Justice Court of Las Vegas Township, in and for the County of Clark, State of Nevada.
By Barbara Hurlbut, Deputy
Date 11-26-07

CASE NO 07PO1673X
DEPT NO. 1

IN THE JUSTICE COURT OF LAS VEGAS TOWNSHIP
COUNTY OF CLARK, STATE OF NEVADA

JAMES C. BARRIER, Applicant,
vs.
BARTHIKIMEW RIZZOLO, Adverse Party.

TEMPORARY ORDER FOR PROTECTION AGAINST STALKING, AGGRAVATED STALKING OR HARASSMENT (NRS 200.591)

Date Issued: NOVEMBER 21, 2007
Expiration Date:
[ ] ______, unless otherwise ordered by the Court
[X] 30 days from the date of service, unless otherwise ordered by the Court

YOU, THE ADVERSE PARTY, ARE HEREBY NOTIFIED that ANY INTENTIONAL VIOLATION OF THIS ORDER IS A CRIMINAL VIOLATION and can result in your immediate arrest or issuance of an arrest warrant. Unless a more severe penalty is prescribed by law for the act that constitutes the violation of the order, a violation of a Temporary Order for Protection Against Stalking, Aggravated Stalking or Harassment is a gross misdemeanor which is punishable by imprisonment in the county jail for not more than one (1) year, or by a fine of not more than $2,000, or both.

Documents

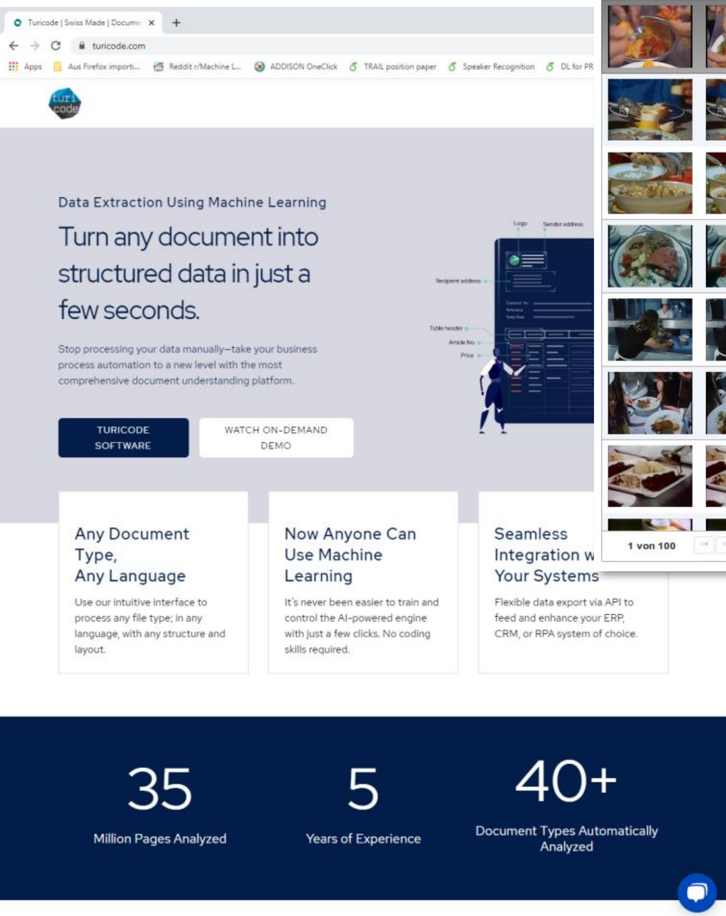
- **Ubiquitous** in human communication and every scenario involving an office
- Somewhat structured for human expert; **unstructured** w.r.t machines
- **Great use case for** various **AI** techniques, including computer vision

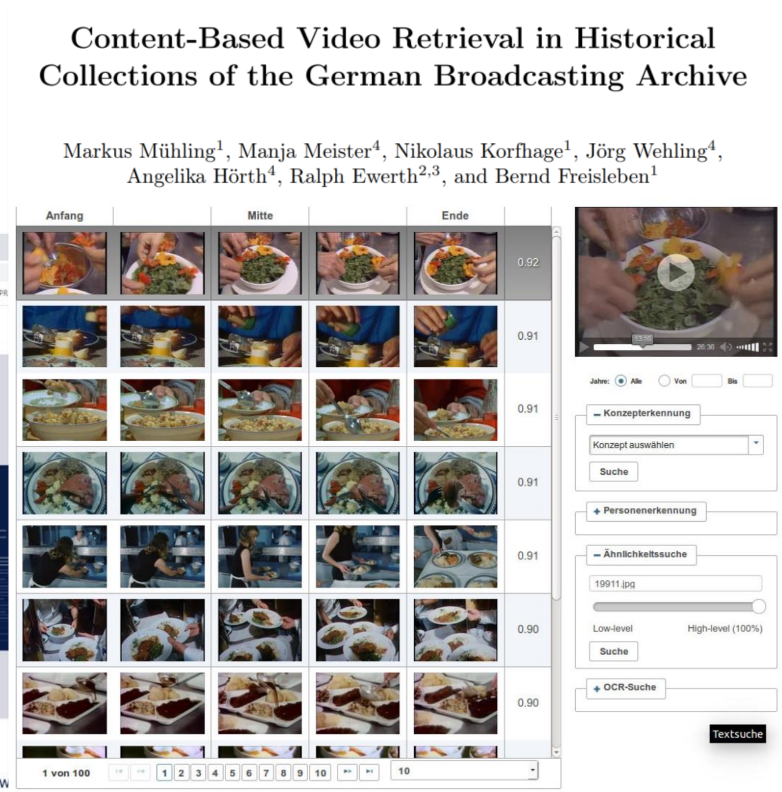
Own scientific community

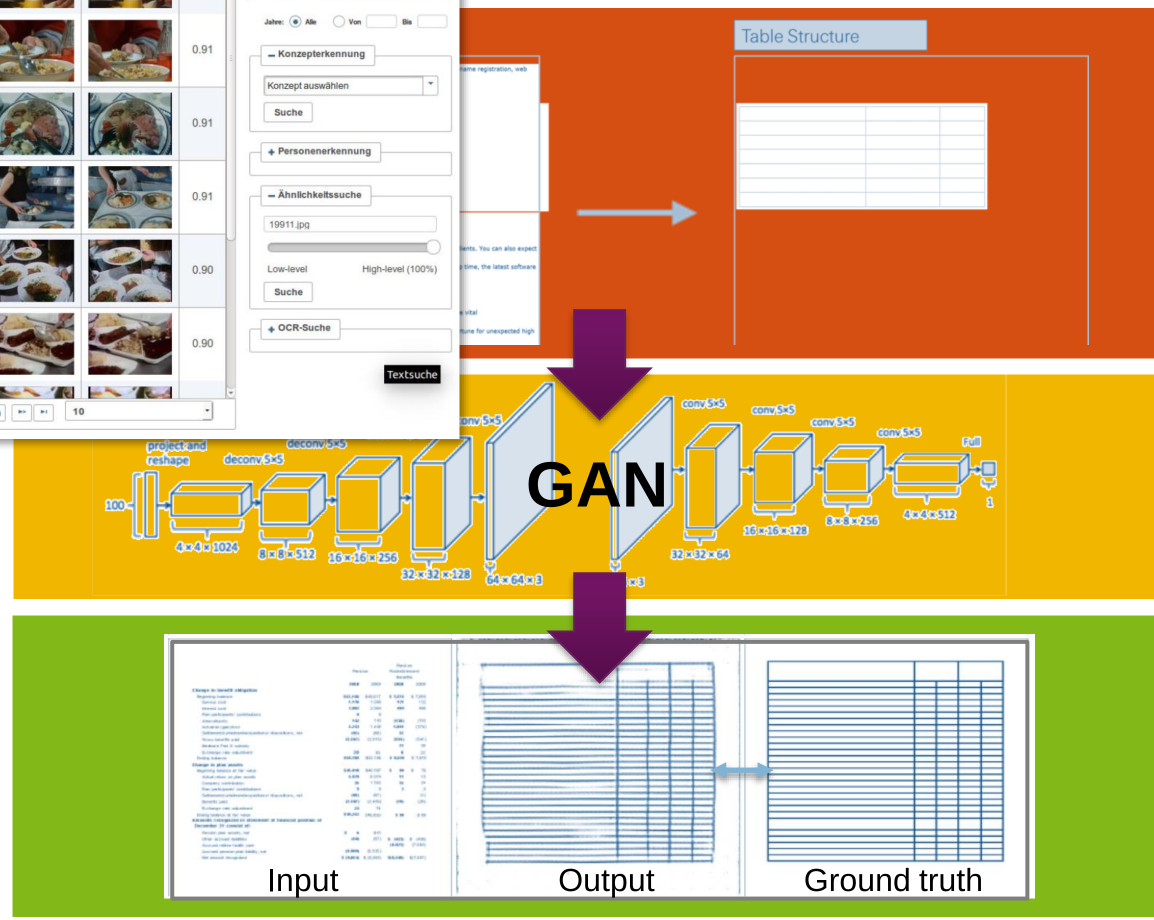
- IAPR's biannual Intl. Conference on Document Analysis & Recognition (ICDAR):
  character & symbol recognition, printed/handwritten text recognition, graphics analysis & recognition, document analysis & understanding, historical documents & digital libraries, document-based forensics, camera & video-based scene text analysis

Examples

Zürcher Hochschule für Angewandte Wissenschaften



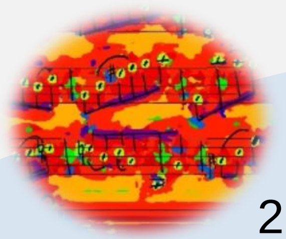




Roadmap

0. Prologue on document analysis

1. Print media monitoring



2. Music scanning

3. Outlook: accessible documents for all

Zürcher Hochschule
für Angewandte Wissenschaften

1. Print media monitoring

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Task


International.

Spionage für den Erzfeind Iran

Israelischer Ex-Minister arbeitete als Agent für die Mullahs. Jetzt droht ihm lebenslänglich

Asylbewerber können bleiben

Nordkoreanischer Diktator zu Besuch in Peking

Challenge


Sein Juniorentrainer Mano Pavesi über unseren WM-Helden Steven Zuber

«Steven hat sich alles selber beigebracht»

Unsere Nati ist gegen Serbien unter Druck

Verlieren verboten!

Transfer Ticker

Liverpool will Yann Sommer!

Nuisance


Das Tages-Horoskop

Liebling der Sterne

15,1 Millionen

Sind Sie der nächste Lotto-König?

Jungfrau Waage Skorpion Schütze Steinbock Wassermann Fische Widder Stier Zwillinge Krebs

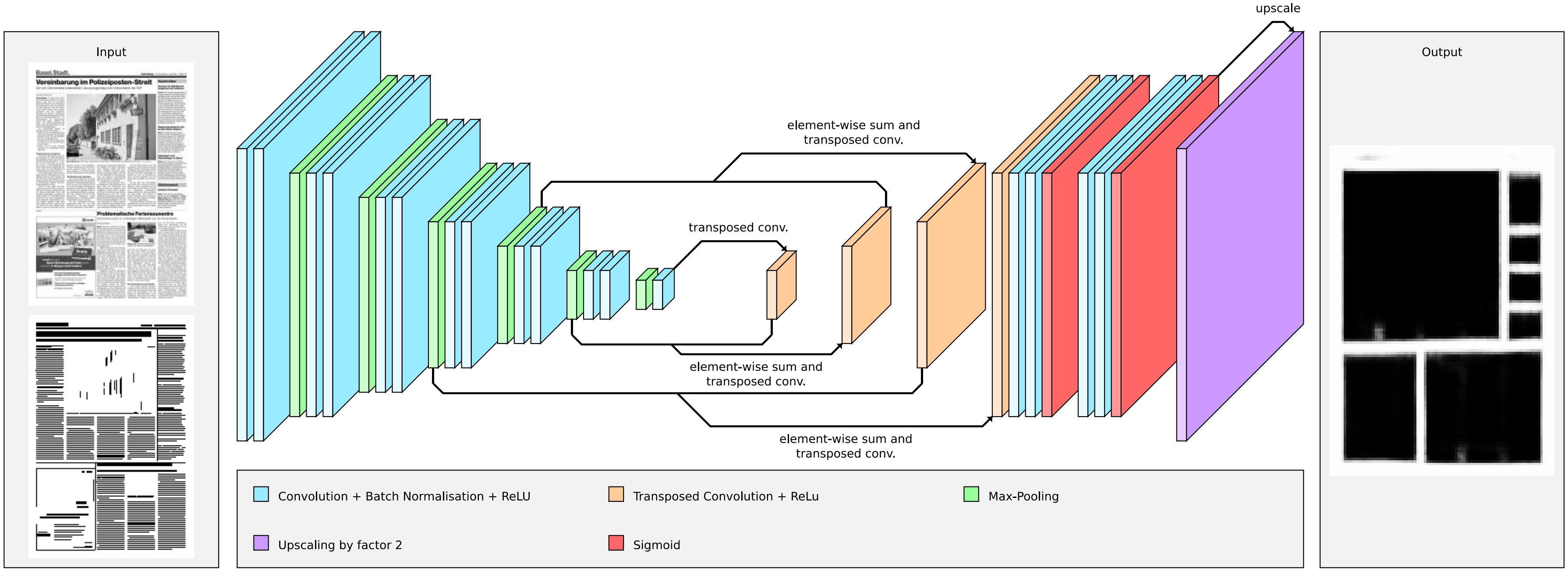
SCHWEDEN-RÄTSEL

SUDOKU

Wochenpreis: 1 x sieben Nächte für 2 Personen, inkl. HP, im ****Seehotel Pilatus Hergiswil im Wert von 3000 Franken!

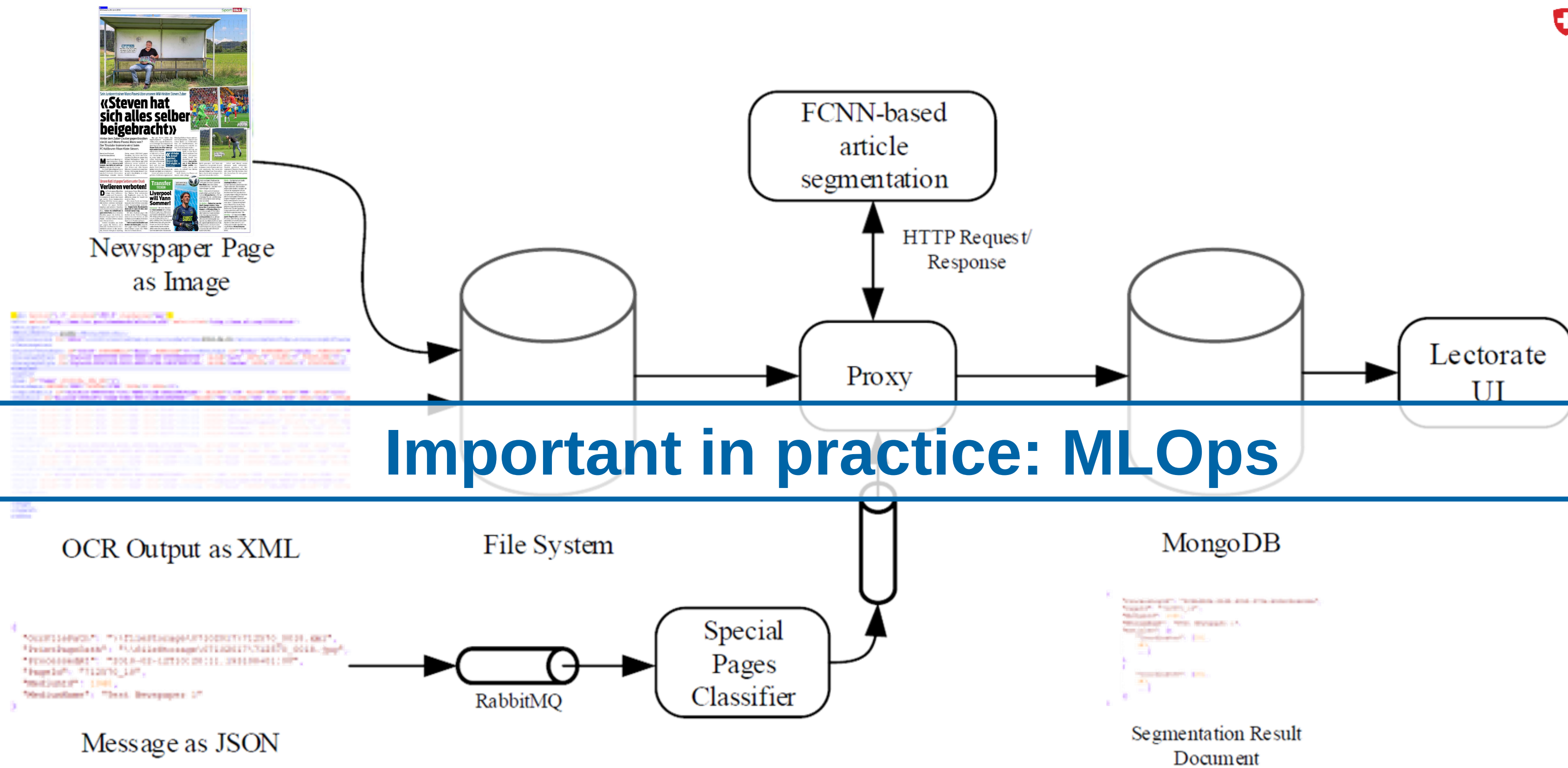
FLOTTENMANÖVER

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

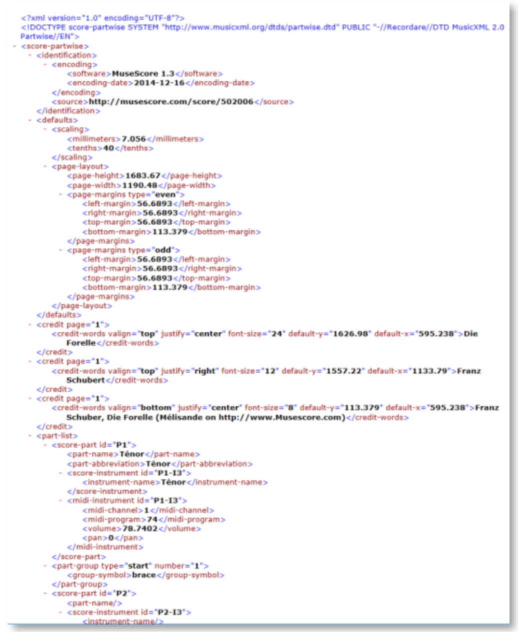
1. Print media monitoring – deployment



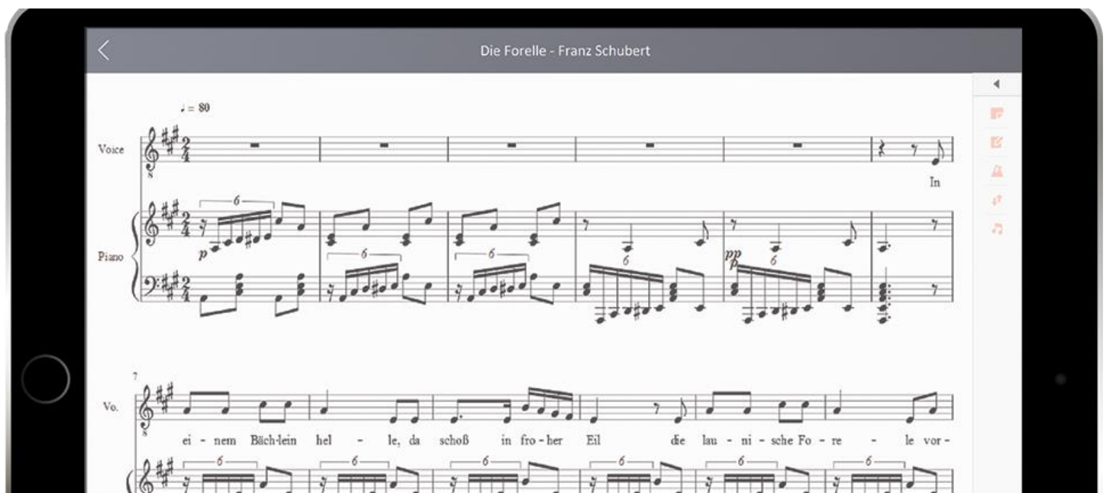
Important in practice: MLOps

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

2. 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/>
```

Zürcher Hochschule
für Angewandte Wissenschaften

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

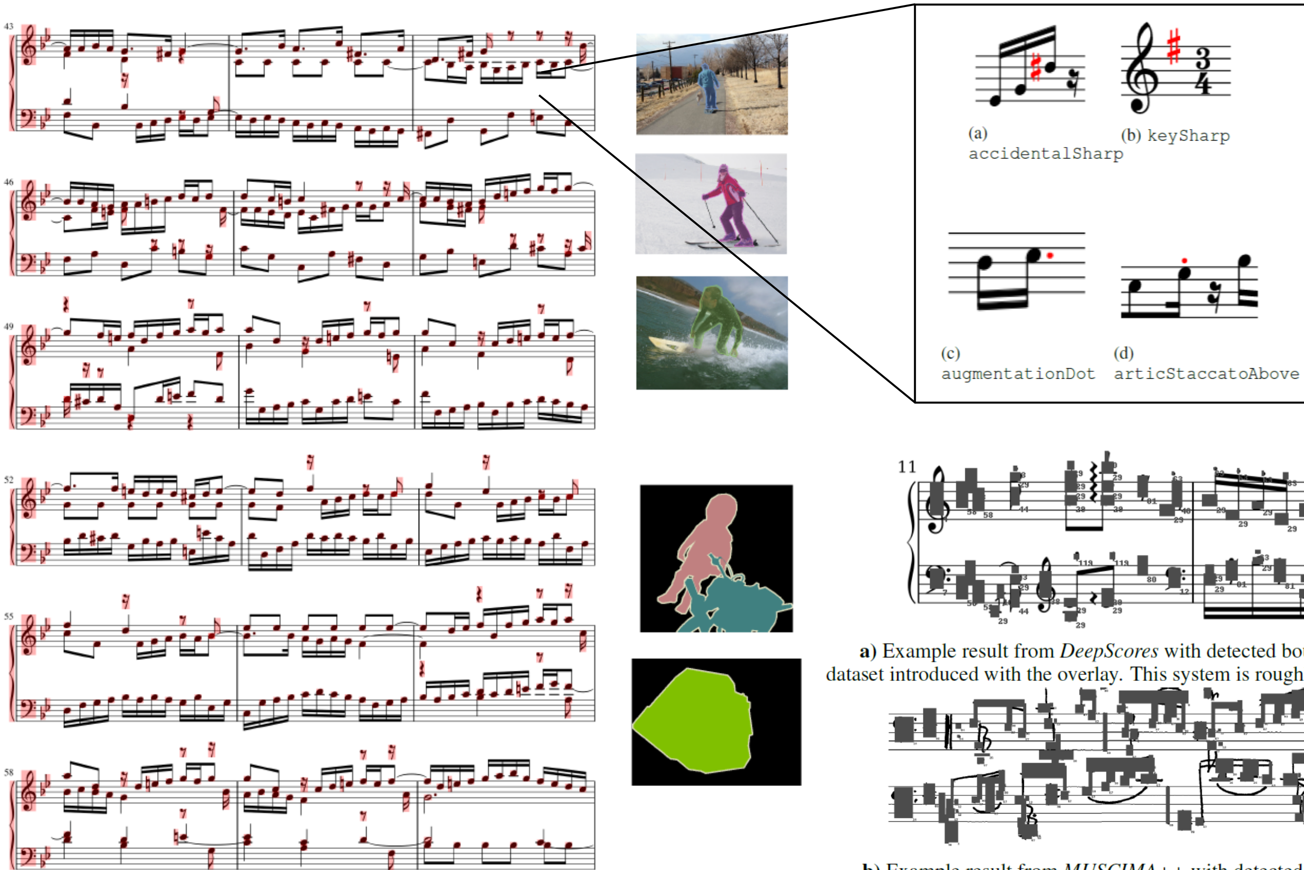
2. Music scanning – challenges & results

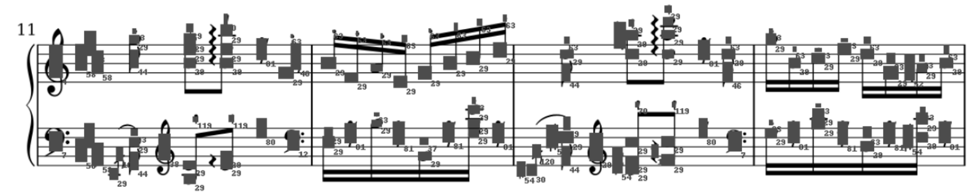
Score Pad

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

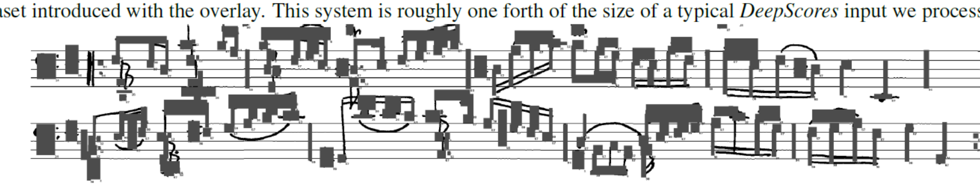
Swiss Confederation

Innosuisse – Swiss Innovation Agency





a) Example result from *DeepScores* with detected bounding boxes as overlays. The tiny numbers are class labels from the dataset introduced with the overlay. This system is roughly one forth of the size of a typical *DeepScores* input we process at once.

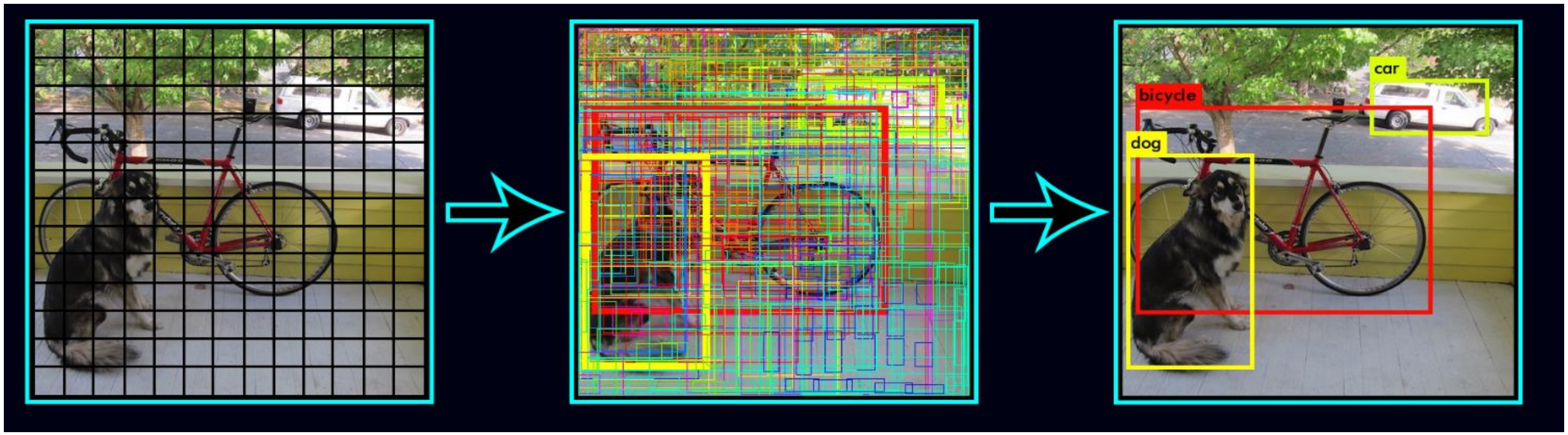


b) Example result from *MUSCIMA++* with detected bounding boxes and class labels as overlays. This system is roughly one half of the size of a typical processed *MUSCIMA++* input. The images are random picks amongst inputs with many symbols.

Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR'2018.
Tuggener, Satyawan, Pacha, Schmidhuber & Stadelmann (2020). *«The DeepScoresV2 Dataset and Benchmark for Music Object Detection»*. ICPR'2020.
Tuggener, Elezi, Schmidhuber & Stadelmann (2018). *«Deep Watershed Detector for Music Object Recognition»*. ISMIR'2018.

Music scanning – methodology (differentiation)
OMR vs state of the art object detectors

YOLO/SSD-type detectors



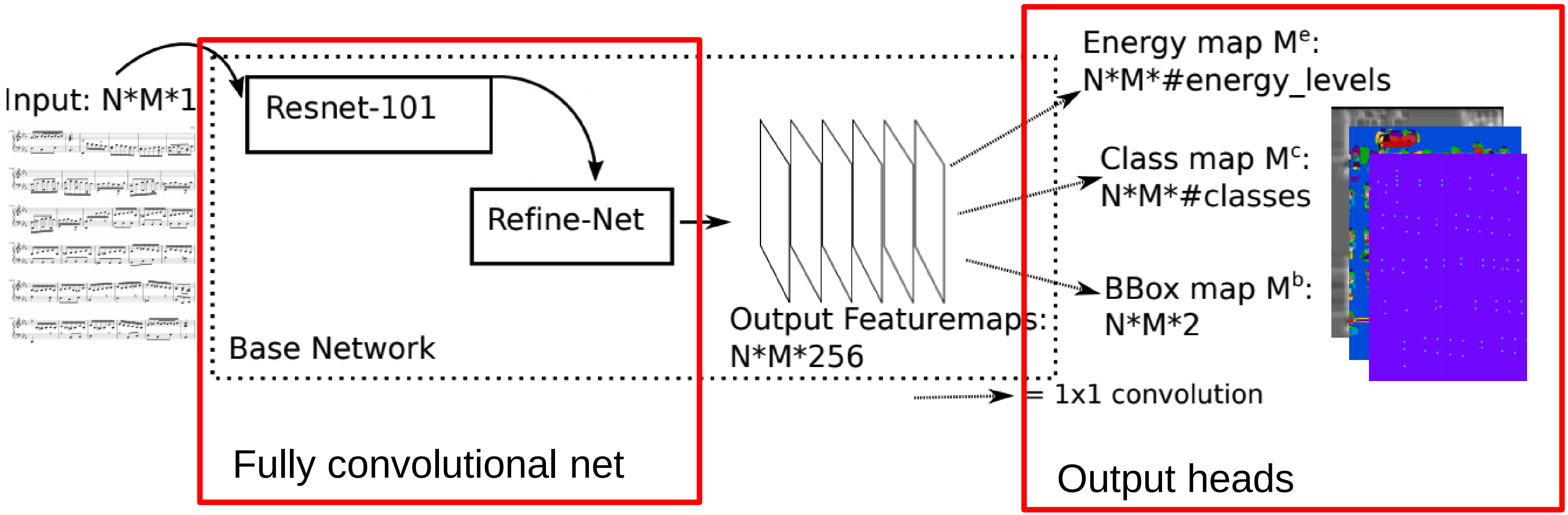
Source: https://pjreddie.com/darknet/yolov2/ (11.09.2018)

R-CNN

- Two-step proposal and refinement scheme
- Very large number of proposals at high resolution needed

Zürcher Hochschule
für Angewandte Wissenschaften

2. Music scanning – methodology (ours)
The deep watershed detector

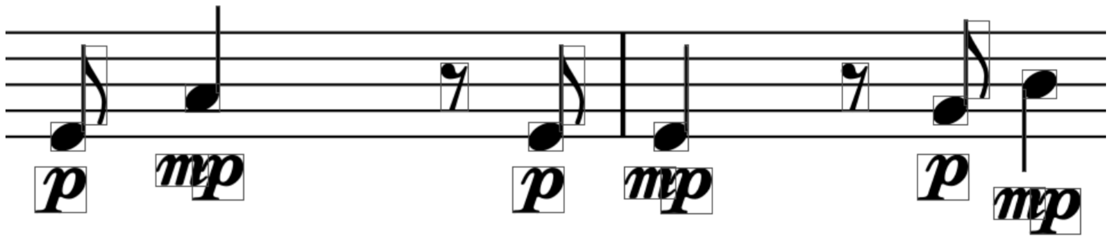


Zürcher Hochschule
für Angewandte Wissenschaften

2. Music scanning – methodology (details)

Output heads of the deep watershed detector

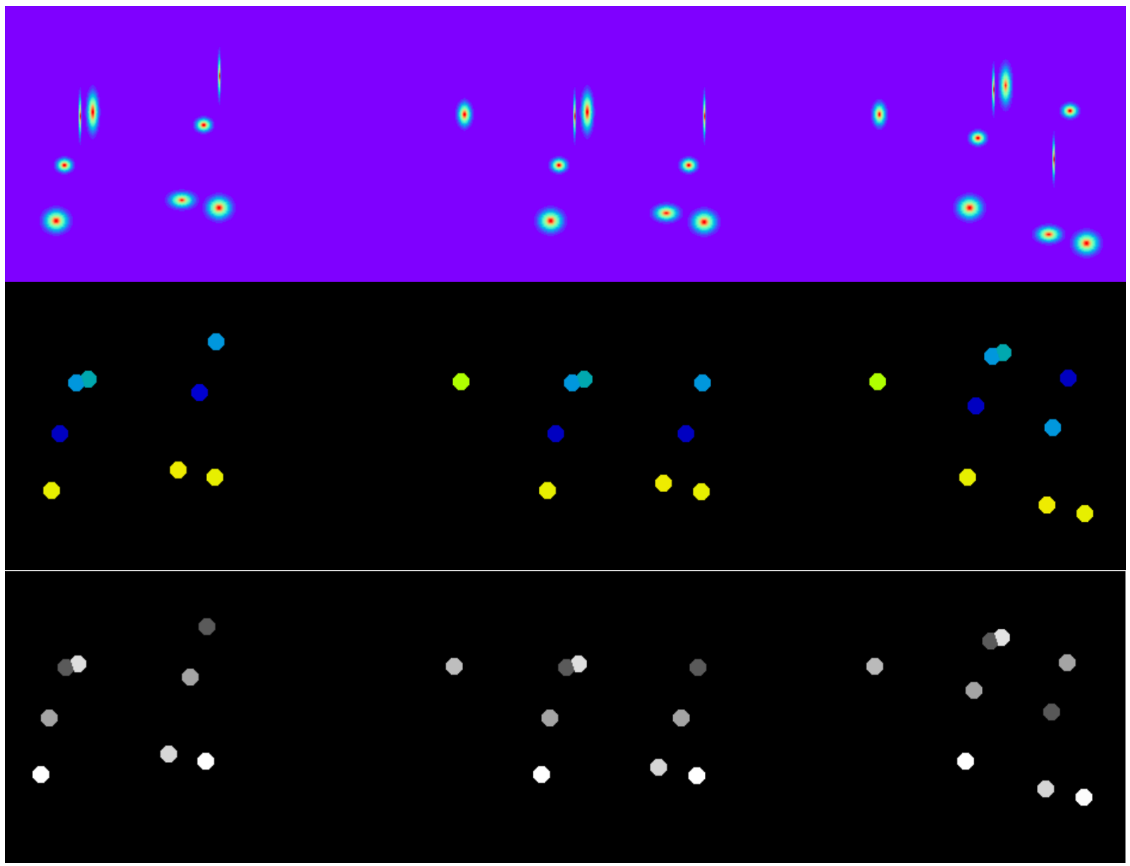
Input

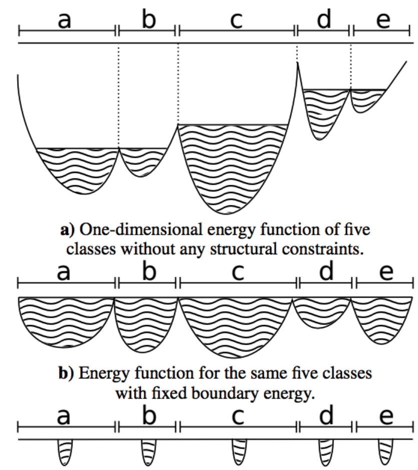


Energy

Class

Size
2D projection





a) One-dimensional energy function of five classes without any structural constraints.

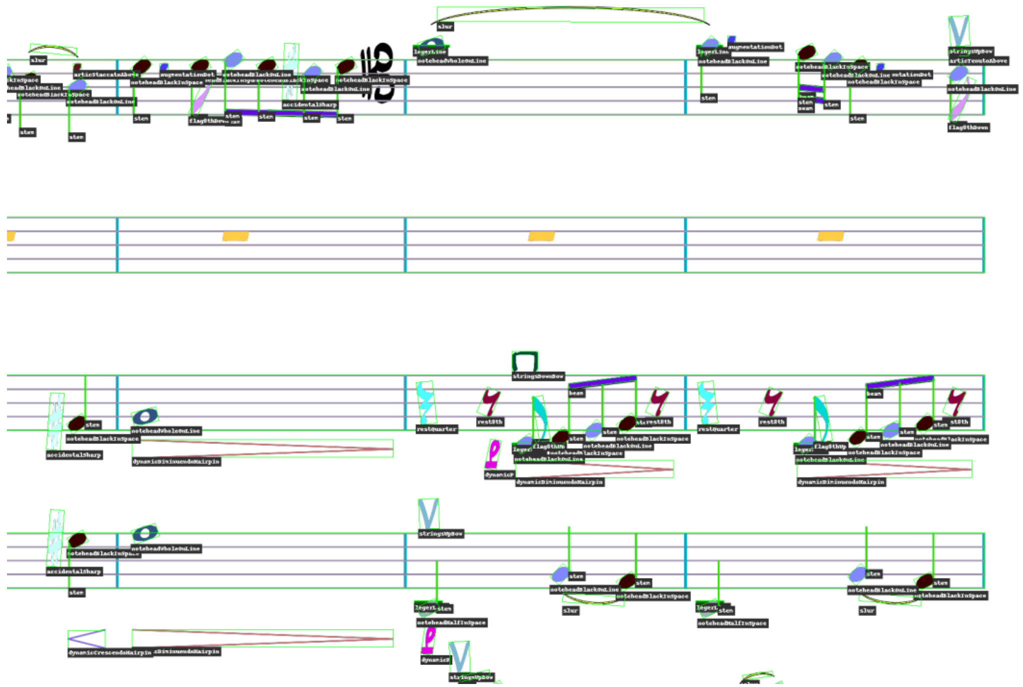
b) Energy function for the same five classes with fixed boundary energy.

c) Energy function for the same five classes this time with small energy markers at the class centers.

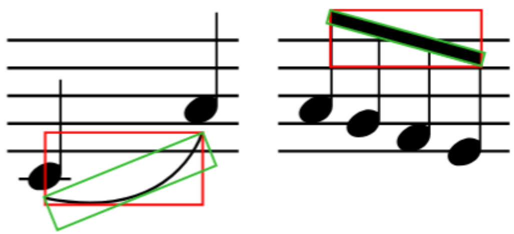
Score Pad
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Innosuisse – Swiss Innovation Agency

2. Music scanning – recent work

1. Extension of the detection alphabet to 135 classes (updated dataset released)



2. Natural incorporation of rotation



2. Music scanning – future work
Dealing with real world noise

Synthetic quality + labels:
perfect quality

Real world quality: print/scan artifacts, wrinkles, dirt, …

Model training

Data distribution shift

Model deployment

Remedy:
Use semi-supervised learning to model distribution change and/or disentangle latent signals

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

3. Outlook – 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)
- **Unsupervised** methods need to be used creatively
- Users & label providers need to be **trained**

Prerequisite: stable data & label acquisition pipeline
Learning from (raw) data is powerful, yet one is fully dependent on what is in that data

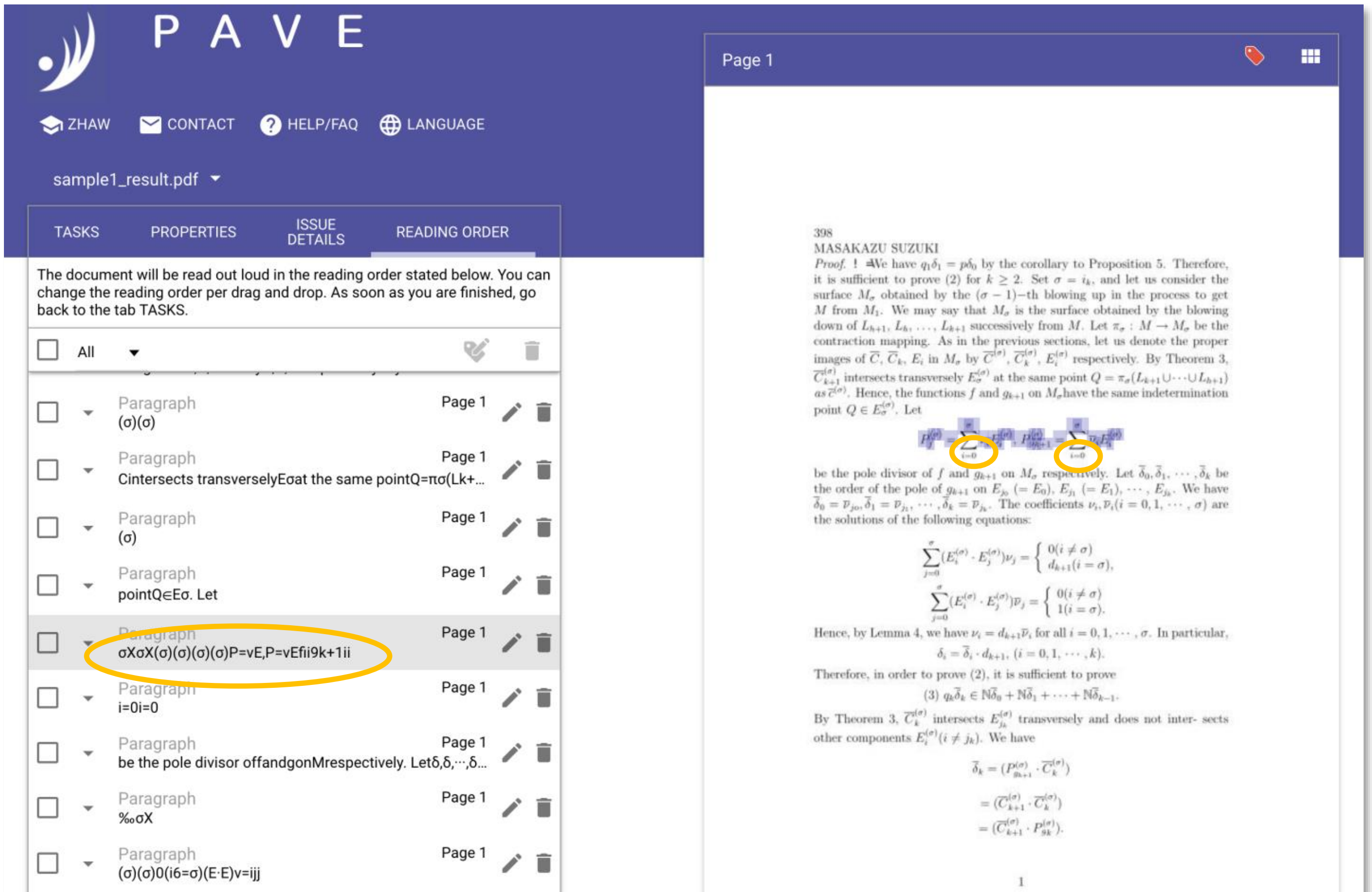
Robustness is important

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

Sufficient condition: lots of tuning

Zürcher Hochschule
für Angewandte Wissenschaften

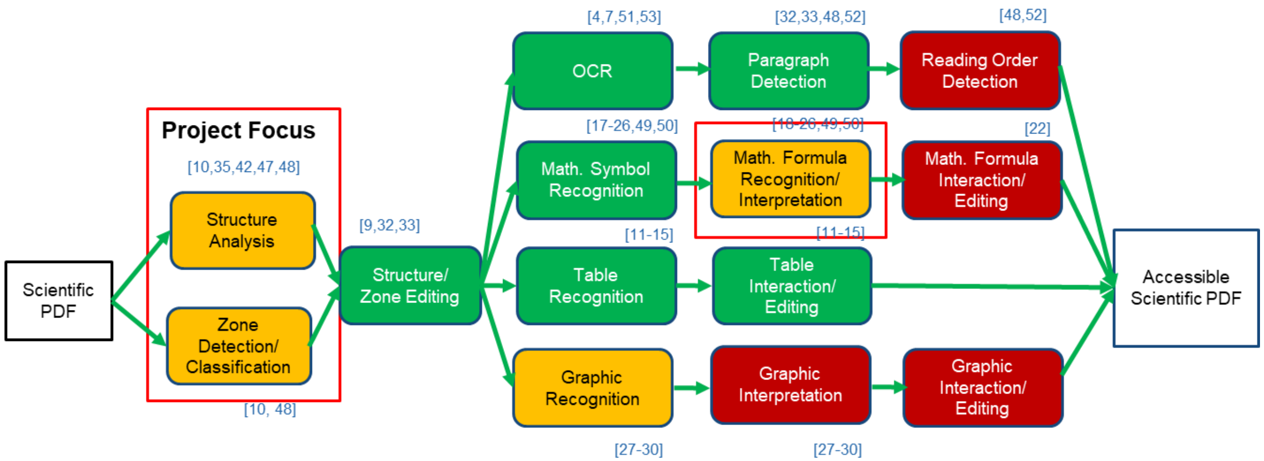
3. Outlook – accessible documents for all



BRIDGE

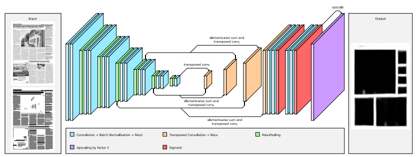
3. Outlook – research approach

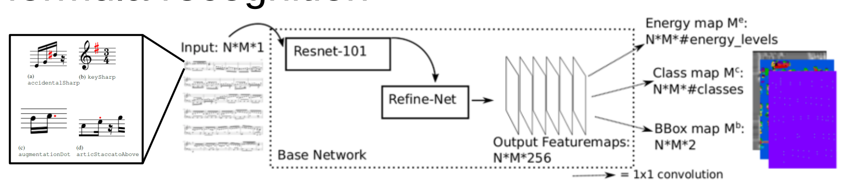
- Goal is **a semi-automatic process** to make scientific documents accessible
- Focus on **structure analysis** & **mathematical formula recognition**



- **Build on previous experience & methods** for newspapers and OMR

structure recognition — formula recognition

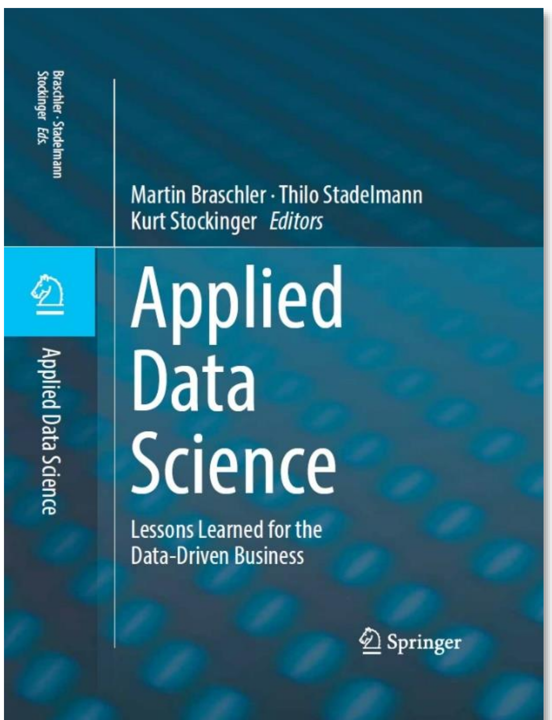




Zürcher Hochschule
für Angewandte Wissenschaften

Contact

- Document analysis is a **very fruitful use case** for Deep Learning (for archives + R&D)
- **Latest research** is **applied** and deployed in «normal» organizations (e.g., libraries)
- It does not need big-, but some **data** (**effort usually underestimated**)
- DL/RL **training** for new use cases **can be tricky** (→ needs experience)



About me:

- Prof. AI/ML, scientific director ZHAW digital
- Email: stdm@zhaw.ch
- Phone: +41 58 934 72 08
- Web: https://stdm.github.io/
- Twitter: @thilo_on_data
- LinkedIn: thilo-stadelmann

Further contacts:

- datalab@zhaw.ch, info.office@data-innovation.org, office-switzerland@claire-ai.org

zh aw Datalab
zhaw.ch/datalab

d•i
data innovation alliance

CLAIRE

APPENDIX

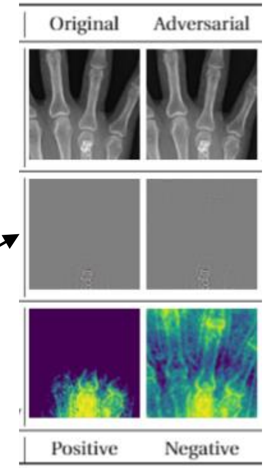
Zürcher Hochschule
für Angewandte Wissenschaften

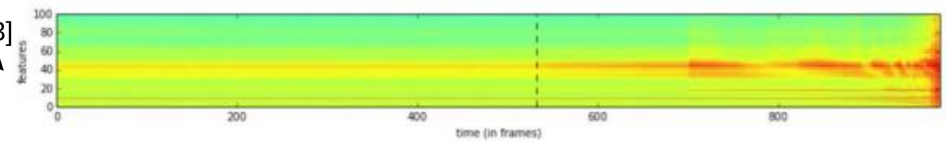
Team AI/ML: Overview (cp. https://stdm.github.io/research/)

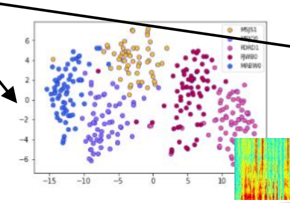
ZHAW School of Engineering, Winterthur, Switzerland

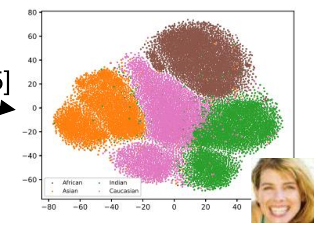
Machine learning-based Pattern Recognition

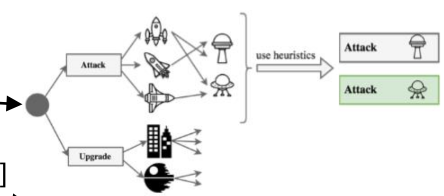
- Robust applications [1] [2] [3]
- Biometrics [4] [5]
- Document Analysis [7] [6]
- Learning to act [1] [10] [8] [9]

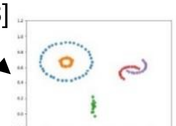


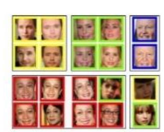


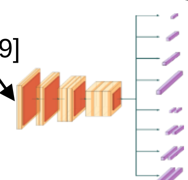


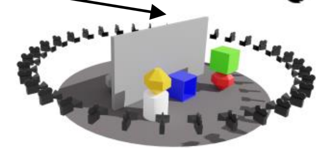










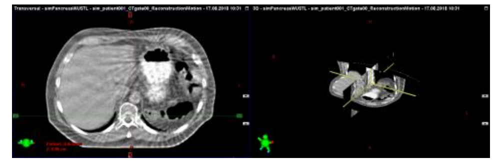


References for overview

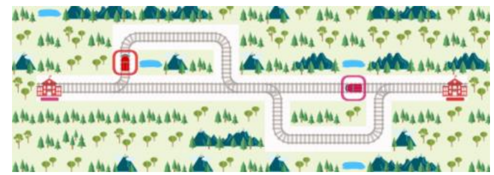
1. Thilo Stadelmann, Mohammadreza Amirian, Ismail Arabaci, Marek Arnold, Gilbert François Duivesteijn, Ismail Elezi, Melanie Geiger, Stefan Lörwald, Benjamin Bruno Meier, Katharina Rombach, and Lukas Tuggener. "**Deep Learning in the Wild**". In: Proceedings of the 8th IAPR TC 3 Workshop on Artificial Neural Networks for Pattern Recognition (**ANNPR'18**), Springer, LNAI 11081, pp. 17-38, Siena, Italy, September 19-21, 2018.
2. Mohammadreza Amirian, Friedhelm Schwenker, and Thilo Stadelmann. "**Trace and Detect Adversarial Attacks on CNNs using Feature Response Maps**". In: Proceedings of the 8th IAPR TC 3 Workshop on Artificial Neural Networks for Pattern Recognition (**ANNPR'18**), Springer, LNAI 11081, pp. 346-358, Siena, Italy, September 19-21, 2018.
3. Thilo Stadelmann, Vasily Tolkachev, Beate Sick, Jan Stampfli, and Oliver Dürr. "**Beyond ImageNet - Deep Learning in Industrial Practice**". In: Martin Braschler, Thilo Stadelmann, and Kurt Stockinger (Editors). **"Applied Data Science - Lessons Learned for the Data-Driven Business"**. **Springer**, 2019.
4. Thilo Stadelmann, Sebastian Glinski-Haefeli, Patrick Gerber, and Oliver Dürr. "**Capturing Suprasegmental Features of a Voice with RNNs for Improved Speaker Clustering**". In: Proceedings of the 8th IAPR TC 3 Workshop on Artificial Neural Networks for Pattern Recognition (**ANNPR'18**), Springer, LNAI 11081, pp. 333-345, Siena, Italy, September 19-21, 2018.
5. Stefan Glüge, Mohammadreza Amirian, Dandolo Flumini, and Thilo Stadelmann. "**How (Not) to Measure Bias in Face Recognition Networks**". In: Proceedings of the 9th IAPR TC 3 Workshop on Artificial Neural Networks for Pattern Recognition (**ANNPR'20**), Springer, LNAI, Winterthur, Switzerland, September 02-04, 2020.
6. Lukas Tuggener, Yvan Putra Satyawan, Alexander Pacha, Jürgen Schmidhuber, and Thilo Stadelmann. "**The DeepScoresV2 Dataset and Benchmark for Music Object Detection**". In: Proceedings of the 25th International Conference on Pattern Recognition (**ICPR'20**), IAPR, Milan, Italy, January 10-15 (online), 2021.
7. Benjamin Meier, Thilo Stadelmann, Jan Stampfli, Marek Arnold, and Mark Cieliebak. "**Fully convolutional neural networks for newspaper article segmentation**". In: Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (**ICDAR'17**). 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto Japan, November 13-15, 2017. Kyoto, Japan: CPS.
8. Benjamin Bruno Meier, Ismail Elezi, Mohammadreza Amirian, Oliver Dürr, and Thilo Stadelmann. "**Learning Neural Models for End-to-End Clustering**". In: Proceedings of the 8th IAPR TC 3 Workshop on Artificial Neural Networks for Pattern Recognition (**ANNPR'18**), Springer, LNAI 11081, pp. 126-138, Siena, Italy, September 19-21, 2018.
9. Lukas Tuggener, Mohammadreza Amirian, Fernando Benites, Pius von Däniken, Prakhar Gupta, Frank-Peter Schilling, and Thilo Stadelmann. "**Design Patterns for Resource-Constrained Automated Deep-Learning Methods**". **AI** section "Intelligent Systems: Theory and Applications" 1(4):510-538, MDPI, Basel, Switzerland, Novemer 06, 2020.
10. Dano Roost, Ralph Meier, Giovanni Toffetti Carughi, and Thilo Stadelmann. "**Combining Reinforcement Learning with Supervised Deep Learning for Neural Active Scene Understanding**". In: Proceedings of the Active Vision and Perception in Human(-Robot) Collaboration Workshop at IEEE RO-MAN 2020 (**AVHRC'20**), online, August 31, 2020.

Outlook: Late-breaking results

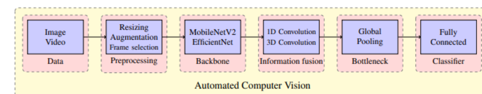
- Medical image analysis: learning to reduce motion artifacts in 3D CT scans



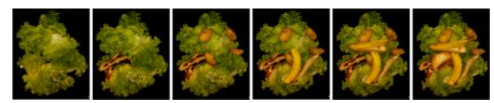
- Learning an artificial communication language for multi-agent reinforcement learning in logistics
  (notable rank in Flatland 2019 competition, best poster award [1])



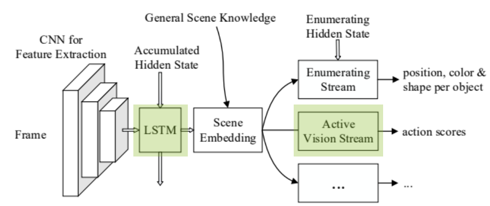
- Automated deep learning
  (top rank in AutoDL 2020 challenge [2])



- Learning to segment and classify food waste in professional kitchens under adversarial conditions [4]



- Improving robotic vision through active vision and combined supervised and reinforcement learning
  (Dr. Waldemar Jucker Award 2020 [3])



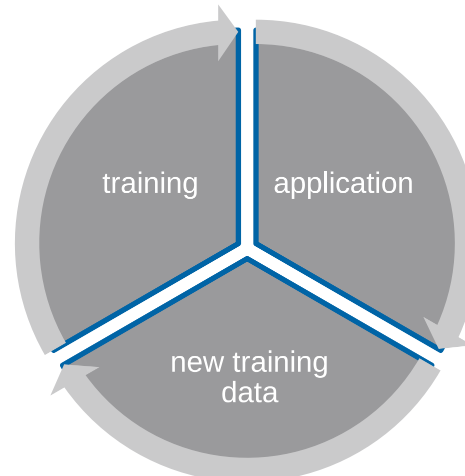
[1] Roost, Meier, Huschauer, Nygren, Egli, Weiler & Stadelmann (2020). *«Improving Sample Efficiency and Multi-Agent Communication in RL-based Train Rescheduling»*. SDS'2020.
[2] Tuggener, Amirian, Benites, von Däniken, Gupta, Schilling & Stadelmann (2020). *«Design Patterns for Resource Constrained Automated Deep Learning Methods».* AI 1(4) 510-538.
[3] Roost, Meier, Toffetti Carughi & Stadelmann (2020). *«Combining Reinforcement Learning with Supervised Deep Learning for Neural Active Scene Understanding».* AVHRC 2020.
[4] Simmler, Sager, Andermatrt, Chavarriaga ,Schilling, Rosenthal & Stadelmann (2021). *«Noisy labels, missing labels: A survey of un-, weakly-, and semi-supervised learning methods for industrial vision applications». SDS* 2021.

1. Lessons learned 1/2

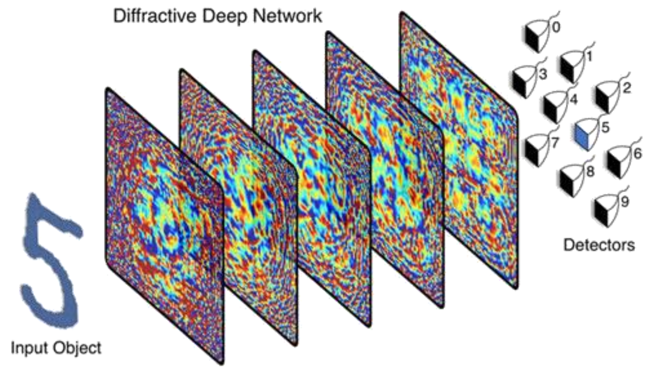
 , …

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.

2. Music scanning – methodology (further details)

Reconstructing labeled bounding boxes

Finding object instances:

1. Set all values of the energy prediction below a certain threshold to zero (watershed step).
2. Perform a connected component analysis on this output --> the center of mass for every connected component is defined to be the position of one detected object.

Predicting the class:

3. For each connected component look up the corresponding class predictions for of all its pixels and use the majority vote.

Predicting object size:

4. For each connected component look up the corresponding size predictions for of all its pixels and compute the mean.

Lessons learned 2/2

, …

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