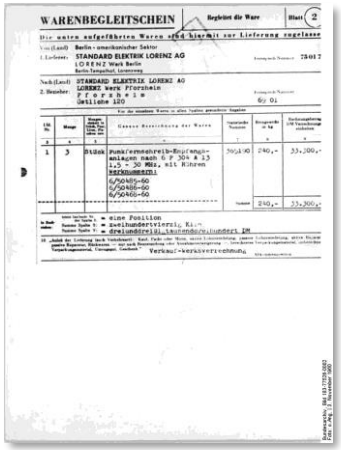
Zürcher Hochschule
für Angewandte Wissenschaften

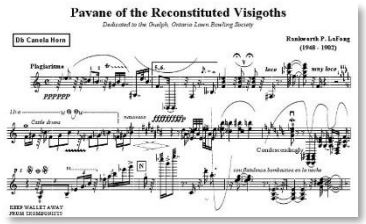
Document Analysis with Deep Learning

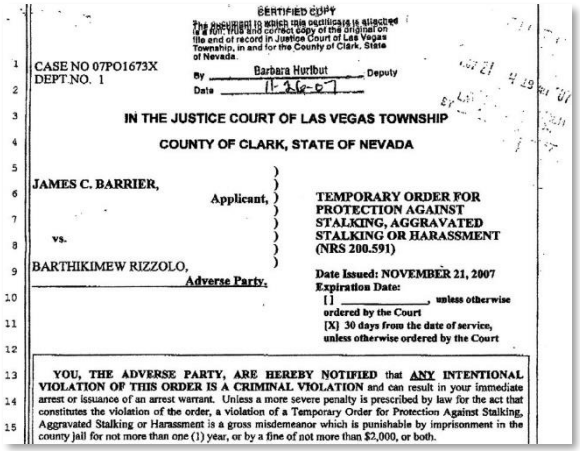
Use Case Talk Series, Zurich, Oct 31, 2018

Thilo Stadelmann

Document recognition?







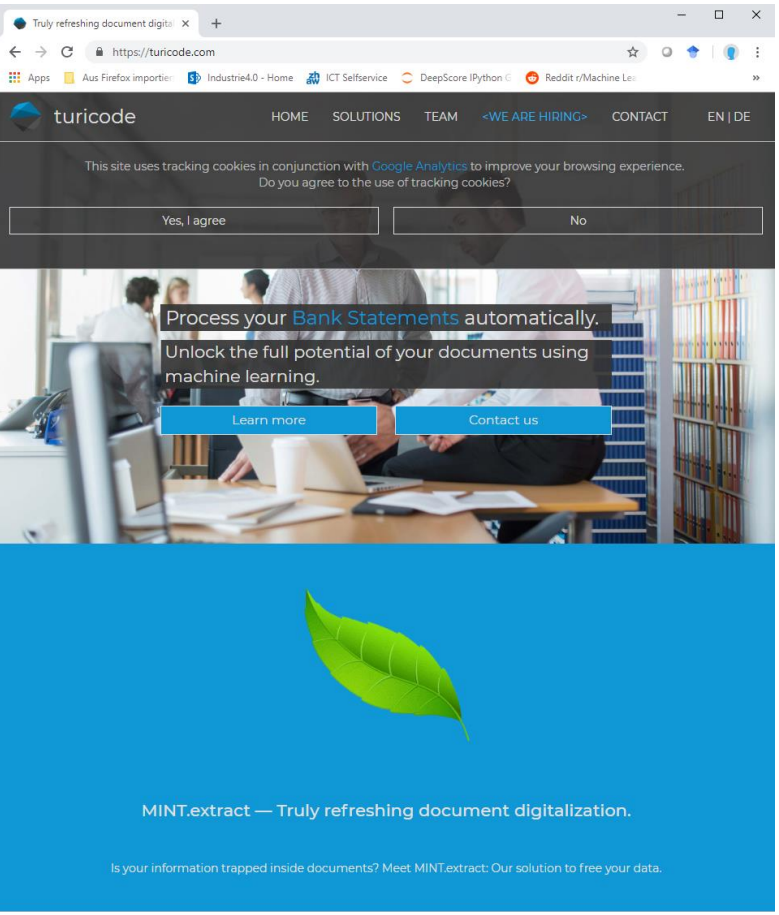
Documents

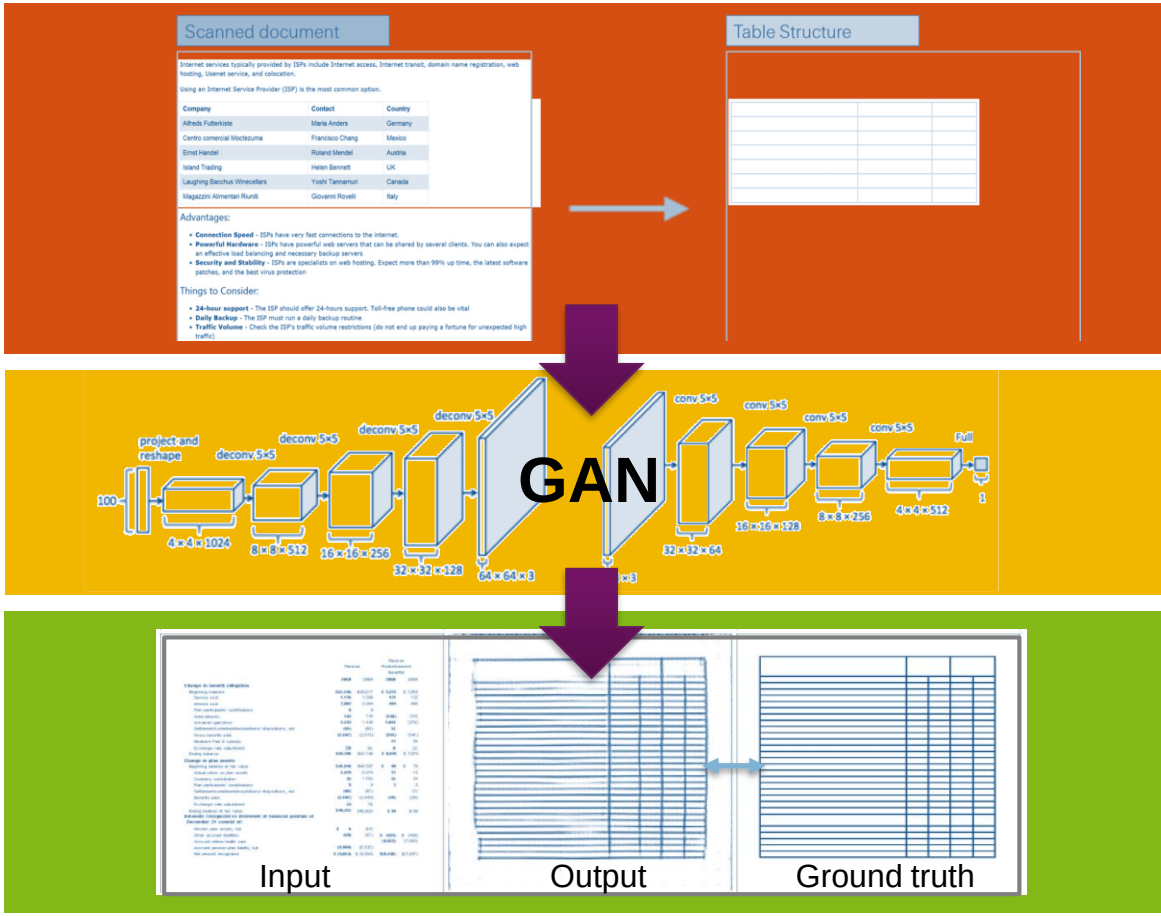
- **Ubiquitous** in human communication and every scenario involving an office
- Somewhat structured for human expert; **unstructured** w.r.t machines
- **Great use case for** various **AI** techniques, including computer vision techniques

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recognition, printed/handwritten text recognition, graphics analysis & recognition, document analysis & understanding, historical documents & digital libraries, document based forensics, camera & video based scene text analysis

Industrial examples

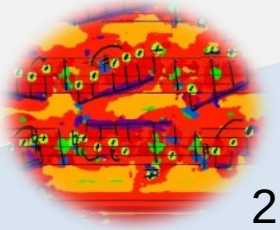




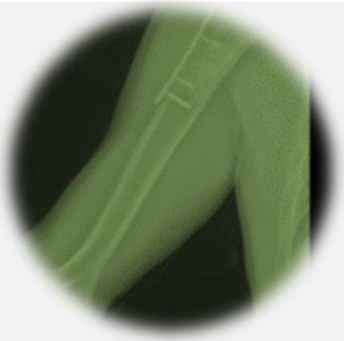
Roadmap

0. Prologue on document analysis

1. Print media monitoring



2. Music scanning



Conclusions

Zürcher Hochschule
für Angewandte Wissenschaften

1. Print media monitoring

Task


International.
Nachrichten
Spionage für den Erzfeind Iran
Israelischer Ex-Minister arbeitete als Agent für die Mullahs. Jetzt droht ihm lebenslänglich
Vermögen beschlagnahmt
Asylbewerber können bleiben
Europäischer Gerichtshof relativiert Abschiebepraxis
Nordkoreanischer Diktator zu Besuch in Peking
Klage von Le Pen abgelehnt

Challenge

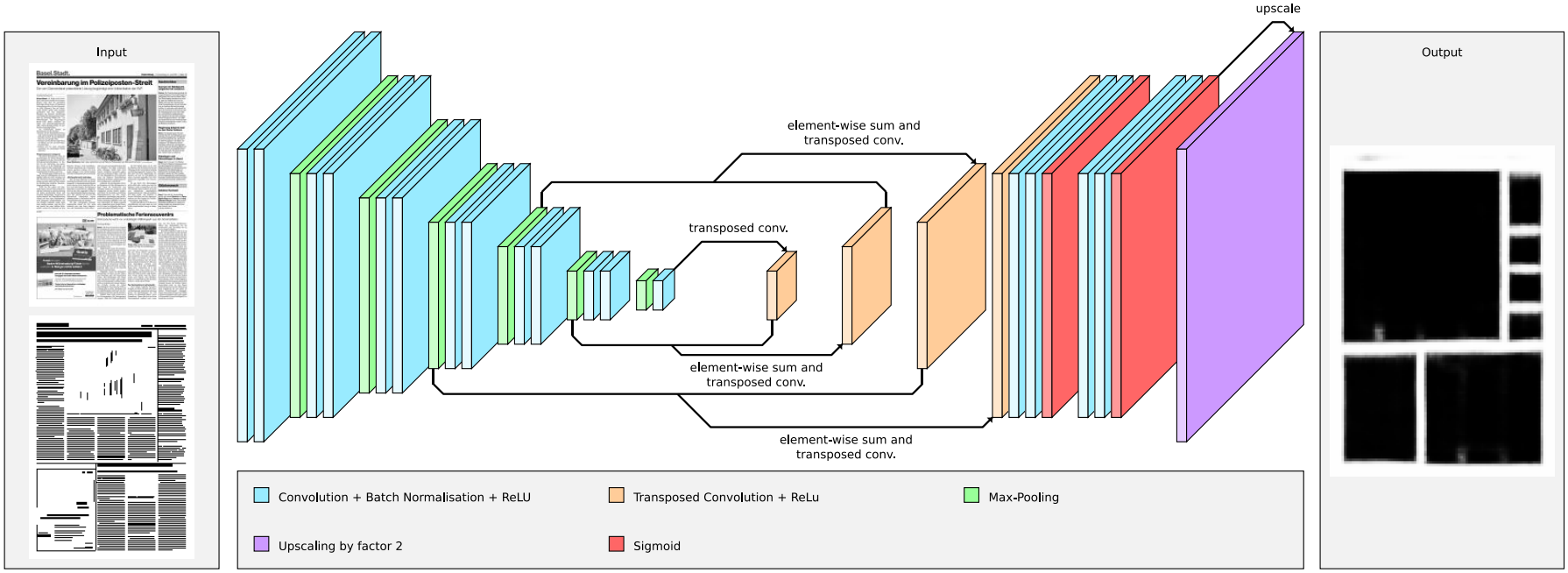

Sport Blick 15
«Steven hat sich alles selber beigebracht»
Verlieren verboten!
Transfer TICKER
Liverpool will Yann Sommer!

Nuisance


Liebling der Sterne
Das Tages-Horoskop
15,1 Millionen
Sind Sie der nächste Lotto-König?
Wochenpreis: 1 x sieben Nächte für 2 Personen, inkl. HP, im ****Seehotel Pilatus Hergiswil im Wert von 3000 Franken!

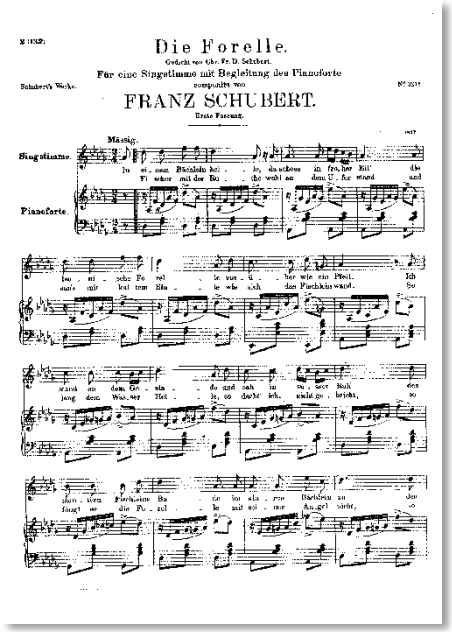
ARGUS DATA INSIGHTS
WISSEN ZUM ERFOLG
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Innosuisse – Swiss Innovation Agency

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.

2. Music scanning



Die Forelle.

Gedicht von Chr. Fr. D. Schubert.

Für eine Singstimme mit Begleitung des Pianoforte

componirt von

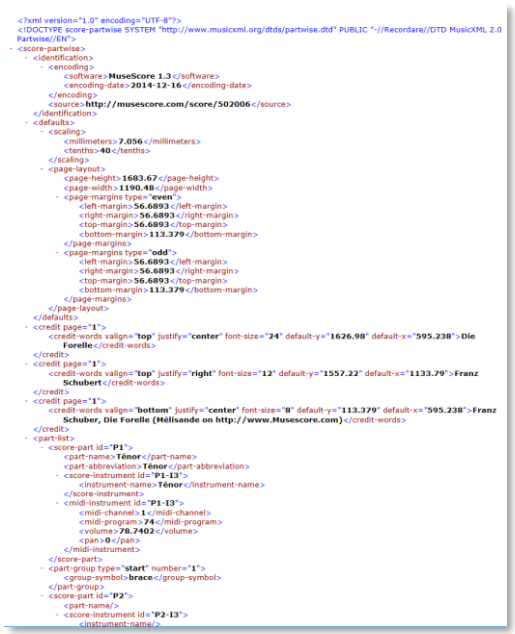
FRANZ SCHUBERT.

Erste Ausgabe.

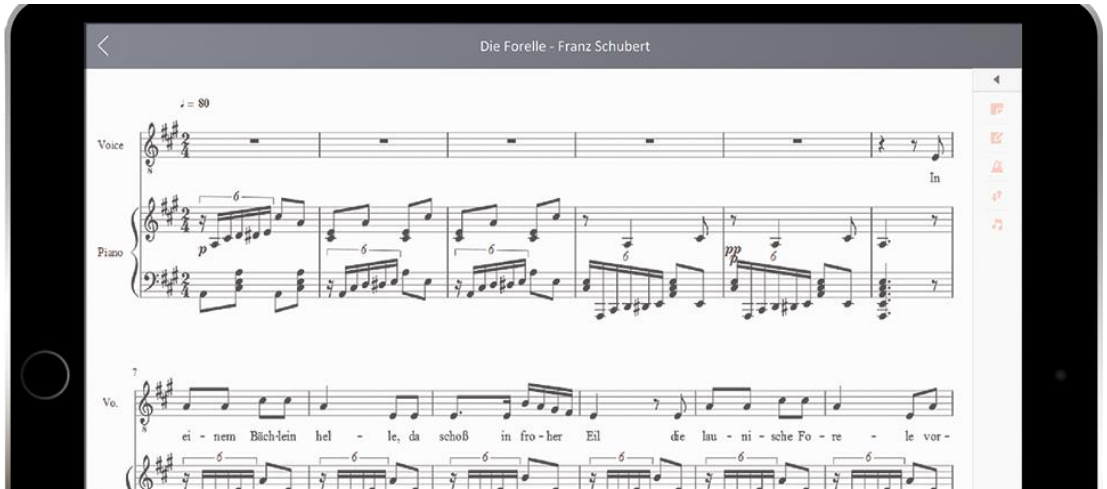
Mässig.

Singstimme.

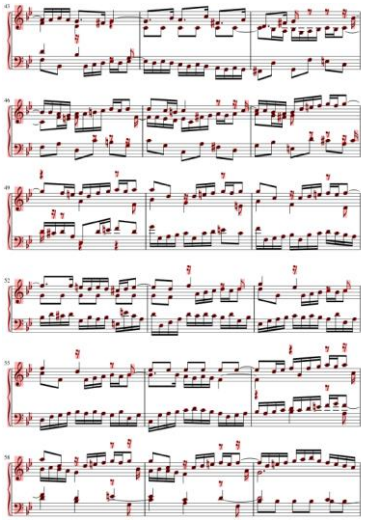
Pianoforte.



```
<?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 (Hélisande 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/>
```



2. Music scanning – challenges & solutions



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

Zürcher Hochschule
für Angewandte Wissenschaften

zh
aw

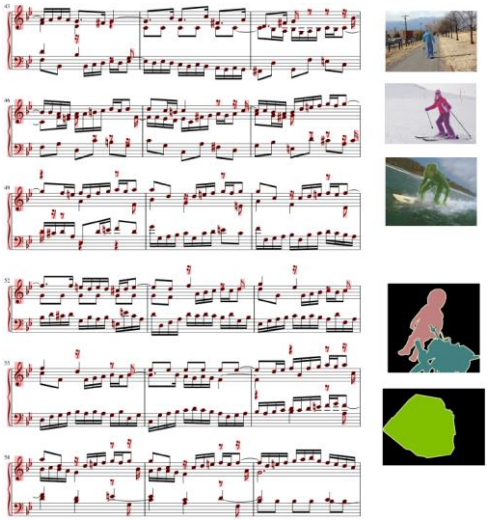
SCOREPAD

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

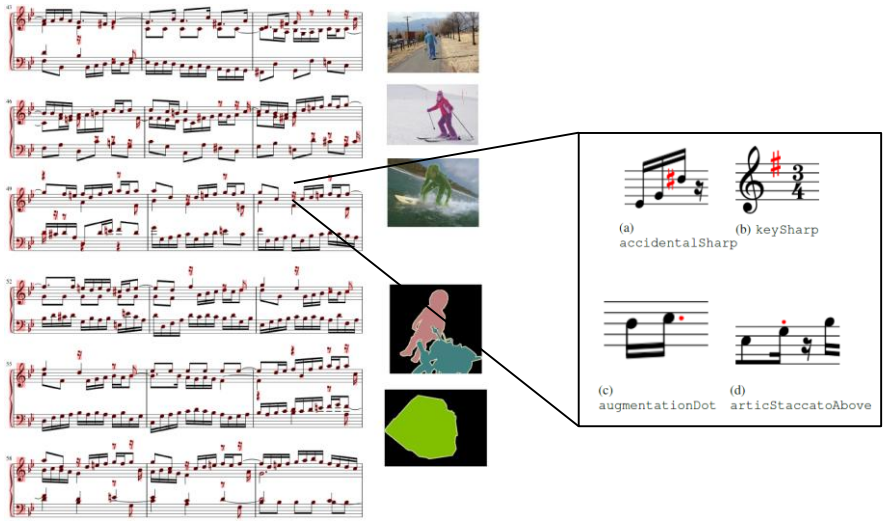
2. Music scanning – challenges & solutions



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

Zürcher Hochschule
für Angewandte Wissenschaften

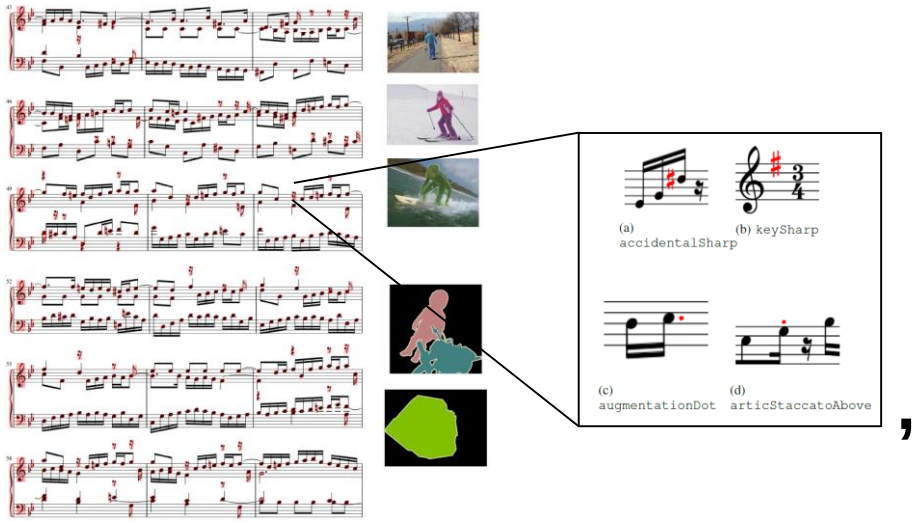
2. Music scanning – challenges & solutions



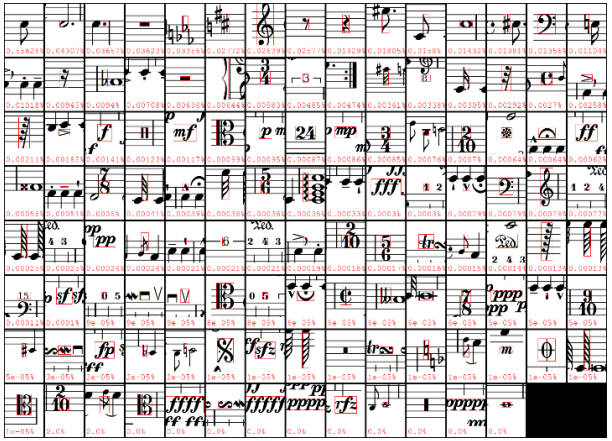
SCOREPAD
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Innosuisse – Swiss Innovation Agency

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

2. Music scanning – challenges & solutions

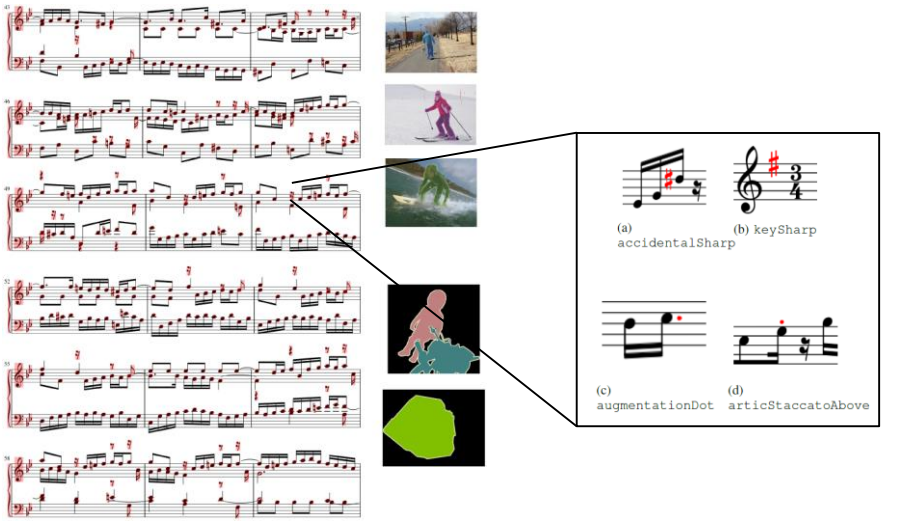


,

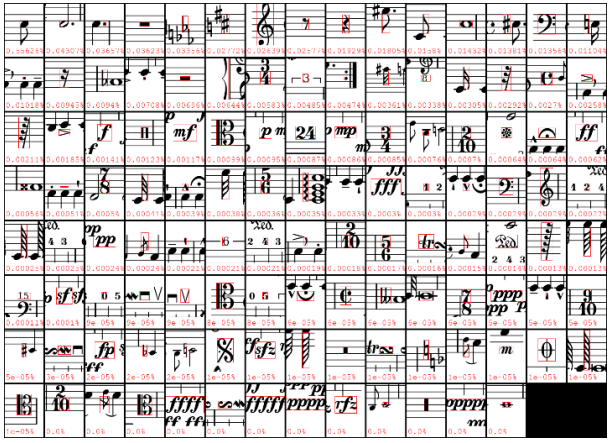


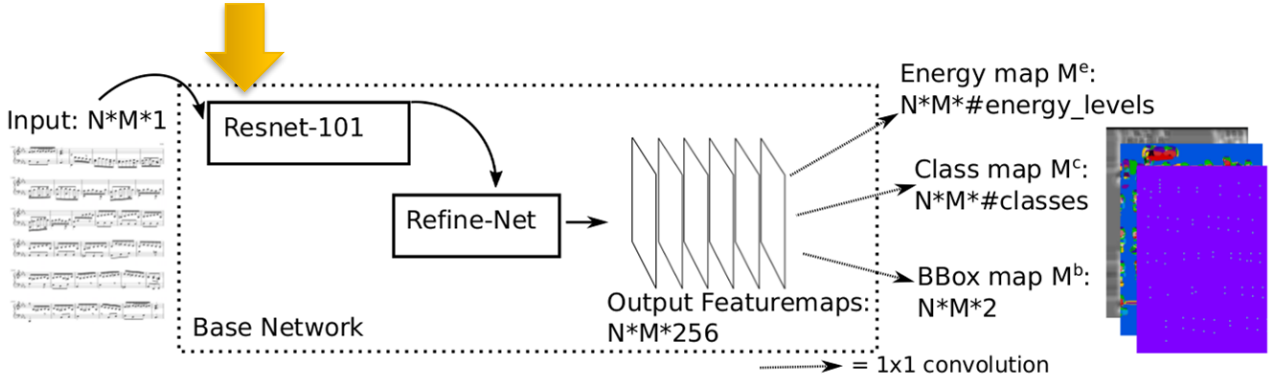
Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR'2018.

2. Music scanning – challenges & solutions



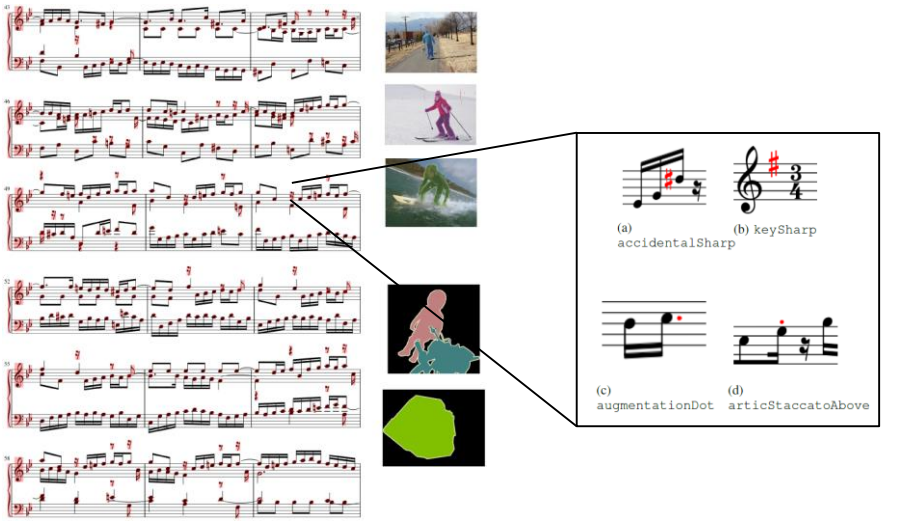
,



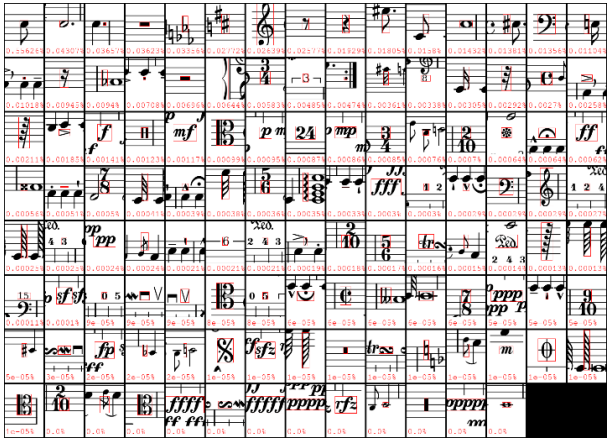


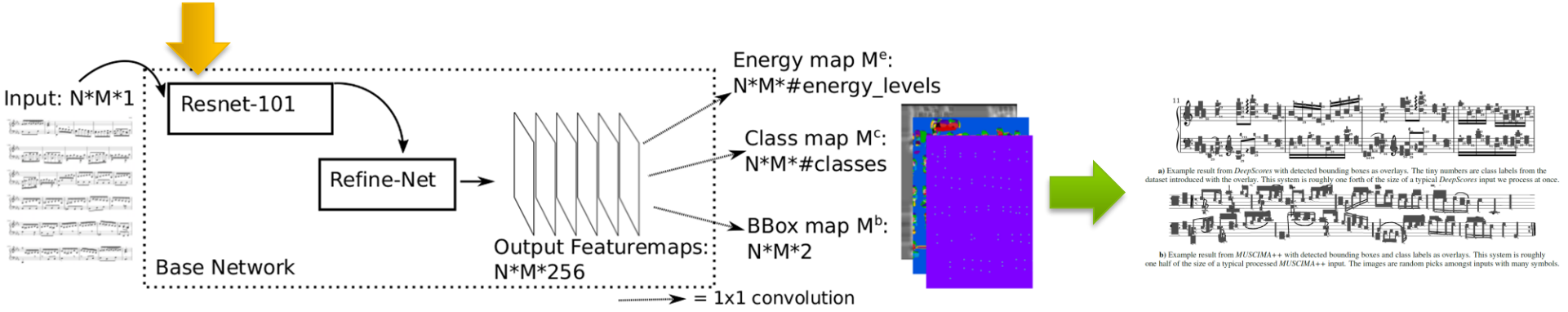
Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR'2018.
Tuggener, Elezi, Schmidhuber & Stadelmann (2018). *«Deep Watershed Detector for Music Object Recognition»*. ISMIR'2018.

2. Music scanning – challenges & solutions



,



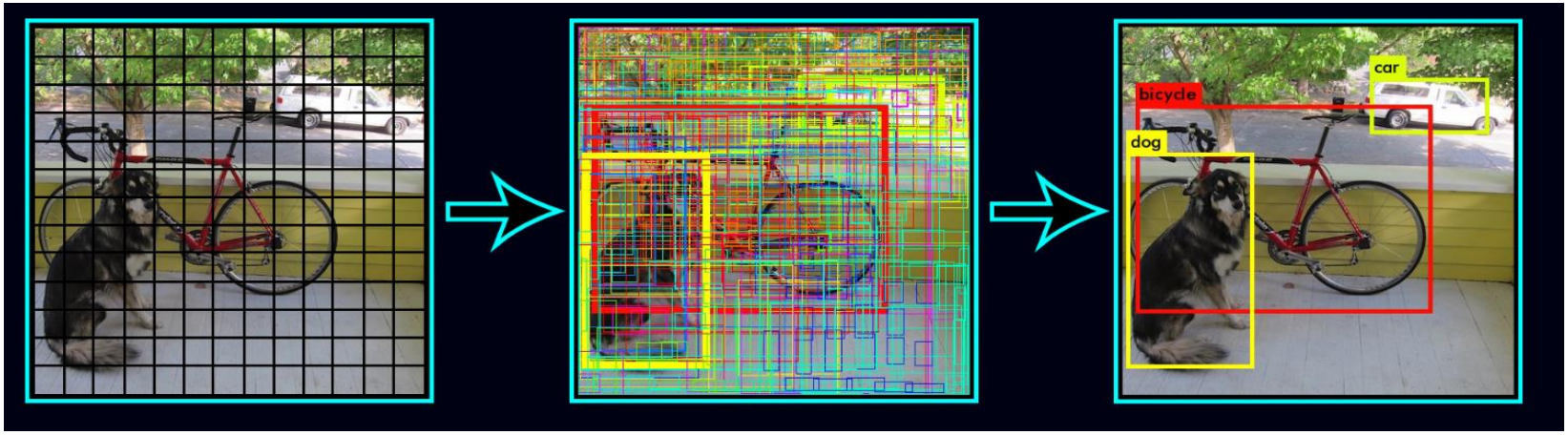


Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR'2018.
Tuggener, Elezi, Schmidhuber & Stadelmann (2018). *«Deep Watershed Detector for Music Object Recognition»*. ISMIR'2018.

2. OMR deep dive
OMR vs state of the art object detectors

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

YOLO/SSD-type detectors



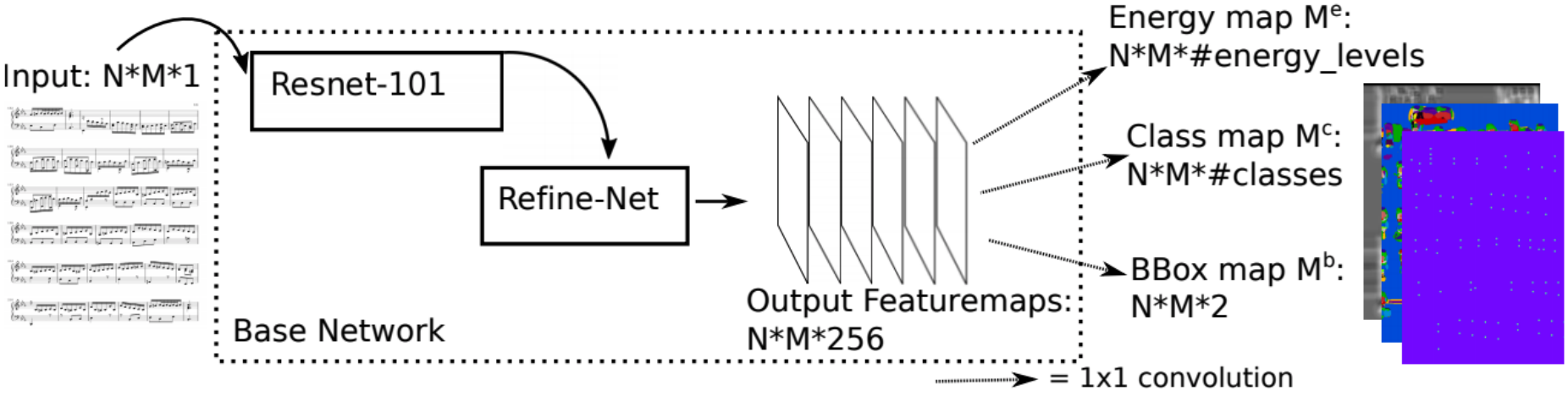
Source: https://pjreddie.com/darknet/yolov2/ (11.09.2018)

R-CNN

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

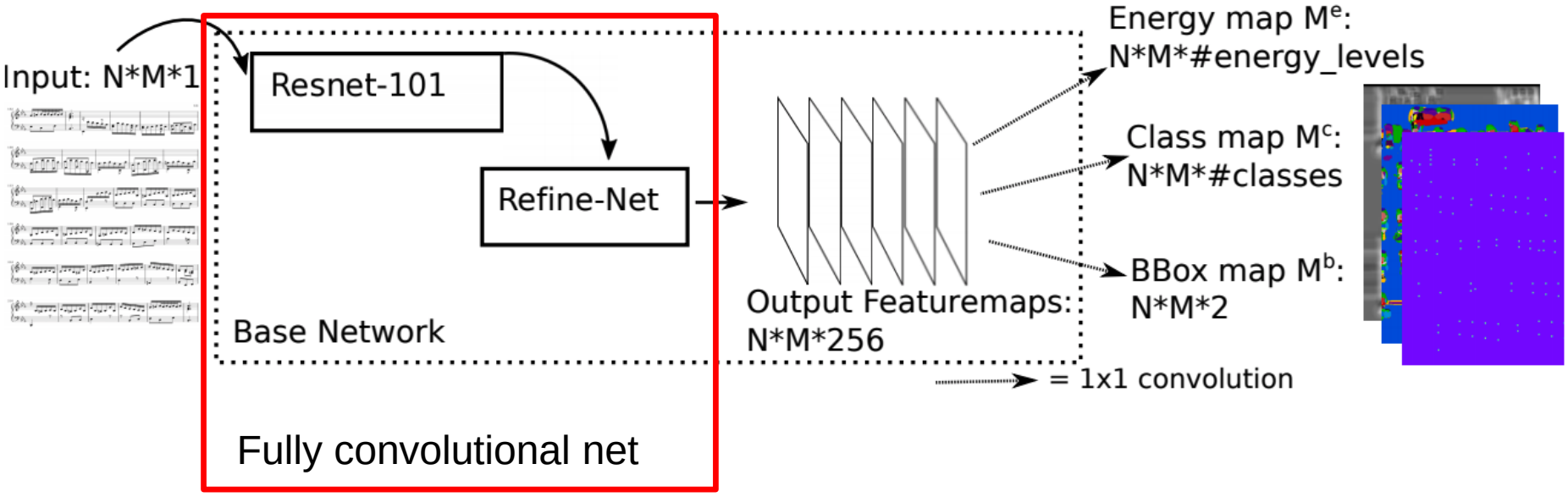
2. OMR deep dive (contd.)

The deep watershed detector



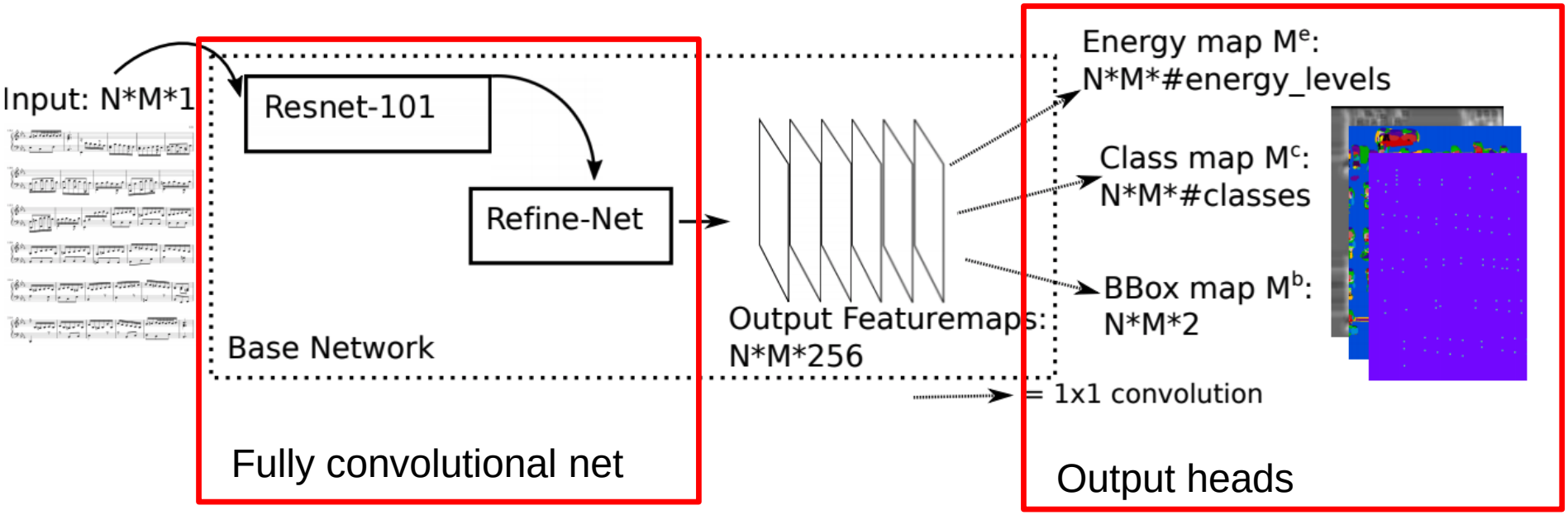
2. OMR deep dive (contd.)

The deep watershed detector

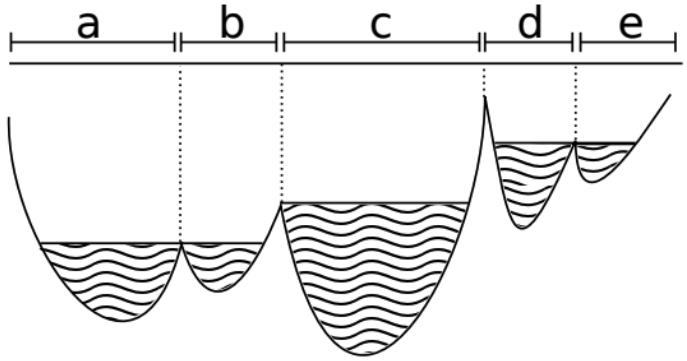


2. OMR deep dive (contd.)

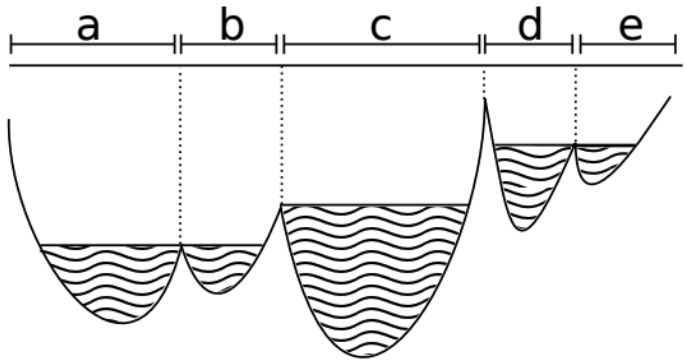
The deep watershed detector



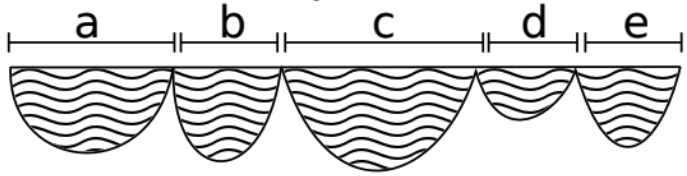
2. OMR deep dive (contd.)
The (deep) watershed transform



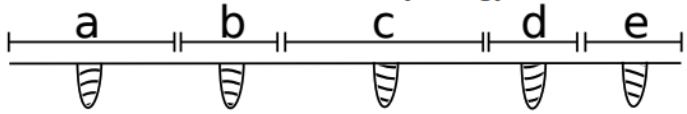
2. OMR deep dive (contd.)
The (deep) watershed transform



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



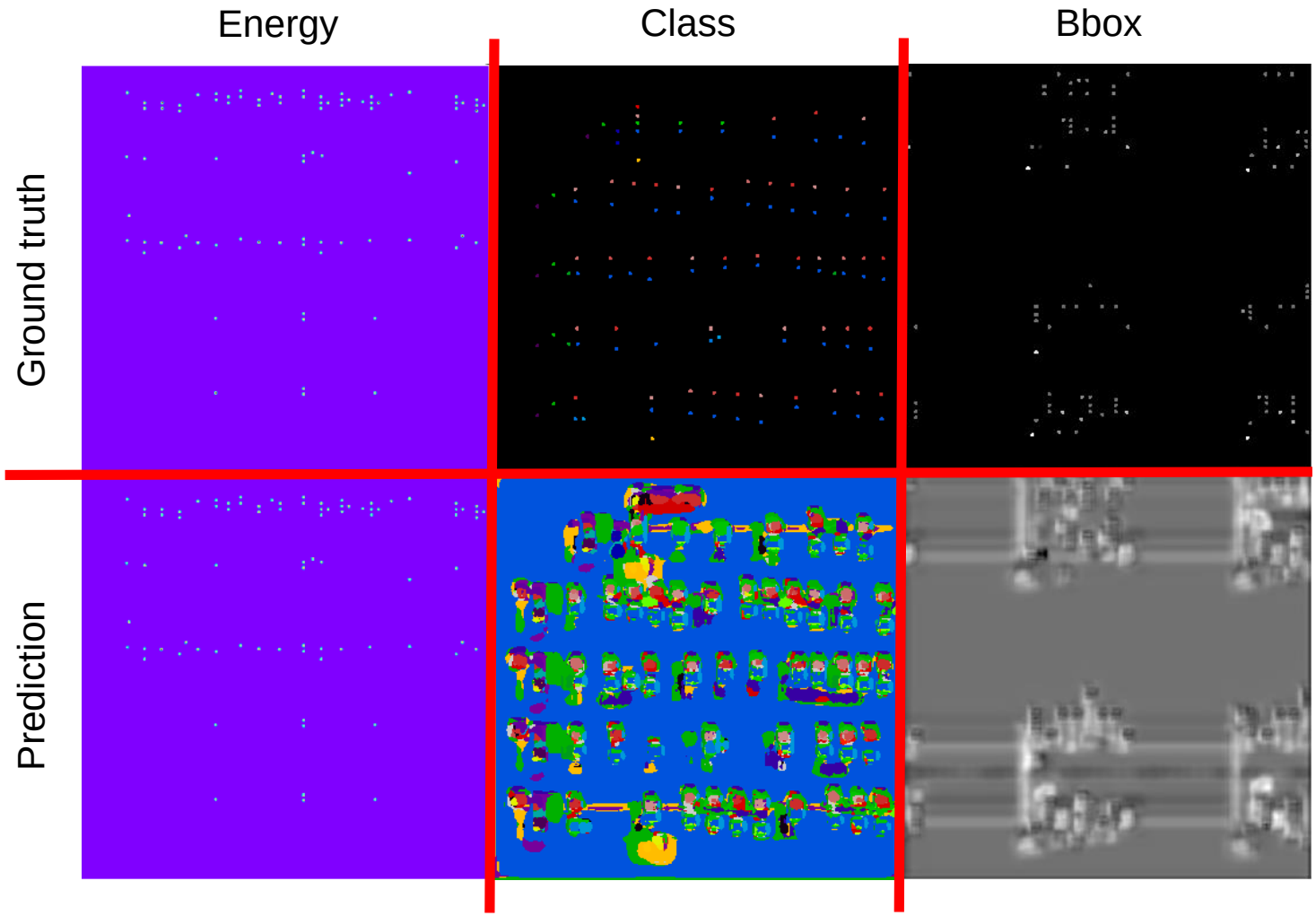
b) Energy function for the same five classes with fixed boundary energy.



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

2. OMR deep dive (contd.)

Output heads of the deep watershed detector



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

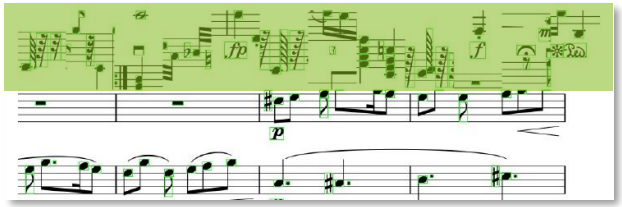
2. Music scanning – industrialization
(Work in progress)

Zürcher Hochschule
für Angewandte Wissenschaften

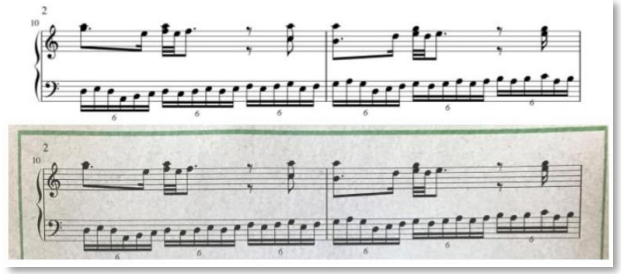
SCOREPAD
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Innosuisse – Swiss Innovation Agency

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.

Conclusions

- Document analysis is a **very fruitful use case** for Deep Learning (for business + R&D)
- **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)

datalab
www.zhaw.ch/datalab

data service
Swiss Alliance for
Data-Intensive Services

On me:

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

On the topics:

- Document analysis: https://stdm.github.io/research/#recent
- Data+Service Alliance: www.data-service-alliance.ch
- Collaboration: datalab@zhaw.ch

➔ Happy to answer questions & requests.

Zürcher Hochschule
für Angewandte Wissenschaften

APPENDIX

ML @ Information Engineering Group

Institute of Applied Information Technology, ZHAW School of Engineering

Zürcher Hochschule für Angewandte Wissenschaften

zh aw





Machine learning-based Pattern Recognition

- Robust Deep Learning
- Voice Recognition
- Document Analysis
- Learning to Learn & Control