Deep Watershed Detector & Music Object Recognition

Deep Learning Day 2018
Friday, 14th September 2018

Lukas Tuggener

Zürcher Hochschule
für Angewandte Wissenschaften

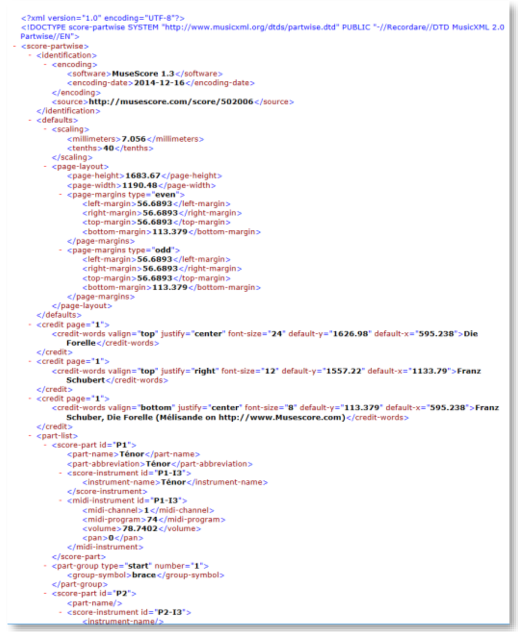
Università della Svizzera italiana

datalab
www.zhaw.ch/datalab

Contents

- **Why music scanning ?**
- **Why build a custom detection system ?**
- **How does it work ?**
- **How does it *really* work?**

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

SCOREPAD

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

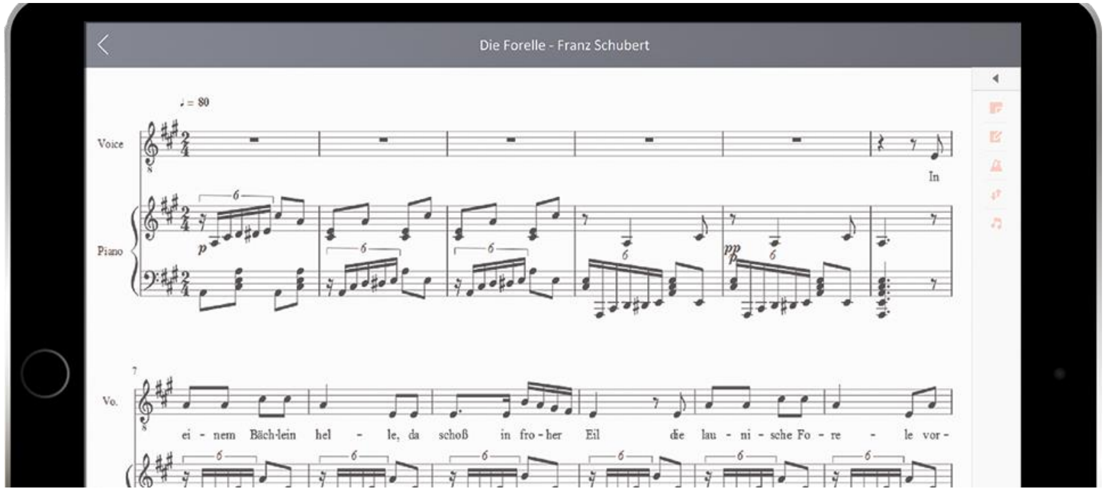
Innosuisse – Swiss Innovation Agency

Pdfs

Scans

Photos

Antique / Handwriting



Page turning

Transposing

Orchestra synchronization

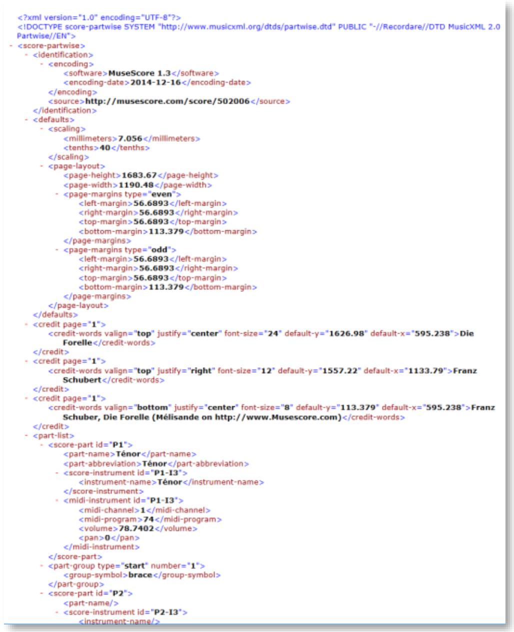
…

Music scanning

Zürcher Hochschule
für Angewandte Wissenschaften

Music object recognition

Semantic reconstruction



Rendering Software

Audio Processing

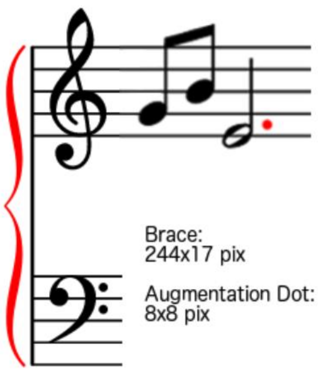
….

Music object recognition – challenges

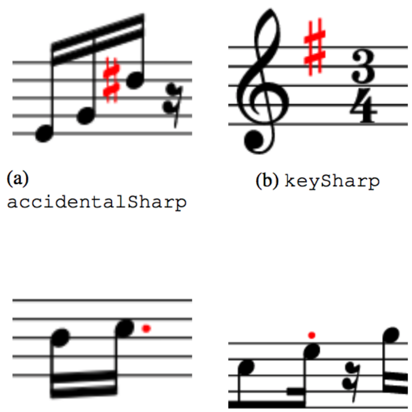
Data availability
No dataset available at the time large enough for DL

Object size & frequency, image size
Next slide

Size imbalance



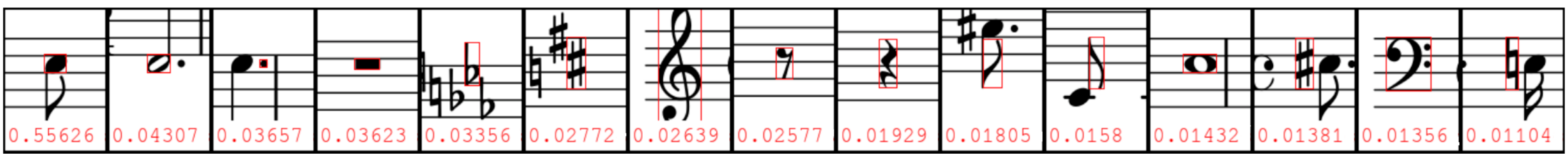
Context dependency



(a) accidentalSharp (b) keySharp

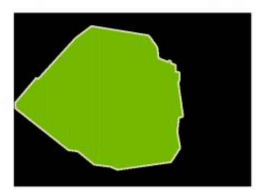
(c) augmentationDot (d) articStaccatoAbove

Class imbalance
(top 15 of 118 classes)



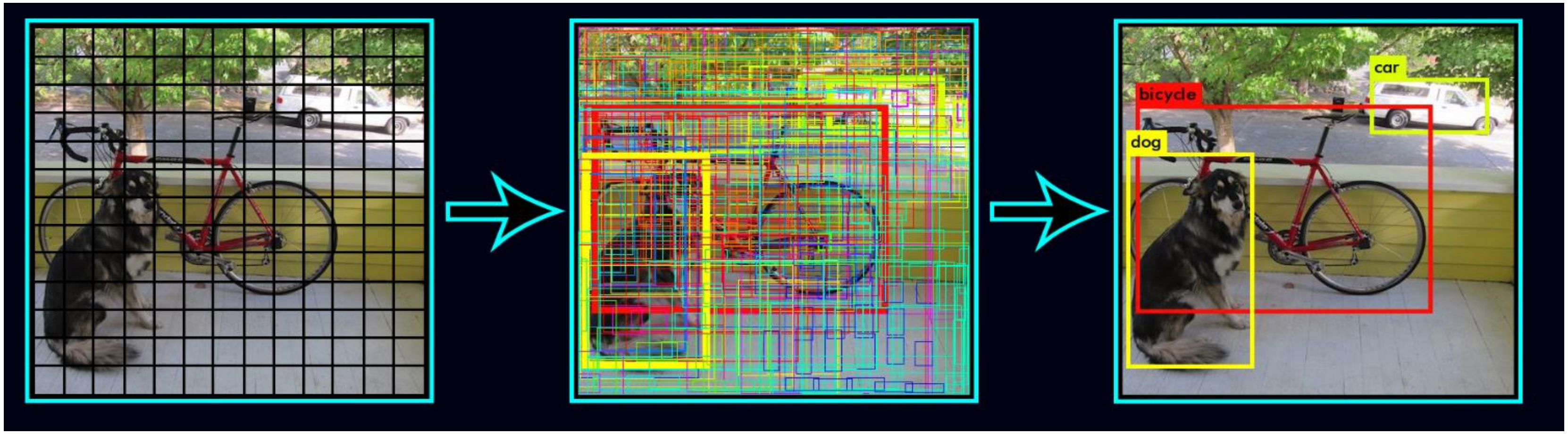
Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR'2018.

Mor vs Natural Images



Mor vs state of the art object detectors

YOLO/SSD-type detectors

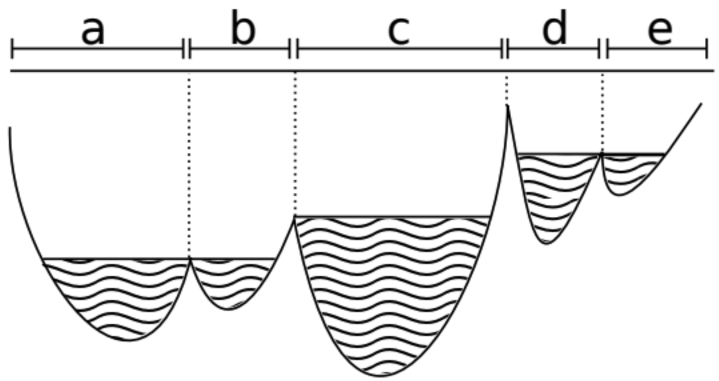


Taken from: https://pjreddie.com/darknet/yolov2/ on 11.9.2018

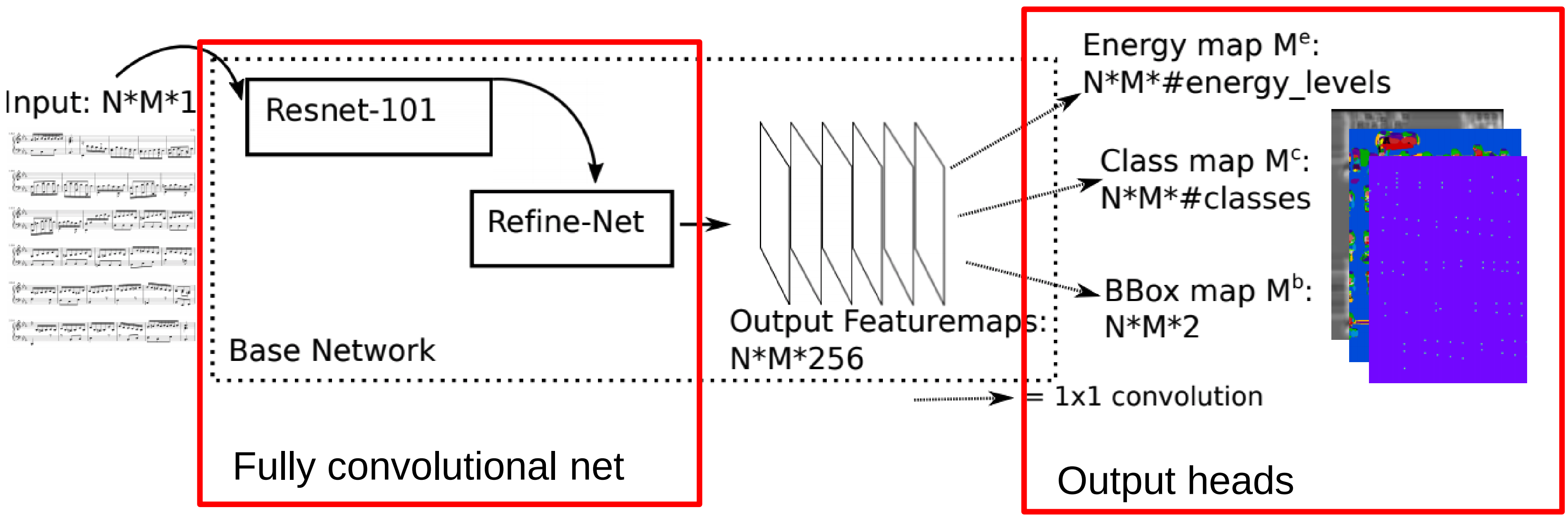
R-CNN

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

The deep watershed transform

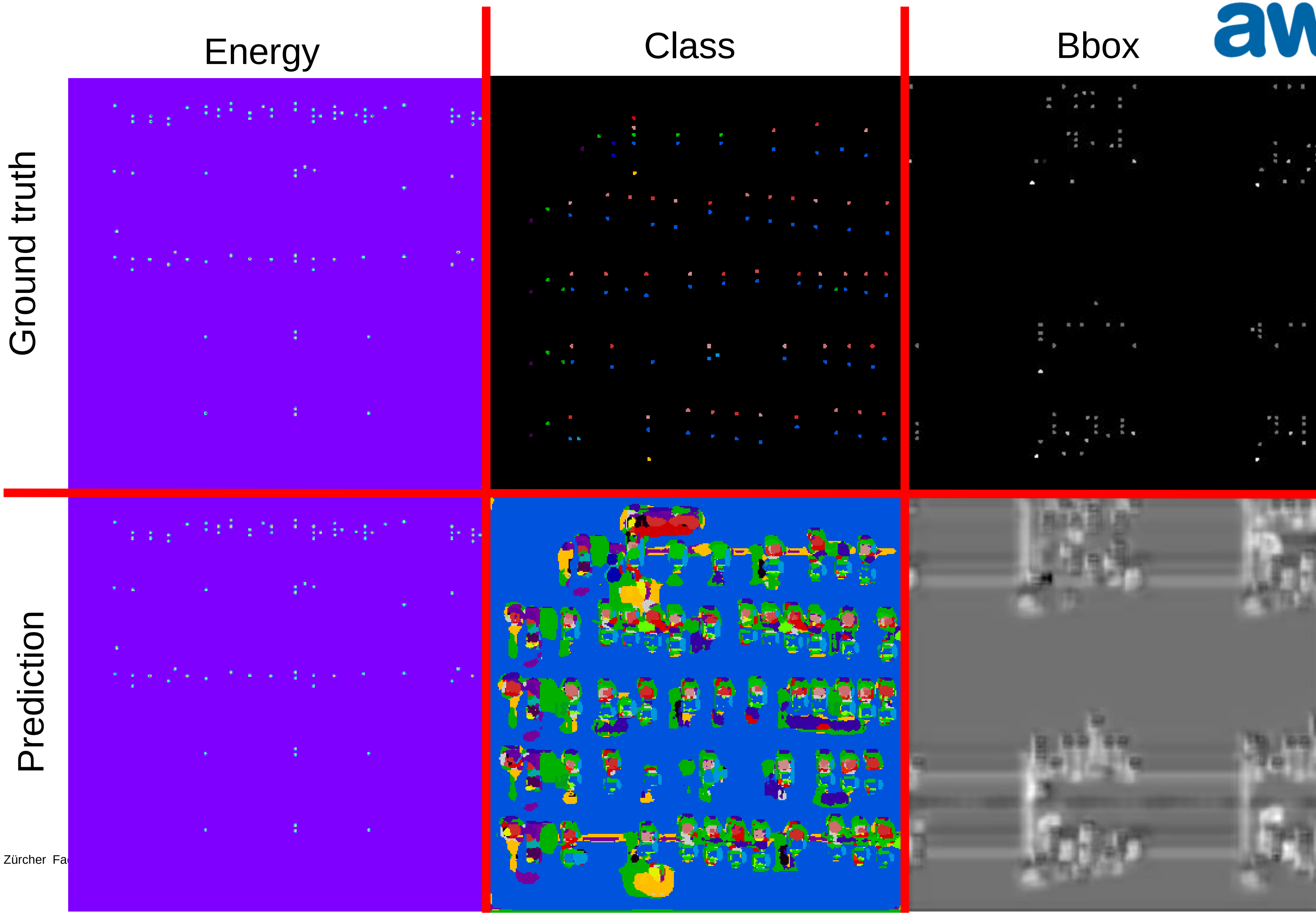


The deep watershed detector



Zürcher Hochschule
für Angewandte Wissenschaften

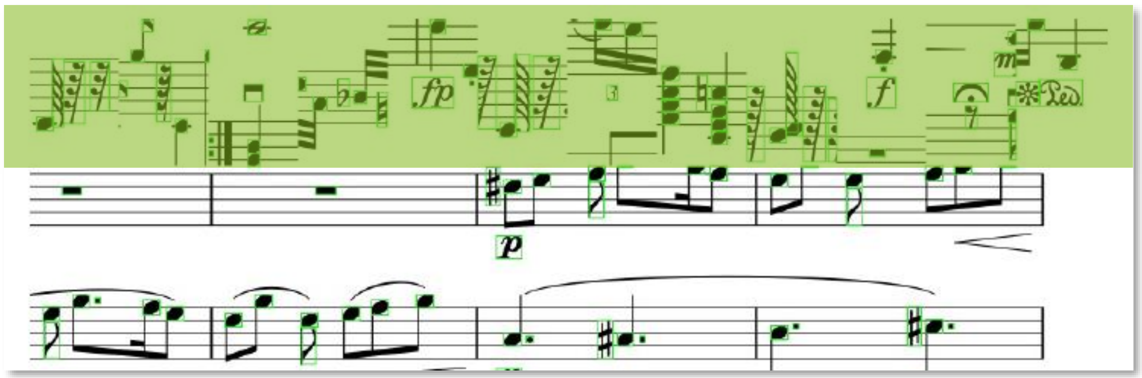
The deep watershed detector

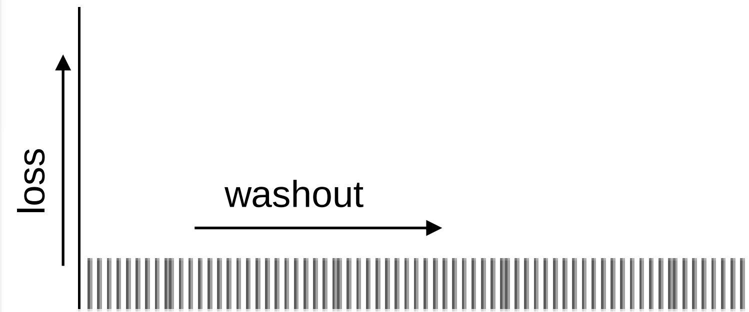


Zürcher Hochschule
für Angewandte Wissenschaften

Tweaks and improvements

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





2. Put additional effort (and compute) into hyperparameter **tuning** and **longer training**

Elezi, Tuggener, Pelillo & Stadelmann (2018). *«DeepScores and Deep Watershed Detection: current state and open issues»*. WoRMS @ ISMIR'2018

Current results

Ours:

DeepScores: 46.7%

State of the art:

	mAP (%)		
	DeepScores	MUSCIMA++	Capitan
Faster R-CNN	19.6	3.9	15.2
RetinaNet	9.8	7.7	14.5
U-Net	24.8	16.6	17.4

Pacha, Hajic, Calvo-Zaragoza (2018). «*A Baseline for General Music Object Detection with Deep Learning*». Appl. Sci. 2018, 8, 1488, MDPI.

Ongoing and future work

- Extend the model capabilities to non-synthetic data.

mAP: 47.5%

- More sophisticated balancing and stability tricks.
- Move to other tasks (natural images ….)

Closing Remarks

- Data is Key
  - Gathering it can be very expensive
  - Behavior outside training distribution is completely unpredictable
- The deep watershed detector can outperform state of the art
- A lot of the performance is in fine-tuning and engineering

On me:
- Doctoral Student ZHAW / USI
- lukas.tuggener@zhaw.ch
- 058 934 47 33
- https://tuggeluk.github.io/

Download DeepScores:
- https://tuggeluk.github.io/downloads/

DWD Code:
- https://github.com/tuggeluk/DeepWatershedDetection

Happy to answer questions & requests.

Università della Svizzera italiana

datalab www.zhaw.ch/datalab

Initial results

Class	AP@$\frac{1}{2}$	Class	AP@$\frac{1}{4}$
rest16th	0.8773	tuplet6	0.9252
noteheadBlack	0.8619	keySharp	0.9240
keySharp	0.8185	rest16th	0.9233
tuplet6	0.8028	noteheadBlack	0.9200
restQuarter	0.7942	accidentalSharp	0.8897
rest8th	0.7803	rest32nd	0.8658
noteheadHalf	0.7474	noteheadHalf	0.8593
flag8thUp	0.7325	rest8th	0.8544
flag8thDown	0.6634	restQuarter	0.8462
accidentalSharp	0.6626	accidentalNatural	0.8417
accidentalNatural	0.6559	flag8thUp	0.8279
tuplet3	0.6298	keyFlat	0.8134
noteheadWhole	0.6265	flag8thDown	0.7917
dynamicMF	0.5563	tuplet3	0.7601
rest32nd	0.5420	noteheadWhole	0.7523
flag16thUp	0.5320	fClef	0.7184
restWhole	0.5180	restWhole	0.7183
timeSig8	0.5180	dynamicPiano	0.7069
accidentalFlat	0.4949	accidentalFlat	0.6759
keyFlat	0.4685	flag16thUp	0.6621

Current results

Zürcher Hochschule
für Angewandte Wissenschaften

APPENDIX