Deep Watershed Detector & Music Object Recognition

Deep Learning Day 2018
Friday, 14th September 2018

Lukas Tuggener
Contents

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

Pdfs
Scans
Photos
Antique / Handwriting

Page turning
Transposing
Orchestra synchronization
...
Music scanning

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

Class imbalance
(top 15 of 118 classes)

Context dependency

Mor vs Natural Images
Mor vs state of the art object detectors

YOLO/SSD-type detectors

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
The deep watershed detector
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**

Current results

Ours:

DeepScores: 46.7%

State of the art:

<table>
<thead>
<tr>
<th></th>
<th>DeepScores</th>
<th>MUSCIMA++</th>
<th>Capitan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faster R-CNN</td>
<td>19.6</td>
<td>3.9</td>
<td>15.2</td>
</tr>
<tr>
<td>RetinaNet</td>
<td>9.8</td>
<td>7.7</td>
<td>14.5</td>
</tr>
<tr>
<td>U-Net</td>
<td>24.8</td>
<td>16.6</td>
<td>17.4</td>
</tr>
</tbody>
</table>

Ongoing and future work

- Extend the model capabilities to non-synthetic data.
- More sophisticated balancing and stability tricks.
- Move to other tasks (natural images ....)

mAP: 47.5%
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.
Initial results

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