Music Document Layout Analysis through Machine Learning and Human Feedback

Jorge Calvo-Zaragoza, Ké Zhang, Zeyad Saleh, Gabriel Vigliensoni, Ichiro Fujinaga

Schulich School of Music
McGill University, Montréal (Canada)

12th IAPR International Workshop on Graphics Recognition (Nov 2017)
Introduction
Introduction

- Music archives and libraries preserve music over the centuries
- Large amounts of content in symbolic format are required for computational analysis
- Manual transcription from source implies a high cost
Introduction

- Music archives and libraries preserve music over the centuries
- Large amounts of content in symbolic format are required for computational analysis
- Manual transcription from source implies a high cost
- **Automatic transcription systems for ancient scores become valuable tools**
Introduction

Optical Music Recognition (OMR)

- From score image to symbolic encoding
Optical Music Recognition (OMR)

- From score image to symbolic encoding
Optical Music Recognition (OMR)

- Several interdisciplinary steps
Introduction

- Most document-processing stages focus on *content separation*:
Most document-processing stages focus on *content separation*.
Introduction

- Most document-processing stages focus on *content separation*:
Introduction

► Most document-processing stages focus on *content separation*:
Introduction

- Poor generalization of hand-crafted strategies
- Music documents have a high level of heterogeneity
Introduction

Framework

- Machine learning framework for music document processing
- Regardless of the specific characteristics of the source
- Detection of the different layers at the same time
Framework
Framework

Pixelwise classification approach

- Categorization of each pixel within the input image

- Allows detecting small and thin elements present in music notation
Framework

- Machine learning for avoiding hand-crafted procedures
Framework

- Machine learning for avoiding hand-crafted procedures
- We make use of Convolutional Neural Networks (CNN)
  - Great performance in image-related tasks
  - Good generalization
Framework

Pixelwise classification

- Straightforward approach: classify every single pixel of the input image

\[ I(x, y) \rightarrow \{ \text{background, staff line, symbol, text, ...} \} \]
Framework

Pixelwise classification

» Ground-truth example\(^1\)
  » One page \(\sim\) 30 million pixels

\(^1\)Salzinnes Antiphonal manuscript (CDM-Hsmu M2149.14)
Framework

Pixelwise classification

- CNN is provided with the surrounding region of the pixel to be classified
Framework

Pixelwise classification

- Estimation of a probability for each possible category
Framework

Performance

- Example over piece of test document

<table>
<thead>
<tr>
<th>Original</th>
<th>Ground-truth</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="original image" /></td>
<td><img src="image2.png" alt="ground-truth image" /></td>
</tr>
<tr>
<td>Prediction</td>
<td>Mislabeled</td>
</tr>
<tr>
<td><img src="image3.png" alt="prediction image" /></td>
<td><img src="image4.png" alt="mislabeled image" /></td>
</tr>
</tbody>
</table>
Framework

Generalization

- Relevant issue
  - How to approach a new archive?
Framework

Generalization

- Relevant issue
  - How to approach a new archive?
    - Human-aided workflow
Human-aided workflow
Human-aided workflow
Human-aided workflow

Pixel.js

- Web-based tool for pixel-level annotation
Human-aided workflow

Pixel.js

- Web-based tool for pixel-level annotation

- More information at 14:00, stay tuned!
Human-aided workflow

Preliminary user-centered evaluation

- Labeling one whole page (~24 million pixels) of a new document type
- Reduction from 30 to 18 hours with the human-aided approach
Conclusions
Conclusions

Summary

- Generalizable music document analysis with machine learning
- Human-aided workflow for a new type of document
Conclusions

Future work

- Integration with the rest of the OMR workflow
- Efficient strategies for the classification stage
Conclusions

Future work

- Integration with the rest of the OMR workflow
- Efficient strategies for the classification stage
  - Image-to-image pixel-wise classification
Conclusions

Future work

- Image-to-image pixelwise classification
  - Classify a whole region at the same time
  - Fully-Convolutional Neural Network

- Similar accuracy but much higher efficiency
Thank you!
Music Document Layout Analysis through Machine Learning and Human Feedback

Jorge Calvo-Zaragoza, Ké Zhang, Zeyad Saleh, Gabriel Vigliensoni, Ichiro Fujinaga

Schulich School of Music
McGill University, Montréal (Canada)

12th IAPR International Workshop on Graphics Recognition (Nov 2017)