Pixelwise classification
for music document analysis

Jorge Calvo-Zaragoza

Center for Interdisciplinary Research in Music Media and Technology
Schulich School of Music
McGill University, Montréal (Canada)

SIMSSA Workshop XII (Aug 2017)
Introduction
Introduction

- Music archives and libraries preserve music over the centuries
- Computational tools for music analysis are of great interest
Introduction

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

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

Optical Music Recognition (OMR)

- From score image to symbolic encoding
Introduction

Optical Music Recognition (OMR)

- From score image to symbolic encoding
Introduction

Optical Music Recognition (OMR)

- Several interdisciplinary steps

![](image)
Introduction

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

- Poor generalization of the existing 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
Convolutional Neural Networks

- Series of hierarchical transformations (convolutions)
- Transformations not fixed but learned through training
- Less dependent on human intervention
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

- To train the CNN we need ground truth
  - Documents whose categories have been correctly separated
Framework

Pixelwise classification

- Ground-truth example\(^1\)
- One page \(~\) 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

![Image of musical notation with symbols, background, and staff regions labeled: Symbol, Background, Staff]
Framework

Pixelwise classification

- Estimation of a probability for each possible category
Framework

Pixelwise classification

- Relevant issues
Framework

Pixelwise classification

- Relevant issues
  - Ground truth creation
Framework

Pixelwise classification

- Relevant issues
  - Ground truth creation
    - Pixel.js
Framework

Pixel.js

- Web-based tool for ground truth creation
Framework

Pixelwise classification

- Relevant issues
  - Ground truth creation
  - Pixel.js
Framework

Pixelwise classification

- Relevant issues
  - Ground truth creation
    - Pixel.js
  - Computational cost
Framework

Pixelwise classification

- Relevant issues
  - Ground truth creation
    - Pixel.js
  - Computational cost
    - Image-to-image approach
Framework

Image-to-image classification

- Image-to-image pixelwise classification
  - Classify a whole region at the same time

- We need to split the document into patches of equal size
Framework

Image-to-image classification

- Similar accuracy
- Much more efficient (from several hours to few minutes)
- Usually needs a bigger training set
Deployment
Deployment

General use

- Full workflow for a new type of document
  - Ground-truth creation with Pixel.js
  - Model training and document processing as Rodan jobs
Deployment

Resources

- Training models: very slow, need of high-performance computing
- Classification: fast with the image-to-image approach
Deployment

DEMO
Conclusions
Conclusions

Summary

- Generalizable music document analysis with machine learning
- Research on effective and efficient strategies
- Usability through Rodan framework
Conclusions

Future work

- Integrate with the rest of the OMR workflow
- Make efforts towards faster adaptation to new document types
  - Efficient ground truth creation with Pixel.js
  - Study of model adaptation techniques
Thank you!
Pixelwise classification for music document analysis

Jorge Calvo-Zaragoza

Center for Interdisciplinary Research in Music Media and Technology
Schulich School of Music
McGill University, Montréal (Canada)

SIMSSA Workshop XII (Aug 2017)