Towards human-aided document processing for Optical Music Recognition

Jorge Calvo-Zaragoza

PRHLT Research Center, Universitat Politècnica de Valencia
Schulich School of Music, McGill University

SIMSSA Workshop XIV
May 28, 2018

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada
Canada
DDMAL DISTRIBUTED DIGITAL MUSIC ARCHIVES & LIBRARIES LAB
Schulich School of Music École de musique Schulich
McGill
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

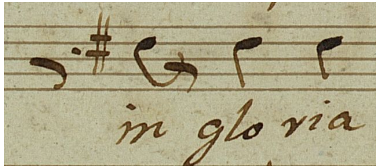
Introduction

- Digital encoding of music enables a wide range of applications
- Many written music remains only in physical format
- Typesetting music represents a costly endeavor
- Optical Music Recognition can be seen as the key to increasing the number of available encoded music sources

Introduction

- Optical Music Recognition (OMR) is the field that studies how to make computers capable of reading music
- Difficulties of OMR
  - Music notation is complex
  - Music manuscripts are highly heterogeneous
    - Document conditions
    - Sheet organization
    - Notational systems

Introduction

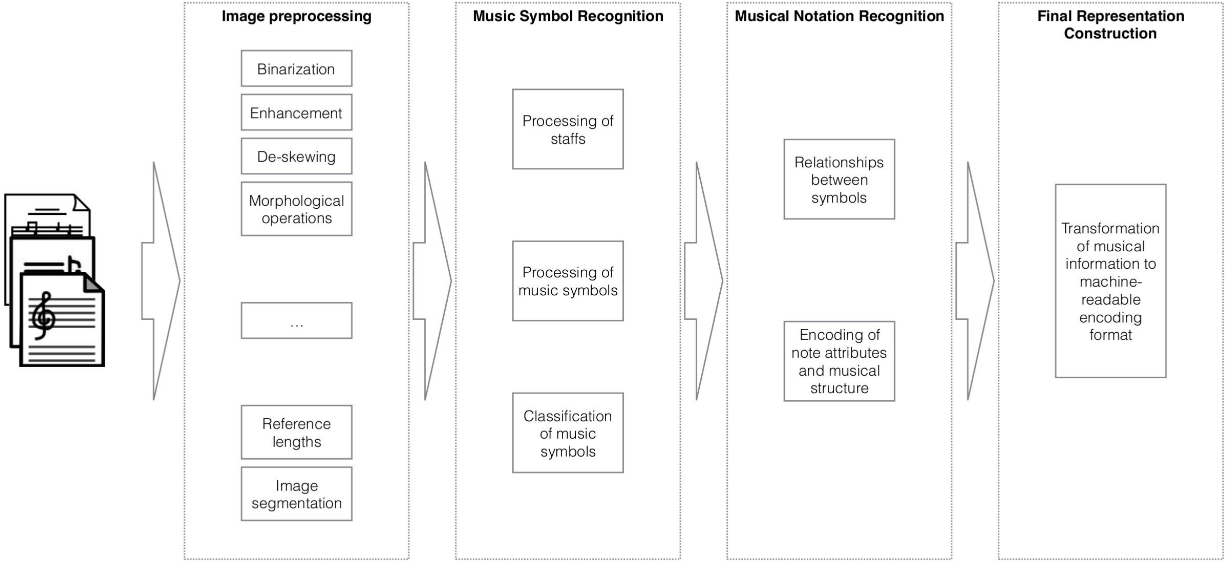


Introduction

- The “universal OMR” may be out of reach
- General OMR workflow
  - Document processing
  - Music symbol recognition
  - Notation assembly
  - Encoding

Framework

Optical Music Recognition workflow



Framework

Optical Music Recognition workflow

- Sequential stages: errors are propagated
- Each stage should be checked before going on
- It is necessary to involve the user in the process
  - Human-aided Optical Music Recognition workflow

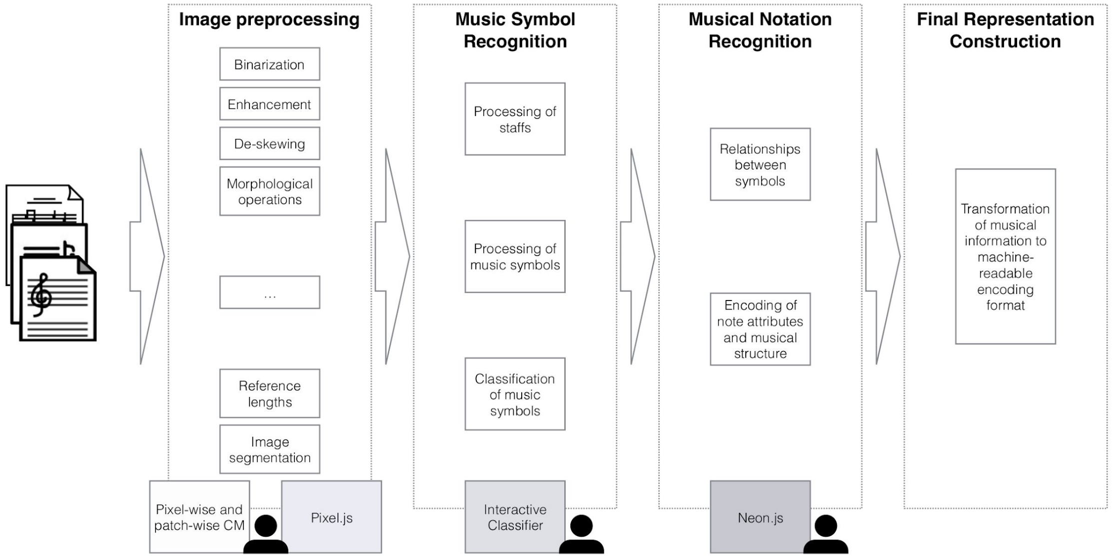
Introduction

Human-aided Optical Music Recognition workflow

- Involve the user in the process to guide the computer
- User is necessary: take the most out of it
- The OMR stages should not be fixed, but allow adaptation

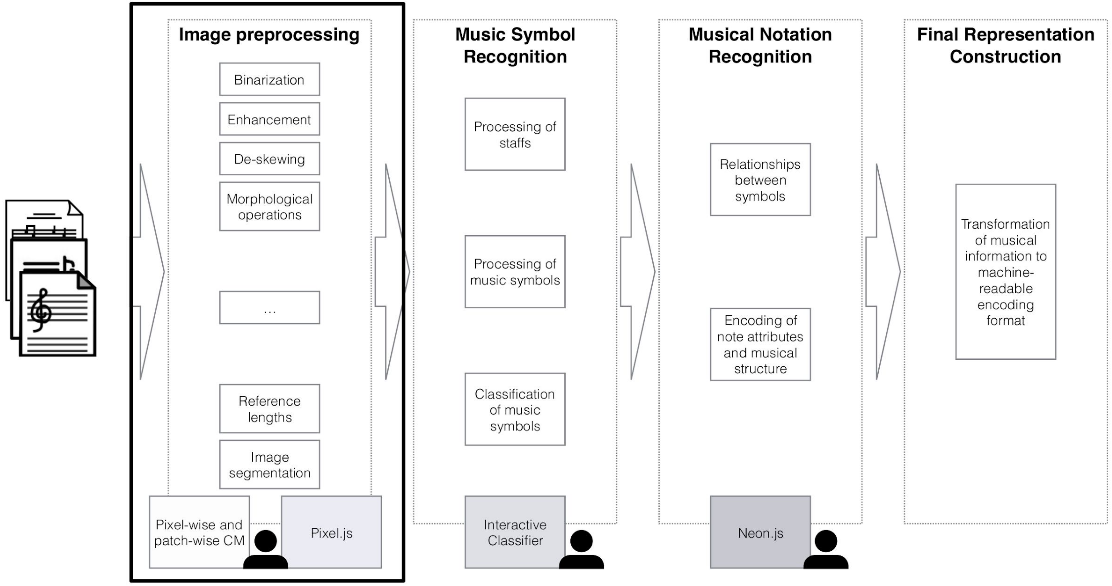
Framework

Human-aided Optical Music Recognition workflow



Framework

Human-aided Optical Music Recognition workflow



Framework

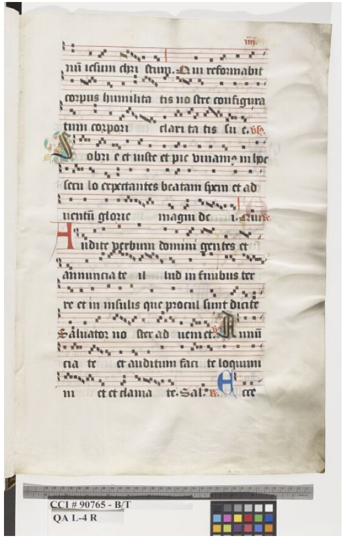
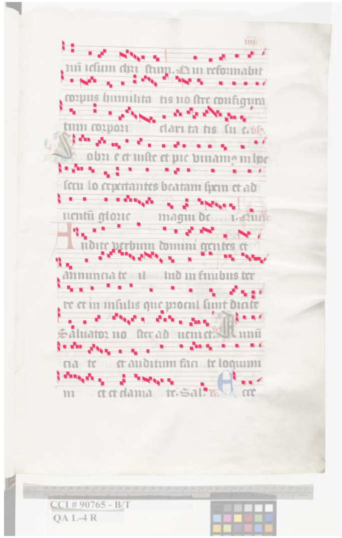
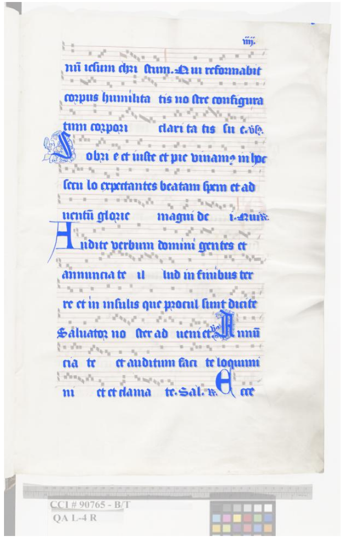
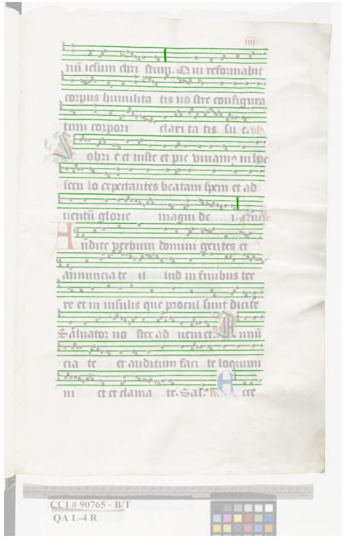
Document processing

- Separate the content of the document into its constituent layers

Framework

Document processing

- Separate the content of the document into its constituent layers

Human-aided music document analysis

Core processes

- Machine Analysis
- Human Teaching

Human-aided document analysis

Machine Analysis

- Avoid hand-crafted procedures that exploit specific characteristics
- We need models that learn to do the task
- This naturally leads to machine learning techniques
  - Ground-truth data is necessary for the models to be trained

Human-aided document analysis

Machine Analysis

- Our workflow requires detection at pixel level
- Pixelwise Classification Method (CM) with Convolutional Neural Networks
  - The surrounding region of each pixel contains enough discriminative information
  - The network is trained from a large number of examples for each category
  - It learns the regularities in these examples and creates a model out of the data
  - Once a model is trained, it is used to classify new examples

Human-aided document analysis

Machine Analysis: Convolutional Neural Networks

- Convolutional Neural Networks represent the state of the art in computer vision and image processing tasks
- Hierarchy of filters (convolutions) that process an image to predict a label
- Filters are not fixed but learned through a training process
- Feature extraction is not necessary

Human-aided document analysis

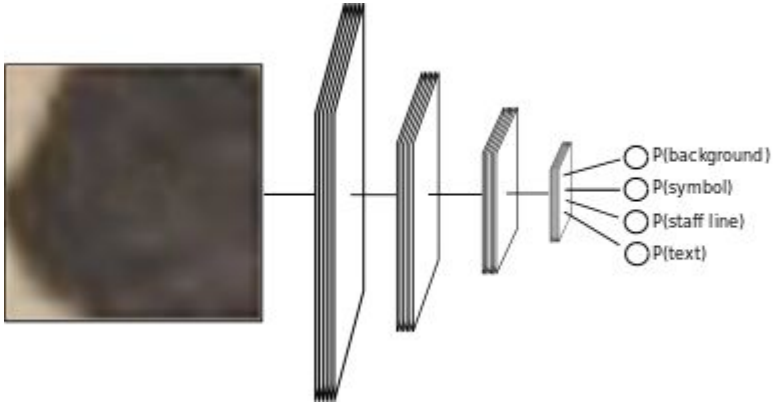
Machine Analysis

- Pixelwise Classification Method (CM) with Convolutional Neural Networks

Human-aided document analysis

Machine Analysis

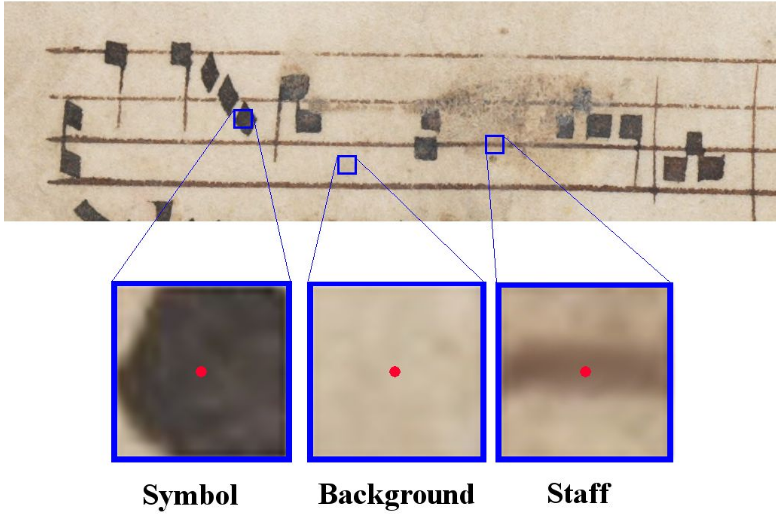
- Pixelwise Classification Method (CM) with Convolutional Neural Networks



Human-aided document analysis

Machine Analysis

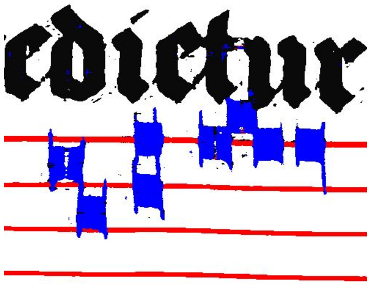
- Pixelwise Classification Method (CM) with Convolutional Neural Networks



Human-aided document analysis

Machine Analysis

- Pixelwise Classification Method (CM) with Convolutional Neural Networks



Human-aided document analysis

Machine Analysis

- Pixelwise CM
    - Advantages
        - Learning-driven model
        - Good performance
        - Learning from limited ground-truth data
    - Disadvantages
        - High temporal cost

Human-aided document analysis

Machine Analysis

- The temporal cost of the pixelwise CM represents a troublesome bottleneck
- Patchwise CM with Auto-Encoders
  - Replace the CNN classifier by class-wise (Convolutional) Auto-Encoders
  - Filters learn an image-to-image prediction
  - Process a complete sub-image (patch) in a single step

Human-aided document analysis

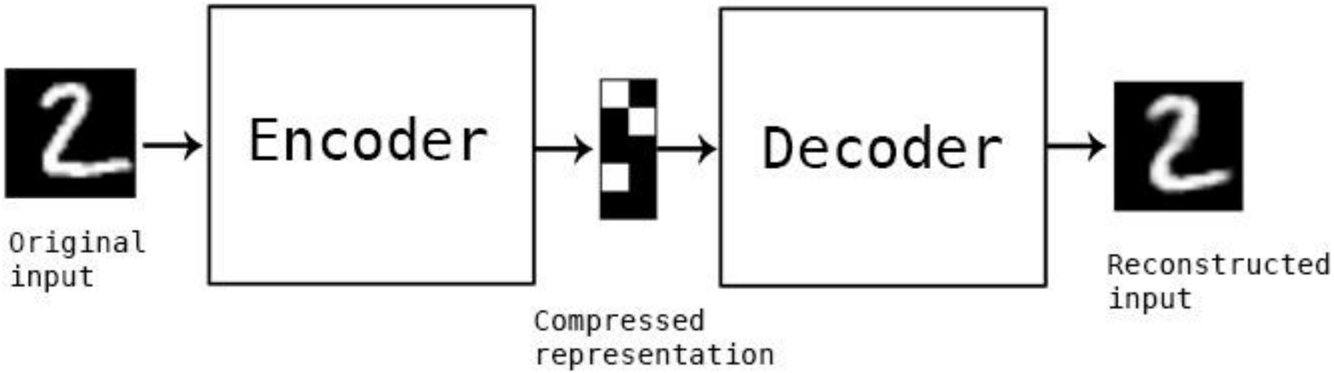
Machine Analysis: Auto-Encoders

- Auto-encoders

Human-aided document analysis

Machine Analysis: Auto-Encoders

- Auto-encoders



Human-aided document analysis

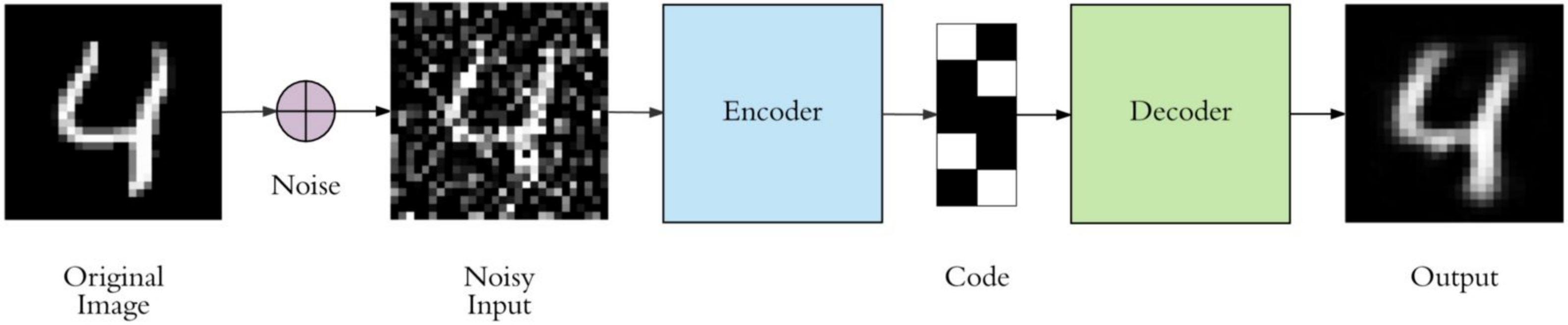
Machine Analysis: Auto-Encoders

- Denoising auto-encoders

Human-aided document analysis

Machine Analysis: Auto-Encoders

- Denoising auto-encoders



Human-aided document analysis

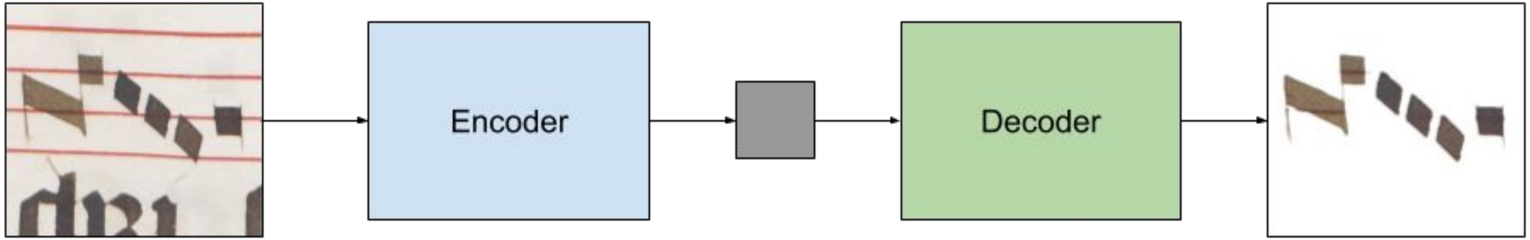
Machine Analysis: Auto-Encoders

- Selectional auto-encoders

Human-aided document analysis

Machine Analysis: Auto-Encoders

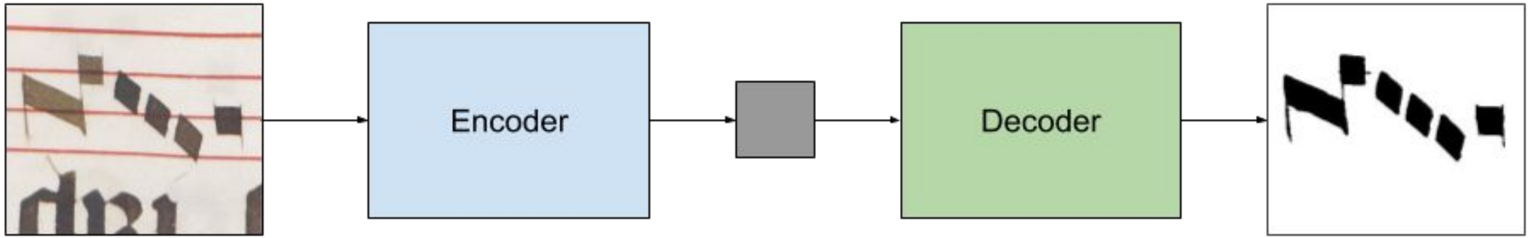
- Selectional auto-encoders



Human-aided document analysis

Machine Analysis: Auto-Encoders

- Selectional auto-encoders



Human-aided document analysis

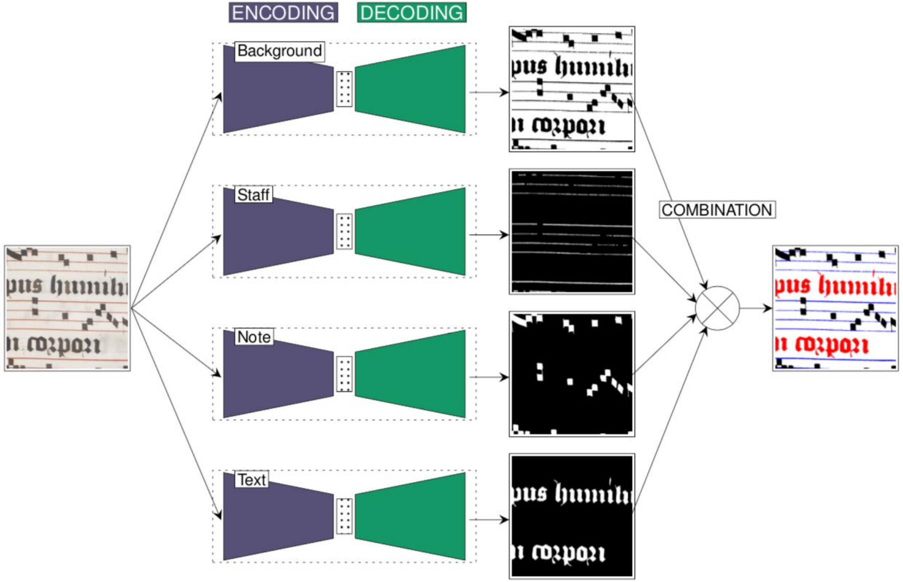
Machine Analysis: Auto-Encoders

- Class-wise selectional auto-encoders for document processing

Human-aided document analysis

Machine Analysis: Auto-Encoders

- Class-wise selectional auto-encoders for document processing



Human-aided document analysis

Machine Analysis: Auto-Encoders

- Patchwise CM
  - Advantages
    - Fast document processing
    - Class-wise independent detection
  - Disadvantages
    - Demanding ground-truth data

Human-aided document analysis

Machine Analysis

	Accuracy	Time per page
Pixelwise CM	~ 90 %	~ 6 hours
Patchwise CM	~ 92 %	~ 1 minute

Human-aided document analysis

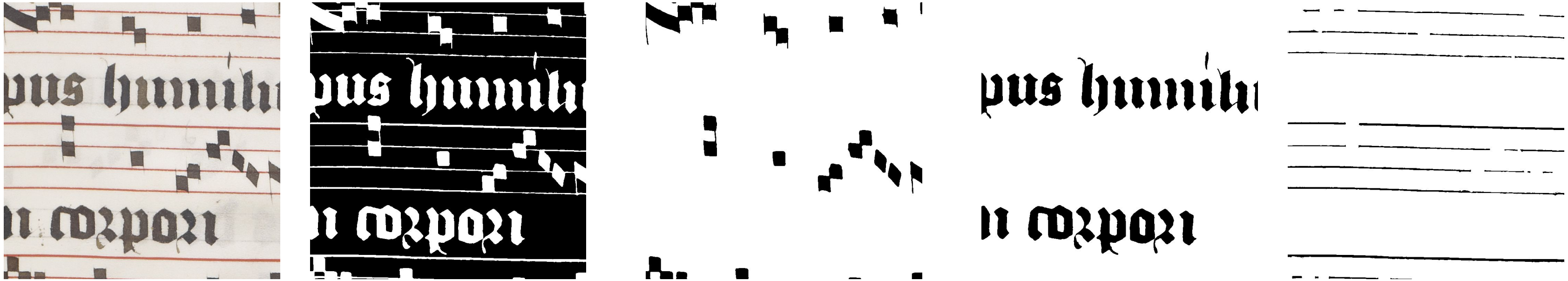
Human Teaching

- User is in charge of teaching what needs to be done
- In practice: manual separation of document layers to create ground-truth data

Human-aided document analysis

Human Teaching

- User is in charge of teaching what needs to be done
- In practice: manual separation of document layers to create ground-truth data



Human-aided document analysis

Human Teaching

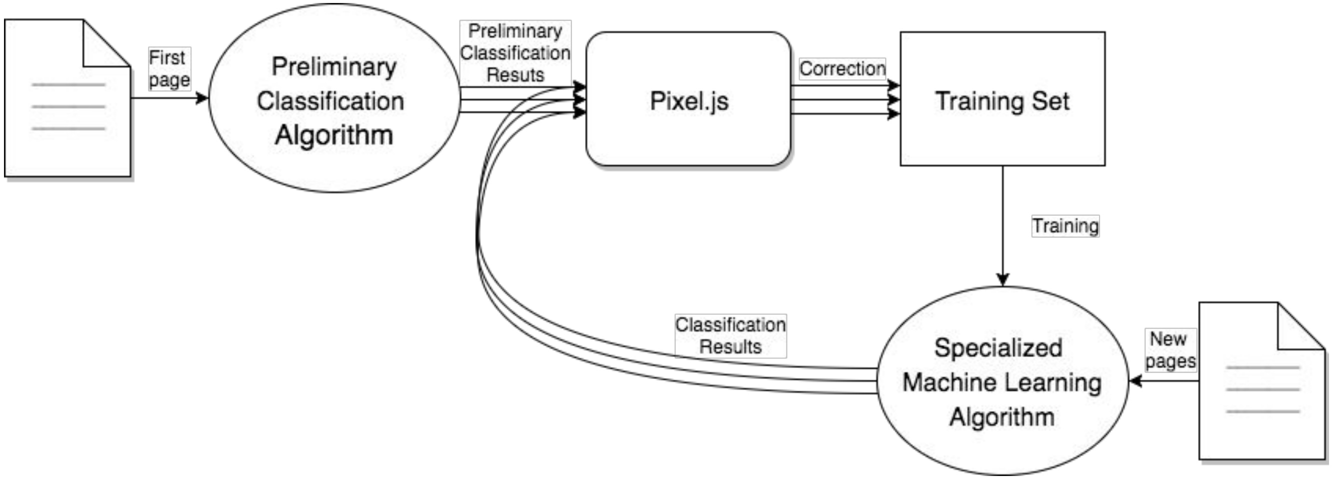
- Models need manuscript-specific training
  - Some cross-manuscript adaptation is possible but not reliable
- It is necessary to integrate the annotation tool into the workflow
- Development of Pixel.js

Human-aided music document analysis

Overview

Human-aided music document analysis

Overview



Human-aided music document analysis

Specialization

- Model re-training is costly
  - Training neural networks requires time
  - Other (adaptive) models are less accurate and slower in classification
  - Trade-off among adaptiveness, efficiency, and accuracy
  - Experiments are to be carried out
- Straightforward solution: assume asynchronization
  - User may do other duties while the machine is learning

Conclusions

Summary

- Universal OMR is not feasible in the short term
- Generic workflow can be assumed
- Human-aided OMR is appealing
- Users provide guidance to the system wherever necessary
    - Correct remaining errors
    - Continuous teaching to improve future performance

Conclusions

Summary

- The document analysis stage is the first task to address in the OMR workflow
- New issues to take into account
  - Learning-driven models
  - Annotations tools
  - Processing time

Conclusions

Discussion

- Towards general and adaptive OMR workflow
- Users do not need technical knowledge to provide guidance
- Lower performance bound than manuscript-specific OMR systems

Conclusions

Discussion

- Environment for Machine Pedagogy: learning how to teach the computer
- Relevance of the user
    - Initial model selection
    - Identification of promising ground-truth data

Conclusions

Future work

- Primary goal: to reduce human effort
  - Domain adaptation techniques
  - Improving accuracy of the models
  - Make the workflow be as friendly as possible

Thank you!

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada
Canada
DDMAL DISTRIBUTED DIGITAL MUSIC ARCHIVES & LIBRARIES LAB
Schulich School of Music
École de musique Schulich
McGill
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology