# Towards Self-Learning Optical Music Recognition

## Teaching the computers to read music scores

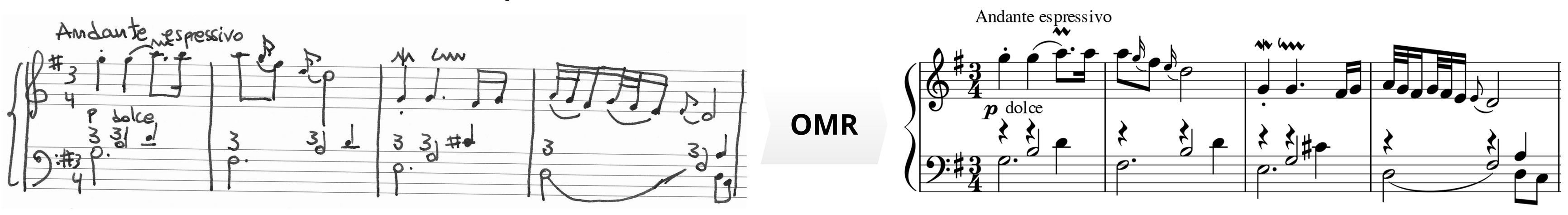

**Alexander Pacha, Horst Eidenberger**

alexander.pacha@tuwien.ac.at, horst.eidenberger@tuwien.ac.at

Interactive Media Systems, Institute of Software Technology and Interactive Systems, TU Wien, Vienna, Austria


## Introduction and Goals

Optical Music Recognition (OMR) is the branch of artificial intelligence that aims at automatically recognizing and understanding musical scores to enable a machine to comprehend music [1].




(Handwritten) music scores — Machine-readable music scores

**The goal of this research is to teach machines to read music scores**

- Humans process information in a hierarchical way, using both top-down and bottom-up mechanisms with all information available
- Mimic human behavior by applying deep learning techniques
- Evaluate, whether the machine can be trained end-to-end [7,8] on an extensive datasets, such as MUSCIMA++ [2]
- Train on realistic images of both printed and handwritten scores to gain robustness and see if it can compensate for incomplete information

Challenging samples of printed and handwritten music scores [3]

## Use-Cases

**Support musicians**

- Listen to songs one cannot play
- Educate musicians to learn from examples
- Provide missing accompanying voices to play along
- Create auditory version of handwritten drafts

**Digitize music scores to**

- Preserve ancient musical manu-scripts and make them accessible
- Support re-editing and the creation of derived works
- Enable musicological analysis
- Convert scores to different formats

## Research Questions

Can a machine mimic human behaviour to …

**Detect Scores**
Distinguish between music scores and arbitrary content

**Understand Structure**
Understand the structure of scores, and distinguish foreground from background

**Detect Music Objects**
Detect and locate music symbols in the scores

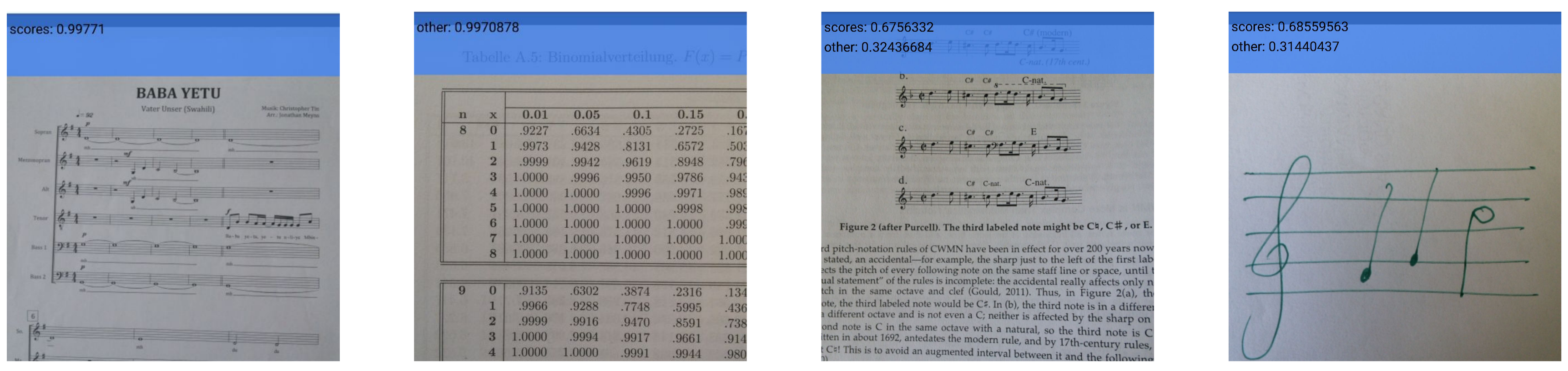
**Understand Relations**
Understand the semantics of detected symbols and their relationships with each other

**Understand Scores**
Fully comprehend the syntax and semantics of music scores

## Experiments and Results

**Experiment 1**

Distinguish music scores from pictures of other things, i.e., train a deep neural network to learn the concept of "what scores look like" on a dataset of 2000 score images and 3500 images depicting something else.

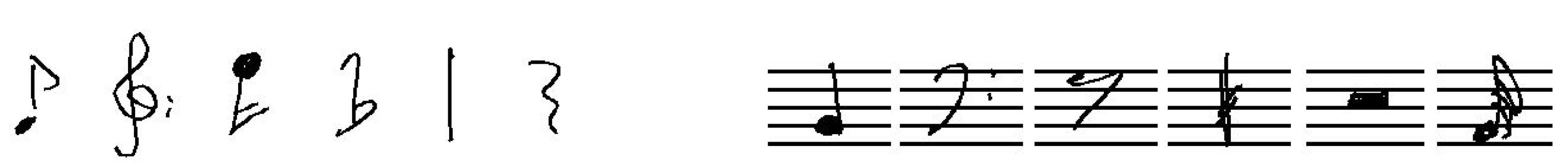


**Results**

Human-level performance (> 98% accuracy) on this simple task. Android app, that performs accurate frame-by-frame classification.

**Experiment 2**

Build a Music Symbol Classifier, i.e., train a deep neural network to perform classification of isolated music symbols on the HOMUS dataset [4] of 15200 handwritten symbols belonging to 32 different classes. The dataset is randomly split per class into 80% training-, 10% validation- and 10% test-data.



Original samples (left) and superimposed with staff lines (right) to create meaningful context

**Results**

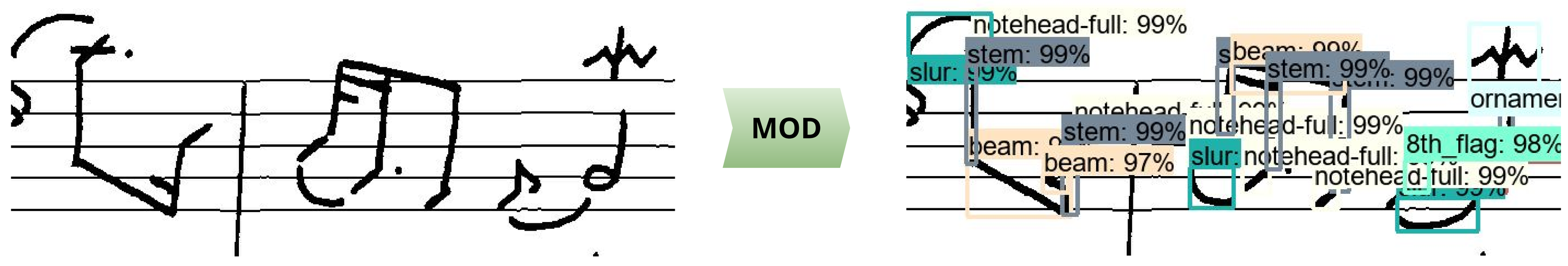
| | Baseline [4] | Pereira [5] | Calvo [6] | Our method | With staff lines |
|---|---|---|---|---|---|
| Precision | 93% | 96.01% | 97.26% | 98.02% | 97.03% |

Hard examples that were misclassified by the machine (left) and by humans (right)

## Current and Future Work

**Experiment 3**

Build fully trainable Music Object Detector by adapting state-of-the-art object detectors and train on handwritten scores of the MUSCIMA++ dataset [2].




Promising first results from trained Music Object Detector

**Future Work**

Build end-to-end trainable neural network to understand the structure of scores (number of staves and bars), relationships of detected symbols and semantic rules of sequences. Evaluation how to integrate machine learning with top-down rulesets, e.g. grammars.

**Challenges**

- Find suitable representations for storing music that allows a neural network to be trained on (MIDI, MusicXML, MEI)
- Collect and unify large enough datasets for deep learning (e.g. IMSLP)
- Identify, adapt and train suitable neural network

## References


[1] A. Rebelo, I. Fujinaga, F. Paszkievicz, A. R. Marcal, C. Guedes, J. S. Cardoso, "Optical music recognition: state-of-the-art and open issues", *International Journal of Multimedia Information Retrieval*, vol. 1, no. 3, pp. 173-190, 2012

[2] J. j. Hajič, P. Pecina, "The MUSCIMA++ Dataset for Handwritten Optical Music Recognition", *Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition*, 2017

[3] A. Fornés, A. Dutta, A. Gordo, J. Lladós, "CVC-MUSCIMA: a ground truth of handwritten music score images for writer identification and staff removal", *International Journal on Document Analysis and Recognition*, vol. 15, pp. 243-251, 2012

[4] J. Calvo-Zaragoza, J. Oncina, "Recognition of Pen-Based Music Notation: The HOMUS Dataset", *22nd International Conference on Pattern Recognition*, 2014

[5] R. M. Pereira, C. E. F. Matos, G. Braz Jr., J. D. S. de Almeida, A. C. de Paiva, "A Deep Approach for Handwritten Musical Symbols Recognition", *WebMedia '16*, 2016

[6] J. Calvo-Zaragoza, A-J. Gallego, A. Pertusa, "Recognition of Handwritten Music Symbols with Convolutional Neural Codes", *14th International Conference on Document Analysis and Recognition*, 2017

[7] J. Calvo-Zaragoza, J.J. Valero-Mas, A. Pertusa, "End-To-End Optical Music Recognition using Neural Networks", *18th International Society for Music Information Retrieval Conference*, 2017

[8] E. van der Wel, K. Ullrich, "Optical Music Recognition with Convolutional Sequence-to-Sequence Models", *18th International Society for Music Information Retrieval Conference*, 2017

[9] A. Pacha, H. Eidenberger, "Towards a Universal Music Symbol Classifier", *12th IAPR International Workshop on Graphics Recognition*, 2017