# A NEURAL APPROACH FOR FULL-PAGE OPTICAL MUSIC RECOGNITION OF MENSURAL DOCUMENTS

Francisco J. Castellanos - Jorge Calvo-Zaragoza - Jose M. Inesta

Department of Software and Computing Systems, University of Alicante, Spain (fcastellanos@dlsi.ua.es)

# 1. INTRODUCTION

- Optical Music Recognition (OMR): process to digitally encode music notation from an image without human intervention.
- The countless number of music documents encourages the development of OMR to exploit this cultural heritage.
- Typical OMR researches focus on individual tasks, thus reporting partial results.
  - There is not knowledge about the interaction between the dif-

## 2. CURRENT STATE

- End-to-end approaches allows to process a staff-region image to retrieve its music-symbol sequence.
  - However, it is not able to tackle a whole music score image with multiple staves.
  - It requires a previous staff-region retrieval.
- *Selectional Auto-Encoders* (SAE) can be used to extract individual staves. It has successfully been applied in other similar OMR tasks

ferent OMR steps within the traditional workflow.

like layout analysis.

#### 3. FRAMEWORK



\*The staff-retrieval from  $\mathcal{B}$  is performed by a connected-component analysis.

### 4. Staff-retrieval



Example from SEILS with 80.5% of IoU. Ground truth: blue Predicted: yellow.

Average histogram of staves predicted and ordered by IoU. Only some *false positives* were yielded. No *false negatives*.

#### 5. END-TO-END

Data			SEILC
Training staves	Test staves		JEIL5
Real scenario			
GT	Pred.	$16.8 \pm 3.7$	$5.2 \pm 1.4$
Pred.	Pred.	$14.8 \pm 3.6$	$4.4\pm0.5$
GT+Pred.	Pred.	$11.5 \pm 2.2$	$3.7\pm0.8$
Reference			
GT	GT	$13.2 \pm 1.1$	$4.4 \pm 1.2$
GT+Pred.	GT	$10.8 \pm 1.1$	$3.6\pm0.9$

\*Results in terms of SER (%) with average  $\pm$  std. deviation format.

### 6. CONCLUSIONS

- The staff-retrieval precision is not the most important issue.
- The end-to-end model is better trained assuming a real staff retrieval.
- Combining predicted and ground-truth staves provides the best results.
- Our approach allows transcribing reliably the music content with minimum human effort.
- We plan to experiment with more complex manuscripts like polyphonic scores in Western modern notation.

#### ACKNOWLEDGMENT

This work was supported by the Spanish Ministry HISPAMUS project TIN2017-86576-R, partially funded by the EU.

First author also acknowledges the support from "Programa I+D+i de la Generalitat Valenciana" through grant ACIF/2019/ 042.



Presented at the 21th International Society for Music Information Retrieval Conference, Montréal, Canada. 11-15 October 2020

<sup>\*</sup>SAE trained with images of  $512 \times 512$  px.