End to End Neural Optical Music Recognition of Monophonic Scores

Jorge Calvo-Zaragoza
PRHLT Research Center,
Universitat Politecnica de Valencia
Valencia, Spain
jorge@gmail.com

David Rizo
Instituto Superior de Ensenanzas Artisticas
Alicante, Spain
drizo@dlsi.ua.es

Abstract

Optical Music Recognition is a field of research that investigates how to computationally decode music notation from images. Despite the efforts made so far, there are hardly any complete solutions to the problem. In this work, we study the use of neural networks that work in an end-to-end manner. This is achieved by using a neural model that combines the capabilities of convolutional neural networks, which work on the input image, and recurrent neural networks, which deal with the sequential nature of the problem. Thanks to the use of the the so-called Connectionist Temporal Classification loss function, these models can be directly trained from input images accompanied by their corresponding transcripts into music symbol sequences. We also present the Printed Images of Music Staves (PrIMuS) dataset, containing more than 80,000 monodic single-staff real scores in common western notation, that is used to train and evaluate the neural approach. In our experiments, it is demonstrated that this formulation can be carried out successfully. Additionally, we study several considerations about the codification of the output musical sequences, the convergence and scalability of the neural models, as well as the ability of this approach to locate symbols in the input score.

Keywords: *Optical Music Recognition, end-to-end recognition, Deep Learning, music score images*

1. Introduction

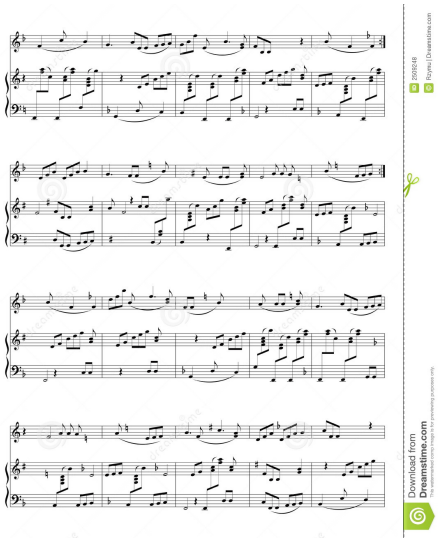
Despite the great advantages of its development, OMR is far from being totally reliable as a black box, as current optical character recognition[1] or speech recognition technologies do. For that, there is example of commercial software. Commercial software is constantly being improved by fixing specific problems from version to version. In the scientific community, there are hardly any complete approach for its solution[1][2]. Traditionally, this has been motivated because of the small sub-tasks in which the workflow can be divided. Simpler tasks such as staff-line removal,[2] symbol localization and classification, or music notation assembly, have so far represented major obstacles.

First	Pre-processing of music score images
Second	Staff-line removal
Third	Symbol classification
Fourth	Detection, classification, and interpretation

Representative summary of previous works in OMR research

2. Background

We believe that the problem to progress in OMR for CWMN lies in the complexity involved in correctly modeling the composition of musical symbols. Unlike these hand-engineered multi-stage approaches, we propose a holistic strategy in which the musical notation is learned as a whole using machine learning strategies. However, to reduce the complexity to a feasible level, we do consider a first initial stage in which the image is pre-processed to find and separate the different staves of the score. Staves are good basic units to work on, analogously to similar text recognition where a single line of text is assumed as input unit. Note that this is not a strong assumption as there are successful algorithms for isolating staves, as mentioned above. Then, the staff can be addressed as a single unit instead of considering it as a sequence of isolated elements that have to be detected and recognized independently. This also opens the possibility to boost the optical recognition by taking into account the musical context which, in spite of being extremely difficult to model entirely, can certainly help in the process. Thus, it seems interesting to tackle the OMR task over single staves in an holistic fashion, in which the expected output is directly the sequence of musical symbols present in the image.



Example of music sheet

$$\lim_{n\to\infty} 2*n+1 $$

$$\sum_{n=0}^{10} n + 2 $$

$$\int_{}^{} f(x) dx $$

References

1. Keil, K., Ward, J.A; Applications of RISM data in digital libraries and digital musicology, Int. J. Digit. Libr, 2017
2. Roland, P; In Proceedings of the First International Conference on Musical Applications Using XML, 2002