Recognizing Printed Melodies using Object Detection and a Convolution Neural Network

Report on the Lecture Machine Learning using MATLAB by Dr. Hanhe Lin

Onur Kocahan and Friedrich Hartmann

University of Konstanz

1 Introduction

The goal of Optical Music Recognition (OMR) is to reproduce a sequence of document-written musical notation by a machine. Since the language of musical notation contains hundreds of symbols, this problems becomes really complex.

Our aim is to to treat only a fragment of the musical notation language, namely the fragment of American and Europe folk music. It has a huge restriction on the alphabet, since it does not contain chords, for example, and is more appropriate of our use case. Because we want to apply OMR to an application that photocopies a tune of a song-book and transforms it later to an audio file. This can be further used to learn how to sing this tune.

Available OMR Datasets such as DeepScores (Tuggener, 2021) contains around 30 000 000 sheets of written music with close to a hundred millions of small objects. This would really blow up our project. That's why we generate our own Dataset.

We further follow the general framework to OMR that contains according to Rabela et al. (2012) the following steps.

1. image pre-processing
2. recognition of the musical symbols;
3. reconstruction of the musical information in order to build a logical description of musical notation; and
4. construction of a musical notation model to be repre-sented as a symbolic description of the musical sheet.

OMR really has difficulties on point 2 (Pacha et al., 2018). The authors applied Faster R-CNN, RetinaNet and U-Net on various datasets. The performance on DeepScores, which is not hand-written and therefore easier be trained, is not promising as the mean average precisions of the intersection over union ratios are 19.6%, 9.8%, and 24.8%, respectively.

An other approach by Metaj and Magnolfi (2021) is more promising, but has really good results around 95% only for a selection where the bounding boxes have at least an intersection over union ratio of 0.50.

That's why we invent a new approach on the object detection that is not based on a neural net.

2 Musical Notation Background

foo

3 Training Data Sampling

foo

4 Object Detection

foo

5 Image Classification

foo

6 Post Processing

foo

Bibliography

Metaj, S. and Magnolfi, F. (2021). MNR_MUSCIMA_Notes_Recognition. `https://github.com/StivenMetaj/MNR_MUSCIMA_Notes_Recognition`.

Pacha, A., jr. Hajič, J., and Calvo-Zaragoza, J. (2018). A baseline for general music object detection with deep learning. *Applied Sciences*.

Rabela, A., Fujinaga, I., Paszkiewicz, F., Marcal, A. R. S., Guedes, C., and Cardoso, J. S. (2012). Optical music recognition: state-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*.

Tuggener, L. (2021). DeepScores. `https://tuggeluk.github.io/deepscores/`.