2017 14th IAPR International Conference on Document Analysis and Recognition

Towards a Universal Music Symbol Classifier

Alexander Pacha
Institute of Software Technology and Interactive Systems
TU Wien
Vienna, Austria
alexander.pacha@tuwien.ac.at

Horst Eidenberger
Institute of Software Technology and Interactive Systems
TU Wien
Vienna, Austria
horst.eidenberger@tuwien.ac.at

***Abstract*—Optical Music Recognition (OMR) aims to recognize and understand written music scores. With the help of Deep Learning, researchers were able to significantly improve the state-of-the-art in this research area. However, Deep Learning requires a substantial amount of annotated data for supervised training. Various datasets have been collected in the past, but without a common standard that defines data formats and terminology, combining them is a challenging task. In this paper we present our approach towards unifying multiple datasets into the largest currently available body of over 90000 musical symbols that belong to 79 classes, containing both handwritten and printed music symbols. A universal music symbol classifier, trained on such a dataset using Deep Learning, can achieve an accuracy that exceeds 98%.**

***Index Terms*—Optical Music Recognition, dataset, classification, deep learning**

I. Introduction

Optical Music Recognition (OMR) is an area of document analysis that aims to automatically understand written music scores [1]. Given an image of musical scores, an OMR system attempts to recognize the content and translate it into a machine-readable format such as MusicXML.

Music symbol classification is the subtask of OMR, where isolated symbols are assigned with class labels. In this work we present the first attempt of building a universal music symbol classifier, that is capable of classifying music symbols regardless of whether they are well printed or just handwritten. To build such a classifier, we propose a data-driven approach. Therefore, we developed tools that can unify multiple datasets into a single large dataset on which the universal music symbol classifier can be trained. In our test setup, we were unifying seven datasets into a collection of over 90000 samples, belonging to 79 classes.

II. Datasets

For training a universal music symbol classifier, we tried to obtain the largest possible dataset that contains both printed and handwritten symbols. We did so by combining the following publicly available datasets:

- The Handwritten Online Musical Symbols (HOMUS) dataset [2] contains 15200 samples of isolated music symbols of 32 different classes.
- The MUSCIMA++ dataset [3] is the largest available dataset that contains detailed annotations for the underlying CVC-MUSCIMA dataset [4] of handwritten music scores. More than 55000 complete symbols can be extracted from the music symbol primitives.
- The group of Rebelo et al. collected at least three different datasets [5], containing more than 15000 printed music symbols.
- The group of Fornés et al. collected a dataset of approximately 4100 images of handwritten symbols [6] depicting accidentals and clefs.
- The Audiveris OMR dataset[1] is a small dataset of four images of scores, along with annotations of 400 printed symbols in those images.
- The Printed Music Symbols dataset[2] is a new dataset created by us, in which we collected more than 200 printed music symbols of 36 different classes.
- The OpenOMR dataset[3] is the last included dataset, that contains 500 printed music symbols of seven different classes.

The resulting dataset contains more than 74000 handwritten and more than 16000 printed symbols, with a substantial amount of inter-class variation.

III. Uniting the Datasets

A. Selecting classes and resolving ambiguities

Modern musical notation knows over 100 different symbol classes, with some classes being more present, like quarter notes or G clefs, whereas other classes are rarely used or just used for specific instruments like glissando or breath marks. Apart from selecting which classes to include into the dataset (ideally all of them), one has to deal with ambiguous class names. E.g. a quarter note may also be called quaver or a G clef is also referred to as Treble clef. To resolve this issue, a common terminology is selected and all aliases and variations are mapped to those names. The actual names are secondary, as long as the schema is clear. We follow the naming conventions of the HOMUS dataset and map all other names to their respective counterparts or to similar class names if they did not exist in the HOMUS dataset.

Besides class names, symbols themselves can be ambiguous too. Although having the same visual appearance, they might resolve to different semantics depending on the context (e.g.

[1] https://github.com/Audiveris/omr-dataset-tools
[2] https://github.com/apacha/PrintedMusicSymbolsDataset
[3] https://sourceforge.net/projects/openomr/

2379-2140/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDAR.2017.265

tie vs. slur vs. phrase mark or staccato vs. dot of a dotted note). This ambiguity can not be resolved when working with isolated symbols outside of a context which determines the class. Therefore, all ambiguous symbols are placed in a unifying super-class such as *Dot* or *Whole-Half-Rest*.

B. Joining different levels of decomposition

Some creators of OMR systems suggest to decompose music symbols into individual primitives (e.g. note-heads, stems, numbers, letters) and combine them in a later stage, whereas others choose to work with entire sets of symbols that might consist of multiple smaller units (e.g. eighth-note, 2/4-time). This decision can be made for notes, accidentals, numbers, and letters. While some primitives form a class on their own (e.g. flat or sharp), others do not (e.g. stem, flag). Datasets with different conventions are at least partially incompatible. To integrate them nevertheless, a decision has to be made for each type, whether to exclude samples, use primitive symbol classes, preprocess primitives into compound symbols or enumerate all variants of combining primitives (e.g. 2/4-time, 3/4-time, 6/8-time, ...). To lose as little data as possible when joining the mentioned datasets, we propose a mixed approach: notes only appear as compound classes which require preprocessing in some cases, time signatures are enumerated and key signatures consisting of multiple flats or sharps are excluded with only their primitives being considered.

C. Tools for the automatic unification

We have built tools that are capable of automatically downloading all datasets and processing them. As images are the input for music symbol classification in OMR, all other representations have to be processed to obtain images: Our *HOMUS image generator* allows to render textual descriptions into symbol images and the *MUSCIMA++ image generator* creates symbol images from the underlying masks. The *image extractor* for the Audiveris OMR dataset takes annotations and extracts sub-images that contain individual symbols while the *image inverter* converts the white-on-black images from the Fornés dataset to black-on-white images. Finally, the entire dataset can be obtained and split into a training-, validation-, and test-set by calling a single script, the *training dataset provider*.

IV. Building A Universal Music Classifier

A universal music classifier should be able to recognize all sorts of music symbols, regardless of whether they are handwritten or printed. Deep neural networks, especially convolutional neural networks offer a convenient, yet powerful way of solving computer vision tasks like the one at hand [7]. Therefore, we aim to build such a classifier by training a convolutional neural network on the presented dataset. Extending it to other notations is possible by adding a respective dataset. To the best of our knowledge, no such work has been done before.

V. Discussion and Conclusion

By providing tools for easily obtaining and merging multiple datasets, we believe that building a universal music symbol classifier can be reduced to the training of a suitable deep neural network. We evaluated this thesis by training various networks on the presented dataset and our preliminary results are promising with an error rate below 2% and over 98% precision and recall on an unseen test-set containing 10% of the data[4]. Our next step is to analyze the results and build a music symbol object detector on top of the classifier.

The united dataset is not perfect and currently suffers from being somewhat unbalanced with some classes having fewer than 10 instances while others have more than 1000, with the quarter note alone having almost 18000 samples. This poses a problem to any classifier that optimizes for accuracy on this dataset, as it might just learn the underlying distribution and simply ignore the classes with the fewest samples. Therefore, there is a need to gather more samples from classes with insufficient instances. Furthermore, our dataset has the following limitations:

- It currently contains modern notation symbols only.
- Some datasets have one dedicated class for non-recognizable symbols, including text fragments and dynamics. We incorporated that container class and store symbols in there, that currently do not fit our categorization as opposed to discarding them. In the next version, some symbols will be extracted from this container and put into their appropriate classes.
- Despite their prominence, beamed notes are currently underrepresented, because most underlying datasets do not contain any or decompose them into primitives that can not be joined easily.

To have the greatest possible impact, we publish all tools under a liberal MIT license along with a list of other OMR datasets at https://apacha.github.io/OMR-Datasets/.

References

[1] A. Rebelo, I. Fujinaga, F. Paszkiewicz, A. R. Marcal, C. Guedes, and 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. Calvo-Zaragoza and J. Oncina, "Recognition of pen-based music notation: The HOMUS dataset," in *2014 22nd International Conference on Pattern Recognition*, Aug 2014, pp. 3038–3043.

[3] J. j. Hajič and P. Pecina, "In search of a dataset for handwritten optical music recognition: Introducing MUSCIMA++," *arXiv preprint arXiv:1703.04824*, vol. 1, pp. 1–16, 2017.

[4] A. Fornés, A. Dutta, A. Gordo, and 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 (IJDAR)*, vol. 15, no. 3, pp. 243–251, 2012.

[5] A. Rebelo, G. Capela, and J. S. Cardoso, "Optical recognition of music symbols," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 13, no. 1, pp. 19–31, 2010.

[6] A. Fornés, J. Lladós, and G. Sánchez, *Old Handwritten Musical Symbol Classification by a Dynamic Time Warping Based Method*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 51–60.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, insight.

[4] https://github.com/apacha/MusicSymbolClassifier