AN ACCURATE OPEN-SOURCE SOLO MUSICAL INSTRUMENT CLASSIFIER

Hanna Yip
The Spence School
New York, NY

Rachel M. Bittner
Music and Audio Research Lab, New York University
New York, NY

ABSTRACT

We present an open-source solo instrument classifier trained on 18 common instrument classes. Despite using a simple Random Forest classifier and MFCC-based features, our classifier achieves 96% frame-level test accuracy by training on a dataset of almost 7000 tracks. The pre-trained model and source code are released in an open source library and the model is used in an interactive web application.

1. INTRODUCTION

Automatic solo musical instrument classification is a useful task for a number of audio applications such as automatically organizing a large collection of recordings. This research addresses the problem of labeling solo instrument recordings. While there is a great deal of previous work on solo instrument classification, it is difficult to find an open source implementation that utilizes modern software practices and performs well. Unlike proprietary systems, open source applications can be customized to meet specific user needs. Open-sourcing the code allows many independent developers to work on a project, permitting a vast peer review process that addresses security threats quickly and ensures reliability.

2. RELATED WORK

Instrument classification research in MIR has been moving towards using neural networks. One research team in particular applied deep convolutional networks on the classification of 8 instruments and achieved an accuracy of 74% [6]. Another recent work using deep neural networks showed an accuracy of around 82% for 11 classes [4]. While deep convolutional networks have the potential of avoiding handcrafted features, the goal of our research was to build a high-performing model that can be embedded in other applications.

Google recently released AudioSet [1], a large-scale dataset of manually annotated sound clips from YouTube, but has not yet released pre-trained models [3] for predicting tags on new audio examples.

[1] `https://research.google.com/audioset/ontology/musical_instrument_1.html`

3. METHODS

In this work we train a Random Forest multi-class classifier on 18 instrument classes common in popular music: piano, electric piano, synthesizer, violin, cello, acoustic guitar, clean electric guitar, distorted electric guitar, electric bass, drum set, auxiliary percussion, female singer, male singer, clarinet, flute, trumpet, saxophone, and banjo. The training data consists of solo instrument recordings from the MedleyDB [1] dataset and the Philharmonia library [2].

We use MFCCs [5] with 40 coefficients, as well as their first and second order differences as input features (120 total features per frame) with a frame length of 11.6 ms. Before computing the features, silence is removed from the audio files in the training set, and the features in the training set are normalized to have 0 mean and unit variance. The hyper-parameters of the model are tuned using a randomized search and cross-validating over the training set.

4. RESULTS

We tested the performance of our Random Forest classifier on a holdout set containing over 1000 examples, and the confusion matrix is shown in Figure 3. Overall predictions for a full example are computed by calculating the mode of the frame-level prediction array. We see that overall the classifier performs very well, with quite high probabilities along the diagonal (>0.94), and with reasonable confusions. The largest confusions occur between male vs. female singer, electric piano vs. synthesizer, and drum set vs. auxiliary percussion.

5. APPLICATIONS

Instclf, available for download on Github [3], is a web application that implements the classifier. On the website, the user can play live recording of an instrument and the computer attempts to identify the instrument present in the audio. The website is currently being used for classifying instruments in new multitracks [2] as well as recordings from Bandhub, which is an online music collaboration

[2] `http://www.philharmonia.co.uk/explore/sound_samples`
[3] `https://github.com/hmyip1/instclf`

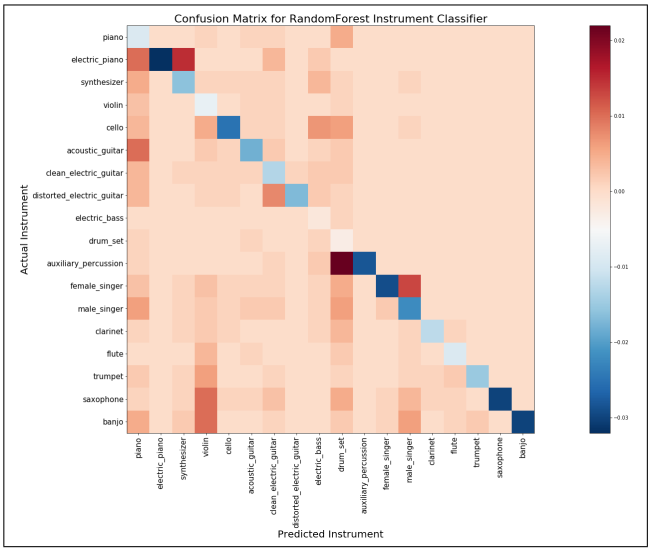

Figure 1. Confusion Matrix for the Random Forest Classifier. To view the contrast in matrix values, the diagonal of the matrix is plotted as the true diagonal - 1, giving e.g. a probability of 0.94 a value of -0.6 .

website that contains thousands of solo instrument recordings. These recordings, when automatically classified and labeled, are useful for MIR research. Eventually, the classifier can be applied as a plugin for various Digital Audio Workstations, such as Protools.

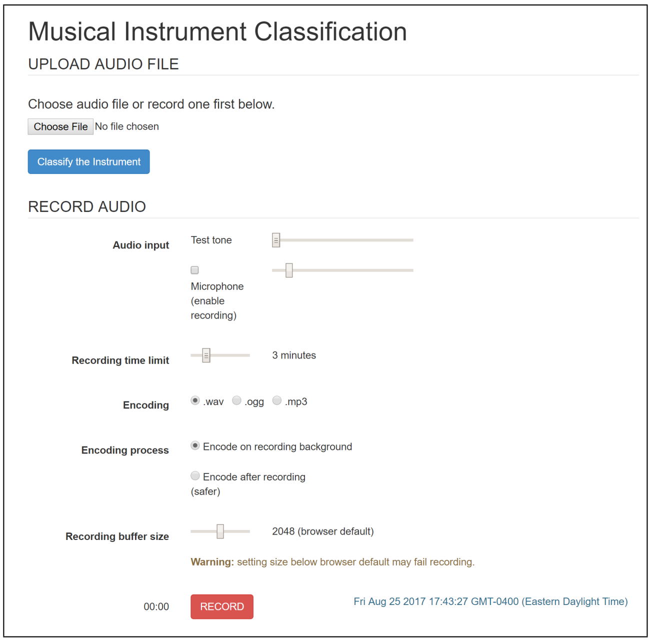

Figure 2. Classify instrument in an uploaded or recorded audio file.

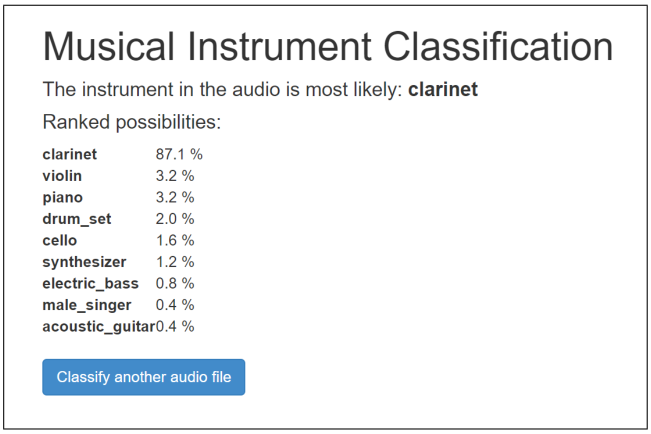

Figure 3. Example results of classification.

6. REFERENCES

[1] Rachel M Bittner, Justin Salamon, Mike Tierney, Matthias Mauch, Chris Cannam, and Juan Pablo Bello. MedleyDB: A multitrack dataset for annotation-intensive mir research. In *ISMIR '14*, 2014.

[2] Rachel M Bittner, Julia Wilkins, Hanna Yip, and Juan P. Bello. Medleydb 2.0: New data and a system for sustainable data collection. In *Proceedings of the ISMIR '16 Late Breaking and Demo Papers*, 2016.

[3] Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron Weiss, and Kevin Wilson. Cnn architectures for large-scale audio classification. 2017.

[4] Peter Li, Qian Jiyuan, and Tian Wang. Automatic instrument recognition in polyphonic music using convolutional neural networks. In *arXiv preprint, 1511.05520,*, 2015.

[5] Beth Logan. Mel frequency cepstral coefficients for music modeling. In *ISMIR '00*, 2000.

[6] Vincent Lostanlen and Carmine-Emanuele Cella. Deep convolutional networks on the pitch spiral for music instrument recognition. In *arXiv preprint arXiv:1605.06644*, 2016.