Tuning a convolutional neural network for classical artist recognition

Sri Datta Budaraju
budaraju@kth.se

Peter Mastnak
mastnak@kth.se

Nik Vaessen
vaessen@kth.se

Laura Cros Vila
lcros@kth.se

Abstract

In this report we propose a Deep Learning approach for classical artist recognition in piano music samples. Our deep neural network interprets the music in the form of spectrograms and leverages the power of Convolutions to learn the low level features and the Long Short Term Memory cells (LSTM) to learn the temporal features of the music. This enables our model to not only understand the notes but also the melody.

1 Introduction

Automatic music classification has been an interesting subject of machine learning. There has been a lot of research in music genre classification, but not so much in composer classification. Being able to classify high-dimensional, stylistically similar compositions is hard even for humans. I this project, we make use of Convolutional Neural Networks (CNNs), which are highly used for image classification as it exploits the correlation between close pixels. When it comes to audio signals, temporal locality of the samples has to be taken into account. We approach this by making use of spectrogram treated as images. Representing audio as a spectrogram allows convolutional layers to learn global structure and recurrent layers to learn temporal structure.

1.1 Related work

Many researchers have tried to approach the problem of classical composer classification. Most of them, though, base their work on MIDI files, which is a Musical Instrument Digital Interface file that contains no audio data. Some of the previous research done based on audio files are the following:

- A neural network for composer classification[5]. A neural network approach that uses CNNs and LSTM, having as input images of the short-time Fourier transform (STFT) of the audio samples.
- Music Artist Classification with Convolutional Recurrent Neural Networks[6]. This study adapts the Convolutional Recurrent Neural Network (CRNN) architecture proposed for audio tasks in prior work[1] to establish a baseline for artist classification with deep learning.
- Musical Composer Identification through Probabilistic and Feedforward Neural Networks[3]. In this paper, Maximos A. Kaliakatsos-Papakostas, Michael G. Epitropakis and Michael N. Vrahatis utilize Probabilistic Neural Networks for the construction of a "similarity matrix" of different composers and analyze the Dodecaphonic Trace Vector's ability to identify a composer through trained Feedforward Neural Networks.
- Modeling Music as Markov Chains:Composer Identification[4]. In this paper, Yi-Wen Liu proposed that music can be thought of as random processes, and music style recognition can be thought of as a system identification problem. Then, a general framework for modeling music using Markov Chains is used to identify the composers.

Table 1: This table shows the names of the composers whose audio was selected from the maestro [2] dataset. For each composer it shows the number of unique compositions present in the audio, the total duration in seconds of the audio, and the title of one of the compositions present in the audio.

composer name	num titles	duration (s)	example composition title
Alexander Scriabin	26	19268	Fragilite, Op.51
Claude Debussy	41	21507	L'isle joyeuse, L. 106
Domenico Scarlatti	29	6170	Sonata in G Major, K. 455
Franz Liszt	91	72959	Rhapsodie Espagnole, S. 254
Franz Schubert	96	128959	Sonata in A minor, D. 845 (Complete)
Frédéric Chopin	107	94389	Preludes, Book II, L 123, VI. Général Lavine - eccentric
Johann Sebastian Bach	93	49560	Prelude and Fugue in G-sharp Minor, WTC I, BWV 863
Joseph Haydn	29	17005	Sonata in D Major, Hob. XVI:24
Ludwig van Beethoven	96	99358	Sonata No. 21, Op. 53, "Waldstein", I. Allegro con brio
Robert Schumann	31	56596	Sonata No. 2 in G Minor, Op. 22
Sergei Rachmaninoff	50	23986	Etude-Tableaux Op. 39 No. 5
Wolfgang Amadeus Mozart	35	20149	Sonata K. 457

To classify composers through audio signals, there are two clear approaches: the use of probabilistic models[3][4] and the use of convolutional neural networks with some kind of recurrence[5][6][1].

2 Methods

2.1 Data

The method proposed below were tested on the maestro dataset [2], a collection of 1282 recordings of 857 unique piano compositions spanning 61 different composers. The audio originate from performances of contestants in the International Piano-e-Competition and were recorded over a time-period of 10 years. The audio samples vary in sampling rate (44.1 to 48 kHz) but are consistently stored as 16-bit PCM stereo. To have a significant amount of variation in the compositions, a subset of 12 composers was selected. Table 1 shows the composer name, the number of distinct compositions of each composer present in the data set, the total length of the audio of each composer, and an example title of a composition.

To test whether the network learned features specific to a composer instead of remembering whether a song belongs to a specific composer, the data was split into a training, validation and test set such that a particular recording of a composition is unique to one split [1]. Each audio recording was transformed into 128-length mfcc features by using a frame length of 1024 samples (~20 milliseconds) and a fast-fourier transform window of length of 2048. Each recording was then split into consecutive network inputs by taking a window of 256 features (~5 seconds). Thus, each network input has a dimension of size 128x256.

2.2 Neural Network

Our neural network is mainly comprised of three Convolutional blocks, an LSTM cell and two Dense layers which are followed by a Softmax to provide the classification. The Convolutional block is made of a Uni Directional Convolutional layer, a Batch Normalisation, a ReLU activation followed by a pooling and dropout layers as depicted in the figure 2. The intuition behind this architecture is to first extract low level features and an LSTM to take into account the temporal information. These both parts together give us the final features which are given to dense layers to learn to associate them to corresponding artists. The model is built using Keras and Tensorflow.

[1] Instructions to reproduce the particular split train/val/test split used can be found at `https://github.com/nikvaessen/pianition`

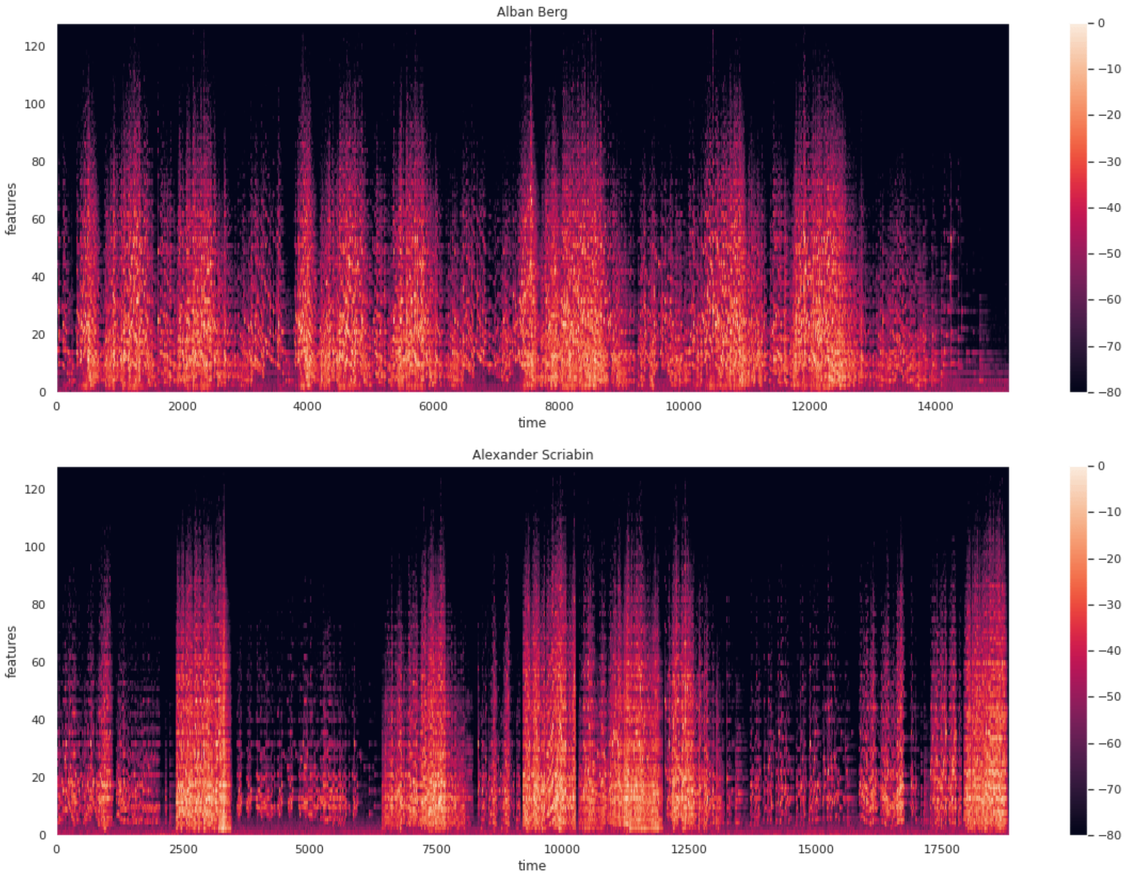

Figure 1: Spectrogram of the music samples

3 Experiments

4 Results

5 Discussion & Conclusion

References

[1] Keunwoo Choi, György Fazekas, Mark Sandler, and Kyunghyun Cho. Convolutional recurrent neural networks for music classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2392–2396. IEEE, 2017.

[2] Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. Enabling factorized piano music modeling and generation with the MAESTRO dataset. In *International Conference on Learning Representations*, 2019.

[3] Maximos A Kaliakatsos-Papakostas, Michael G Epitropakis, and Michael N Vrahatis. Musical composer identification through probabilistic and feedforward neural networks. In *European Conference on the Applications of Evolutionary Computation*, pages 411–420. Springer, 2010.

[4] Yi-Wen Liu and Eleanor Selfridge-Field. Modeling music as markov chains: Composer identification, 2002.

[5] Gianluca Micchi. A neural network for composer classification. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*, 2017.

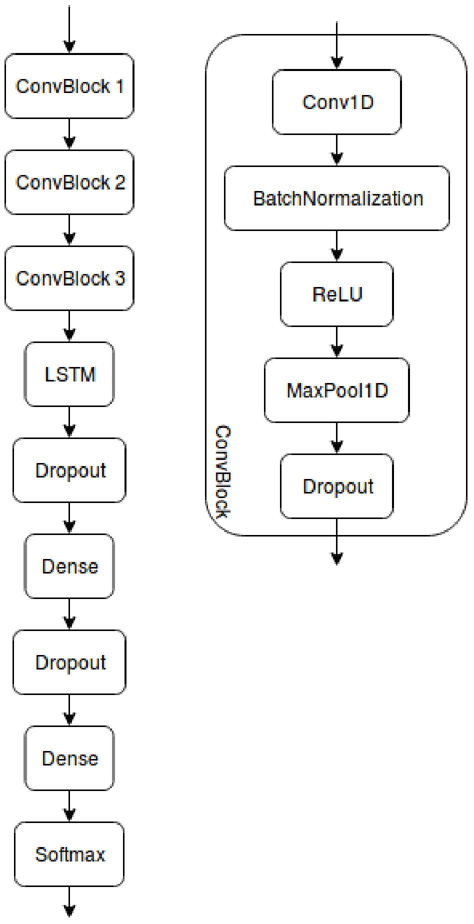

Figure 2: Overview of the model architecture

[6] Zain Nasrullah and Yue Zhao. Music artist classification with convolutional recurrent neural networks. *arXiv preprint arXiv:1901.04555*, 2019.