Music Super-Resolution

Şut George-Mihai

3rd-year undergraduate, Computer Science
Babeş-Bolyai University
georgesut@yahoo.com

Abstract—Audio super-resolution refers to the task of increasing the sampling rate of an audio signal by training a neural net to produce outputs whose sampling rate is higher by a specific factor (x2, x4, x6 etc.).

I. INTRODUCTION

In this paper, the goal is to investigate on whether a GAN can be trained with low-resolution audio data given as an input to produce super-resolution audio (i.e a reconstructed high-resolution audio signal). The point of the model is to predict the samples which are missing from the audio signal, which in this case will consist of short, downsampled pieces of music collected from a publicly available API. The project has been inspired by image super-resolution and especially by time-series super-resolution, which in essence operates with the same methods for generating the training dataset by downsampling high-resolution data and making use of a generative model to reconstruct a signal. (Kuleshov, Enam, and Ermon [2], Hetherly [1])

The process of audio super-resolution using neural nets (also called bandwidth extension) is explained in Kuleshov, Enam, and Ermon [2] which states that the goal is to reconstruct a low-resolution signal with a sample rate R_1 into a high-resolution signal with a greater sample rate R_2. The paper clarifies the concept by giving a simple example of a 4 KHz signal being upsampled through audio super-resolution to a 16 KHz signal by a factor of 4. The audio signal is encoded into a spectrogram which displays the frequencies contained in the signal and the sound intensity in decibels. For the model, a bottleneck-type architecture has been used, containing residual connections between pairs of layer b and layer $B - b + 1$. The first part of the network is responsible for downsampling data, whereas the second part upsamples it. However, the experiment conducted on the MagnaTagATune dataset in Kuleshov, Enam, and Ermon [2] shows that the limitation imposed by a reduced variety of data and lack of computing power leads to mediocre results on this bottleneck-ResNet network.

REFERENCES

[1] Jeffrey Hetherly. *Using Deep Learning to Reconstruct High-Resolution Audio*. 2017. URL: https : / / blog . insightdatascience . com / using - deep - learning - to - reconstruct-high-resolution-audio-29deee8b7ccd.

[2] Volodymyr Kuleshov, S. Zayd Enam, and Stefano Ermon. *Audio Super Resolution using Neural Networks*. 2017. arXiv: 1708.00853 [cs.SD].