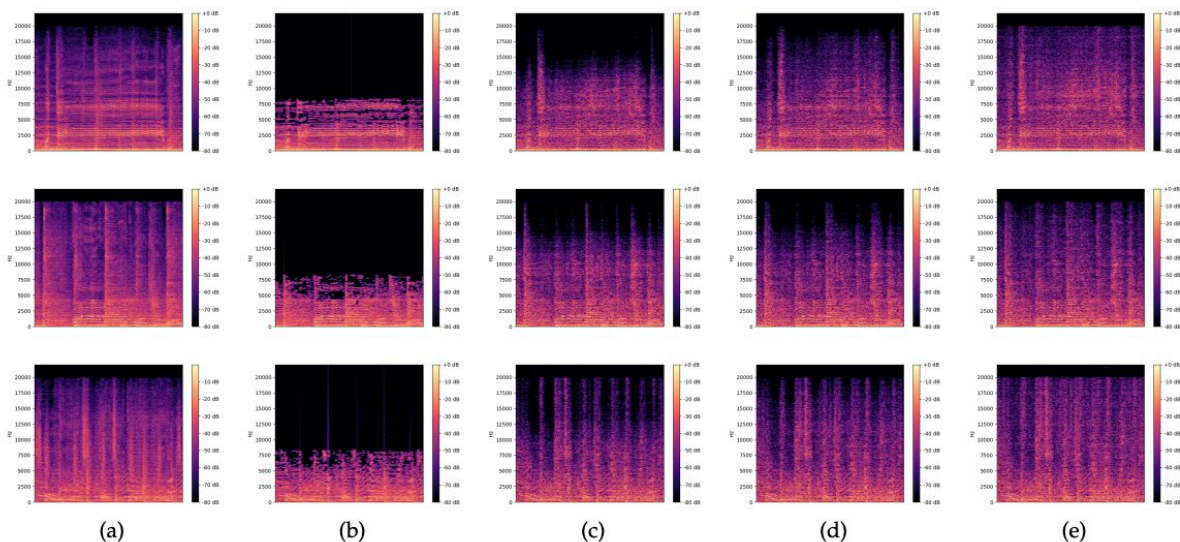


Using a GAN architecture to restore heavily compressed music files

August 31 2022, by Ingrid Fadelli



Spectrograms of (a) original audio excerpts, (b) corresponding 32kbit/s MP3 versions, and (c), (d), (e) restorations with different noise z randomly sampled from $N(0, I)$. Credit: Lattner & Nistal.

Over the past few decades, computer scientists have developed increasingly advanced technologies and tools to store large amounts of music and audio files in electronic devices. A particular milestone for music storage was the development of MP3 (i.e., MPEG-1 layer 3) technology, a technique to compress sound sequences or songs into very small files that can be easily stored and transferred between devices.

The encoding, editing and compression of media files, including PKZIP, JPEG, GIF, PNG, MP3, AAC, Cinepak and MPEG-2 files, is achieved using a set of technologies known as codecs. Codecs are compression technologies with two key components: an encoder that compresses files and a decoder that decompresses them.

There are two types of codecs, the so-called lossless and lossy codecs. During decompression, lossless codecs, such as PKZIP and PNG codecs, reproduce the exact same file as original files. Lossy compression methods, on the other hand, produce a facsimile of the original file that sounds (or looks) like the original but takes up less storage space in [electronic devices](#).

Lossy audio codecs essentially work by compressing digital audio streams, removing some data, and then decompressing them. Generally, the difference between the original and decompressed file is hard or impossible for humans to perceive.

When lossy codecs use high compression rates, however, they can introduce impairments and perceptibly alter [audio signals](#). Recently, [computer scientists](#) have been trying to overcome this limitation of lossy codecs and enhance the quality of compressed files using deep learning techniques.

Researchers at Sony Computer Science Laboratories (CSL) have recently developed a new deep learning method to enhance and restore the quality of heavily compressed songs and [audio recordings](#) (i.e., [audio files](#) that were compressed by lossy codecs with high compression rates). This method, introduced in a paper pre-published on arXiv, is based on [generative adversarial networks](#) (GANs), machine learning models in which two neural networks "compete" to make increasingly accurate or reliable predictions.

"Many works have tackled the problem of audio enhancement and compression artifact removal using [deep learning techniques](#)," Stefan Lattner and Javier Nistal wrote in their paper. "However, only a few works tackle the restoration of heavily compressed audio signals in the musical domain. In this study, we test a stochastic generator for a generative adversarial network (GAN) architecture for this task."

Like other GANs, the model created by Lattner and Nistal is comprised of two separate models, known as the "generator (G)" and the "critic (D)". The generator receives an excerpt of an MP3-compressed musical audio signal, represented through a spectrogram (i.e., a visual representation of an audio signal's spectrum frequencies).

The generator continuously learns to produce a restored version of this original signal, which is lower in size. Meanwhile, the GAN architecture's critic component learns to distinguish between the original, high-quality files and restored versions, thus spotting differences between them. Ultimately, the information gathered by the critic is used to improve the quality of the restored files, ensuring that the music or audio data present in the restored files is as faithful as possible to that in the original.

Lattner and Nistal evaluated their GAN-based architecture in a series of tests, which were aimed at determining whether their model could improve the quality of the MP3 inputs and generate compressed samples that are of higher quality and closer to an original file than those created by other baseline models for audio compression. Their results were highly promising, as they found that the model's restorations of heavily compressed MP3 files (16 kbit/s and 32 kbit/s) were typically better than the original compressed files, as they sounded better to expert human listeners. When using weaker compression rates (64 kbit/s mono), on the other hand, the team found that their model achieved slightly worse results than the baseline MP3 [compression](#) tools.

"We perform an extensive evaluation of the different experiments utilizing objective metrics and listening tests," Lattner and Nistal said. "We find that the models can improve the quality of audio signals over the MP3 versions for 16 and 32 kbit/s and that the stochastic generators are capable of generating outputs that are closer to the original signals than those of the deterministic generators."

As part of their study, the researchers also showed that their architecture could successfully generate and add realistic high-frequency content that improved the audio quality of compressed songs. The generated content included percussive elements, a singing voice producing sibilants or plosives (i.e., "s" and "t" sounds), and guitar sounds.

In the future, the model they created could help to reduce the size of MP3 music files significantly without altering their content or creating easily perceivable errors. This could have significant implications for the storage and transmission of music on both streaming apps (e.g., Spotify, Apple Music, etc.) and modern electronic devices, including smartphones, tablets and computers.

More information: Stefan Lattner et al, Stochastic Restoration of Heavily Compressed Musical Audio Using Generative Adversarial Networks, *Electronics* (2021). DOI: 10.3390/electronics10111349 , www.mdpi.com/2079-9292/10/11/1349 . On *Arxiv*: arXiv:2207.01667v1 [cs.SD], arxiv.org/abs/2207.01667

© 2022 Science X Network

Citation: Using a GAN architecture to restore heavily compressed music files (2022, August 31) retrieved 23 April 2024 from <https://techxplore.com/news/2022-08-gan-architecture-heavily-compressed-music.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.