Using Subpixel Interpolation and Deep Learning Convolution Models To Compress Domain-Specific Audio Waveforms

Nayak, Nikhil (School: Sunset High School)

Audio streaming has become more prevalent in daily life through music, phone calls, and online meetings, among other applications. However, effectively streaming high-quality audio poses a challenge in areas with low internet bandwidth. While current solutions exist, traditional codecs like MP3 or AAC use deterministic algorithms for compression, limiting their effectiveness. An Artificial Intelligence-based audio codec would allow for complex feature recognition in audio, allowing for higher fidelity compression. The Pixl audio codec, developed in this project, applies Artificial Intelligence, Convolutional Neural Networks, and Subpixel Interpolation to faithfully compress and decompress audio at ultra-low bandwidths (4 kbps), allowing for the same quality of compression as MP3 at a tenth of its size. Pixl allows for sequential encoding, acting as an interposer for an existing encoding pipeline (e.g. MP3, AAC, FLAC, OPUS) to increase compression rates without sacrificing audio fidelity. The algorithm uses Sinc interpolation to downsample the original waveform. Subsequently, sets of Convolutional and Subpixel layers are used to reconstruct the original waveform. A perceptual audio metric (CDPAM) is used to compare the original and reconstructed waveforms during training and inference to determine the effectiveness of the codec. The Pixl codec can be applied to various use cases, from high fidelity audio compression in developing nations with limited network access to applications where data storage is a premium. Currently, the Pixl codec supports compression at 44.1KHz from 2-48 kbps bitrates. If further researched, extensible versions of the Pixl codec would be created, allowing for compression for various sample rates and bitrates.

Awards Won:

Third Award of $1,000

Central Intelligence Agency: First Award: $1000 award