Day 29: Generative Model for Audio Series

WaveGAN

Designed specifically for generating raw audio waveforms

Time series generation

Unlike images (2D data - width × height), audio waveforms are 1D sequences over time

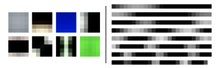


Figure 1: First eight principal components for 5x5 patches from natural images (**left**) versus those of length-25 audio slices from speech (**right**). Periodic patterns are unusual in natural images but a fundamental structure in audio.

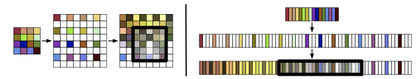
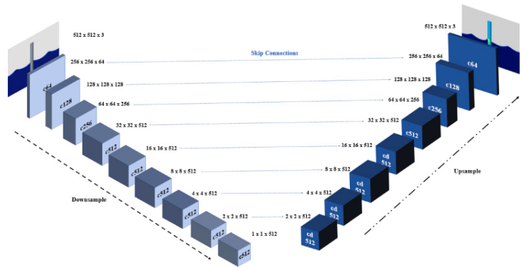


Figure 2: Depiction of the transposed convolution operation for the first layers of the DCGAN (Radford et al., 2016) (**left**) and WaveGAN (**right**) generators. DCGAN uses small (5x5), two-dimensional filters while WaveGAN uses longer (length-25), one-dimensional filters and a larger upsampling factor. Both strategies have the same number of parameters and numerical operations.

1D Convolutions instead of 2D convolutions (used for images). Because audio is sequential

The generator upsamples random noise through several layers of 1D transposed convolutions to produce a waveform

The discriminator performs downsampling through multiple layers to distinguish between real and fake waveforms.



Audio Length: WaveGAN typically generates short, fixed-length audio clips

WaveGAN is particularly useful for:

- Music Generation: Generating short pieces of music or sound effects.
- Sound Design: Producing environmental sounds, such as ambient noises or effects
- Speech Synthesis: Creating short voice clips.

Project: Generating Audio

What are we doing?

- Training a pre-built WaveGAN model

Let's Build!