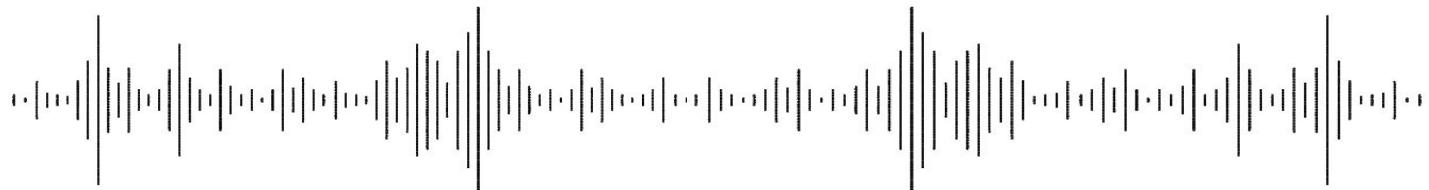




DEEP GENERATIVE DENOISING

Acoustic Guitar Processing with Autoencoders

Reid Nelson



Audio Data

- 2664 acoustic guitar samples 
- 5012 noise samples 
- 2664 mixed samples 
- 3 seconds each



Audio Data Difficulties

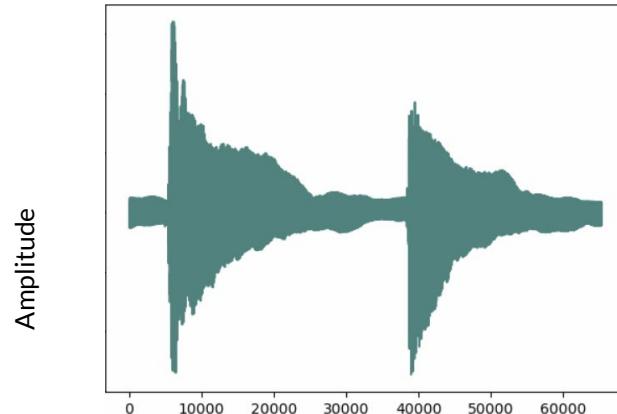
- 01 Representing a continuous signal with discrete data
- 02 Perception vs quantifiable measures
- 03 Complex and non-linear interaction (interference) between sound sources
- 04 Time-Frequency Resolution Trade-off

SPECTROGRAMS

- Frequency over time
- Derived using STFT
- Amplitude shown by color intensity like a heatmap
- Complex Valued 2D array output usually modeled with 2-3 real valued

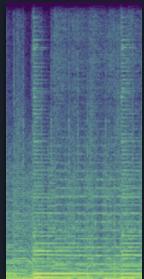
WAVE FORMS

- Amplitude over time
- Most common audio visualization
- Very difficult to discern complex frequency composition
- Single real valued 1D Array captures frequency, amplitude and phase



AUTOENCODER ARCHITECTURE

Input: Noisy Spectrogram



Encoder

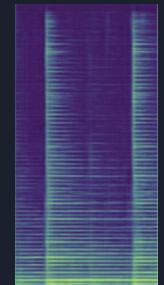
2D Convolution Layers
Stride of 2 (downsampling)

Skip Connection

Skip Connection

Latent Space
(vector length 256)

Output: Denoised Spectrogram



Decoder

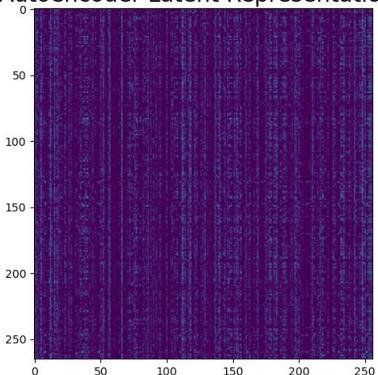
2D Transposed Convolution Layers
Stride of 2 (upsampling)

Autoencoder (U-Net)

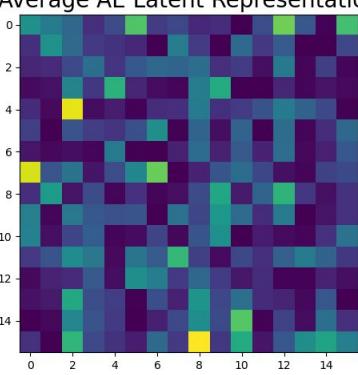
With skip connections

- Simple MSE Loss
- Can have gaps in the latent space making them less effective for purely generative tasks

Autoencoder Latent Representations



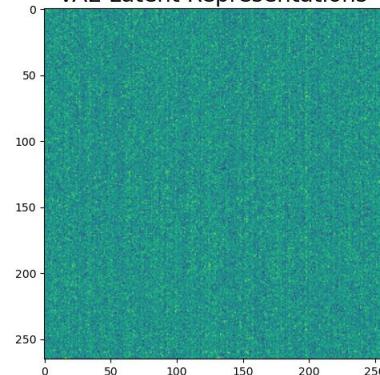
Average AE Latent Representation



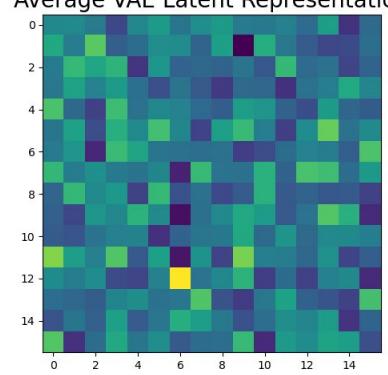
Variational Autoencoder

- Adds penalty to loss function to make the latent dimension resemble a standard normal distribution
- When trained well, they handled unseen data very well
- Can be difficult to train

VAE Latent Representations



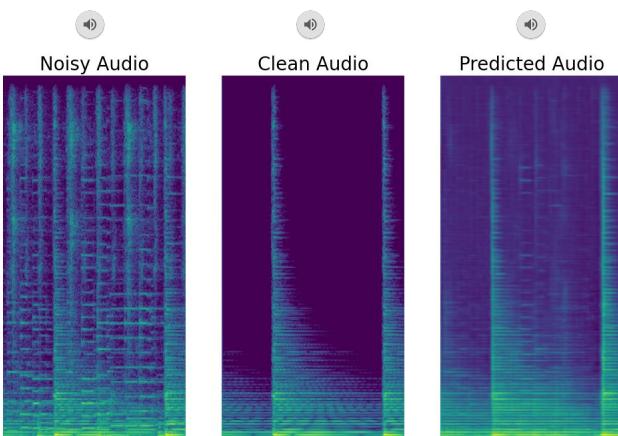
Average VAE Latent Representation



Autoencoder (U-Net)

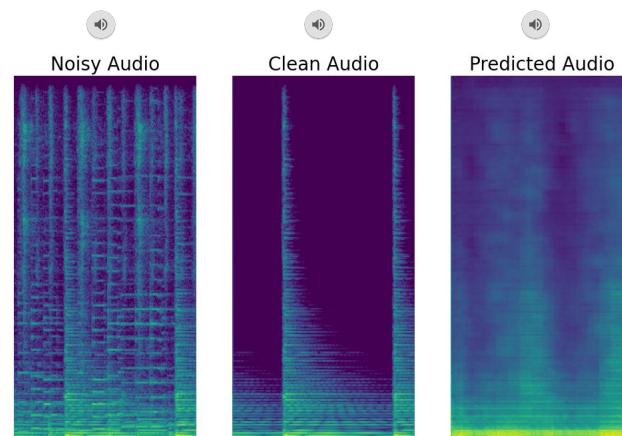
With skip connections

- Significant clarity and performance is gained with skip connections
- Higher testing mse than vae, but much better reconstructed audio



Variational Autoencoder

- Sometimes loses definition due to incentivization to estimate average signals
- In this case the testing mse was actually lower for the vae



Future Applications

VST Plugins

- Intelligent Noise Gates
- Clipped Audio Restoration
- Post-Processing + Live





Future Application

SPEAKER-SPECIFIC SPEECH ENHANCEMENT

- Hearing Aids and Cochlear Implants





Thank you!

QUESTIONS?

