Audio Enhancement Using Simple Feed Forward Networks

By Daniel Nissani

Background

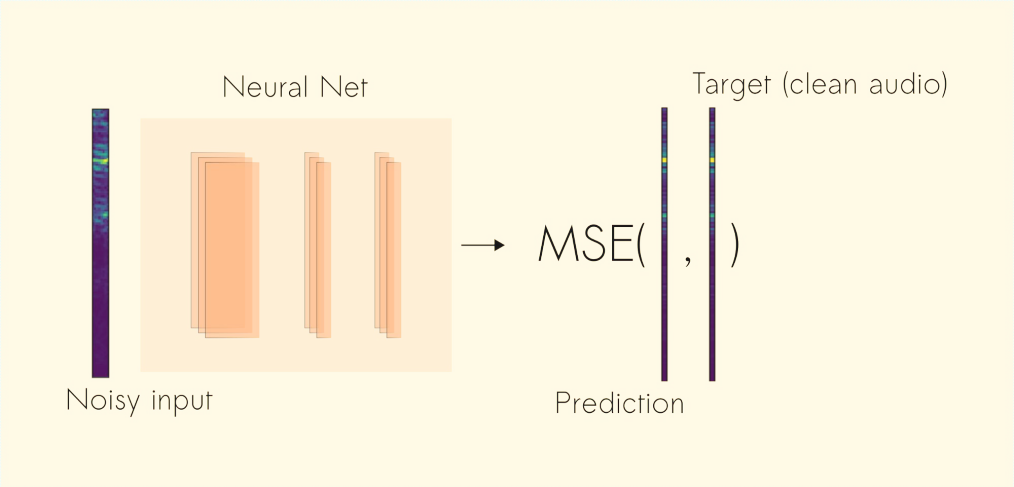


Image taken from:
https://medium.com/better-programming/how-to-build-a-deep-audio-de-noiser-using-tensorflow-2-0-79c1c1aea299

Dataset

Clean Signal

-Mozilla Common Voice

-Used only the valid samples

Noise

-Urban8k

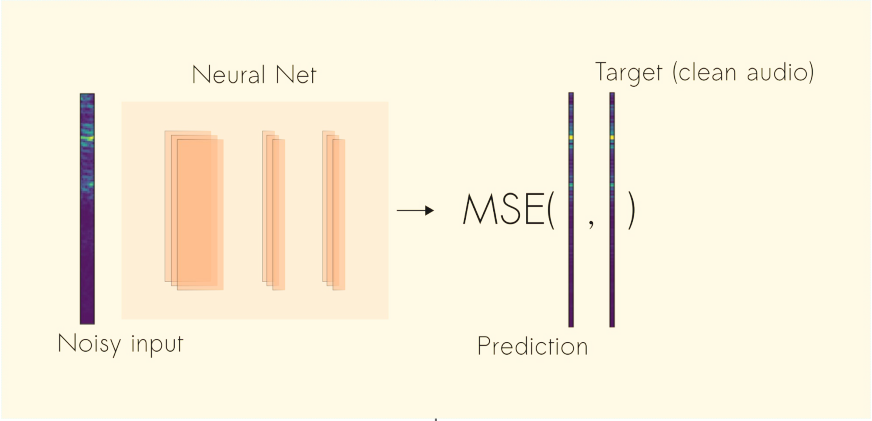
-Used only playground background noise

Model Description

Some assumptions:

1) W is the MLT window matrix
2) D is the DCT matrix
3) s is the signal
4) N = 128
5) 50% overlap

DWs



$WD^{-1}s\_hat$

Model Description

	Activation Functions
Two Layer Network	Tanh
Three Layer Network	SELU, GELU
Four Layer Network	ELU, SELU, GELU

Results

AMSE = Average MSE

Evaluation results on validation set

	AMSE noisy vs clean	AMSE output vs noisy	AMSE output vs clean
Two Layer Network	0.00426	0.00239	0.00201
Three Layer Network	0.00426	0.00215	0.00173
Four Layer Network	0.00426	0.00225	0.00163

Complications

1) Lack of GPU
2) Spectrogram non-invertible
3) Matlab vs Python implementations

Next Steps

1) Explore approximations of spectrogram inversion
2) Explore using an encoder-decoder networks with convolution

Thanks!

Any questions?

Error Analysis

Original:

Two Layer Network:

Three Layer Network:

Four Layer Network: