THE COOPER UNION
ALBERT NERKEN SCHOOL OF ENGINEERING

A Deep Partitioned Autoencoder for De-Noising Live Audio

by

Ethan Lusterman

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Engineering

September 2016

Professor Sam Keene, Advisor

THE COOPER UNION FOR THE ADVANCEMENT OF SCIENCE AND ART

ALBERT NERKEN SCHOOL OF ENGINEERING

This thesis was prepared under the direction of the Candidate's Thesis Advisor and has received approval. It was submitted to the Dean of the School of Engineering and the full Faculty, and was approved as partial fulfillment of the requirements for the degree of Master of Engineering.

Dean, School of Engineering   Date

Prof. Sam Keene, Thesis Advisor   Date

Ack

Abstract

Contents

List of Figures

Table of Nomenclature

1 Introduction

Advances in smartphone technology have led to smaller devices with more powerful audio hardware, allowing for common consumers to make higher quality recordings. However, recorded speech and music are subject to noisy conditions, often hampering intelligibility and listenability. The goal of denoising audio recordings is to improve intelligibility and perceived quality. A variety of applications of audio denoising exist, including listening to a recording of a band or an artist's live performance in a noisy crowd, or listening to a recorded conversation or speech under noisy conditions.

A common technique for denoising involves the use of deep neural networks (DNN). [PARIS] Advances in parallel graphics processing units (GPU) and in machine learning algorithms have allowed for training deeper networks faster, utilizing more hidden layers with more neurons.

Prior work in denoising audio has involved access to noise-free training data. Since common consumers do not often have access to clean audio, we seek to denoise without the use of clean audio.

In this thesis, we compare several neural network architectures and problem scenarios, ranging from data input types, level of noise, depth of network, training objectives, and more. In Chapter 2, we present background information on machine learning and neural networks as well as prior work in audio denoising. In Chapter 3, we detail all considered network architectures. In Chapter 4, we compare results from different data inputs, levels of noise, network architectures, and training objectives and discuss methods of evaluation. Finally, we make conclusions and recommendations for future work in Chapter 5.

2 Background

2.1 Machine Learning Overview

2.1.1 Classification and Regression

2.1.2 Supervised and Unsupervised Learning

2.1.3 Overfitting and Curse of Dimensionality

2.1.4 Loss functions and Regularization

2.1.5 Gradient Stuff?

2.2 Neural Networks

2.2.1 Dense Layer

2.2.2 Convolutional Layer

2.2.3 Nonlinearity Choice

3 System Description