A Convolutional Neural Network Approach to Speaker Identification

Ethan Swistak

November 27, 2020

1 Overview

Convolutional Neural Networks have achieved much success in a variety of computer vision tasks such as image classification, tracking, and segmentation. The presence of two dimensional structure in images has allowed convolutional neural networks to significantly outperform other, more connected networks while reducing the required training time given the presence of much fewer trainable parameters. However, they have been less used in the processing of other types of signal information such as speech processing. In this assignment, we will attempt to apply a ConvNet to the task of identifying a speaker in a short audio clip. While a speech signal in a single time step lacks the two dimensional structure than convolutional neural networks have been used to identify, the addition of a time dimension allows a second dimension over which this two dimensional structure can be identified. In addition, the presence of noise in the input signal can make identifying specific areas of the input signal to focus on for speaker identification challenging. Noise sources such as birds chirping in a forest, cars driving by, and background conversation can make the filtering out of ambient background noise challenging. It is for this reason that we additionally propose the addition of a self-attention layer which allows the network to focus on specific features of the audio sample to perform speaker identification.

2 Timetable

Table 1: Timetable for Deep Learning Architectures Assignment

Index	Date	Deliverable
1	November 5, 2020	Organization & Team Building
2	November 19, 2020	Brief introduction of topic (3 min)
3	December 10, 2020	Written Report of Findings(12-15 pages)
4	December 17, 2020	Oral Presentation of Topic(20 mins)
5	February 11, 2020	Written Report of Experimental Results(15-20 pages)
6	February 18, 2020	Oral Presentation of Project Results(20 min)

3 Dependencies

Table 2: List of Dependencies for Assignment

Purpose	Tool	Description
Dataset generated from YouTube with labeled speakers names	VoxCelebA Dataset	A dataset of ap
Machine Learning Framework	Tensorflow	A machine lear
IO	Scipy.io	Provides a num
Preprocessing	Scipy.signal	Allows for appl