



Audio Sample Organization Using a Hybrid Neural Network

Clark Allen (10855990@uvu.edu)

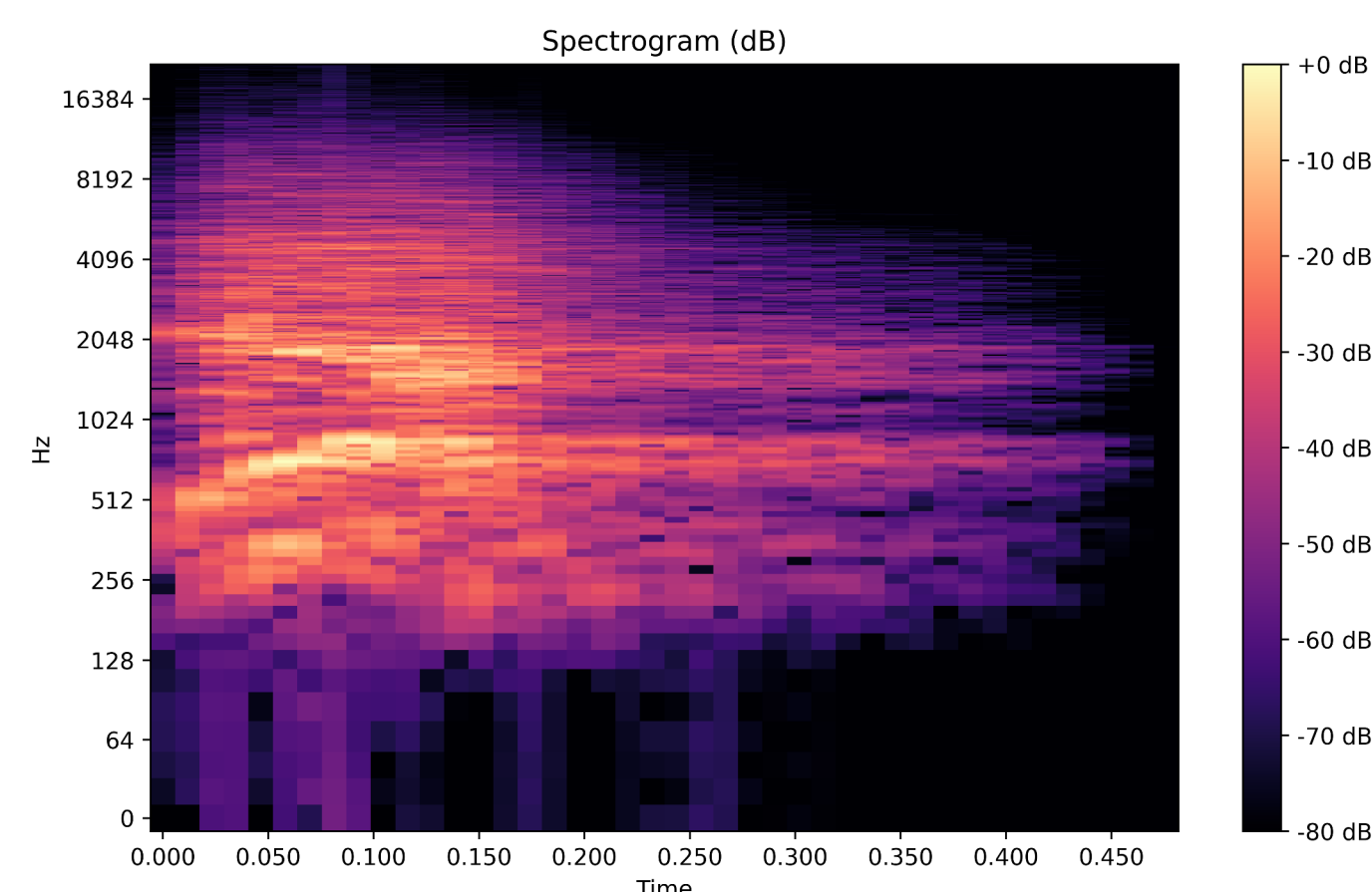
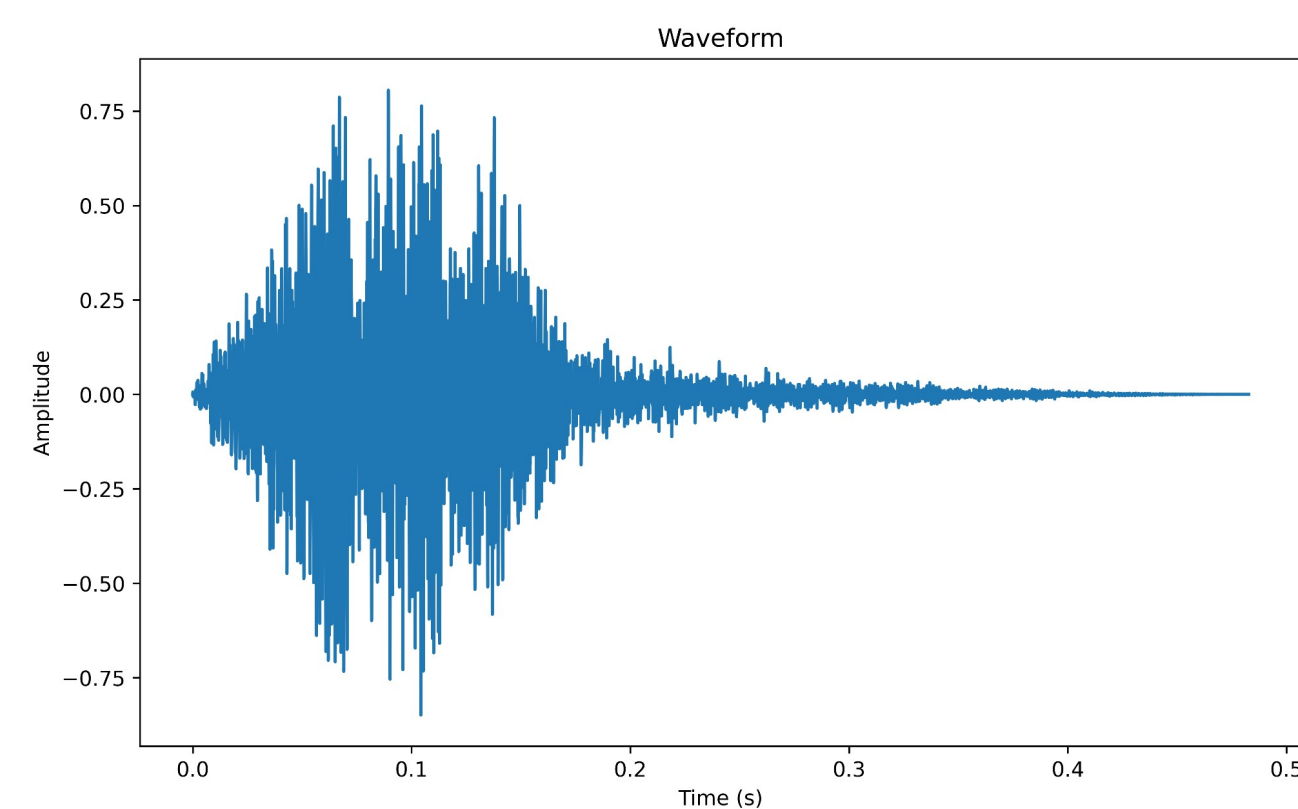
Utah Valley University

Introduction

In music production, producers often find themselves will sprawling disorganized sample libraries. Which can disrupt the creative workflow when looking for a specific sound.

Objective

Design a tool to classify and sort audio samples into an organized file structure using machine learning.

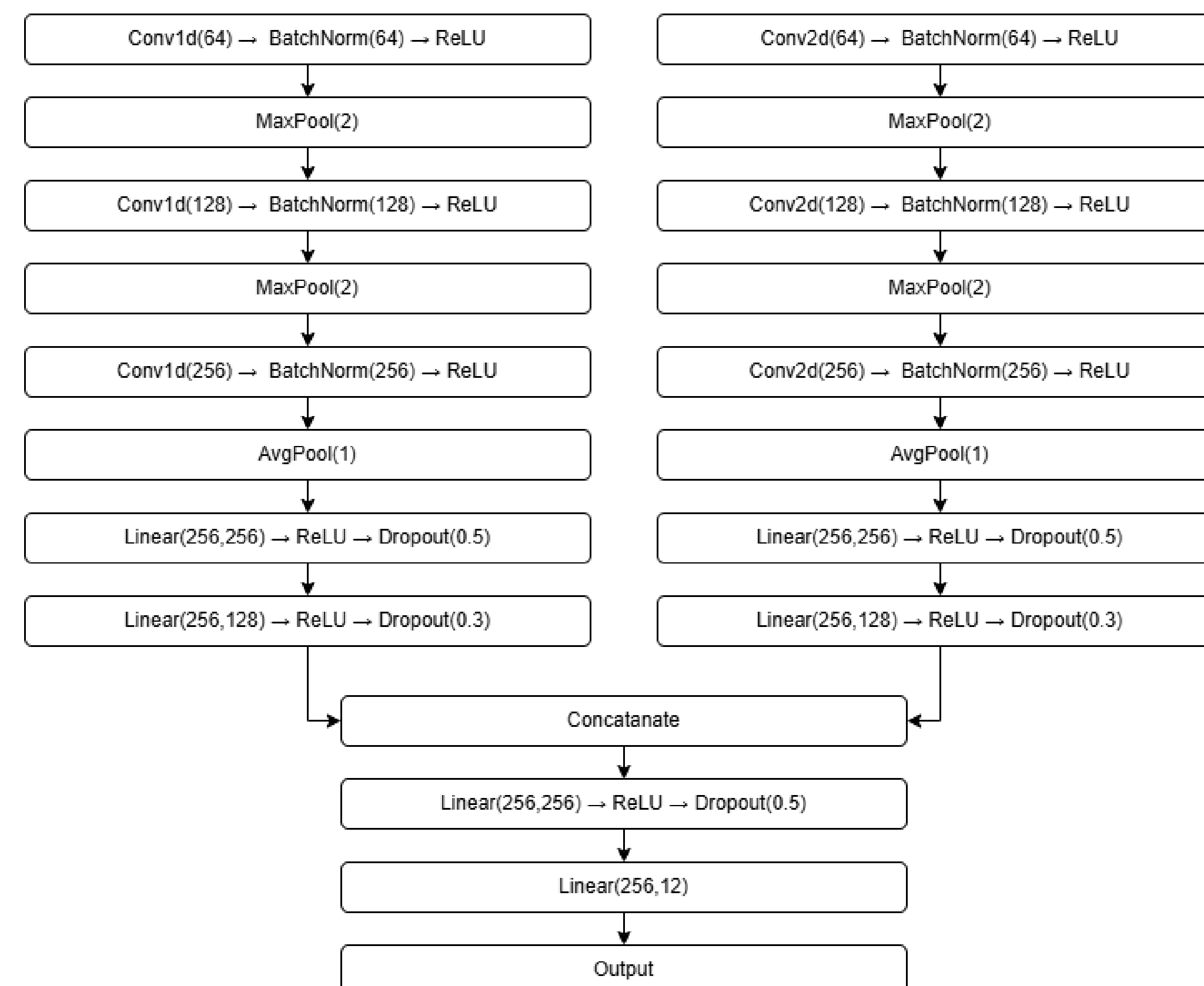


Methods

Utilizing a 2-branch hybrid convolutional neural network. The first processes the waveform with 1-dimensional convolutions, the second processes a Mel spectrogram using 2-dimensional convolution layers.

Trained on 15,000 audio samples across the following 12 classes.

- 808s
- Claps
- Crashes & Cymbals
- Drum Loops
- FX
- Closed Hats
- Open Hats
- Kicks
- Melody Loops
- Misc Percussion
- Snares
- Vocals



Results

Test accuracy of 85.70% across 6,000 samples.

Confusion Matrix

808s	482	5	0	0	2	0	0	1	7	2	2	0
claps	0	453	8	0	4	0	0	0	0	13	19	4
crashes_cymbals	0	0	443	0	3	9	45	0	0	0	0	0
drum_loops	0	7	4	440	8	1	6	6	14	6	9	0
fx	1	3	26	19	354	0	2	0	34	29	2	30
hihats_closed	0	1	6	9	0	447	24	0	0	11	1	0
hihats_open	0	0	74	0	0	9	412	0	0	1	1	1
kicks	2	0	0	0	18	0	0	480	0	1	0	0
melody_loops	6	0	1	21	30	0	0	1	428	3	1	10
percs	0	16	22	8	22	13	9	2	3	334	56	13
snares	0	15	0	1	3	0	0	0	0	28	450	3
vocals	0	1	1	10	31	0	0	0	35	2	1	418
True Label	808s	claps	crashes_cymbals	drum_loops	fx	hihats_closed	hihats_open	kicks	melody_loops	percs	snares	vocals
	808s	claps	crashes_cymbals	drum_loops	fx	hihats_closed	hihats_open	kicks	melody_loops	percs	snares	vocals

Predicted Label

Bibliography

1. XXXXXXXX
2. XXXXXXXX
3. XXXXXXXXX
4. XXXXXXXXX