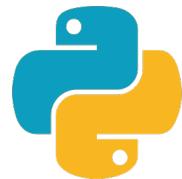


Sax | Not Sax

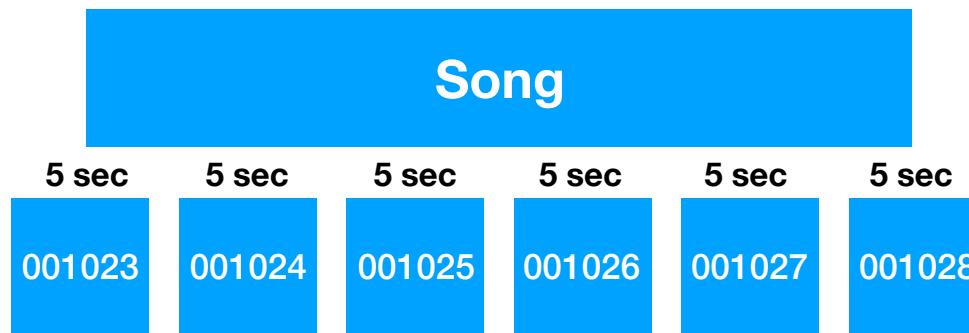
Instrument Detection in Digital Audio
Using Convolutional Neural Networks

by David Luther

Creating and Storing Audio Samples



Python
+
Librosa



How to label 24,000 audio samples?

Welcome to data labeling para X David

audio sample labeler.com

Audio Sample Labeler

Who, What, Why, etc.

My name is [David Luther](#), and I've spent most of my adult life as a professional musician. Recently, I decided to transition into the world of data science. For a current project, I'm trying to train a neural net to recognize whether a song contains saxophone, vocals, piano, or none of the three. To do this, I need a whole bunch of properly-labeled data, which is where you come in. I've chopped a number of songs into 5-second clips, and using this site, you can listen to an unlabeled sample, select the proper labels, and then submit them to the database. Fun for you and incredibly helpful for me.

Directions

Play the loaded sample, then select Foreground, Background, or None for each of the three elements. Foreground means a solo instrument or vocalist. Background means a piano in a band, a sax in a section, etc. There are samples for which the most appropriate selection for each element will be None.

For the sake of this project, Piano will include acoustic piano only. No Rhodes, Wurlitzer, other electric pianos, organ, synth, etc.

When a button is selected for each element, the Submit button will activate. Double-check your choices, then click to submit your labels to the database and load another sample. Repeat as much as you'd like

When To Skip?

Are you guessing? **This is not a test, and guessing will do more harm than good.** If you don't know or can't tell, hit Skip instead. Also do so if the sample is mostly silence, or you only hear a fraction of one of the three elements at the beginning or end of the sample.

Ready to go? Just a click away...

Get Labeling

How to Train a Neural Net?



How to Train a Neural Net?

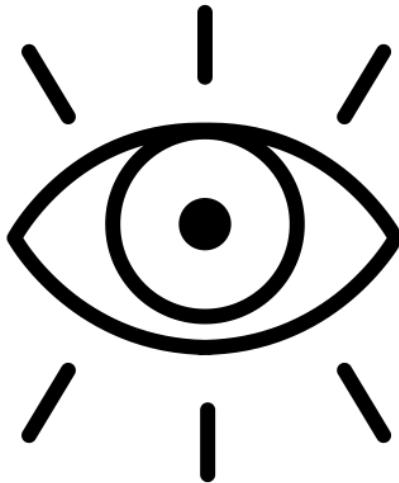
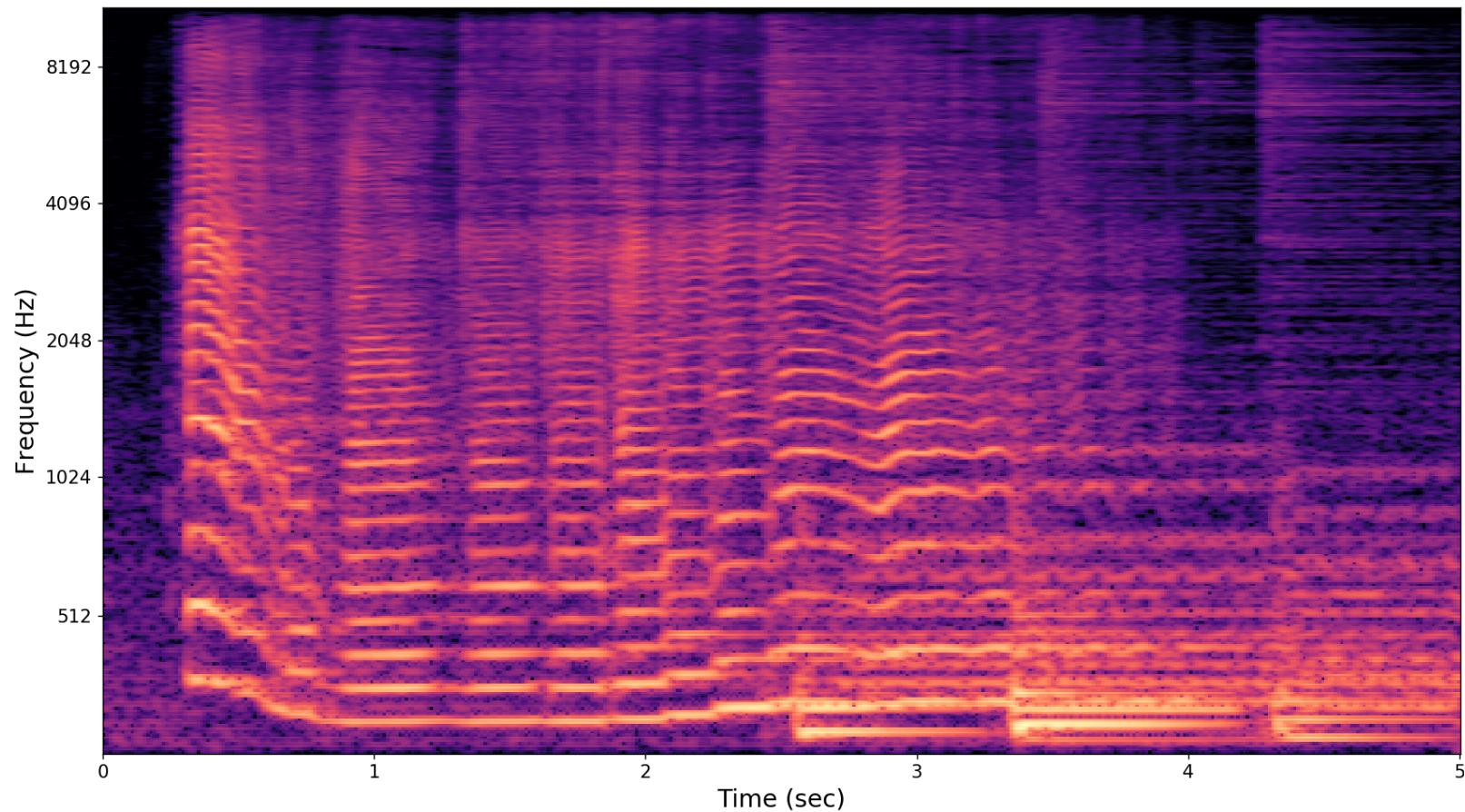


Image Recognition

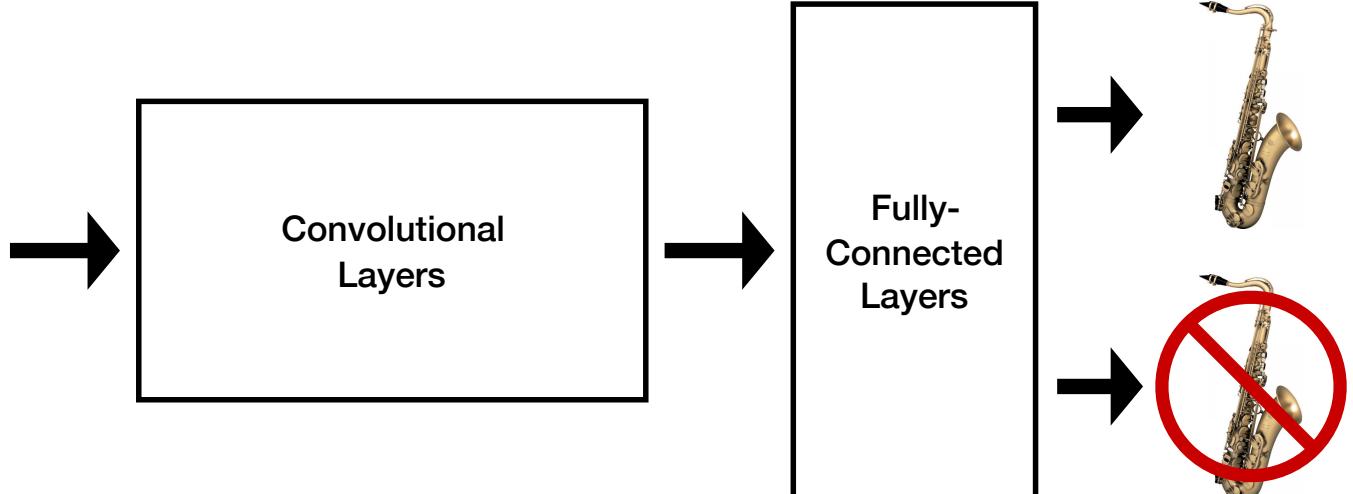
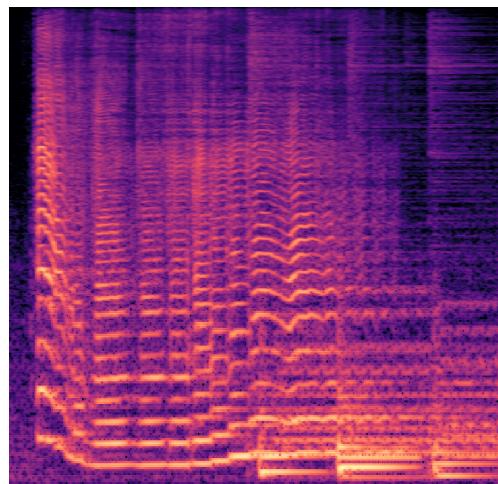
Spectrograms



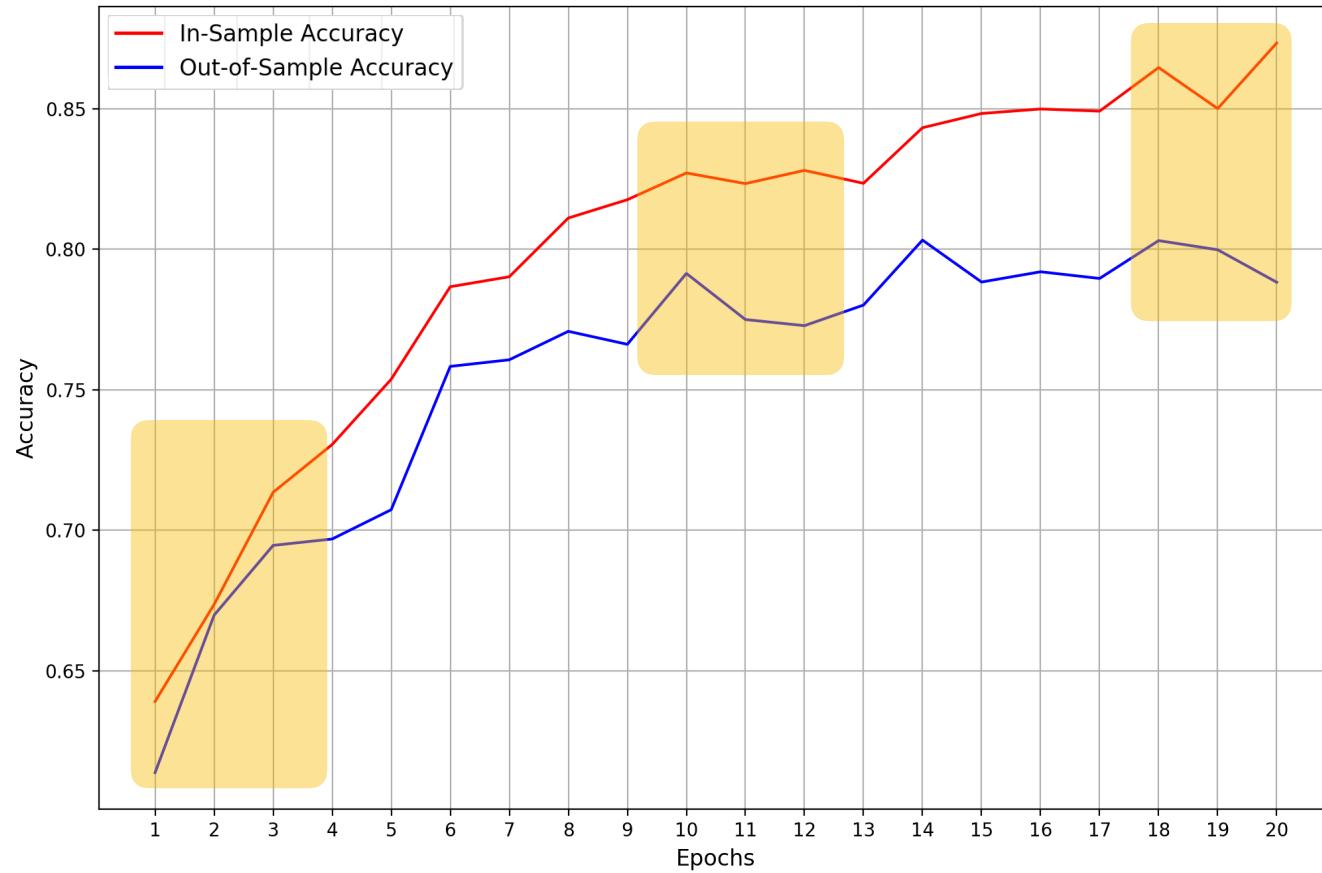
Training the CNN



Generate
spectrogram
from dataset



Model Accuracy by Training Epoch



Train/Test Scores

Score	Train	Test

Where did the model miss?

Sample 001294

Prediction: **sax**

Ground truth: **no sax**

Eddie Van Halen guitar solo

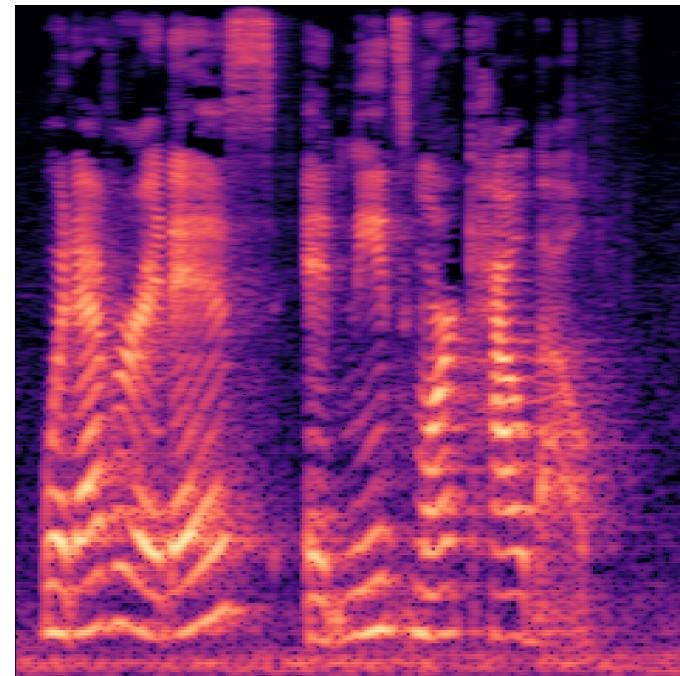
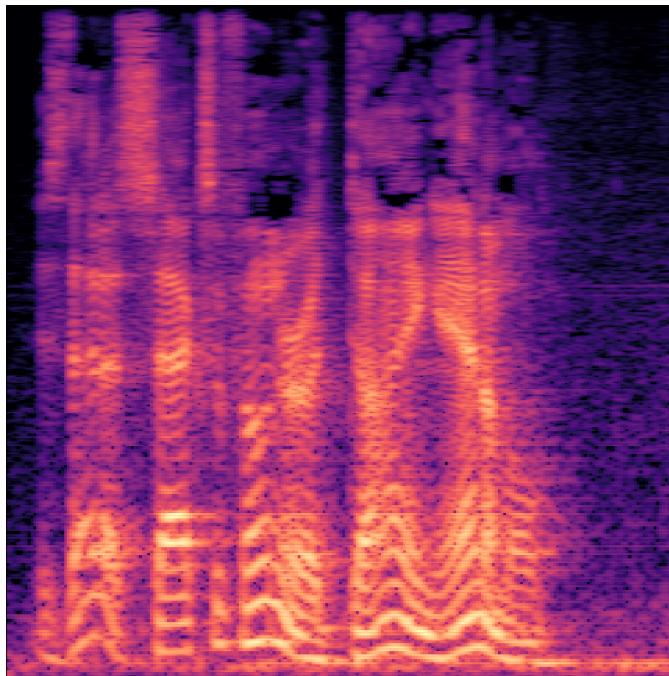
intro to “Mean Street”, *Fair Warning*

Next Steps

1. Try with more labeled data
2. Tune CNN architecture to increase accuracy, address overfitting
3. Move to GPUs accelerate optimization process

Possible Application

Voice Recognition





David Luther

✉ davidrluther@gmail.com
LinkedIn: linkedin.com/in/davidrluther
GitHub: github.com/davidluther

Thank You!

