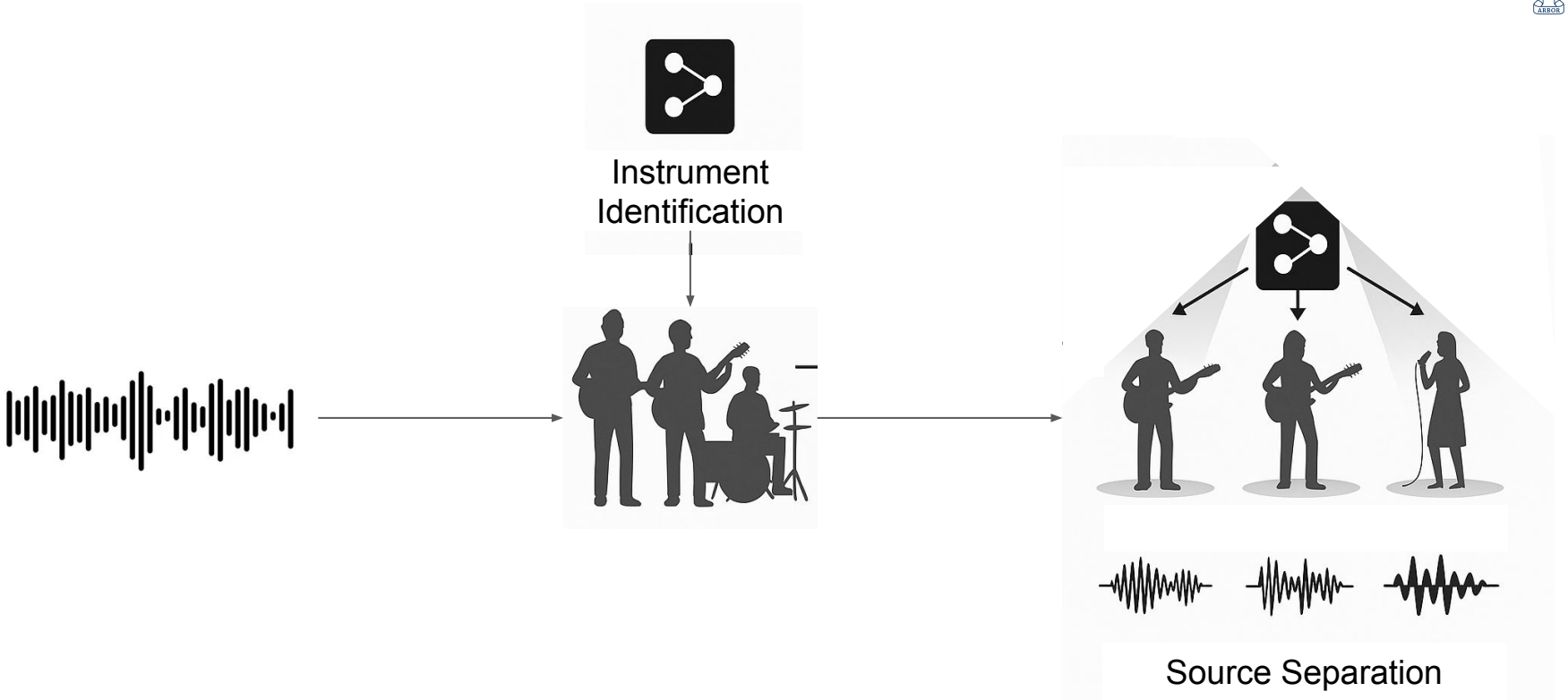
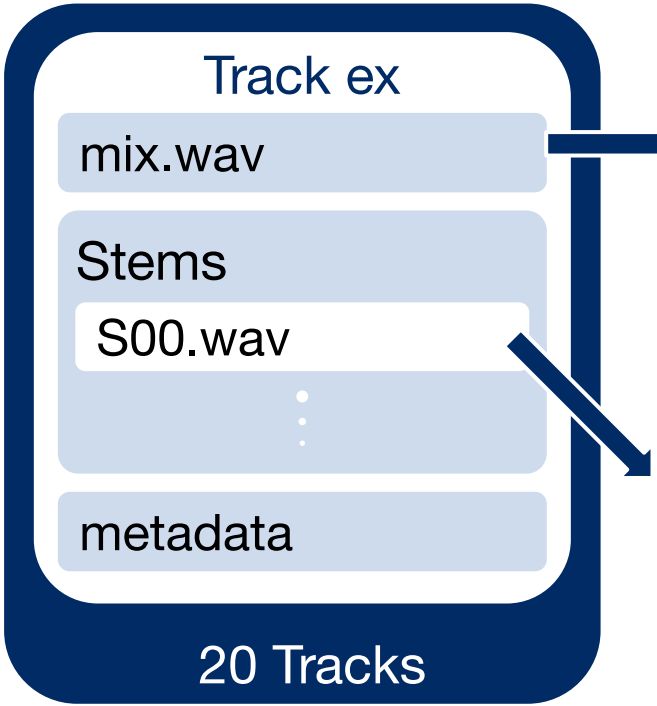
Multi-label Instrument Identification Using a CNN - InstruNET

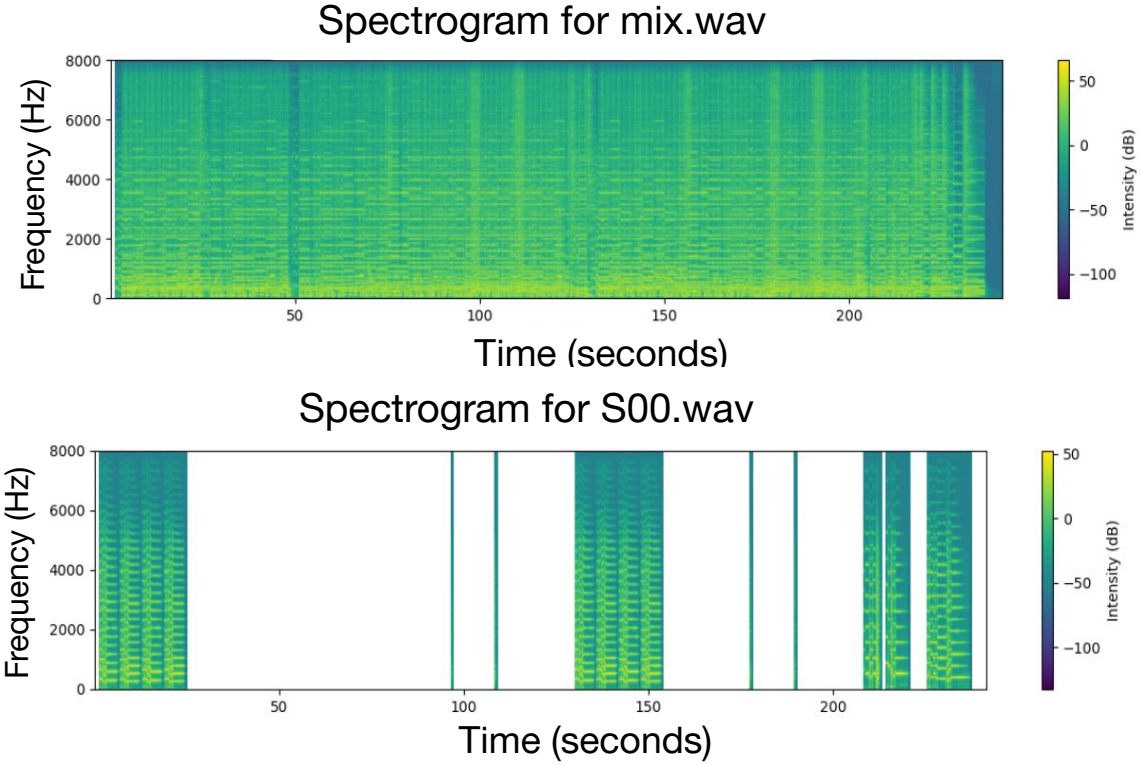
Daphne Baron, Amsal Gilani, Hanhee Lee

Motivation

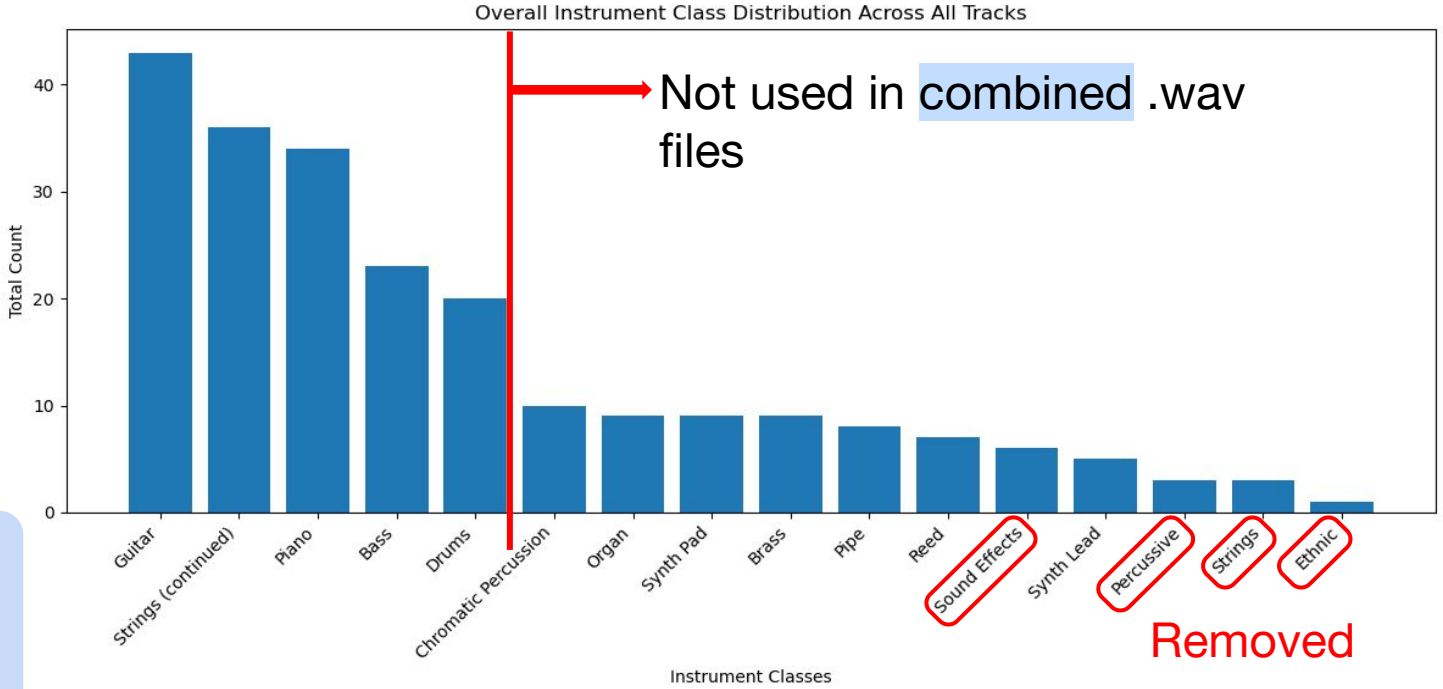


BabySlakh [1]


Track ex
mix.wav
Stems
S00.wav
metadata
20 Tracks

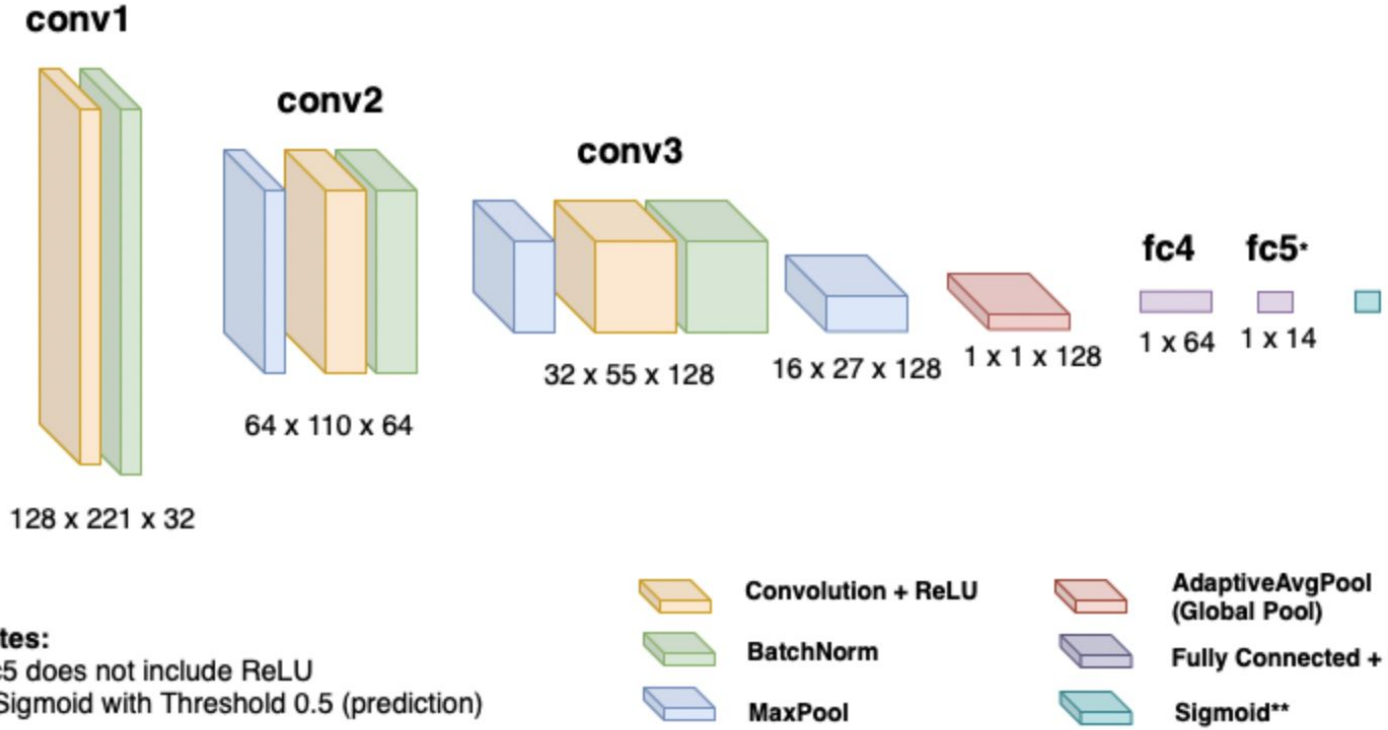

Spectrogram for mix.wav
Frequency (Hz)
8000
6000
4000
2000
0
50
100
150
200
Time (seconds)
Intensity (dB)
50
0
−50
−100
Spectrogram for S00.wav
Frequency (Hz)
8000
6000
4000
2000
0
50
100
150
200
Time (seconds)
Intensity (dB)
50
0
−50
−100

Data Exploration and Preprocessing



Instrument Specific Folders

InstruNET Architecture



Hyperparameter Optimization

Kernel Size

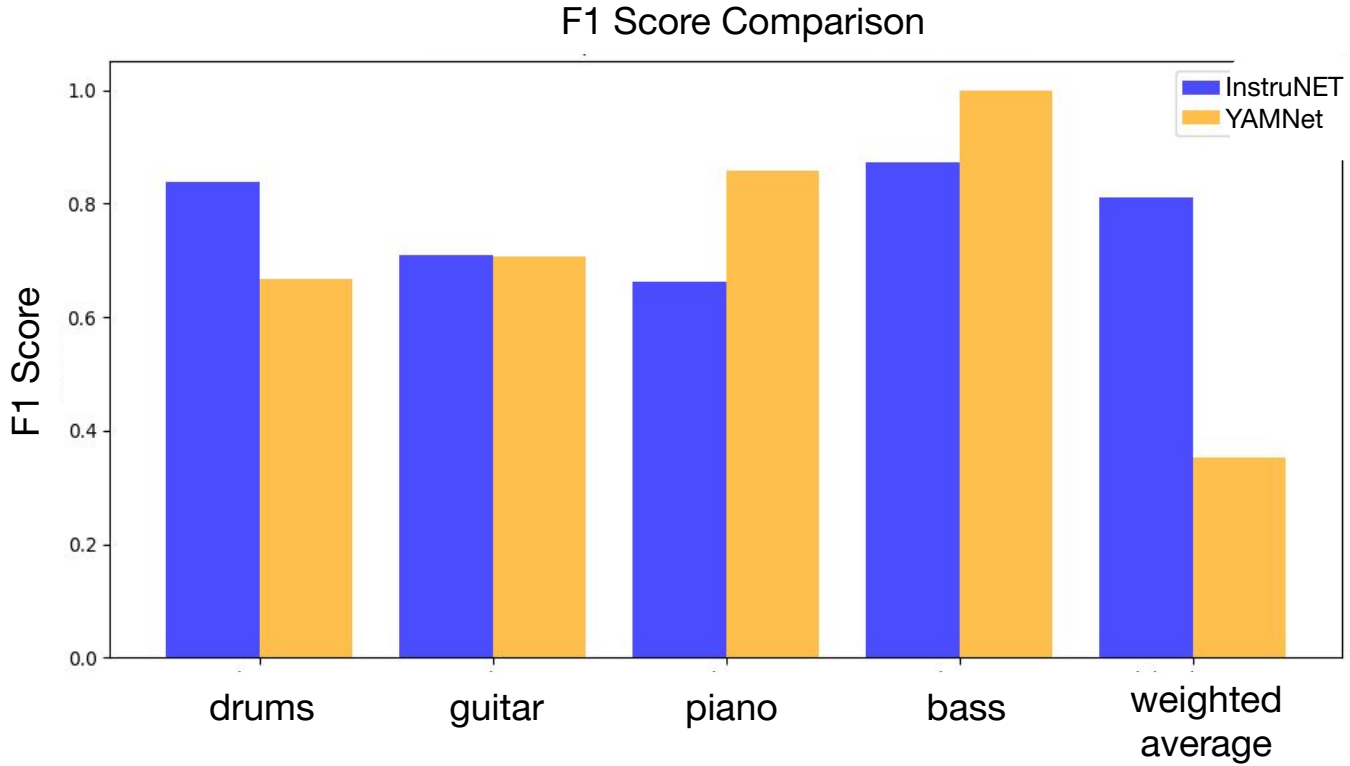
Dropout Percentage	1	3	5
0.2	59.36	59.21	65.2
0.3	56.29	**77.19**	67.11
0.4	69.59	76.46	65.79

Trained with 5% of data and 10 epochs

Results ~ 78.27% Accuracy

Comparison with YAMNet [2][3]

YAMNet ~ 73.81% accuracy:



Works Cited

[1] E. Manilow, G. Wichern, P. Seetharamanand J. Le Roux, “BabySlakh”. Zenodo, Oct. 20, 2019. Available: 10.5281/zenodo.4603870

[2] Tensorflow, “Models/research/audioset/yamnet at master · Tensorflow/models,” GitHub, https://github.com/tensorflow/models/tree/master/research/audioset/yamnet (accessed Mar. 11, 2025).

[3] “Sound classification with YAMNet: tensorflow hub,” TensorFlow, https://www.tensorflow.org/hub/tutorials/yamnet (accessed Mar. 11, 2025).