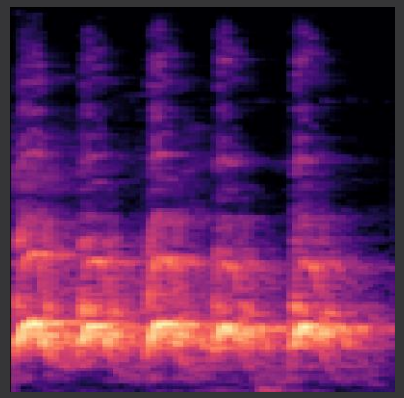
Urban Sound8k Dataset

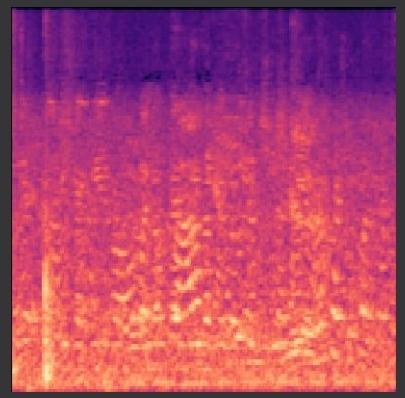
Deep Learning Project

Audio Classification

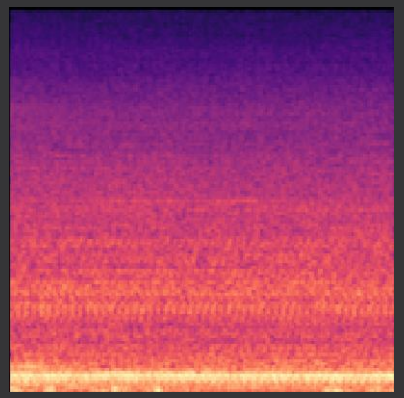
Class Overview



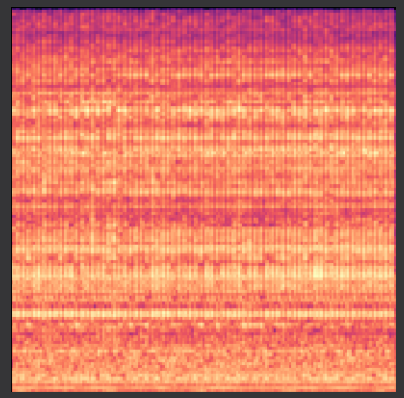
Dog Bark



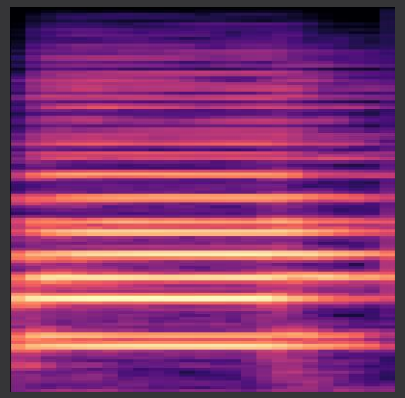
Air Conditioner



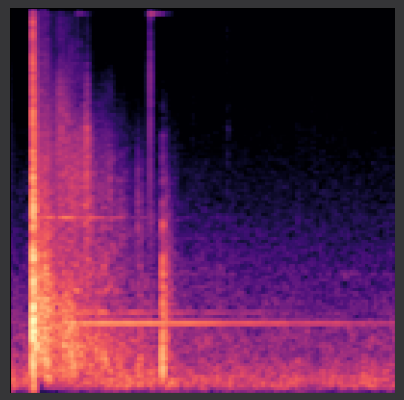
Engine Idling



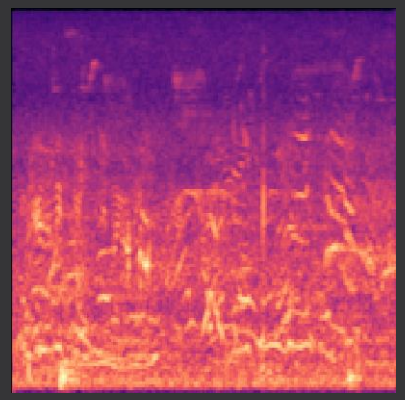
Jackhammer



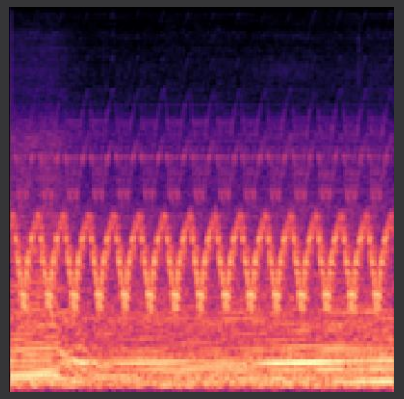
Car Horn



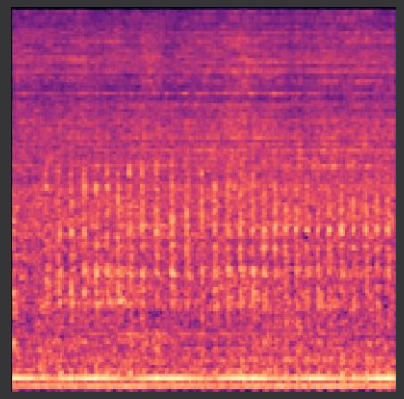
Gun Shot



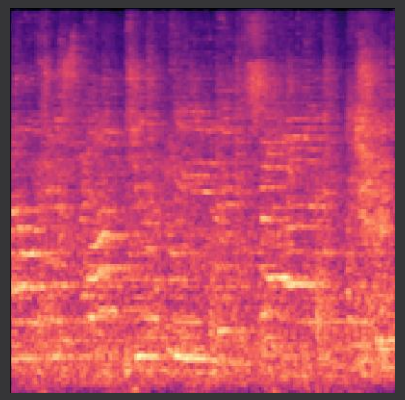
Children Playing



Siren



Drilling



Street Music

UrbanSound Pipeline

8372 Samples
Train/Test/Eval : 60/20/20

Feature Extraction

PyAudio Analysis
Mid –Term Features

Fully
Connected

Deep Audio Features
Mel-Spectrogram
Librosa

CNN

(Default 0.05 Window-Hop length * Sample Rate)

Fully Connected Architecture

Activation Function : ReLU

Drop Out = 0.2

Batch Normalization

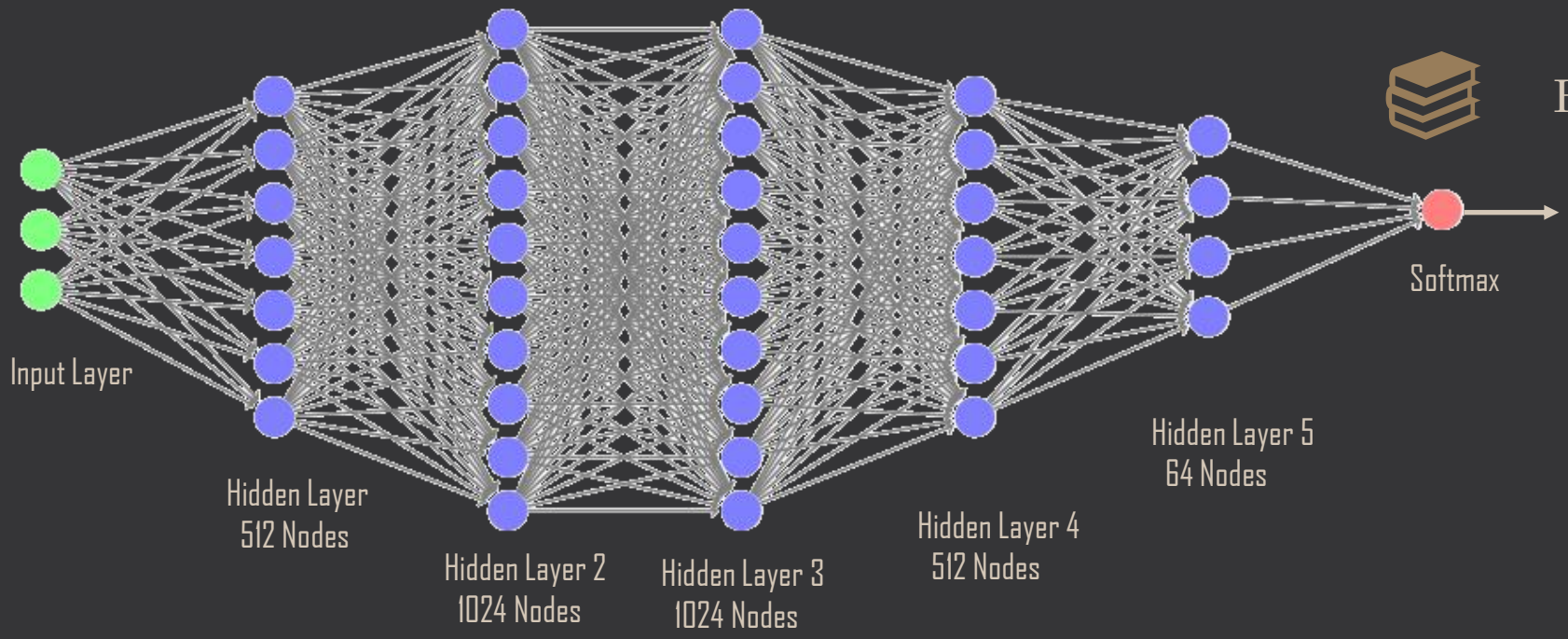
Epochs: 159

Learning Rate= 0.001

Early Stopping : Accuracy

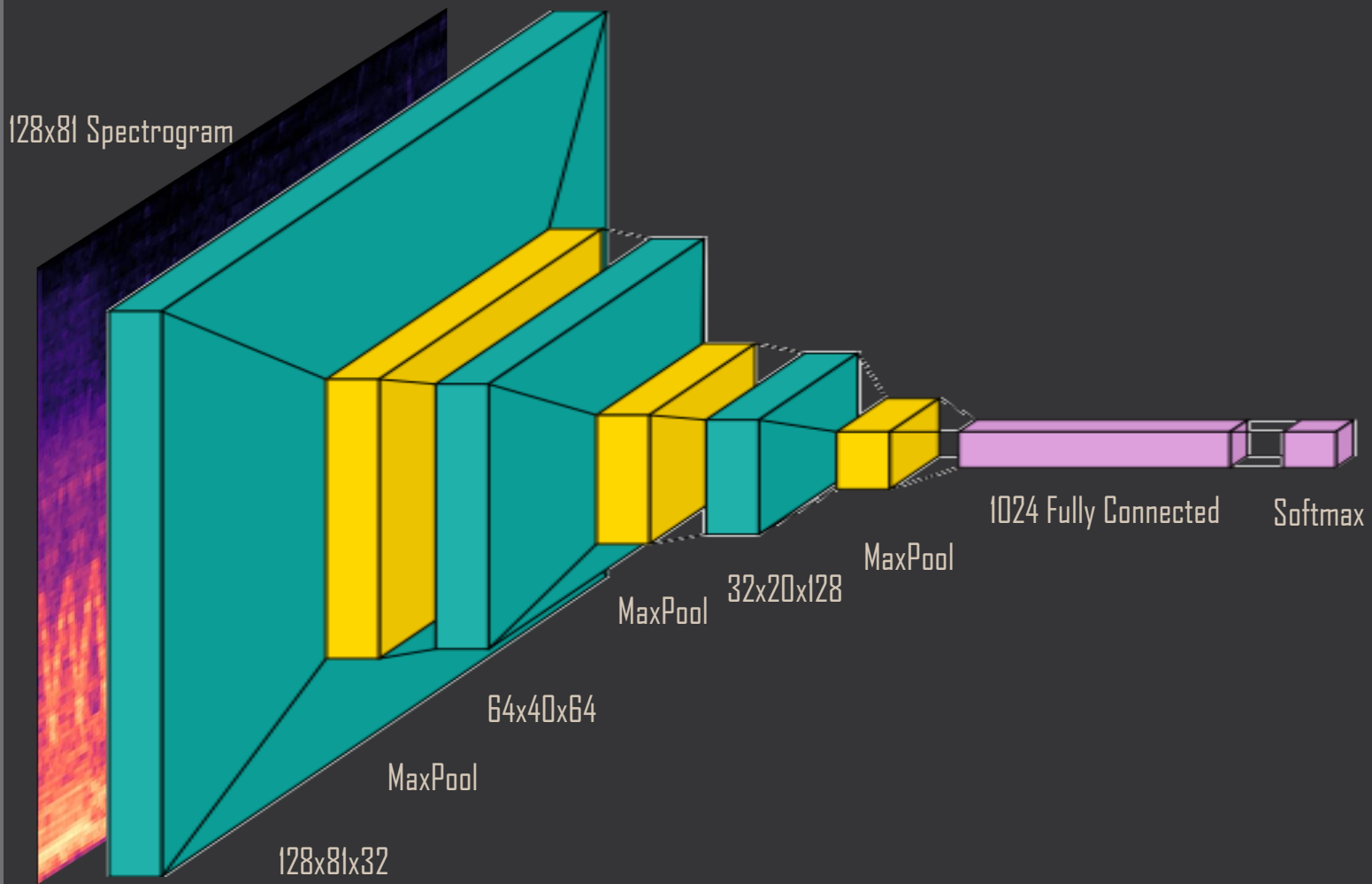
Patience : 15

Batch Size: 32



CNN Architecture

Activation Function : ReLU



Zero Padding

Learning Rate= 0.001

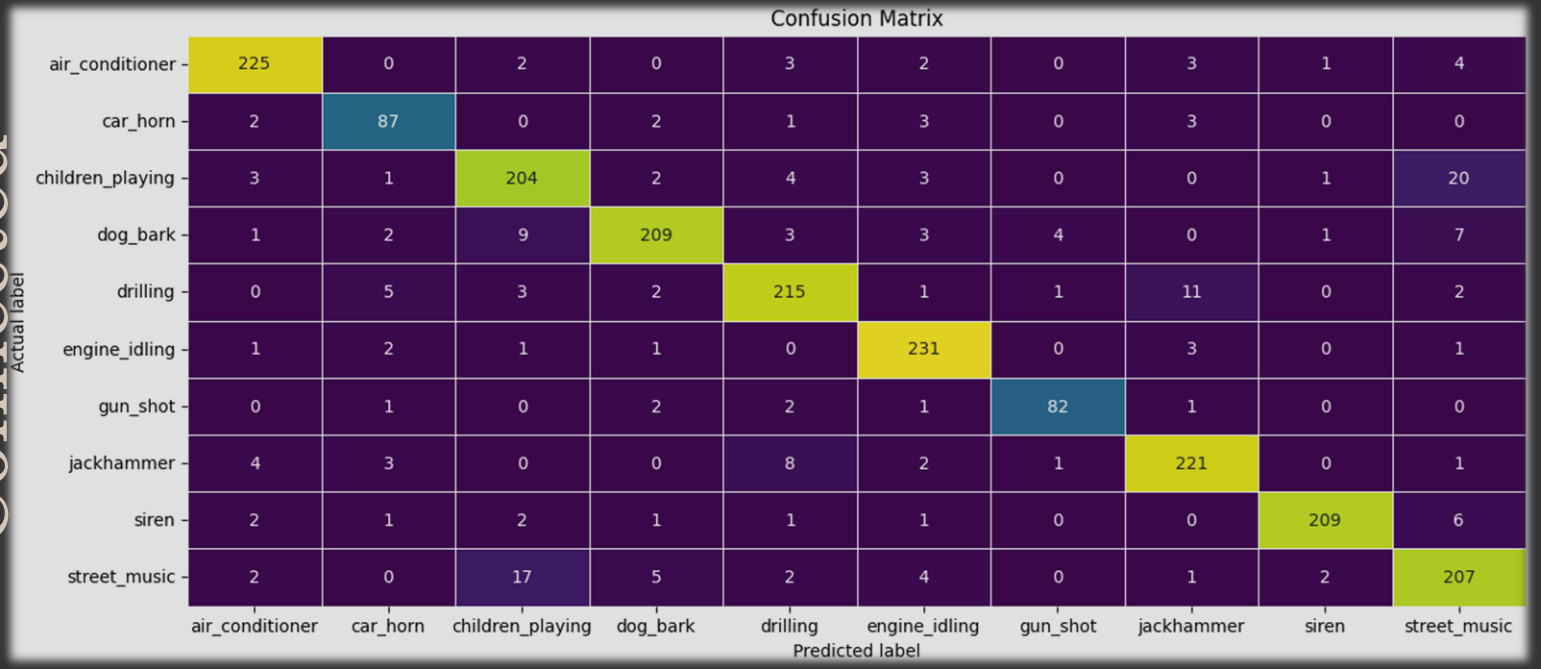
Early Stopping : Accuracy

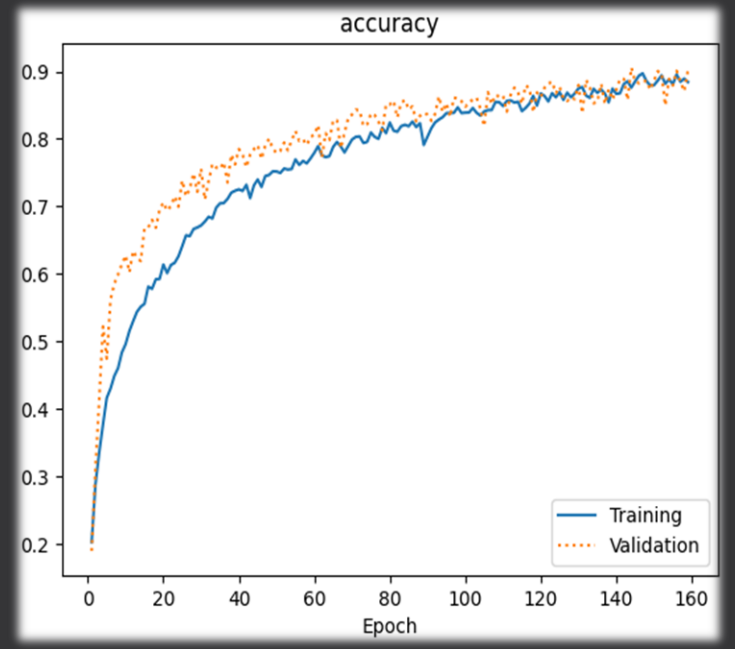
Patience : 15

Epochs: 50

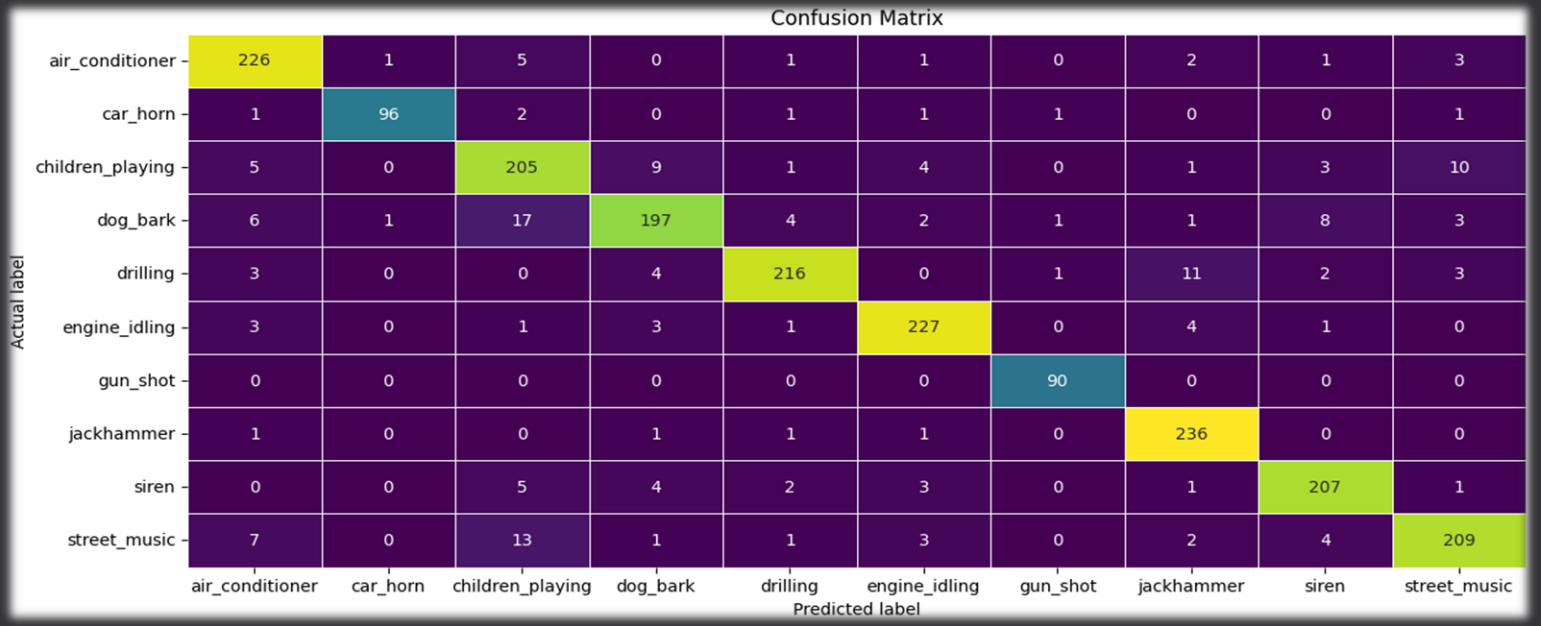
Batch Size: 32

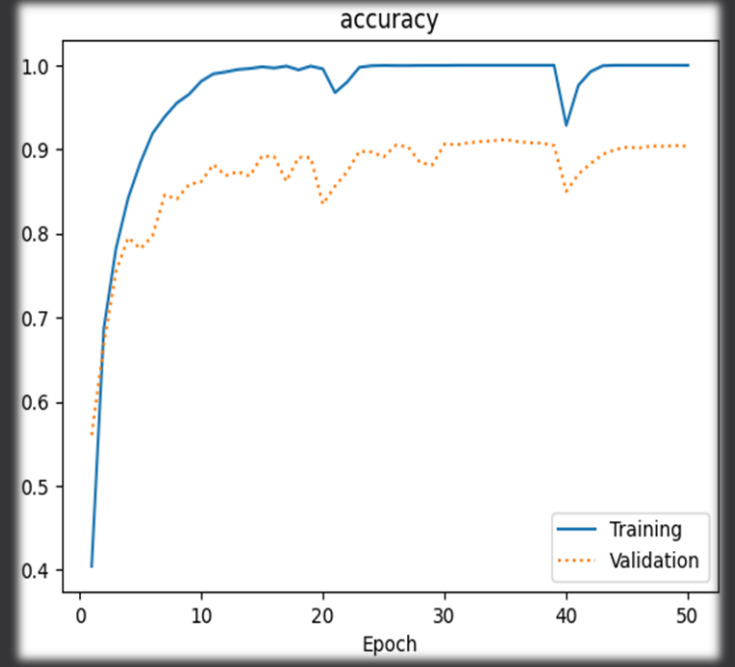
Fully Connected





CNN





AudioSet

Dataset By Google

Over 2 million Samples
+500 classes

Used balanced Eval/Train 22k

Download

- Get Real Labels
- Group into Super Labels

Ask GPT

8 Categories

~~Semantically clustering/ NLP~~

The AudioSet Problem

Long to Set Up

Big Complexity
Imbalance Labels

Labels are not Always True

Try to learn us much as possible

The Audioset CNN Architecture

Activation Function : ReLU

Drop Out = 0.2 (0.3 FN)

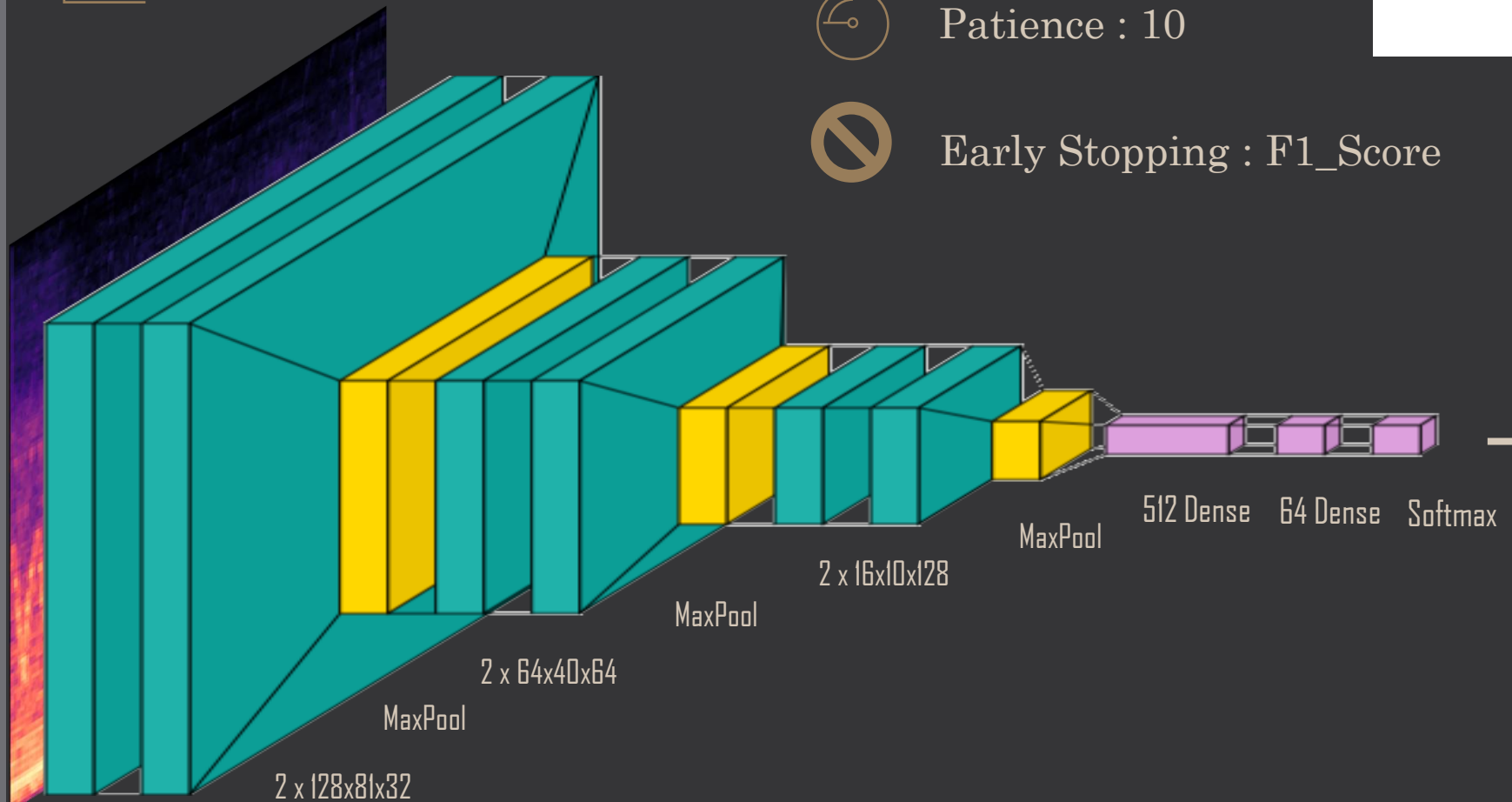
Batch Normalization

Zero Padding

Learning Rate= 0.001

Patience : 10

Early Stopping : F1_Score



Epochs: 61

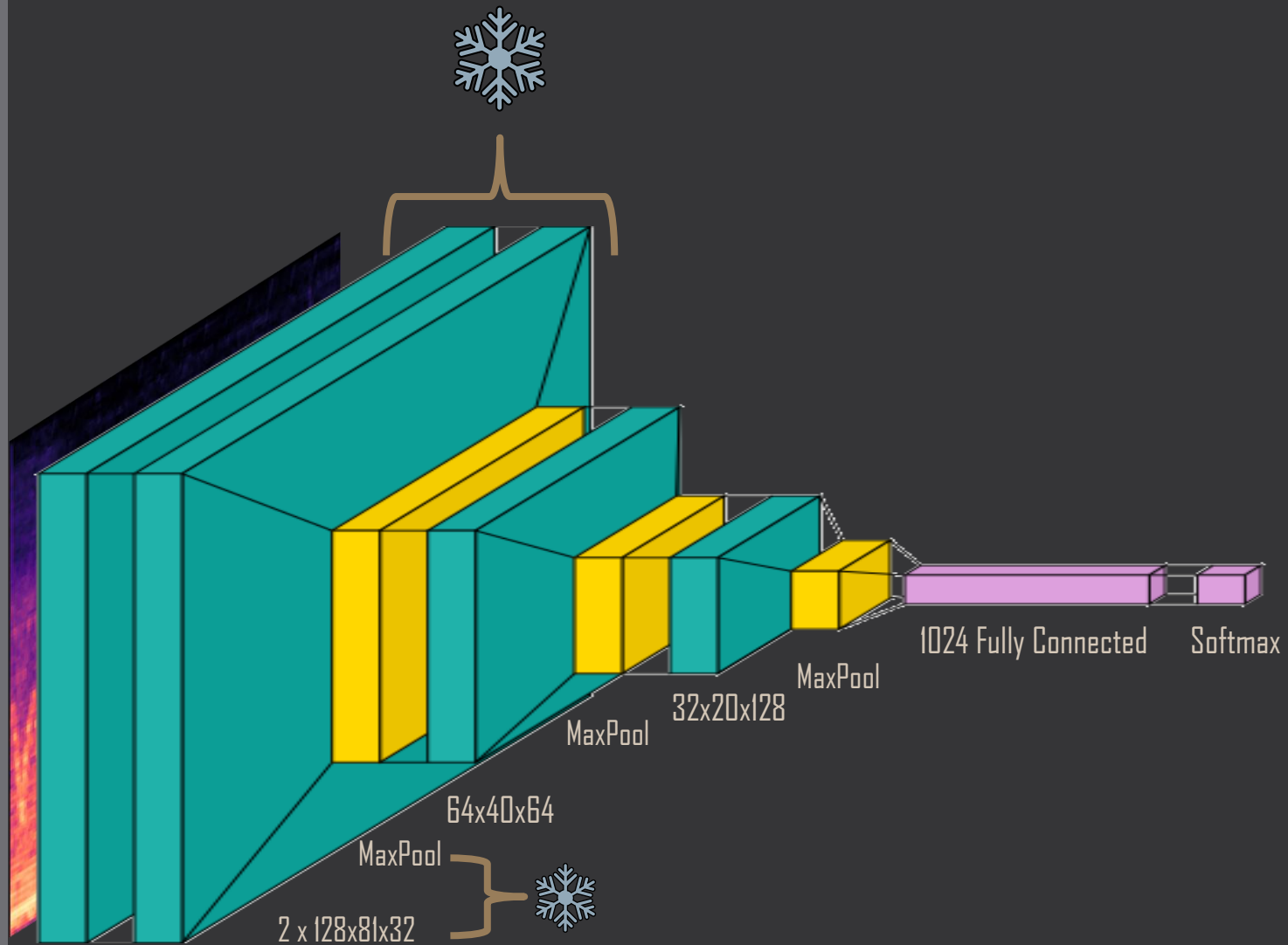
Batch Size: 64

F1:47.11%
Accuracy: 49%

SMO?

Transfer Learning

Activation Function : ReLU



Drop Out = 0.2

Batch Normalization

Zero Padding

Learning Rate= 0.001

Patience : 10

Early Stopping : F1_Score

Epochs: 27 !

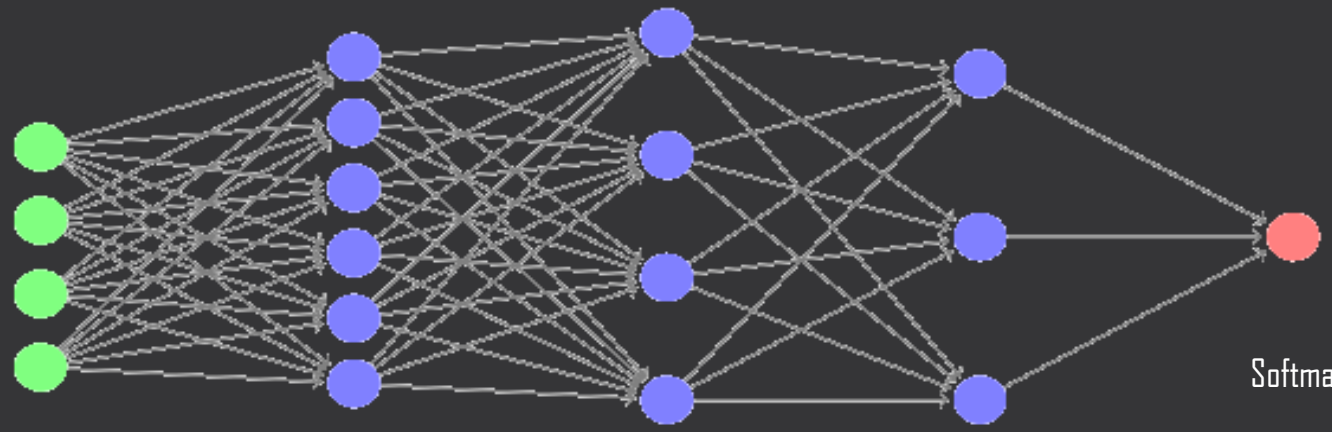
Batch Size: 32

YamNet Architecture

Trained Full Audioset

Produces Embeddings

Averaged The Embeddings



Activation Function : ReLU

Drop Out : 0.2

Batch Normalization

Epochs: 42

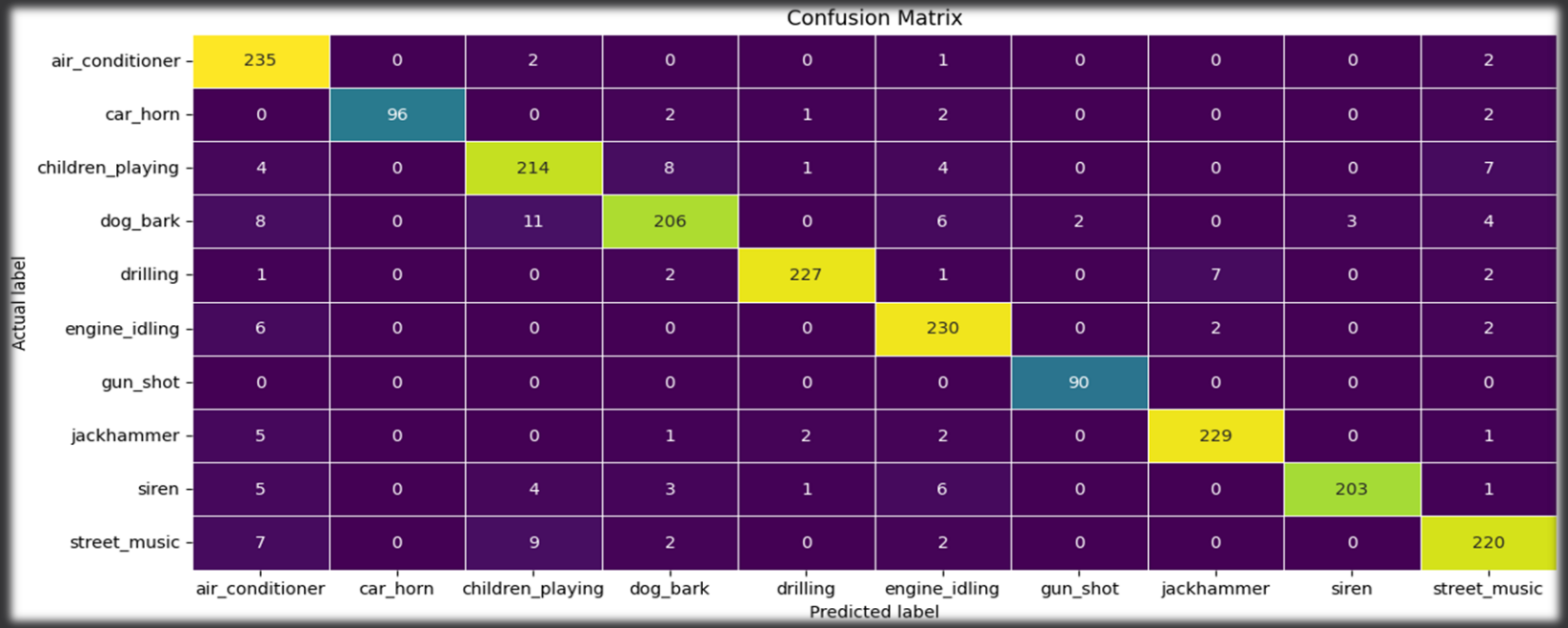
Learning Rate: 0.0005

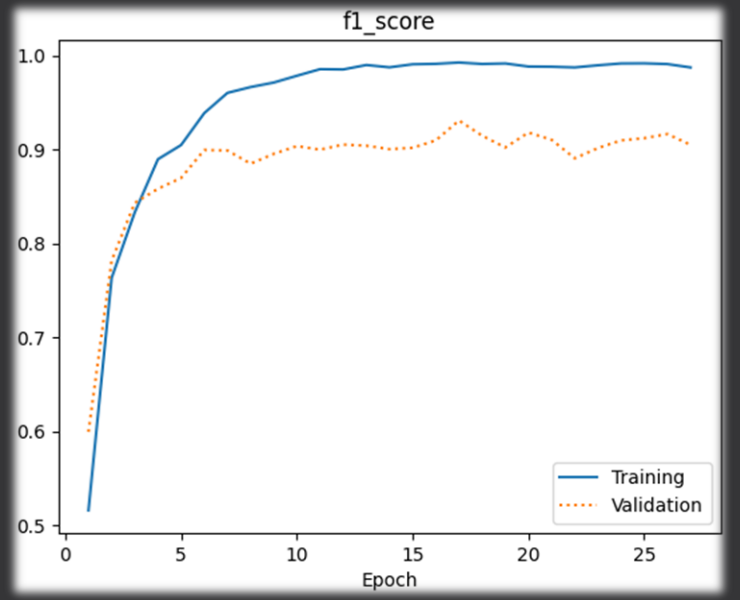
Patience : 15

Early Stopping : F1_Score

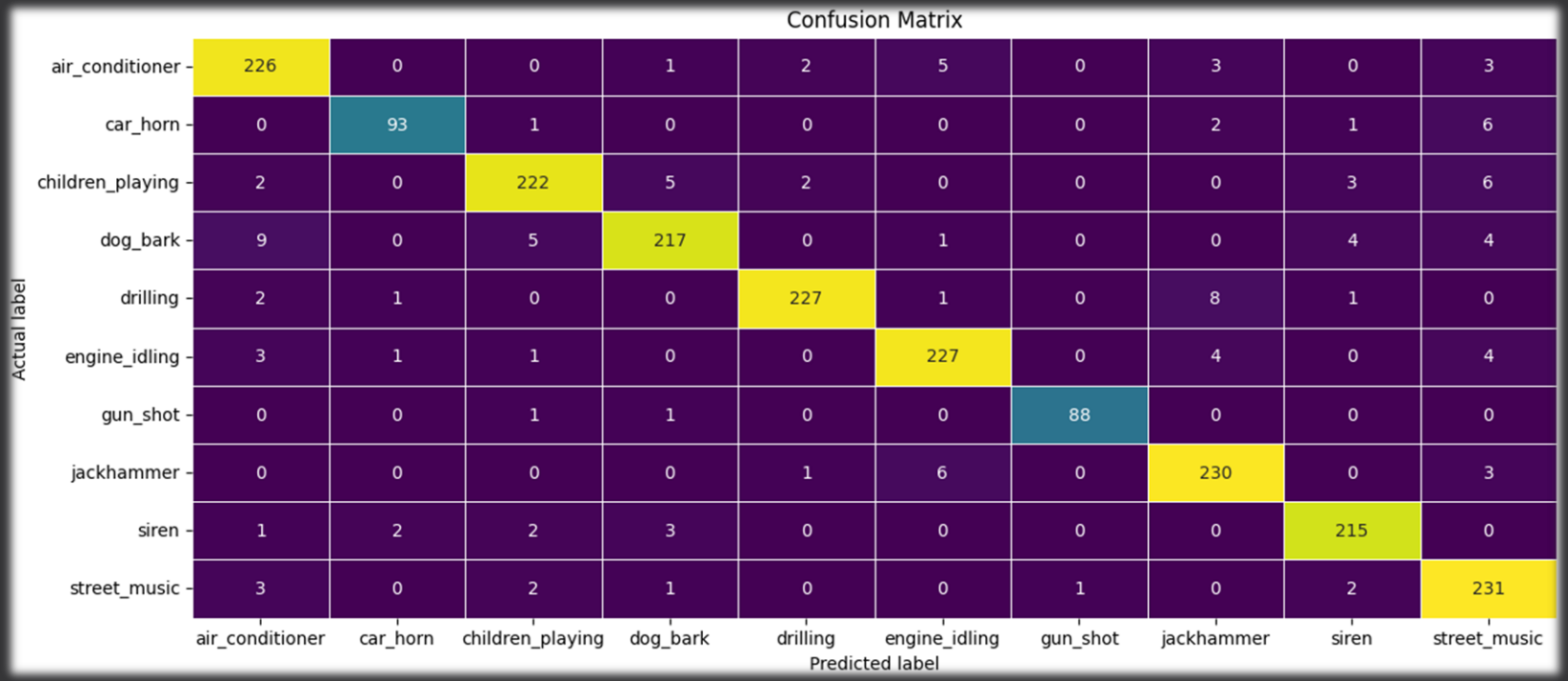
Batch Size: 32

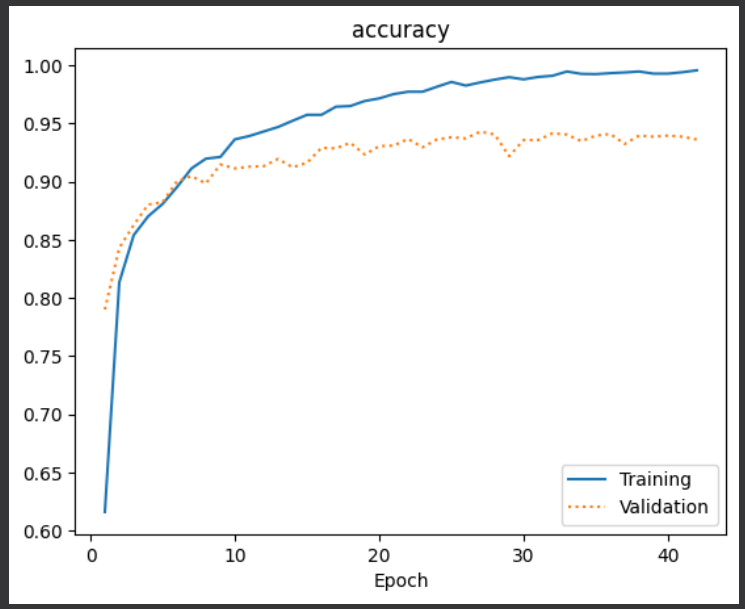
Transferred CNN



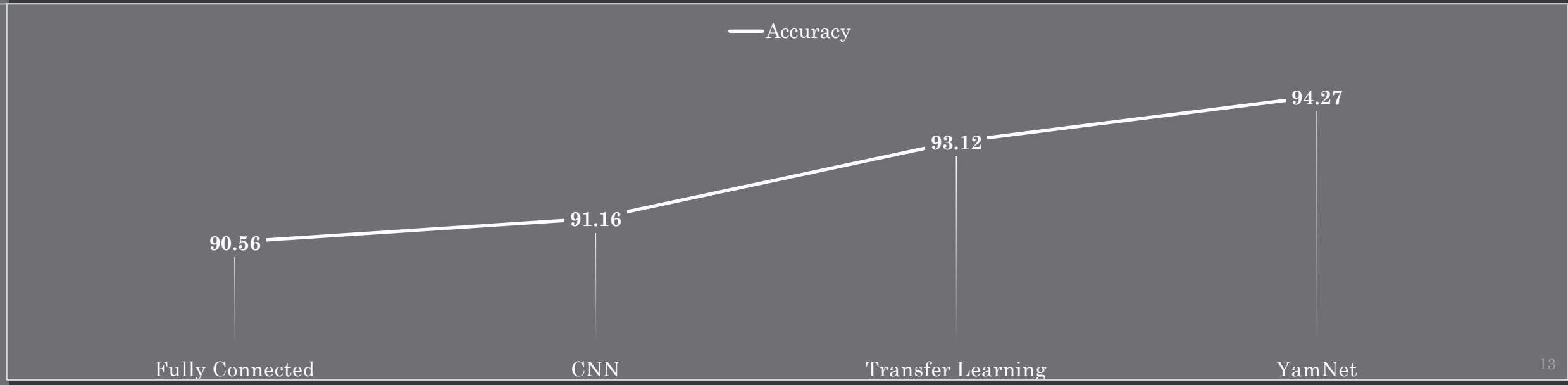


YamNet





	Accuracy	F1_Score	Training Time
Fully Connected	0.9056	0.9057	Fast
CNN	0.9116	0.9113	Fast
Transfer Learning	0.9312	0.9313	Medium
YamNet Embeddings	0.9427	0.9427	Very Fast



Let's See a live Demo

Thank You

Ranches Vince

Raptis Georgios