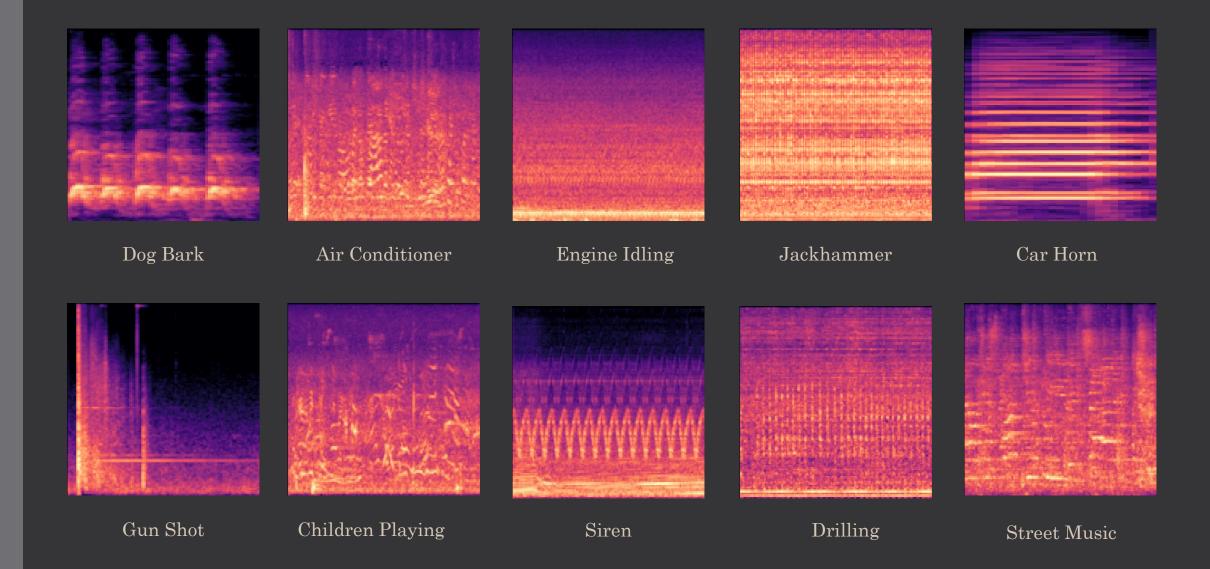
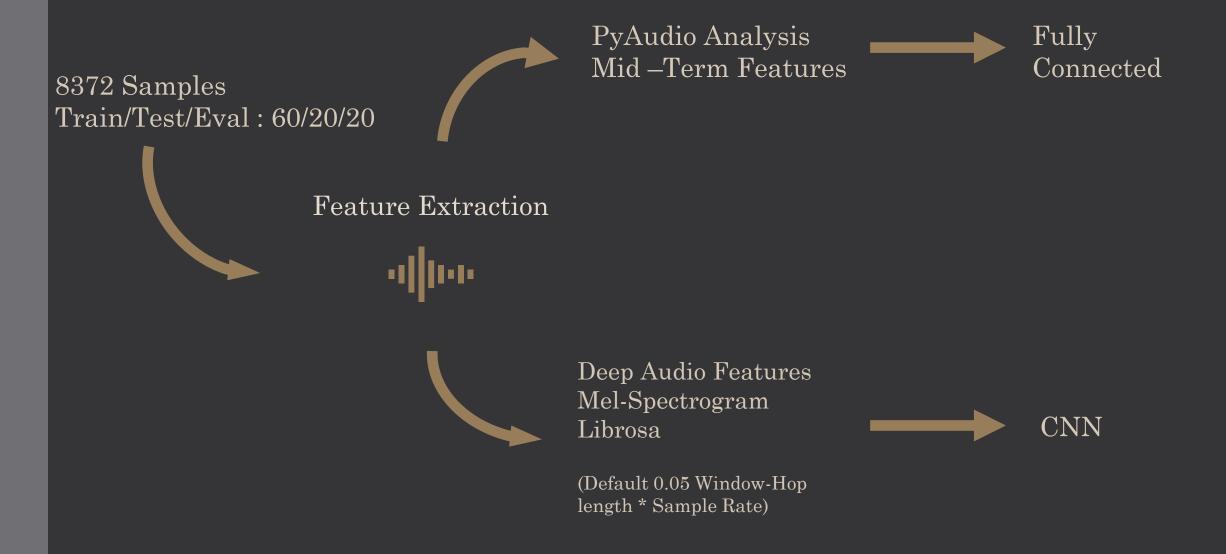


Class Overview



UrbanSound Pipeline



Fully Connected Architecture



Activation Function: ReLU



Drop Out = 0.2

Hidden Layer

512 Nodes

Hidden Layer 2

1024 Nodes



Batch Normalization



Epochs: 159



Learning Rate= 0.001



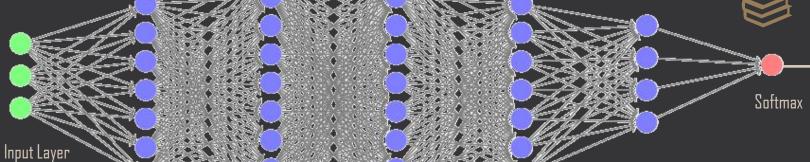
Early Stopping: Accuracy



Patience: 15



Batch Size: 32



Hidden Layer 3

1024 Nodes

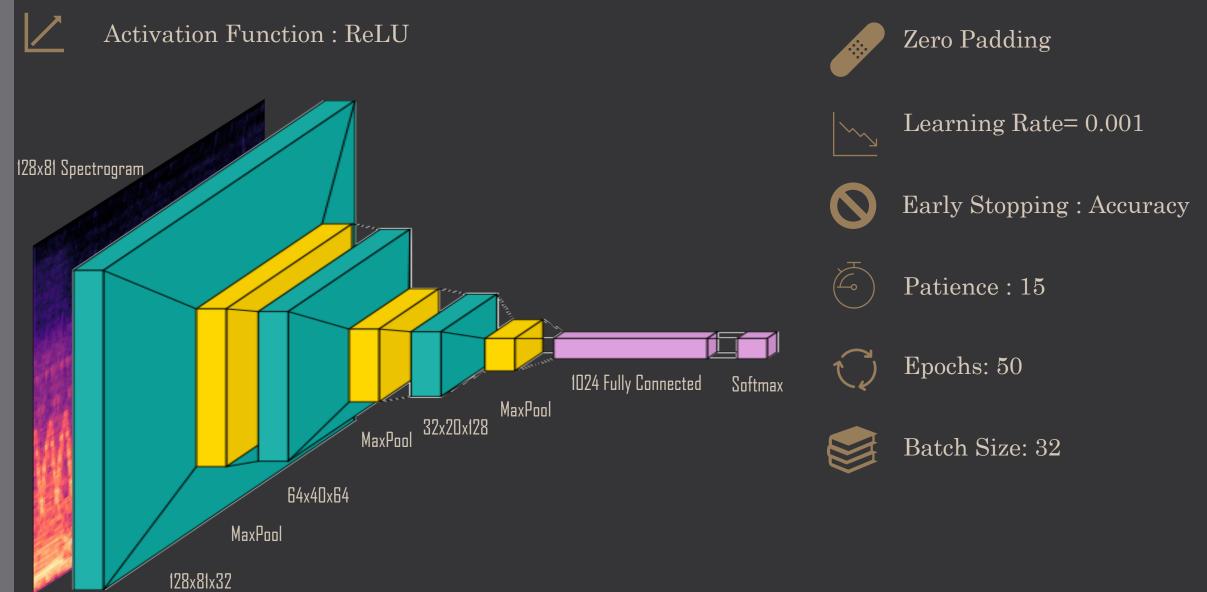
Hidden Layer 5 64 Nodes

512 Nodes

Hidden Layer 4



CNN Architecture

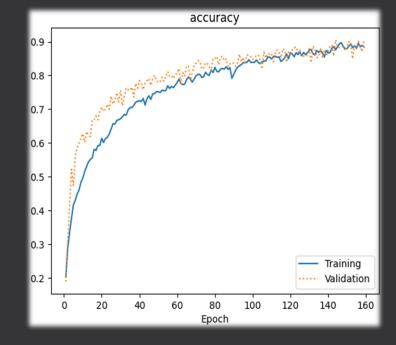


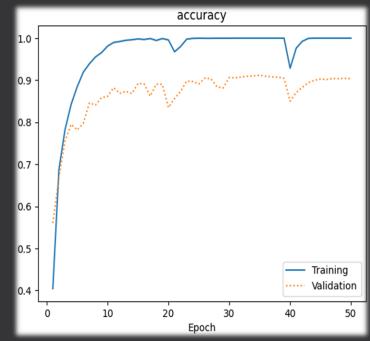
Fully Connected

' /.
()

Confusion Matrix										
air_conditioner -	225	0	2	0	3	2	0	3	1	4
car_horn -	2	87	0	2	1	3	0	3	0	0
children_playing -	3	1	204	2	4	3	0	0	1	20
dog_bark -	1	2	9	209	3	3	4	0	1	7
drilling -	0	5	3	2	215	1	1	11	0	2
drilling -	1	2	1	1	0	231	0	3	0	1
gun_shot -	0	1	0	2	2	1	82	1	0	0
jackhammer -	4	3	0	0	8	2	1	221	0	1
siren -	2	1	2	1	1	1	0	0	209	6
street_music -	2	0	17	5	2	4	0	1	2	207
	air_conditioner car_horn children_playing dog_bark drilling engine_idling gun_shot jackhammer siren street_musi Predicted label									street_music

	Confusion Matrix									
air_conditioner -	226	1	5	0	1	1	0	2	1	3
car_horn -	1	96	2	0	1	1	1	0	0	1
children_playing -	- 5	0	205	9	1	4	0	1	3	10
dog_bark -	- 6	1	17	197	4	2	1	1	8	3
drilling -	3	0	0	4	216	0	1	11	2	3
engine_idling -	. 3	0	1	3	1	227	0	4	1	0
gun_shot -	0	0	0	0	0	0	90	0	0	0
jackhammer -	1	0	0	1	1	1	0	236	0	0
siren -	0	0	5	4	2	3	0	1	207	1
street_music -	7	0	13	1	1	3	0	2	4	209
	air_conditioner	car_horn	children_playing	dog_bark	drilling Predicte	engine_idling	gun_shot	jackhammer	siren	street_music





AudioSet



Dataset By Google



Over 2 million Samples +500 classes



Used balanced Eval/Train 22k



Get Real Labels



Ask GPT



8 Categories



Semantically clustering/ NLP

The AudioSet Problem



Long to Set Up



Big Complexity
Imbalance Labels



Try to learn us much as possible



Labels are not Always True

The Audioset CNN Architecture



Activation Function : ReLU



Drop Out = 0.2 (0.3 FN)

MaxPool

2 x 128x81x32



Batch Normalization



Zero Padding



Learning Rate= 0.001



Patience: 10



Early Stopping: F1_Score



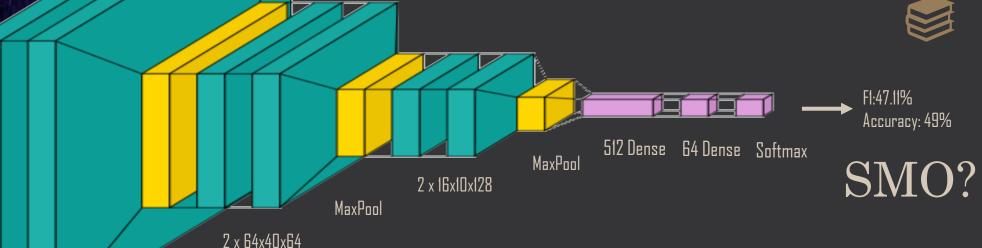
Environmental Sounds

Machine and Tool Sounds

Epochs: 61



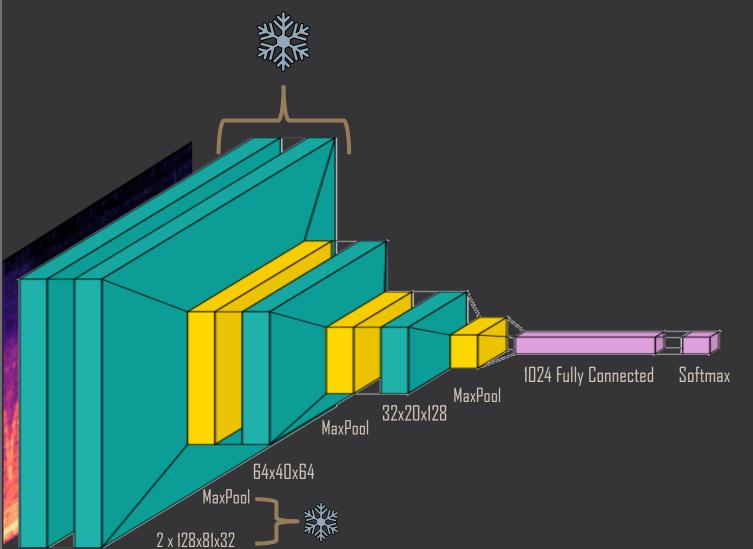
Batch Size: 64



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



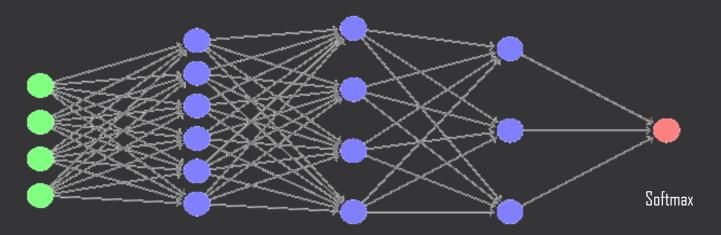
Trained Full Audioset



Produces Embeddings



Averaged The Embeddings



Input Layer Embedings (1.1024)

Hidden Layer 1 2048 Nodes Hidden Layer 2 1024 Nodes Hidden Layer 3 512 Nodes



Activation Function: ReLU



Drop Out: 0.2



Batch Normalization



Epochs: 42



Learning Rate: 0.0005



Patience: 15

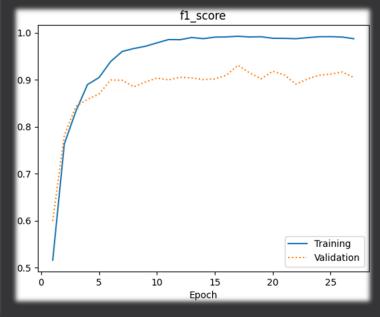


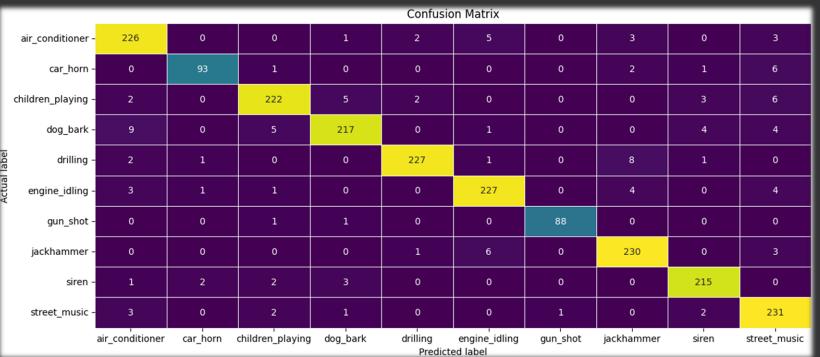
Early Stopping: F1_Score

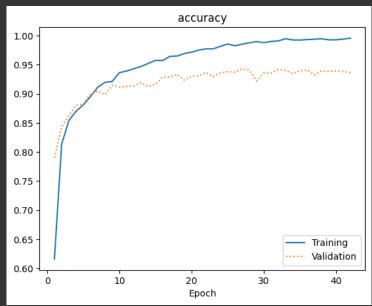


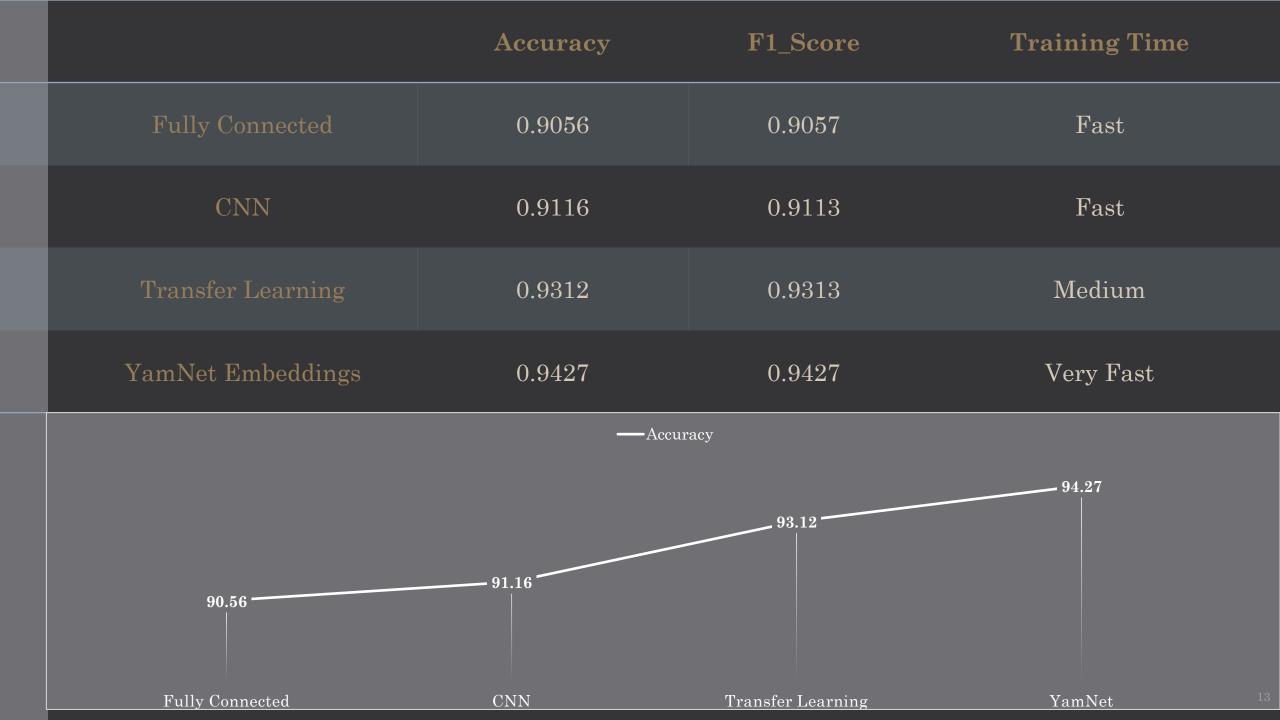
Batch Size: 32

Confusion Matrix										
air_condition	er - 235	0	2	0	0	1	0	0	0	2
car_ho	n - 0	96	0	2	1	2	0	0	0	2
children_playin	g - 4	0	214	8	1	4	0	0	0	7
dog_ba	·k - 8	0	11	206	0	6	2	0	3	4
ape drillin	g - 1	0	0	2	227	1	0	7	0	2
engine_idlin	g - 6	0	0	0	0	230	0	2	0	2
gun_sho	ot - 0	0	0	0	0	0	90	0	0	0
jackhamme	er - 5	0	0	1	2	2	0	229	0	1
sire	n - 5	0	4	3	1	6	0	0	203	1
street_mus	ic - 7	0	9	2	0	2	0	0	0	220
	air_conditioner car_horn children_playing dog_bark drilling engine_idling gun_shot jackhammer siren street_music Predicted label									











Tet's See a live Demo Tet's See a live Demo







Thank You







Ranches Vince



Raptis Georgios