EE603 : Assignment 2

Multiple Events Detection

Name - Sanskar Ghyar

Roll no – 200873

Audio Classification is very important task in the real-world problem. There are various Machine Learning and Deep learning ways to solve this problem.

10000 melspec.npy files were provided in form of X training dataset with shape (1,64,1000) for each file and 10000 corresponding eventroll.npy files as Y label with shape (11,1000) for each file. Eventroll was converted to Multihot vector with: -

```
events_types = {
   0: 'Alarm_bell_ringing',
   1: 'Blender',
   2: 'Cat',
   3: 'Dishes',
   4: 'Dog',
   5: 'Electric_shaver_toothbrush',
   6: 'Frying',
   7: 'Running_water',
   8: 'Speech',
   9: 'Vacuum_cleaner'
}
```

After Data preprocessing shape of each file of X is (64,1000,1) and Y is (10,)

ANN

Layers: -

Layer (type)	Output Shape	Param #
flatten (Flatten)	(None, 64000)	0
dense (Dense)	(None, 32)	2048032
batch_normalization (BatchNormalization)	(None, 32)	128
dense_1 (Dense)	(None, 8)	264
batch_normalization_1 (BatchNormalization)	(None, 8)	32
dense_2 (Dense)	(None, 128)	1152
batch_normalization_2 (BatchNormalization)	(None, 128)	512
dense_3 (Dense)	(None, 10)	1290

Total params: 2,051,410
Trainable params: 2,051,074
Non-trainable params: 336

F1 score = 0.4831923809523809

Precision = 0.5544133333333333

Recall = 0.46749714285714283

Confusion Matrix: -

```
[[1910  190]
 [ 366   34]]
[[1800  434]
 [ 208   58]]
[[2078  138]
 [ 261   23]]
[[1272  539]
 [ 504  185]]
[[2040  119]
 [ 316   25]]
[[2088  129]
 [ 263   20]]
[[1820  303]
 [ 352   25]]
[[1964  230]
 [ 274   32]]
[[  10  117]
 [ 222 2151]]
[[2098  151]
 [ 228   23]]
```

CNN

Layers: -

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 998, 4)	40
batch_normalization (BatchNormalization)	(None, 62, 998, 4)	16
max_pooling2d (MaxPooling2D)	(None, 31, 499, 4)	0
dropout (Dropout)	(None, 31, 499, 4)	0
conv2d_1 (Conv2D)	(None, 29, 497, 16)	592
batch_normalization_1 (BatchNormalization)	(None, 29, 497, 16)	64
max_pooling2d_1 (MaxPooling2D)	(None, 14, 248, 16)	0
dropout_1 (Dropout)	(None, 14, 248, 16)	0
conv2d_2 (Conv2D)	(None, 12, 246, 64)	9280
batch_normalization_2 (BatchNormalization)	(None, 12, 246, 64)	256
max_pooling2d_2 (MaxPooling2D)	(None, 6, 123, 64)	0
dropout_2 (Dropout)	(None, 6, 123, 64)	0
flatten (Flatten)	(None, 47232)	0
dense (Dense)	(None, 128)	6045824
dropout_3 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 10)	1290

Total params: 6,057,362
Trainable params: 6,057,194
Non-trainable params: 168

F1 score = 0.5810761904761905

Precision = 0.8566333333333332

Recall = 0.4636704761904762

Confusion Matrix: -

```
[[2074   26]
 [ 393    7]]
[[2220   14]
 [ 263    3]]
[[2187   29]
 [ 281    3]]
[[1535  276]
 [ 566  123]]
[[2134   25]
 [ 336    5]]
[[2194   23]
 [ 280    3]]
[[2005  118]
 [ 354   23]]
[[2183   11]
 [ 305    1]]
[[   0  127]
 [   7 2366]]
[[2197   52]
 [ 249    2]]
```