Best RALCT Architecture - achieved 94.56% on UrbanSound8K

Model: "model_aug"

Layer (type)	Output Shape	Param #	Connected to
input (InputLayer)	[(None, 220500)]	0	[]
log_mel_spectrogram (LogMelSpectrogram)	(None, 496, 128, 1)	0	['input[0][0]']
tf.signal.mfccs_from_log_mel_s pectrograms (TFOpLambda)	(None, 496, 128, 1)	0	['log_mel_spectrogram[0][0]']
batch_normalization (BatchNorm alization)	(None, 496, 128, 1)	512	['log_mel_spectrogram[0][0]']
batch_normalization_1 (BatchNormalization)	(None, 496, 128, 1)	512	<pre>['tf.signal.mfccs_from_log_mel_sp ectrograms[0][0]']</pre>
tf.concat (TFOpLambda)	(None, 496, 128, 2)	0	<pre>['batch_normalization[0][0]', 'batch_normalization_1[0][0]']</pre>
conv2d (Conv2D)	(None, 496, 128, 32	1632	['tf.concat[0][0]']
batch_normalization_2 (BatchNormalization)	(None, 496, 128, 32	128	['conv2d[0][0]']
max_pooling2d (MaxPooling2D)	(None, 496, 42, 32)	0	['batch_normalization_2[0][0]']
dropout (Dropout)	(None, 496, 42, 32)	0	['max_pooling2d[0][0]']
conv2d_1 (Conv2D)	(None, 496, 42, 64)	18496	['dropout[0][0]']

batch_normalization_3 (BatchNormalization)	(None, 496, 42, 64)	256	['conv2d_1[0][0]']
max_pooling2d_1 (MaxPooling2D)	(None, 165, 14, 64)	0	['batch_normalization_3[0][0]']
dropout_1 (Dropout)	(None, 165, 14, 64)	0	['max_pooling2d_1[0][0]']
conv2d_2 (Conv2D)	(None, 165, 14, 32)	18464	['dropout_1[0][0]']
batch_normalization_4 (BatchNo rmalization)	(None, 165, 14, 32)	128	['conv2d_2[0][0]']
max_pooling2d_2 (MaxPooling2D)	(None, 55, 4, 32)	0	['batch_normalization_4[0][0]']
dropout_2 (Dropout)	(None, 55, 4, 32)	0	['max_pooling2d_2[0][0]']
positional_encoding (Positiona lEncoding)	(None, 55, 4, 32)	7040	['dropout_2[0][0]']
layer_normalization (LayerNorm alization)	(None, 55, 4, 32)	64	['positional_encoding[0][0]']
multi_head_attention (MultiHea dAttention)	(None, 55, 4, 32)	16800	<pre>['layer_normalization[0][0]', 'layer_normalization[0][0]']</pre>
dropout_3 (Dropout)	(None, 55, 4, 32)	0	['multi_head_attention[0][0]']
tfoperatorsadd (TFOpLamb da)	(None, 55, 4, 32)	0	<pre>['dropout_3[0][0]', 'positional_encoding[0][0]']</pre>
layer_normalization_1 (LayerNo rmalization)	(None, 55, 4, 32)	64	['tfoperatorsadd[0][0]']
dense (Dense)	(None, 55, 4, 32)	1056	['layer_normalization_1[0][0]']
dropout_4 (Dropout)	(None, 55, 4, 32)	0	['dense[0][0]']
dense 1 (Dense)	(None, 55, 4, 32)	1056	['dropout_4[0][0]']

tfoperatorsadd_1 (TFOpLa mbda)	(None, 55, 4, 32)	0	['dense_1[0][0]', 'tfoperatorsadd[0][0]']
layer_normalization_2 (LayerNo rmalization)	(None, 55, 4, 32)	64	['tfoperatorsadd_1[0][0]']
<pre>multi_head_attention_1 (MultiH eadAttention)</pre>	(None, 55, 4, 32)	16800	<pre>['layer_normalization_2[0][0]', 'layer_normalization_2[0][0]']</pre>
dropout_5 (Dropout)	(None, 55, 4, 32)	0	['multi_head_attention_1[0][0]']
tfoperatorsadd_2 (TFOpLa mbda)	(None, 55, 4, 32)	0	['dropout_5[0][0]', 'tfoperatorsadd_1[0][0]']
layer_normalization_3 (LayerNo rmalization)	(None, 55, 4, 32)	64	['tfoperatorsadd_2[0][0]']
dense_2 (Dense)	(None, 55, 4, 32)	1056	['layer_normalization_3[0][0]']
dropout_6 (Dropout)	(None, 55, 4, 32)	0	['dense_2[0][0]']
dense_3 (Dense)	(None, 55, 4, 32)	1056	['dropout_6[0][0]']
tfoperatorsadd_3 (TFOpLa mbda)	(None, 55, 4, 32)	0	['dense_3[0][0]', 'tfoperatorsadd_2[0][0]']
flatten (Flatten)	(None, 7040)	0	['tfoperatorsadd_3[0][0]']
dense_4 (Dense)	(None, 32)	225312	['flatten[0][0]']
dropout_7 (Dropout)	(None, 32)	0	['dense_4[0][0]']
dense_5 (Dense)	(None, 10)	330	['dropout_7[0][0]']

Total params: 310,890 Trainable params: 310,122 Non-trainable params: 768