AudioToSpectrogramTransformJit(nn.Module):

Audio Chunks at 16khz
PCM Float 32bit
32000 Samples as Input
(2 seconds window)

Start Transformation

Ensure audio is mono/1D Array

if waveform is shorter than 2 second?

no

yes

Padding for Length

Mel Spectrogram Transformation

Decibel Transformation and Normalize the Spectrogram

Output Shape:
[1, 128, 256]