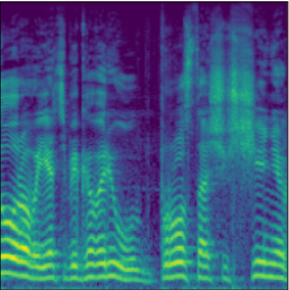


Raw

waveform



Log-mel spectrogram

(1x224x224)

Conv3x3-64

Max pool

Max pool

Conv3x3-128

Conv3x3-256

Conv3x3-256

Max pool

Conv3x3-512

Conv3x3-512

Max pool

Conv3x3-512

Conv3x3-512

Max pool

VGG-11 Feature Extractor

Emotion

classification

FC-25088

FC-2048

FC-512

FC - <CLASSES>

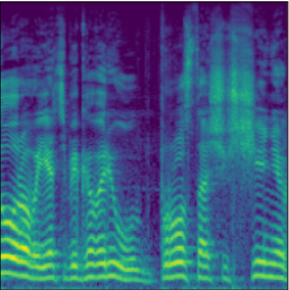
SoftMax

Dropout-0.5



Raw

waveform



Log-mel spectrogram

(1x224x224)

AlexNet Feature Extractor

Emotion

classification

Conv11x11-64

Max pool

Max pool

Conv5x5-192

Conv3x3-384

Conv3x3-256

Max pool

Conv4x4-256

FC-6400

FC-2048

FC-512

FC - <EMOTIONS>

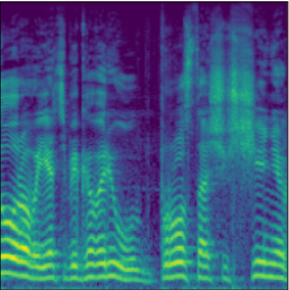
SoftMax

Dropout-0.5



Raw

waveform



Log-mel spectrogram

(1x224x224)

Conv3x3-64

Max pool

Max pool

Conv3x3-128

Conv3x3-256

Conv3x3-256

Max pool

Conv3x3-512

Conv3x3-512

Max pool

Conv3x3-512

Conv3x3-512

Max pool

VGG-11 Feature Extractor

Classification

FC-25088

Dropout-0.5

FC-2048

FC-512

FC - <EMOTIONS>

SoftMax

FC-2048

FC-512

FC - <SPEAKERS>

SoftMax

FC-2048

FC-512

FC - <GENDERS>

SoftMax

Emotion

classification

Speaker

classification

Gender

classification