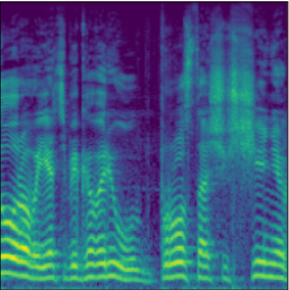


Raw

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

Feature Extraction

Emotion

classification

FC-25088

FC-2048

FC-512

FC - <CLASSES>

SoftMax

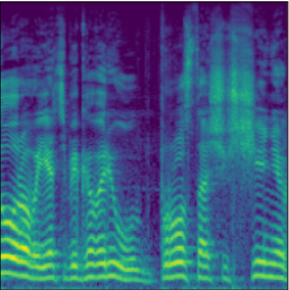
Dropout-0.5

Базовая модель 1.



Raw

waveform



Log-mel spectrogram

(1x224x224)

Feature Extraction

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

Feature Extraction у этой модели – сверточная архитектура, идентичная AlexNet[ссылка].

Используется функция активации ReLU.

Модель обучалась на протяжении 300 эпох, learning rate – 1e-5, алгоритм оптимизации – Adam.

Наборы данных были поделены на тренировочную и тестовую выборки в отношении 4:1.

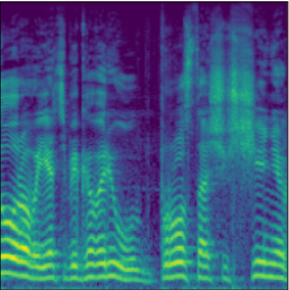
|  |  |  |  |
| --- | --- | --- | --- |
|  | IEMOCAP | RAMAS (descrete) | RAMAS (binary) |
| Validation Loss | 0.808 | 1.561 | 0.375 |
| Validation Accuracy | 0.683 | 0.415 | 0.858 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | IEMOCAP | RAMAS (descrete) | RAMAS (binary) |
| Validation Loss | 0.808 | 1.561 | 0.375 |
| Validation Accuracy | 0.683 | 0.415 | 0.858 |



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

Feature Extraction

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