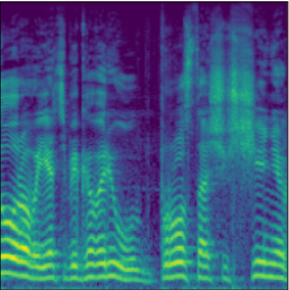


Сигнал

(.wav)



Лог-мел спектрограмма

(1x224x224)

Conv1

MaxPool1

MaxPool2

Conv2

Conv3

Conv4

MaxPool3

Conv5

Conv6

MaxPool4

Conv7

Conv8

MaxPool5

Извлечение признаков

Классификация

эмоций

FC1

FC2

FC3

FC 4

SoftMax

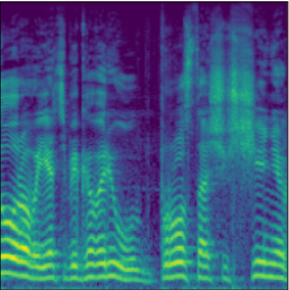
Dropout

Базовая модель 1.



Сигнал

(.wav)



Лог-мел спектрограмма

(1x224x224)

Извлечение признаков

Классификация

эмоций

Conv1

MaxPool1

MaxPool2

Conv2

Conv3

Conv4

MaxPool3

Conv5

FC1

FC2

FC3

FC 4

SoftMax

Dropout

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.

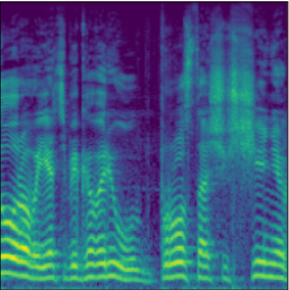
Наборы данных были поделены на тренировочную и тестовую выборки в отношении 4:1.

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



Raw

waveform



Log-mel spectrogram

(1x224x224)

Conv1

MaxPool1

MaxPool2

Conv2

Conv3

Conv4

MaxPool3

Conv5

Conv6

MaxPool4

Conv7

Conv8

MaxPool5

Feature Extraction

Classification

FC1

Dropout

FC2

FC3

FC -4

SoftMax

FC2

FC3

FC 4

SoftMax

FC2

FC3

FC 4

SoftMax

Emotion

classification

Speaker

classification

Gender

classification