

РАСПОЗНАВАНИЕ ЭМОЦИЙ В ЗВУКЕ

Попова А. С., Рассадин А. Г., Пономаренко А. А.

НИУ ВШЭ - Нижний Новгород

Нейроинформатика-2017,
2-6 октября

Москва



Постановка задачи

X – множество всевозможных описаний объектов

Y – множество классов

$y' : X \rightarrow Y$ целевая зависимость, известная только на объектах обучающей выборки

$$X^l = (x_i, y_i)_{i=1}^l, y_i = y'(x_i)$$

Требуется построить алгоритм $a : X \rightarrow Y$,
который аппроксимирует целевую зависимость y'

Мотивация

- Дружественные интерфейсы
- Повышение степени интерактивности
- Робототехника

Набор данных RAVDESS

▶ 1neutral	03-01-01-01-01-01-01.wav
▶ 2calm	03-01-01-01-01-01-02.wav
▶ 3happy	03-01-01-01-01-01-03.wav
▶ 4sad	03-01-01-01-01-01-04.wav
▶ 5angry	03-01-01-01-01-01-05.wav
▶ 6fearful	03-01-01-01-01-01-06.wav
▶ 7disgust	03-01-01-01-01-01-07.wav
▶ 8surprised	03-01-01-01-01-01-08.wav

8 эмоций, 24 актёра, одна фраза произносилась 8 раз
1 класс - 190 экземпляров

70% - обучающая выборка и 30% - тестовая

<http://neuron.arts.ryerson.ca/ravdess/?f=3>

Подход

Предобработка



Звук в изображение



Свёрточная сеть

Подход

Предобработка



Звук в изображение:
мелоспектрограмма



Свёрточная сеть: VGG11, : VGG16

Используемое программное обеспечение

- Python



- Keras



- Librosa



- Pandas&NumPy

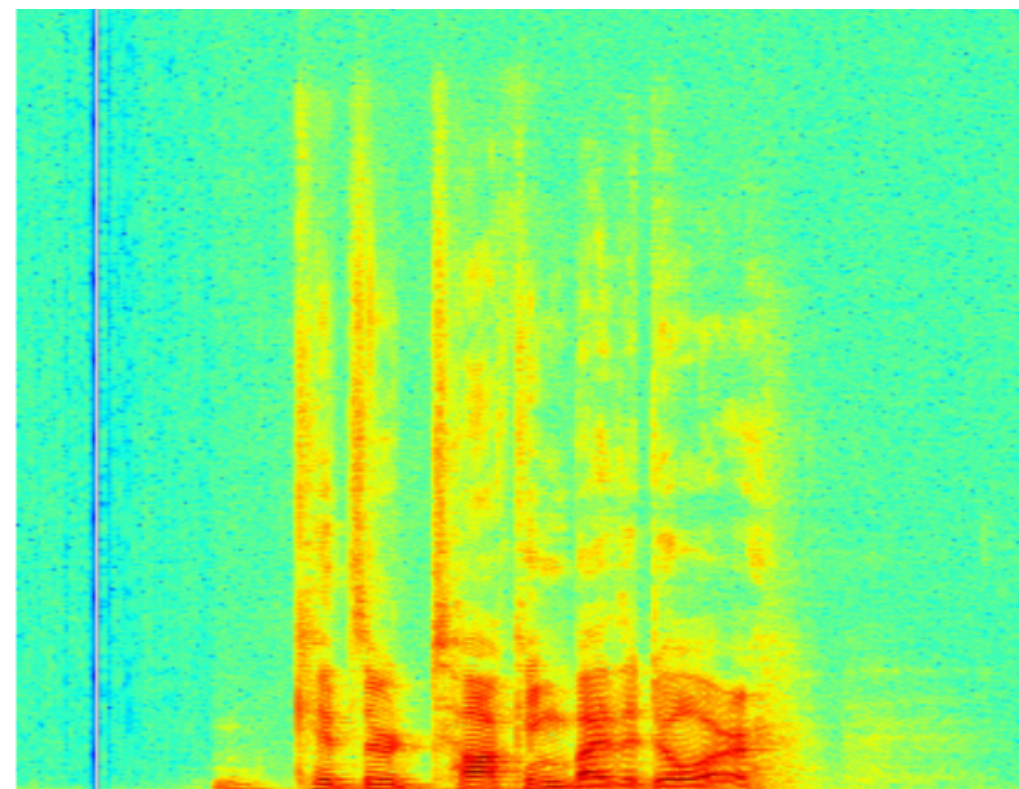
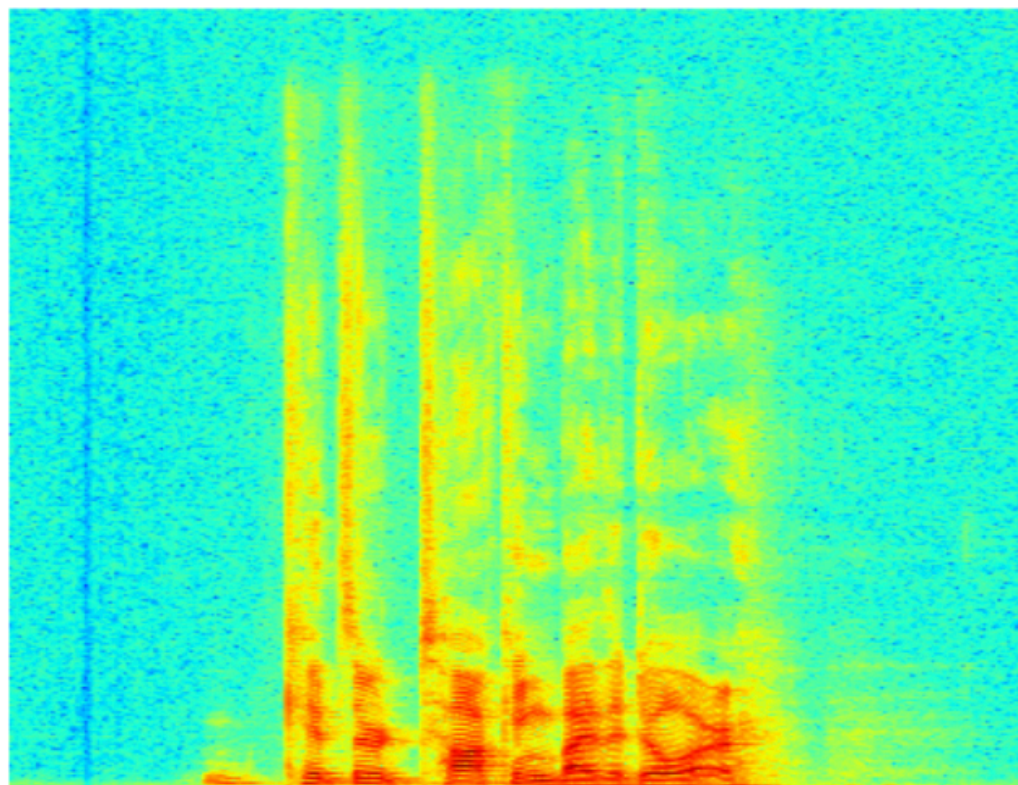


- ScikitLearn



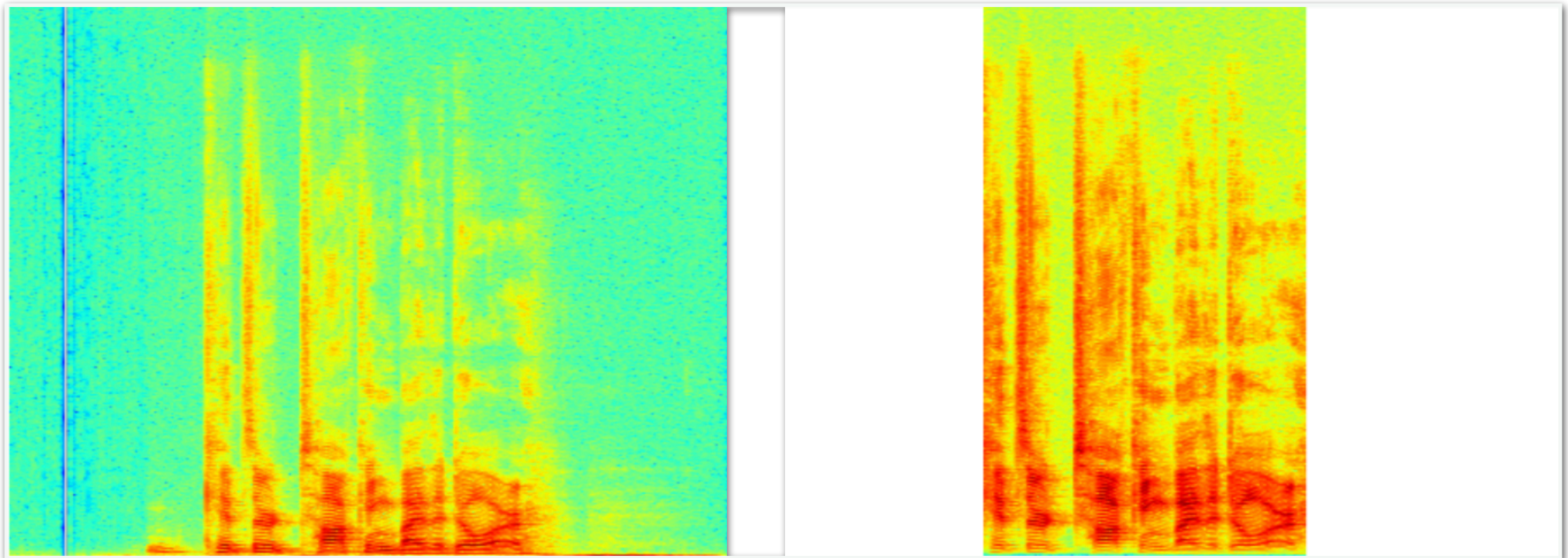
- Matplotlib

Предобработка



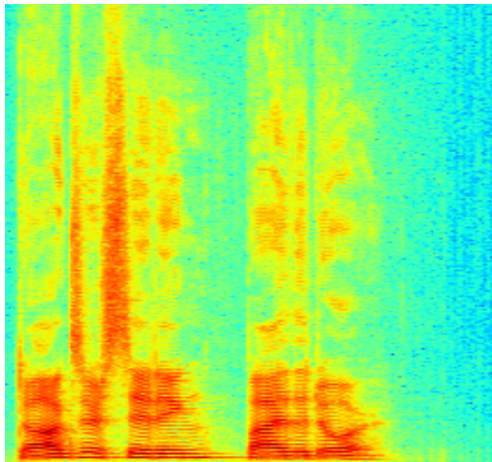
Нормализация громкости

Предобработка

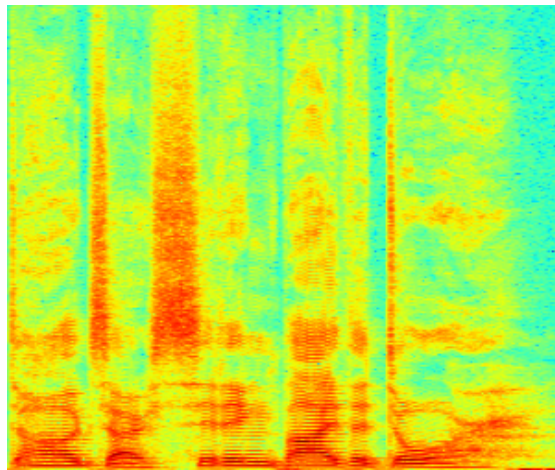


Highpass&Lowpass фильтры, выделение голоса

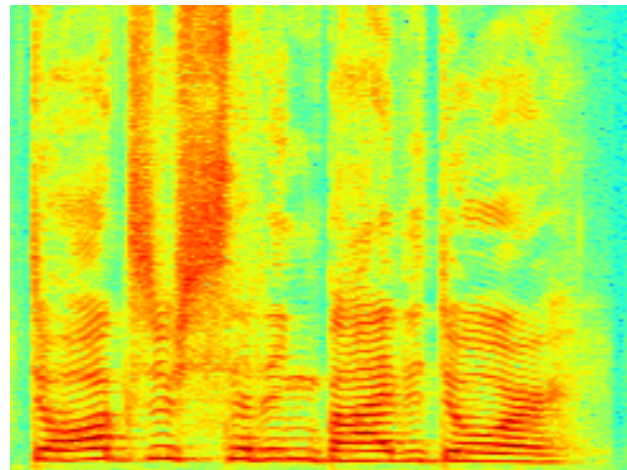
Обучающая выборка



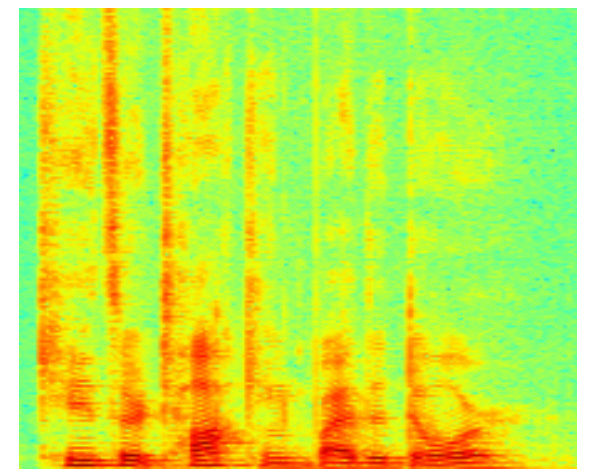
neutral



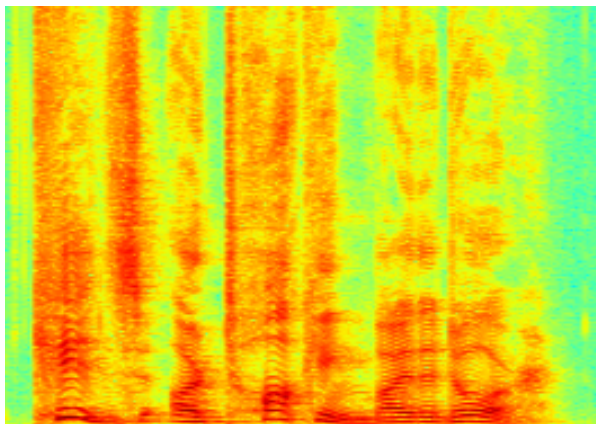
calm



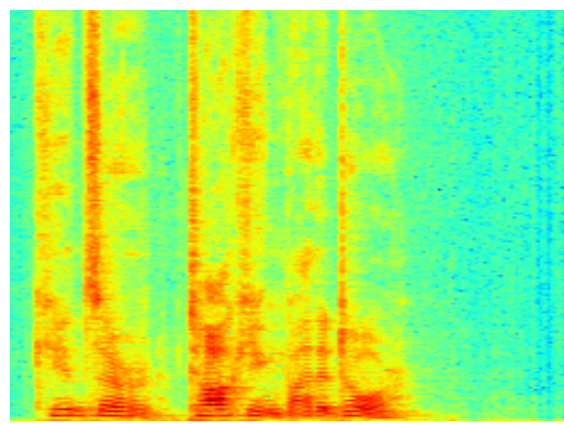
happy



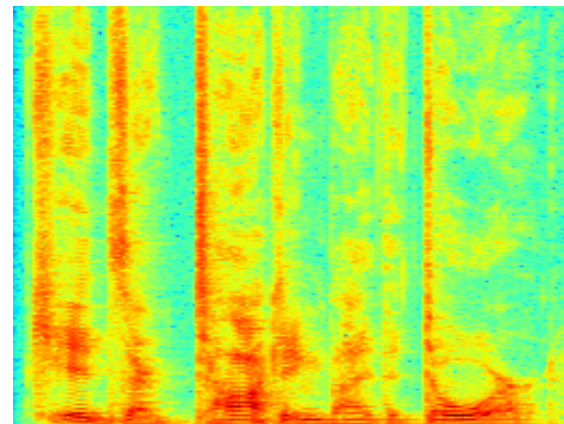
sad



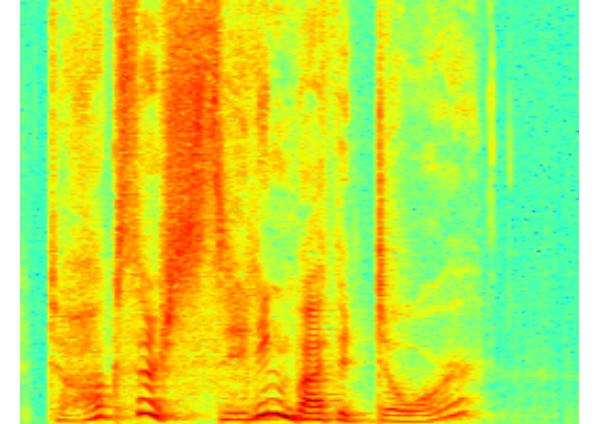
angry



fearful

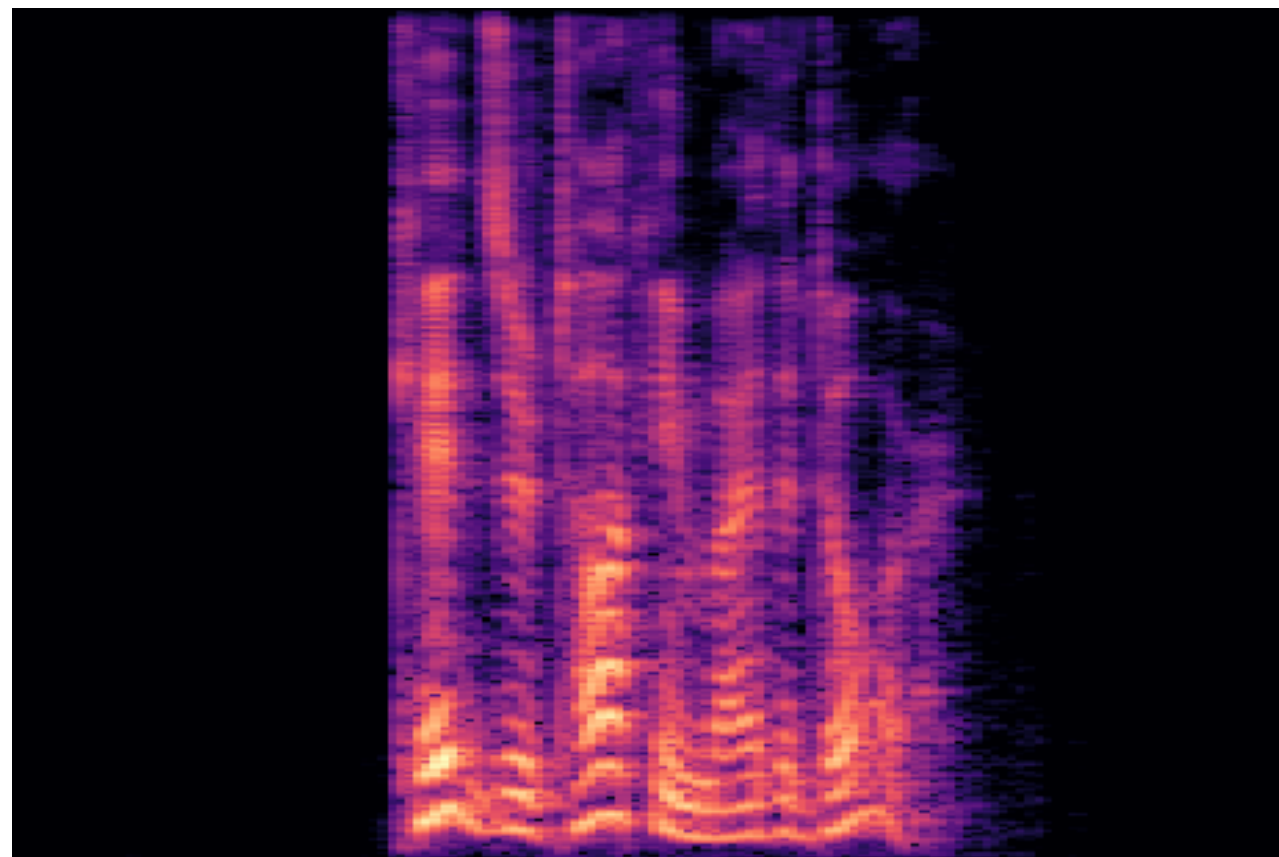
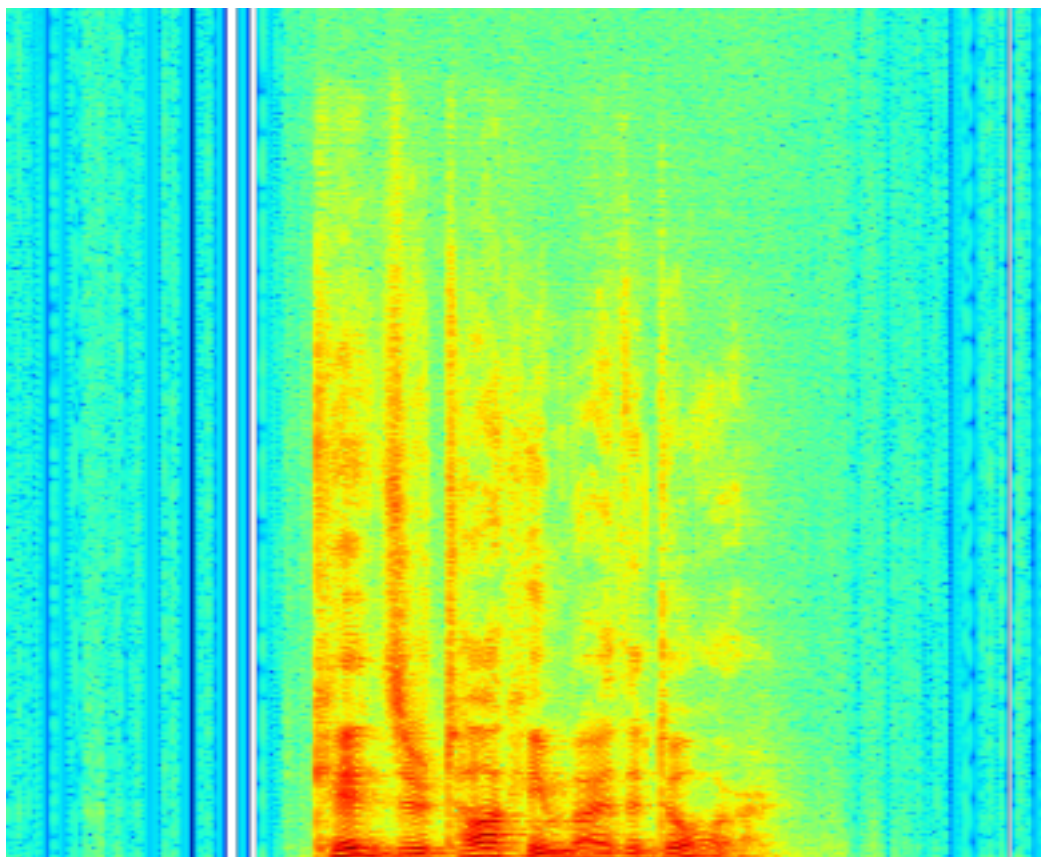


disgust

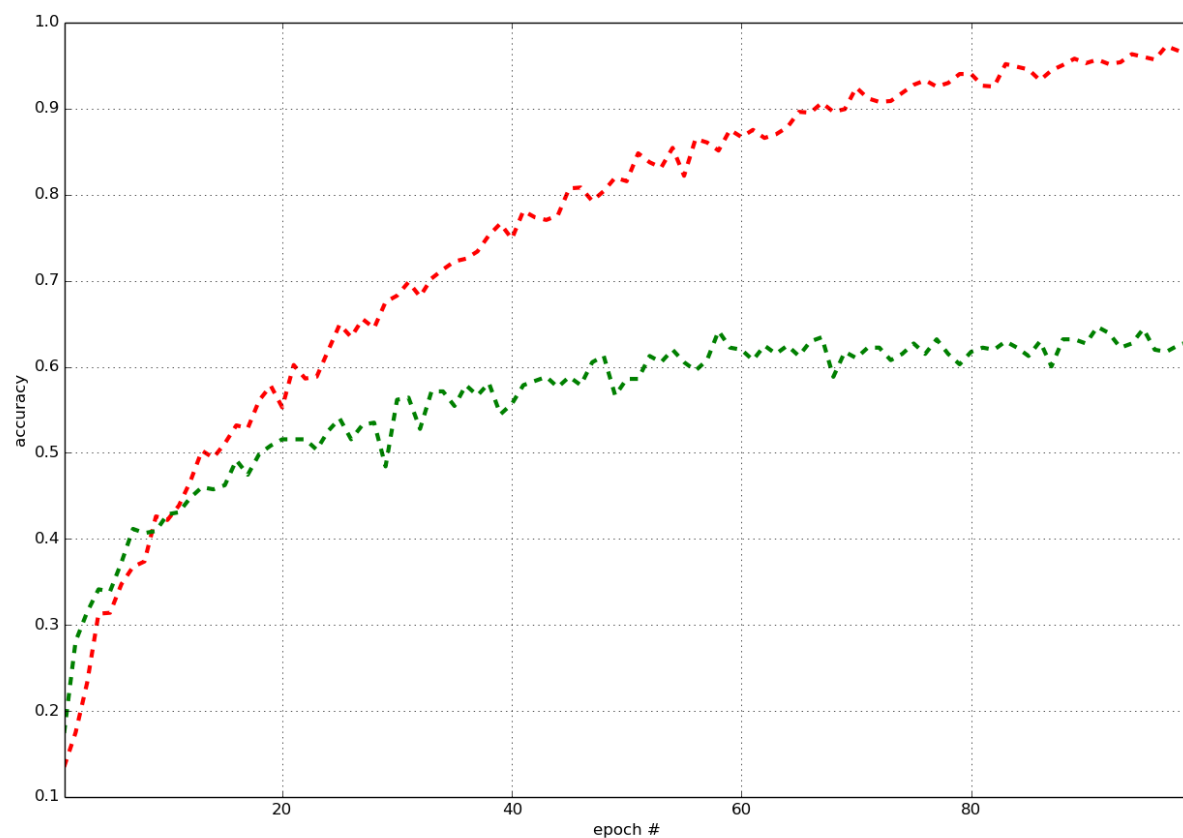


surprised

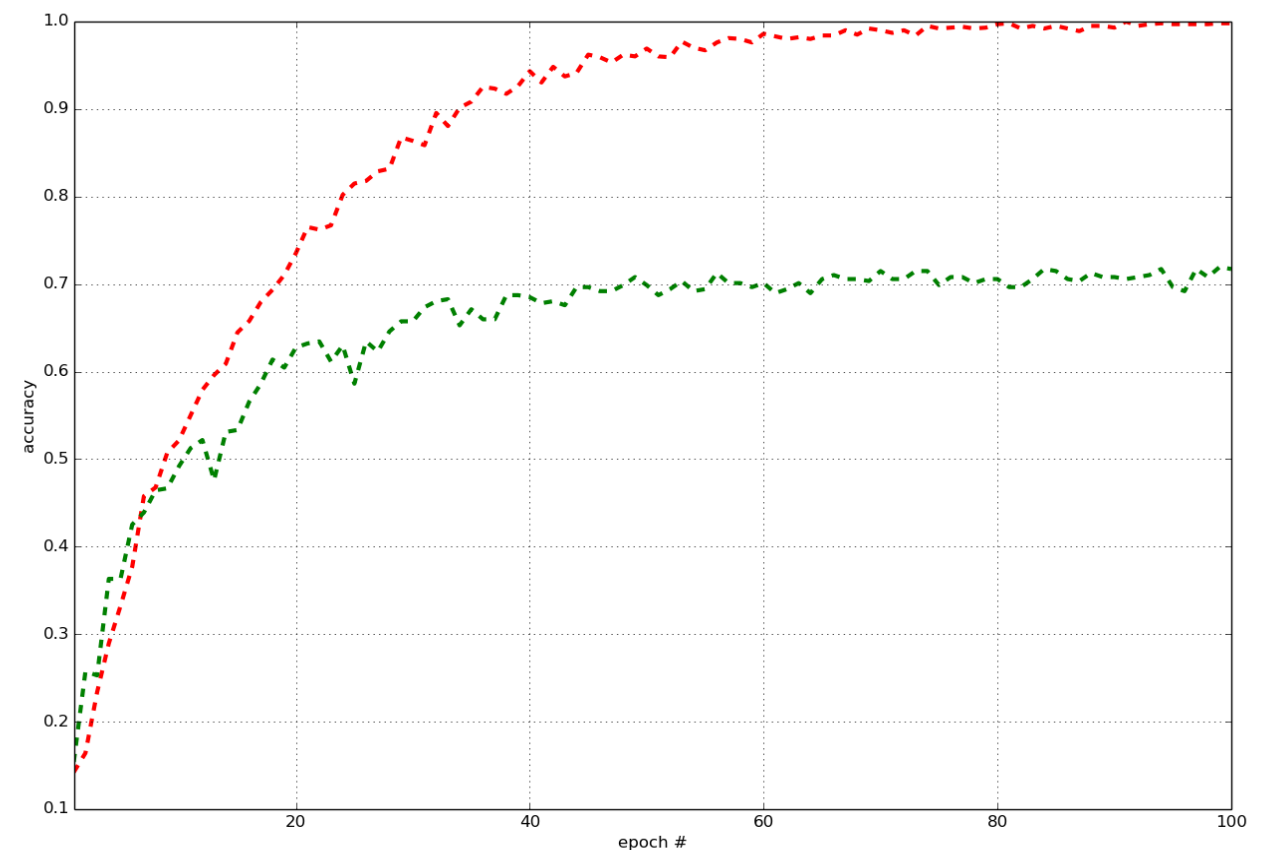
Мелспектрограмма



Точность классификации



VGG-11 + spectrogram



VGG-16 + melspectrogram

Матрица несоответствий

	neutral	calm	happy	sad	angry	fearfull	disgust	surprised	
neutral	21	8	0	0	0	0	0	0	29
calm	7	46	1	1	0	1	2	0	58
happy	0	1	26	7	6	9	5	4	58
sad	0	2	10	31	2	3	9	1	58
angry	0	1	1	0	43	2	5	6	58
fearfull	0	1	3	2	6	34	6	6	58
disgust	0	0	0	3	2	3	49	1	58
surprised	0	0	2	1	1	8	12	12	36
	28	59	43	45	60	60	88	30	

Ошибка предложного подхода 8 классов - ошибка 29%.
Ошибка на 1 класс. **3.625%**

4 класса - ошибка 16,5%. Ошибка на один класс **4.125%**)

[Speech Emotion Recognition Separately from Voiced and Unvoiced Sound for Emotional Interaction Robot, 2007]

Спасибо за внимание

aponomarenko@hse.ru, aponom84@gmail.com, stasysp.96@gmail.com