



EmotiW 2017

Обзор конкурса

- 2 под-конкурса:

- a. *Group-level Emotion Recognition*

- b. *Audio-video Emotion Recognition*

Мы выбрали *Group-level emotion recognition*.

- Данные взяты из набора **Group Affect Database**, содержат изображения свадеб, похорон, вечеринок, совещаний и т.п., которым присвоен один из трех классов: *Positive, Neutral, Negative*.
- Данные поделены на три подмножества: *Train, Validation* и *Test*. Все подмножества доступны для участников, метки доступны только для *Train* и *Validation*. Метрика качества - **classification accuracy**.
- Baseline решение было получено с помощью вычисления дескрипторов CENTRIST для полного изображения + **SVM**. Метрика качества для baseline решения - **52.97%** на валидационном наборе и **53.62%** на тестовом.
- Разрешается ~~нать~~ семь сабмишнов, ручная проверка (с временным лагом) нет публичной *Liderboard*.

Очевидное решение

Обучить (training / fine-tuning) нейросетевую модель напрямую предсказывать класс напрямую по изображению (whole image / image crops)

Недостатки такого решения:

- тональность изображениям передают лица людей (конкретнее - суперпозиция эмоций людей на сцене), а не контекст;
- потеря информации при сжатии изображения до “канонического” размера;
- метка тональности субъективна, обучаться субъективности нельзя.

64+%

Предложенное решение

Идея: научиться достоверно определять (найти подходящие дескрипторы) эмоцию человеческого лица, затем подобрать такую их агрегацию, чтобы перейти от лиц к сцене

- Можно использовать нейросети в качестве дескрипторов, например *Levi et al.*, *ImageNet*, *VGG-Faces*
- Можно использовать embeddings с разных слоев сети, например *fc1*, *fc2* или *average pooling*
- Какой способ агрегации подходит для нейросетевых эмбеддингов? Кажется, что медиана лучше, чем среднее

-
- SVM, PCA, Random Forest, GBT, L1 / L2 нормализация
 - *fine-tuning* не дает прироста
-

67+%

69-70%

Эволюция выбранного решения

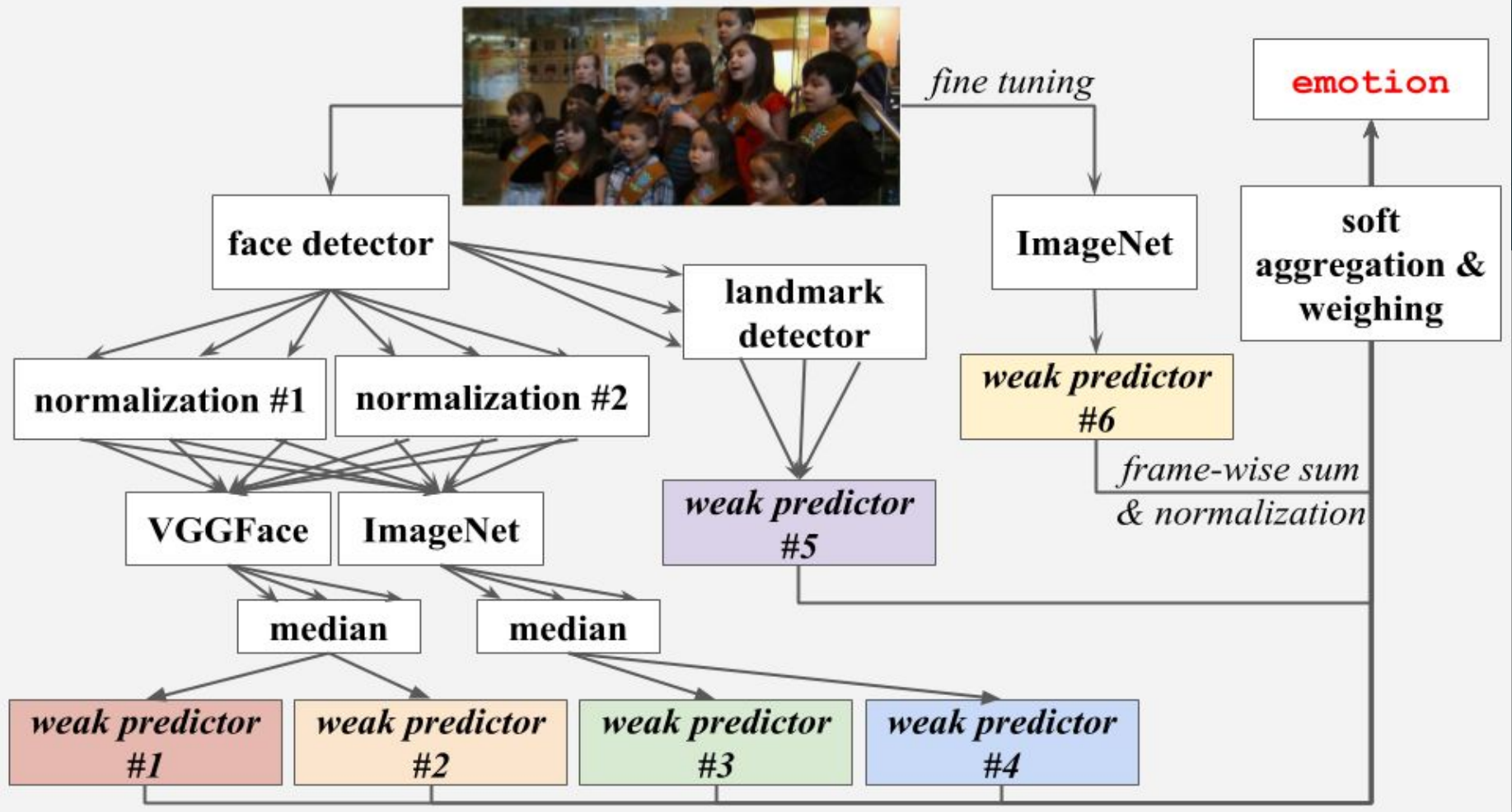
Insights:

- смена 'B' и 'R' каналов может дать прирост;
 - не на всех фото можно найти лица...
 - *ImageNet / VGG-Faces* + RGB / BGR = **ensemble it!**
 - Лицевые расстояния могут описать эмоции? **ensemble it!**
-

72+%

- слабые классификаторы необходимо взвесить;
 - подбор гиперпараметров;
 - End-to-end learning для изображений, где лица не определяются. **ensemble it? YES!**
-

75+%



Validation score vs. Test score

	val set	test set	Описание
submission #1	72+%	78.53%	<i>Полный ансамбль + ручная разметка для пропущенных лиц</i>
submission #2	72+%	78.53%	<i>То же, что и #, но больше деревьев</i>
submission #3	72+%	73.82%	<i>Базовое решение + рандом для пропущенных лиц</i>
submission #4	75+%	77.98%	<i>Базовое решение + end-to-end в ансамбль + end-to-end для пропущенных лиц</i>
submission #5	75+%	75.62%	
submission #6	64+%	62.74%	<i>Алексей попробовал end-to-end</i>

Решения других участников

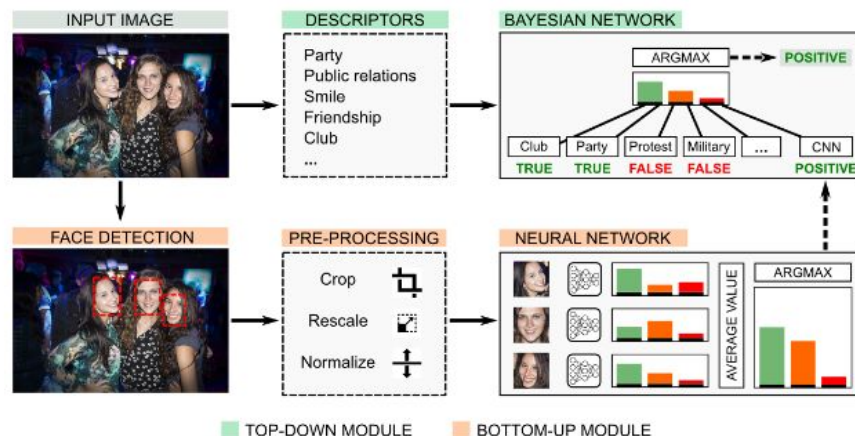
- L. Surace et al. [Emotion Recognition in the Wild using Deep Neural Networks and Bayesian Classifiers](#)

- Bottom-Up:

1. Face detection
2. Face pre-processing
3. CNN forward pass

- Top-Down:

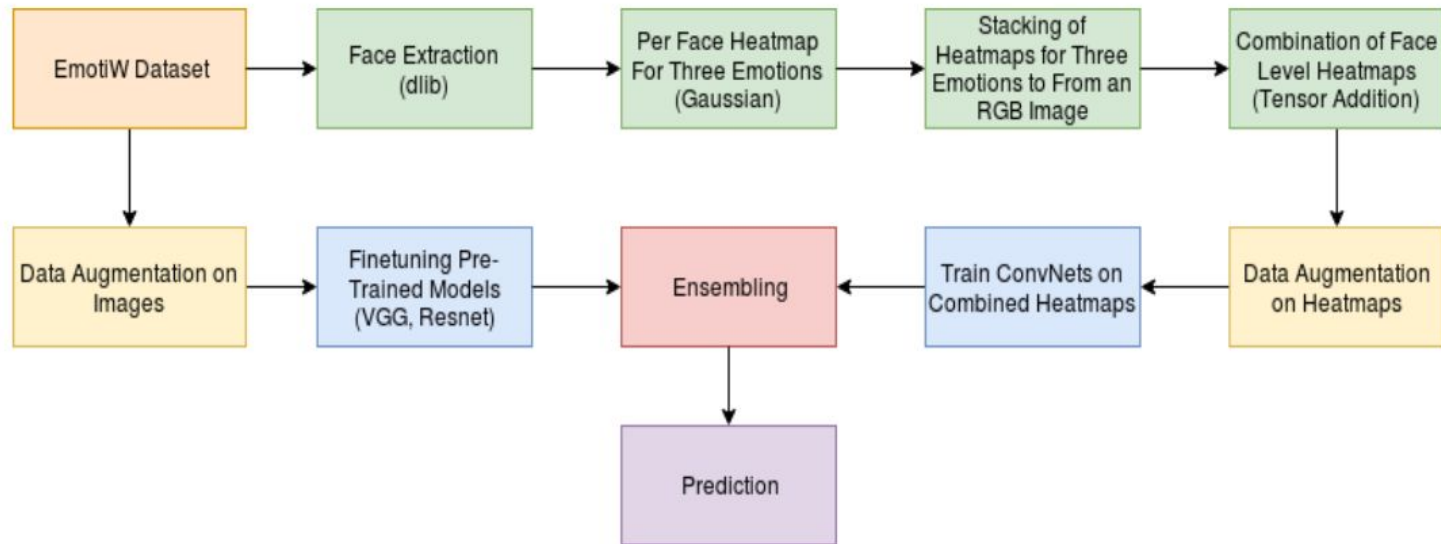
1. Acquiring the scene descriptors
2. Set evidences in the BN
3. Estimate the posterior distribution of the BN



64.68% (test)

Решения других участников

- S. Shamsi et al. [Group Affect Prediction Using Emotion Heatmaps and Scene Information](#)



70.64% (test)

Итоги участия

- 4е место с результатом **75.4%** на валидационном наборе - **22+%** лучше *baseline*, **78.53%** на тестовом наборе - **25+%** лучше *baseline*

Liderboard:

Group-level Emotion Recognition		Audio-video Emotion Recognition	
SIAR	80.89%	ILC	60.34%
UDI-GPB	80.61%	NTechLab	60.03%
BNU	79.79%	BUPT	59.72%

- публикация в сборнике ACM ICMI
- никто не поехал на конференцию в Глазго..

ACM ICMI 2017

EmotiW 2017 challenge @ 17th November 2017

From Individual to Group-Level Emotion Recognition: EmotiW 5.0

Abhinav Dhall, Roland Goecke, Shreya Ghosh, Jyoti Joshi, Jesse Hoey and Tom Gedeon

Keynote talk - Prof. Nadia Berthouze

Multi-modal Emotion Recognition using Semi-supervised Learning and Multiple Neural Networks in the Wild

Dae Ha Kim, Min Kyu Lee, Dong Yoon Choi and Byung Cheol Song

Coffee break

Modeling Multimodal Cues in a Deep Learning-Based Framework for Emotion Recognition in the Wild

Stefano Pini, Olfa Ben Ahmed, Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara and Benoit Huet

Group-Level Emotion Recognition using Transfer Learning from Face Identification

Alexandr Rassadin, Alexey Gruzdev and Andrey Savchenko

Group Emotion Recognition with Individual Facial Emotion CNNs and Global Image Based CNNs

Lianzhi Tan, Kaipeng Zhang, Kai Wang, Xiaoxing Zeng, Xiaojiang Peng and Yu Qiao

Learning Supervised Scoring Ensemble for Emotion Recognition in the Wild

Ping Hu, Dongqi Cai, Shandong Wang, Anbang Yao and Yurong Chen

Group Emotion Recognition in the Wild by Combining Deep Neural Networks for Facial Expression Classification and

Asad Abbas and Stephan K. Chalup

Temporal Multimodal Fusion for Video Emotion Classification in the Wild

Valentin Vielzeuf, Stéphane Pateux and Frederic Jurie

Audio-Visual Emotion Recognition using Deep Transfer Learning and Multiple Temporal Models

Xi Ouyang, Shigenori Kawaai, Ester Goh, Shengmei Shen, Wan Ding, Huaiping Ming and Dong-Yan Huang

Lunch break

Keynote talk 2

A New Deep-Learning Framework for Group Emotion Recognition

Qinglan Wei, Yijia Zhao, Qihua Xu, Liandong Li, Jun He, Lejun Yu and Bo Sun

Emotion Recognition in the Wild using Deep Neural Networks and Bayesian Classifiers

Luca Surace, Massimiliano Patacchiola, Elena Battini Sonmez, William Spataro and Angelo Cangelosi

Emotion Recognition with Multimodal Features and Temporal Models

Shuai Wang, Wenxuan Wang, Jinming Zhao, Shizhe Chen, Qin Jin, Shilei Zhang and Yong Qin

Group-Level Emotion Recognition using Deep Models on Image Scene, Faces, and Skeletons

Xin Guo, Luisa Polania and Kenneth Barner

Multi-Level Feature Fusion for Group-Level Emotion Recognition

Balaji B. and V. Ramana Murthy Oruganti

Challenge result discussion