

**EmotiW 2017**

# **Group-Level Emotion Recognition using Transfer Learning from Face Identification**

**AMD team:**

**Alexandr Rassadin, Alexey Gruzdev, Andrey Savchenko,**

**National Research University *Higher School of Economics***

**2017**

# Agenda

- Introduction
- Face Detection and Facial Landmarks Estimation
- Deep Emotion Descriptors
- Proposed Approach
- Experimental Results
- Participation Results

# Introduction

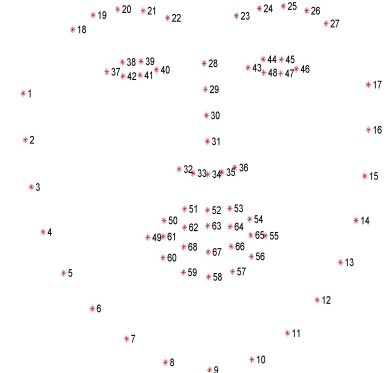
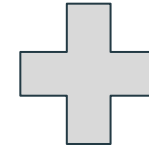
- image-based analysis of human activity and social behavior attracts significant attention in the computer vision area
- faces are more expressive reflection of the human emotions
- increased interest to *group-level* emotion in images last years: crowd analytics for security and social needs

Proposed solution consists of:

- face detection and facial landmarks identification
- transfer learning technique to recognize facial emotion descriptors from model trained on rather different task
- end-to-end CNN-model for emotion recognition from the whole photo

# Face Detection and Facial Landmarks Estimation

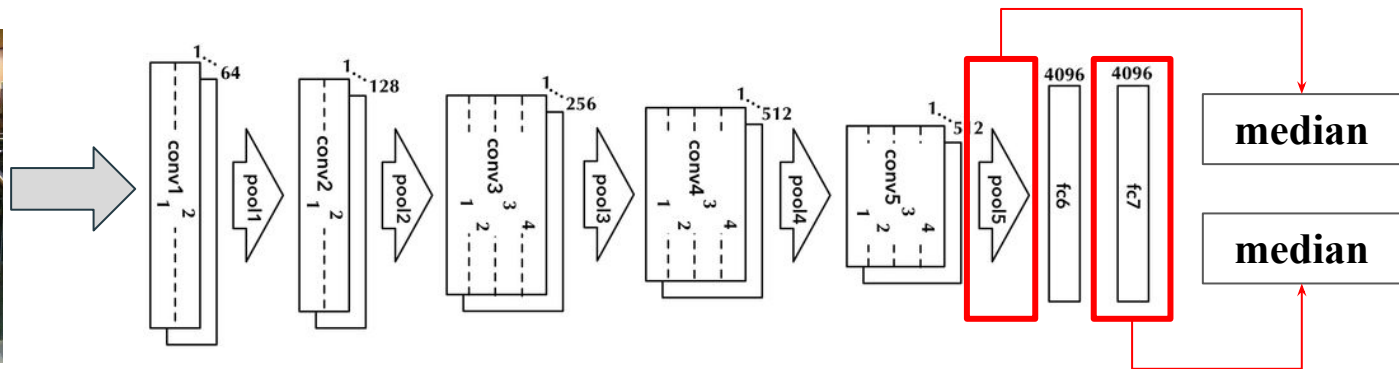
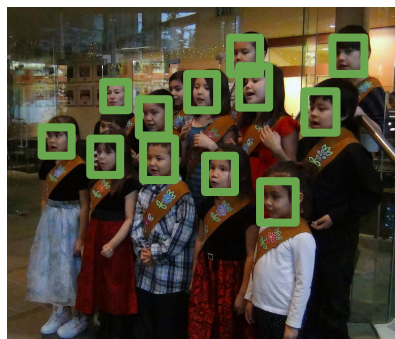
- Viola-Jones cascade classifiers based on the Haar features (**OpenCV** library)
- linear classifiers over an image pyramid in sliding window manner based on the Histogram of Oriented Gradients descriptor (**DLIB** library)
- ResNet-101 CNN trained on WIDER FACE dataset (**TinyFace** detector)
- **DLIB** landmark detector (68 facial landmarks)



# Deep Emotion Descriptors

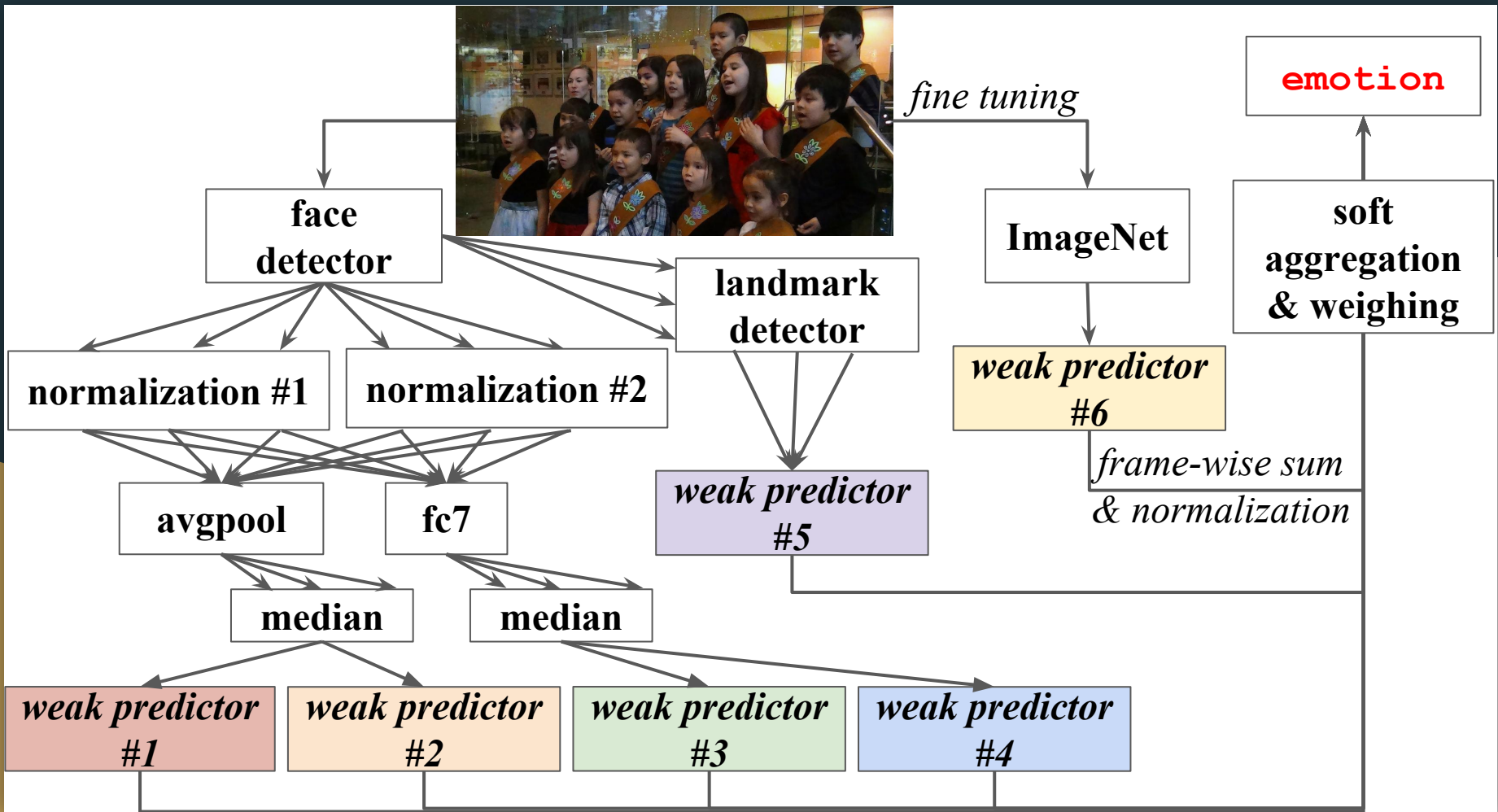
VGG Face neural network pre-trained for face recognition using the large VGG face dataset:

- 512D vector from the *avgpool* layer
- 4096D vector from the *fc7* layer
- both BGR and RGB regimes
- `median` as the transition from the personal descriptors to the group one



# Proposed Approach

- ensemble of four Random Forest classifiers trained on VGGFace *avgpool* and *fc7* features for RGB and BGR representation of the face
- an additional Random Forest classifier was trained on the normalized landmark distances
- fine-tuned a VGG16 model pre-trained on ImageNet dataset for end-to-end emotion classification for the full image (separate classifier for the case when face was not found and the weak classifier in the ensemble)
- weighting weak classifiers in the ensemble based on their accuracy estimated on the validation data



# Experimental Results

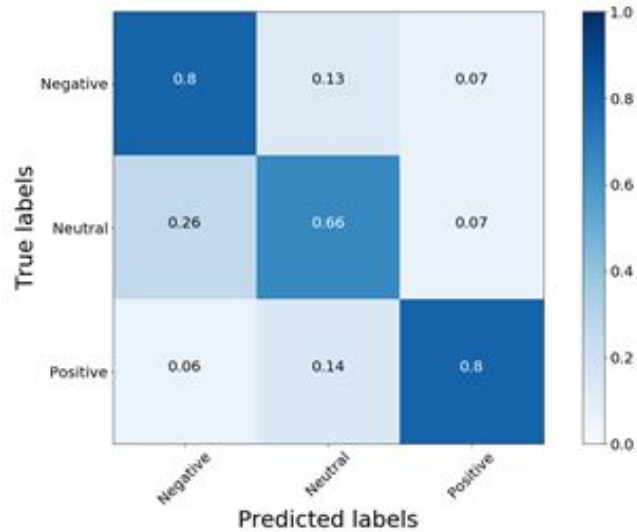
Model	Accuracy
VGG-16 (end-to-end)	64.11%
VGG-19 (end-to-end)	27.64%
VGGFace (end-to-end)	65.42%
VGGFace features, SVM	64.90 %
ResNet-50 (end-to-end)	62.65%
Xception (end-to-end)	60.18%
VGGFace features, Logistic Regression	67.24%
VGGFace features, SVM	65.41%
VGGFace features, MLP (1 hidden layer)	68.09%

Model	Accuracy
VGGFace features (RGB, fc7), Random Forest	67.62%
VGGFace features (BGR, fc7), Random Forest	68.18%
VGGFace features (RGB, avgpool), Random Forest	65.42%
VGGFace features (BGR, avgpool), Random Forest	64.90 %
Landmark features, Random Forest	62.65%
VGG-16 for the whole image	65.89%
Proposed ensemble (except VGG-16 for the whole image)	72.77%
Proposed ensemble, Tiny faces	66.51%
<b>Proposed ensemble</b>	<b>75.39%</b>



# Participation Results

- 4th place with **75.4%** accuracy on validation set - **23+%** better than the *baseline*, **78.53%** on test set - **25+%** better than the *baseline*



Leaderboard:

Group-level Emotion Recognition	
SIAR	80.89%
UDI-GPB	80.61%
BNU	79.79%
<b>AMD</b>	<b>78.53%</b>
AmritaEEE	75.07%

<https://github.com/arassadin/emotiw2017>



**Thank you for your attention!**