Deep Features for First Impression Analysis

July 15, 2016

1 Team details

- ITU SiMit
- Mehmet Aygün
- Istanbul Technical University Computer Engineering Department , +90 543 375 13 63, aygunme@itu.edu.tr
- Rafik Can Mallı, Hazım Kemal Ekenel
- simitlab.itu.edu.tr
- Istanbul Technical University

2 Contribution details

- Title of the contribution Deep features for First Impression Analysis
- Final score Not avaliable
- General method description 80 Different SVR trained using FC7 features of frames and faces.
- References
  - VGG-16 : Very Deep Convolutional Networks for Large-Scale Image Recognition arXiv technical report, 2014
  - VGG-Face : Deep Face Recognition British Machine Vision Conference, 2015
  - Dlib : Davis E. King. Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research 10, pp. 1755-1758, 2009

3 Visual Analysis

3.1 Face Detection Stage

3.1.1 Features / Data representation

We use Dlib library face detector for face detection.

4 Personality Trait recognition from Visual data

4.1 Features / Data representation

We extract features from both faces and frames. For feature extraction we use VGG-16 and VGG-Face models and use their FC7 layer representations.

4.2 Learning strategy

We train these FC7 features using SVR with RBF kernels. For each 5 different Trait we train 2 different model one for frame one for face

4.3 Data Fusion Strategies

We split video two 8 slices and pick 2 frames from each slices. For each slice we train 2 different SVR model. So totaly 5 x 2 x 8 = 80 different SVR model trained. Then all of them fused using averagin each of them.

4.4 Global Method Description

- Which pre-trained or external methods have been used : VGG 16 and VGG-Face models used.
- Only Face or only frame features cannot achieve more score than fused method on validation data.

5 Other details

- Language and implementation details : Debian 8 , Caffe, Dlib, Sklearn libraries used for both training and testing.
- Firstly fc7 features should be extracted both from frame and faces in train and test set. Then 8 different model should perform on faces and 8 different on frames. Then each of the 8 frame and 8 face skores combined in itself. Then each of the 2 different skores should be combined using averaging.
- For trainin it takes 4-5 hours for testing it is approximatly 1 hour.

- Good data, good competitions but web site should be improved. Some times we couldn't do validation.