Hierarchical age estimation using deep CNN features

September 15, 2015

1 Team details

- Team name : UMD
- Team leader name : Azadeh Alavi*
- Team leader address, phone number and email : azadeh@umiacs.umd.edu, +1(240) 581 2498
- Rest of the team members : Rajeev Ranjan*, Amit Kumar*, Sabrina Zhou**, Jun Cheng Chen*, Vishal M. Patel***, and Rama Chellappa*
- Team website URL
- Affiliation : *. University of Maryland, **.(Montgomery Blair High School, Silver Spring, MD), ***. Rutgers University

2 Contribution details

- Title of the contribution : Hierarchical age estimation using deep CNN features
- Final score
- General method description : The method consists of four different modules: 1) Face Detection, 2) Face Alignment, 3) Features for Face recognition, and 4) Regressor for age estimation. We use the face detection algorithm described in [1]. The dlib implementation of [3] was used for face alignment. The features for age estimation were obtained from the $conv_5$ layer of the CNN model described in [2]. These features were used to train a 3-layer CNN regression model with the gaussian loss function defined for the challenge. At test time, the regression model was used to predict the age for the given face verification feature.
- References

[1] Rajeev Ranjan, Vishal M. Patel and Rama Chellappa. A Deep Pyramid Deformable Part Model for Face Detection *CoRR*, abs/1508.04389, 2015.

[2] Jun-Cheng Chen, Vishal M. Patel and Rama Chellappa. Unconstrained Face Verification using Deep CNN Features *CoRR*, abs/1508.01722, 2015.

[3] V. Kazemi, and J. Sullivan. One Millisecond Face Alignment with an Ensemble of Regression Trees *IEEE Conference on Compute Vision and Pattern Recognition (CVPR)*, 2014.

[4] V. Jain and E. Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst, 2010.

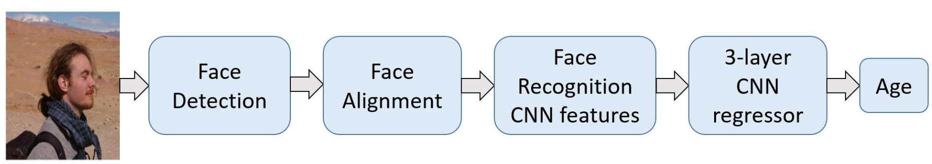
[5] E. Eidinger, R. Enbar, and T. Hassner. Age and Gender Estimation of Unfiltered Faces *IEEE Transactions on Information Forensics and Security*, 2014.

[6] Karl Ricanek Jr and Tamirat Tesafaye. MORPH: A Longitudinal Image Database of Normal Adult Age-Progression *IEEE 7th International Conference on Automatic Face and Gesture Recognition*, 2006.

[7] P. Thukral, K. Mitra, and R. Chellappa. A hierarchical approach for human age estimation *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012.

[8] Wu Tao, P. Turaga, and R. Chellappa Age Estimation and Face Verification Across Aging Using Landmarks *IEEE Transactions on Information Forensics and Security*, 2012

- Representative image / diagram of the method



- Describe data preprocessing techniques applied (if any)

3 Face Detection Stage

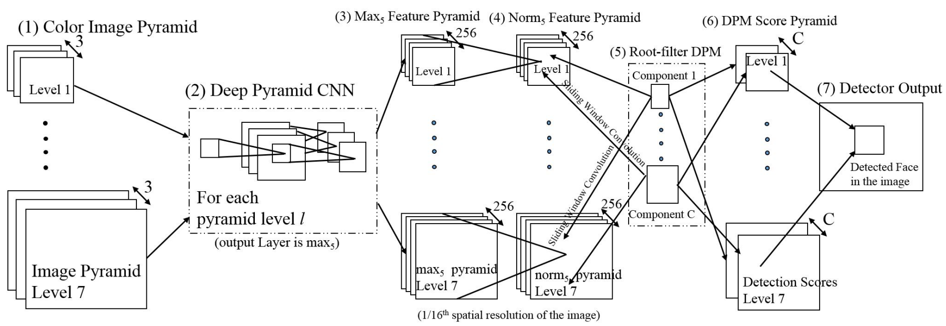
3.1 Features / Data representation

A deep pyramid deformable part model was used for face detection. It consists of multi-resolution deep CNN feature pyramid fed to a DPM.

3.2 Dimensionality reduction

No dimensionality reduction techniques were used.

3.3 Compositional model



3.4 Learning strategy

A root-only DPM was learned using linear SVM on top of deep CNN feature pyramid for FDDB dataset [4].

3.5 Other techniques

A single detection location, which provided maximum SVM score, was considered as candidate face.

3.6 Method complexity

The method takes 0.7s to detect a face on a GPU and 15s on a CPU

4 Face Landmarks Detection Stage

4.1 Features / Data representation

We use the dlib implementation of the face alignment method [3] with an ensemble of regression trees.

4.2 Dimensionality reduction

No dimensionality reduction techniques were used.

4.3 Compositional model

For the compositional model used, please refer to [3]

4.4 Learning strategy

We use the dlib implementation of the face alignment method [3] with an ensemble of regression trees.

4.5 Other techniques

No other techniques were used.

4.6 Method complexity

The method takes one second to compute landmarks for each face.

5 Apparent Age Estimation Stage

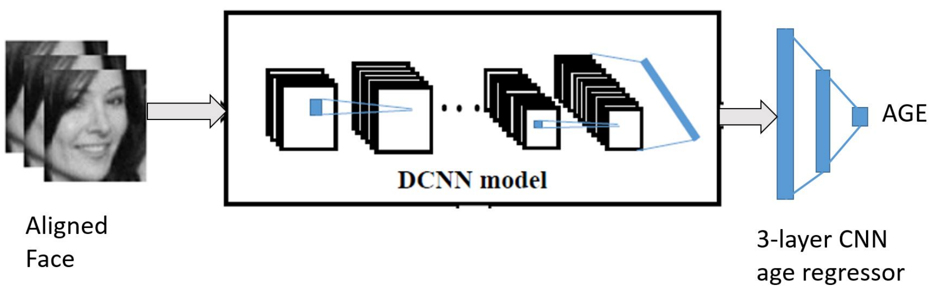
5.1 Features / Data representation

Features from $conv_5$ layer of CNN trained for face verification were used for apparent age estimation.

5.2 Dimensionality reduction

No dimensionality reduction techniques were used.

5.3 Compositional model



5.4 Learning strategy

The pretrained face verification CNN was used for feature extraction as described in [2]. A three layered regression CNN was learned to minimize the gaussian age loss using the training and the validation data from the challenge, as well as data for specific age groups from [5] and [6].

5.5 Other techniques

Three age regressors, each for young-age group, mid-age group and old-age group, were trained separately. The method was inspired by [7]. However, we found that using a CNN based regressor on top of face verification features improves the performance by large margin. A model was trained to predict the age group a face belongs, after which the corresponding age regressor was used to predict the apparent age.

5.6 Method complexity

The method takes one second to compute age for each face on a GPU, and 10s on a CPU.

6 Global Method Description

- Total method complexity : end to end age estimation takes about 3 seconds per image
- Which pre-trained or external methods have been used (for any stage, if any) : The CNN network for face detection is pre-trained with Imagenet classification dataset. The CNN network for face verification features is pre-trained with CASIA dataset.
- Which additional data has been used in addition to the provided ChaLearn training and validation data (at any stage, if any) : Adience OUI [5] and MORPH [6] datasets have been used in addition to the challenge training and validation data.
- Qualitative advantages of the proposed solution : The proposed solution addresses each age group separately. It shows that age forms an important factor in face verification features. Also, a three layered CNN based regressor performs better than a linear regressor or a support vector regressor.
- Results of the comparison to other approaches (if any) : We compared our method with the one described in [8], and obtained more than 20% improvement in the performance.
- Novelty degree of the solution and if is has been previously published : The proposed method is pretty novel in the sense that it leverages the face verification features for predicting age. Also, it uses a CNN based regressor trained with the gaussian loss function, which is appropriate with the challenge data.

7 Other details

- Language and implementation details (including platform, memory, parallelization requirements) : Implementation was done using MATLAB-2014a and CAFFE. The platform used was linux. 4GB memory is required without any parallelization.

- Human effort required for implementation, training and validation? : The method is fully automated with minimal human effort.
- Training/testing expended time? : The total training time with both training and validation data is about 3 hours. The test time is about 3s per image.
- General comments and impressions of the challenge? what do you expect from a new challenge in face and looking at people analysis? : The challenge was conducted really well. It helped us a lot in understanding more about facial factors contributing to age. We expect the new challenge to be a multi-task problem where other face attributes needs to be estimated as well.