Apparent and real age estimation in still images with deep residual regressors on APPA-REAL database

Eirikur Agustsson[1], Radu Timofte[1,2], Sergio Escalera[3,4,6], Xavier Baro[4,5], Isabelle Guyon[6,7], Rasmus Rothe[2]

[1] Computer Vision Lab, D-ITET, ETH Zurich, Switzerland
[2] Merantix GmbH, Berlin, Germany
[3] Dept. Mathematics and Computer Science, UB, Spain
[4] Computer Vision Center, UAB, Barcelona, Spain
[5] EIMT, Open University of Catalonia, Barcelona, Spain
[6] ChaLearn, California, USA
[7] Universite Paris-Saclay, France

FG 2017
Washington, DC.

This work has been partially supported by the ETH General Fund (OK), European Research Council project VarCity (\#273940), a NVIDIA GPU grant, Spanish projects TIN2015-66951-C2-2-R and TIN2016-74946-P (MINECO/FEDER, UE) and CERCA Programme / Generalitat de Catalunya.

Overview

APPA-REAL Database

Apparent vs. Real age prediction

Residual DEX regression

Quantitative Results

Model Visualization

APPA-REAL Database

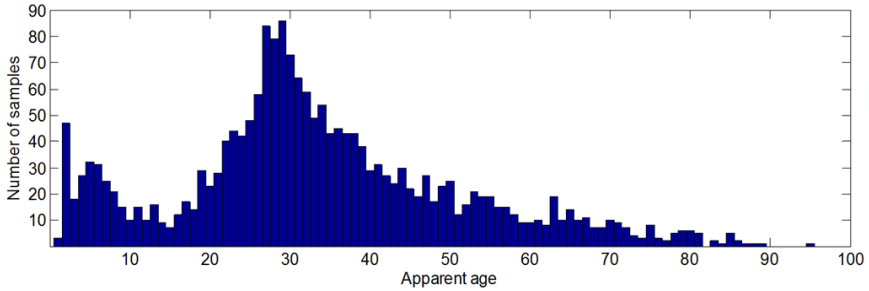
7,591 Faces

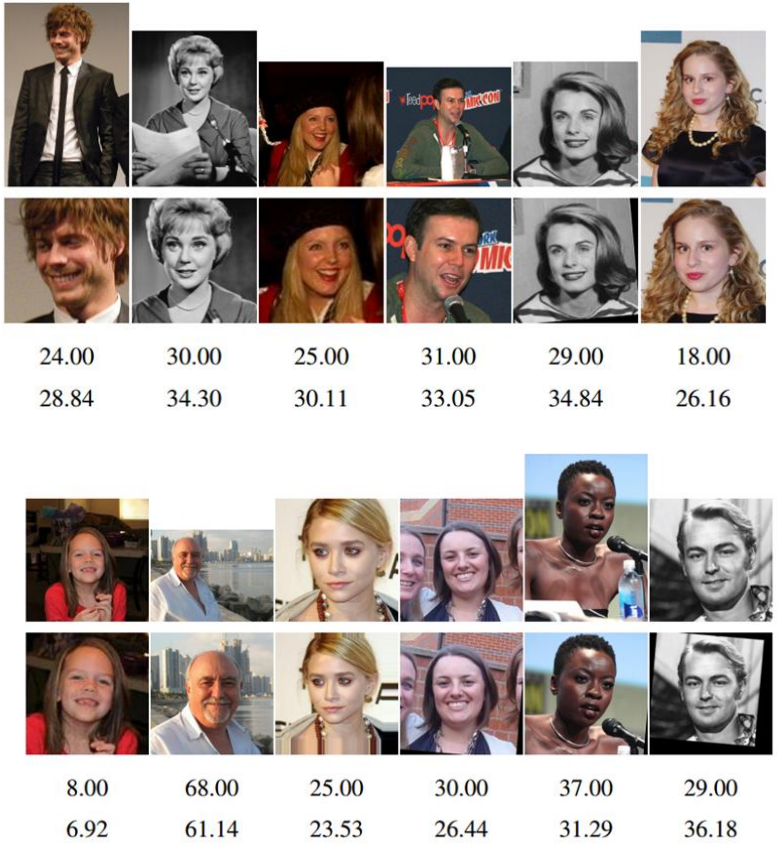
Real Age Labels

260,659 Apparent Age Labels

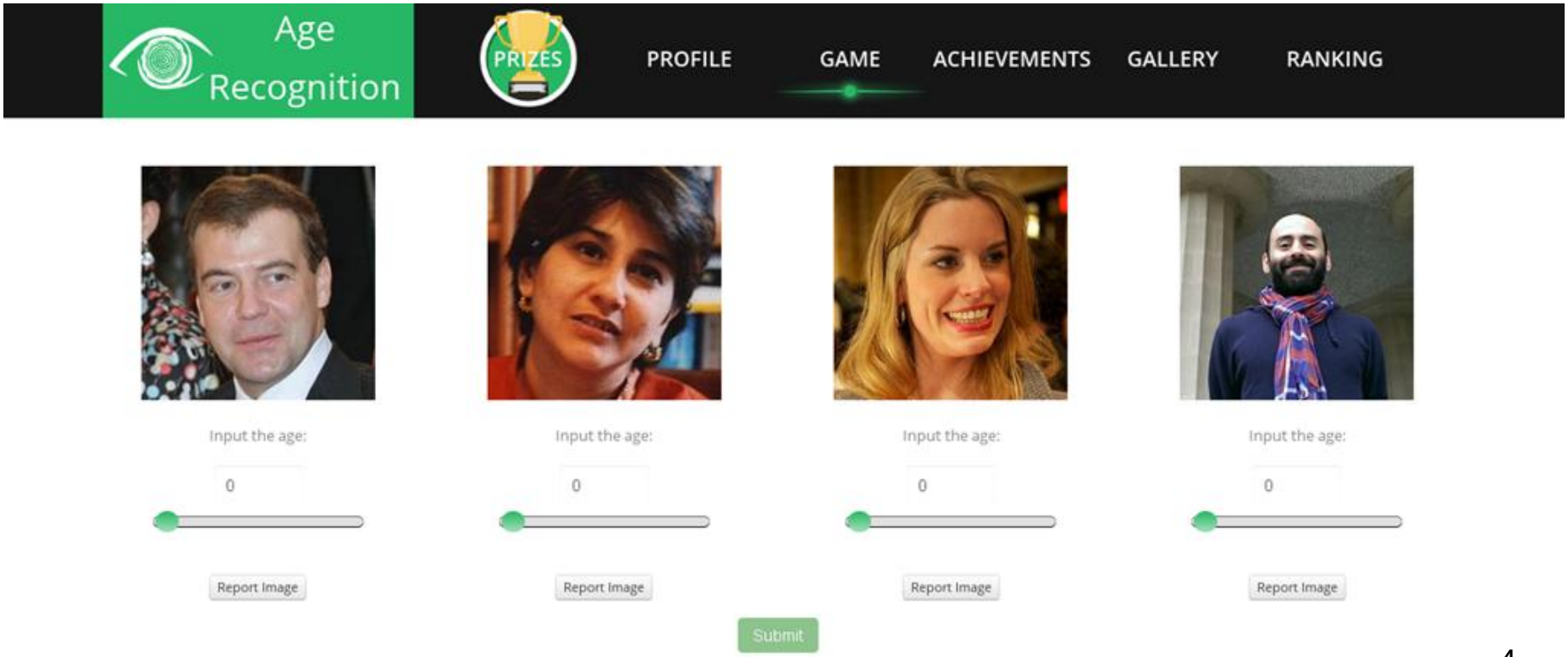
Crowdsourced over the Internet

http://bit.ly/APPA-REAL



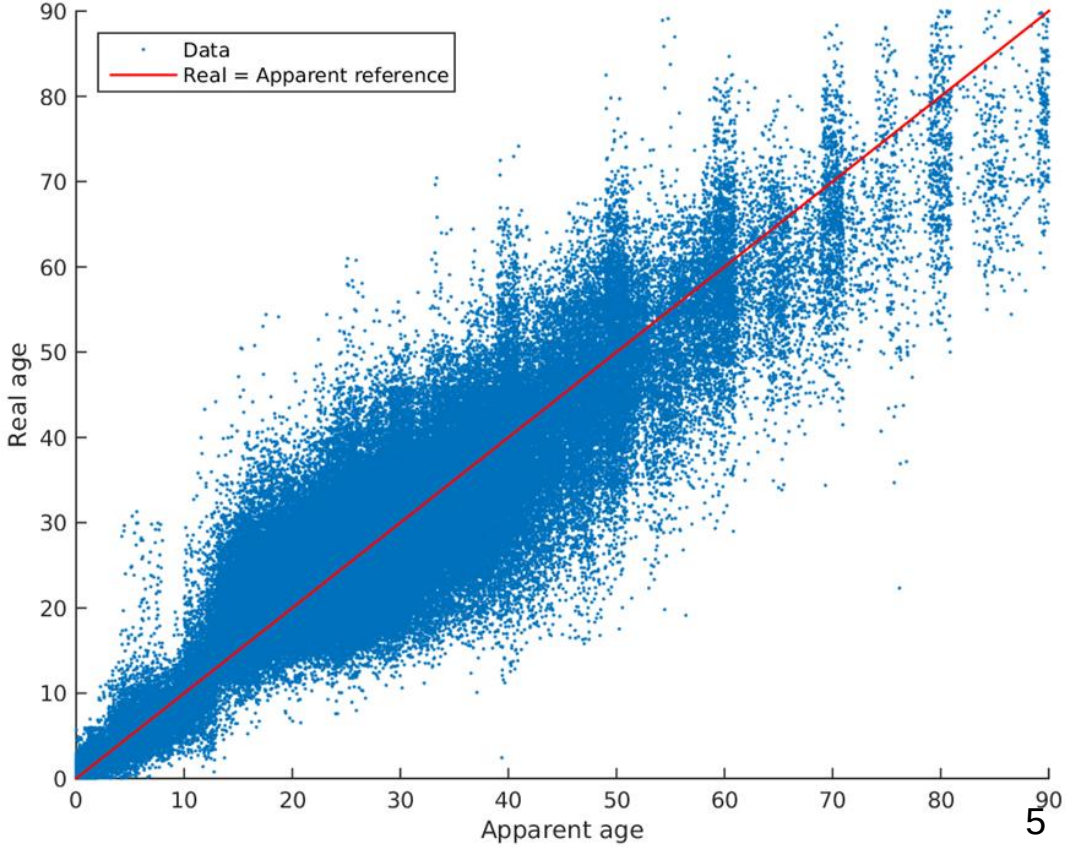


Data Collection Framework



Apparent vs Real Age

Apparent Labels are noisy



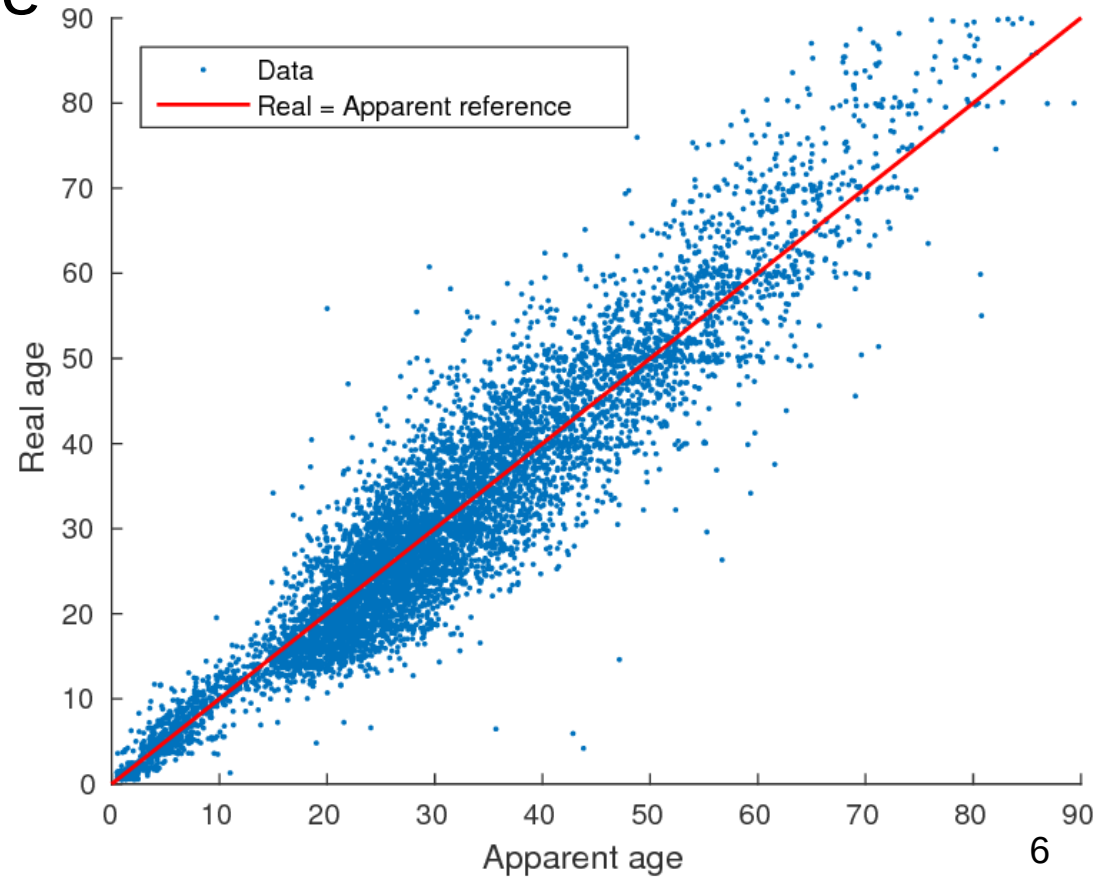
Apparent vs Real Age

Apparent Labels are noisy

Average Apparent age is stable

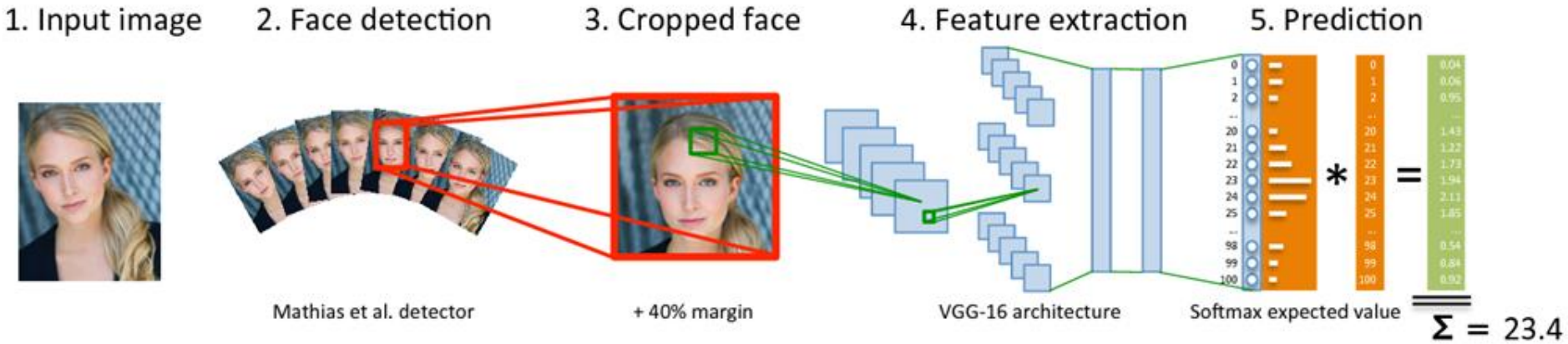
Only depends on the image (with enough ratings)

Can one help predicting the other?



DEX Age Prediction Pipeline

(Rothe *et al.*, ICCV2015, IJCV2016)



- Pre-trained on IMDB-WIKI, 0.5M face images w/ noisy age labels
- Finetuned on APPA-REAL with same learning parameters in all experiments

$$\mathrm{D}(I) = \sum_{j=1}^{Y} p_j(I) y_j$$

Predicting Real Age with Apparent

Real GT is not best predictor of Apparent

"Wisdom of the crowd" outperforms DEX

Apparent DEX marginally outperforms Real DEX with SVR adjustment

Method	MAE Apparent
Apparent GT	0
Real GT	4.573
Apparent DEX	**4.082**
Real DEX	4.513

Method	MAE Real
Real GT	0
Apparent GT (“wisdom of the crowd”)	4.573
Real DEX	5.468
Apparent DEX	5.729
Apparent DEX + SVR	**5.426**

How can we better utilize the

Real and Apparent labels?

Residual DEX

Train regressor D_1 for apparent age

Train regressor D_2 on the residuals:

$$r_i = a_i - D_1(I_i)$$

where a_i is the real age of face image I_i

Final prediction: $D_1(I_i) + D_2(I_i)$

Hope D_2 picks up facial features not captured by apparent age

Results for Real Age Prediction

Best result obtained with
Residual DEX on top of
Apparent DEX

Still a significant gap of ~0.7
years to the
"Wisdom of the crowd"

Method	MAE Real
Real GT	0
Apparent GT (“wisdom of the crowd”)	4.573
Real DEX	5.468
Apparent DEX	5.729
Apparent DEX + SVR	5.426
Apparent + Residual DEX	**5.296**

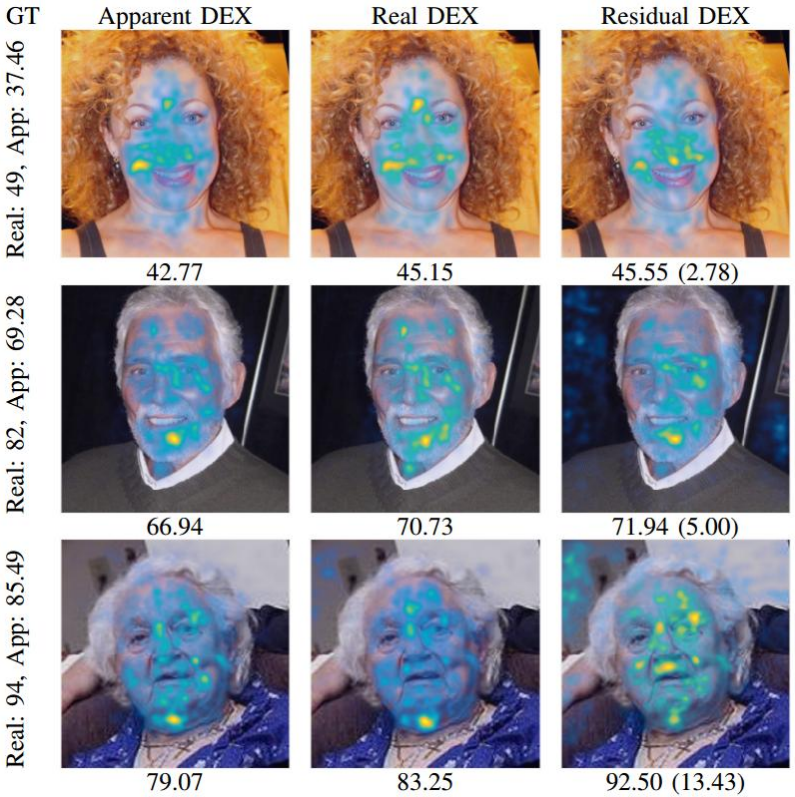
Example Predictions

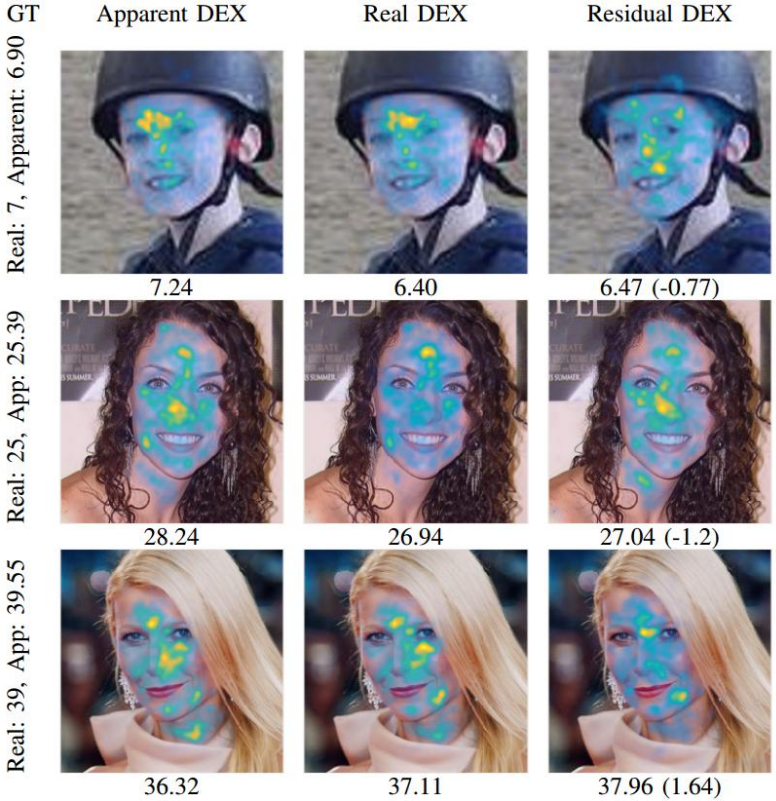
Input						
Cropped Face						
GT Real	24.00	30.00	25.00	31.00	29.00	18.00
GT Apparent	28.84	34.30	30.11	33.05	34.84	26.16
Apparent DEX	26.04	29.28	28.69	30.33	32.76	23.57
Residual DEX	-2.04	0.12	-2.48	-1.04	-1.38	-2.40
Apparent + Residual DEX	24.00	29.40	26.21	29.29	31.38	21.17

Example Predictions

Input						
Cropped Face						
GT Real	8.00	68.00	25.00	30.00	37.00	29.00
GT Apparent	6.92	61.14	23.53	26.44	31.29	36.18
Apparent DEX	4.98	59.32	20.42	24.78	29.26	40.03
Residual DEX	-0.95	3.80	-1.27	-1.84	-1.04	0.69
Apparent + Residual DEX	4.03	63.12	19.15	22.94	28.22	40.72

Where are the models looking?





Conclusions

We proposed and studied APPA-REAL - the first database with both apparent and real age labels.

We showed how apparent age can help when predicting real age.

We proposed Residual DEX for incorporating both apparent and real age labels.

The "Wisdom of the crowd" apparent age prediction sets a new reference that has yet to be outperformed with ML models.