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

Eirikur Agustsson, Radu Timofte, Sergio Escalera, Xavier Baró, Isabelle Guyon, Rasmus Rothe

# 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, Paris, France

***Abstract*— After decades of research, the real (biological) age estimation from a single face image reached maturity thanks to the availability of large public face databases and impressive accuracies achieved by recently proposed methods. The estimation of "apparent age" is a related task concerning the age perceived by human observers. Significant advances have been also made in this new research direction with the recent Looking At People challenges. In this paper we make several contributions to age estimation research. (i) We introduce APPA-REAL, a large face image database with both real and apparent age annotations. (ii) We study the relationship between real and apparent age. (iii) We develop a residual age regression method to further improve the performance. (iv) We show that real age estimation can be successfully tackled as an apparent age estimation followed by an apparent to real age residual regression. (v) We graphically reveal the facial regions on which the CNN focuses in order to perform apparent and real age estimation tasks.**

## I. Introduction

Automated face analysis is a research topic that has received much attention from the Computer Vision and Pattern Recognition communities in the past. Research progress has made us think of problems like face recognition or face detection to be solved for some scenarios. However, several issues of face analysis are still open problems (including the implementation of large scale face recognition/detection methods for real images), in which the community keeps making rapid progress, with the constant improvement of new published methods that push the state-of-the-art. Applications of interest include security and video surveillance, human computer/robot interaction, communication, entertainment, and commerce, while having an important social impact in assistive technologies for education and health.

Computational methods for face analysis are genuinely important in many applications and provide excellent benchmarks for algorithms. The recognition of continuous, natural human faces is very challenging due to the multimodal nature of the visual cues (e.g., movements of lips, facial expressions, eye blinking, etc. ), as well as technical limitations such as spatial and temporal resolution. Furthermore, facial expressions analysis and age estimation are hot topics in the field of Looking at People that serve as additional cues to determine human behavior and mood indicators.

Real age estimation in still images is a difficult task which requires the automatic detection and interpretation of facial features. Age estimation has historically been one of the most challenging problems within the field of facial analysis [31], [10]. It can be very useful for several applications, such as advanced video surveillance, demographic statistics collection, business intelligence and customer profiling, and search optimization in large databases. This field regained interest since 2006 with the availability of large databases like MORPH-Album [32], which increased by a factor of 55 the amount of real age-annotated data. Interestingly, the regression problem is often times turned into a classification problem into age segments, a seemingly easier problem (e.g. [17]). With the increased efficiency of "deep learning", such methods started being adopted since 2013 [23], [21]. However "conventional" methods based on manifold learning [22], support vector machines [14], [13], or related methods [6], [38] remain very popular for real age estimation.

Apparent age estimation is a more recent topic in the field of face and age analysis. Apparent age focuses on how old a subject *looks like*, which may be influenced by several factors, including real age, but also other biological and sociological factors of "aging", resulting sometimes in important departures from the real age. Most of currently available datasets only include real age labels, since collecting data for apparent age is laborious and requires to obtain multiple opinions for each image to capture the subjective and highly variable opinions of the labelers. Consequently, most age estimation papers tackle principally real age. In 2015, a new dataset based on apparent age was published for the Chalearn LAP competition (Round 1 for ICCV2015 [8], and Round 2 for CVPR2016 [9]), only considering apparent age labels.

To the best of our knowledge, most of the computer vision papers addressing the apparent age recognition problem are associated to these two ChaLearn competitions. In the published results (summarized in greater details in the paragraphs that follow), the participants applied face detector approaches, then applied various Deep learning architectures for feature extraction. The final apparent age estimation

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.

combined various strategies using late fusion to obtain the final age prediction. These two contests revealed the real power of deep learning for age estimation (at least for the feature extraction part).

Regarding apparent age recognition in the ICCV 2015 competition, in [33], face detection was performed using [25], and 20 CNN models were applied to the cropped faces. The final value was extracted from 101 softmax-normalized output neurons. In [24], face detection was performed using Boosting+Neural Networks and Face landmark detection using CFAN [41]. They used a GoogleNet model and predictions were based on three cascade CNN (face classification, real age and apparent age). In [18] the authors used a commercial software for face detection. They used a CNN VGG model, using a fusion of regressors for age prediction (lasso, global and local quadratic regress regressor, and random forest). Finally, in [42] the authors used Face++ [16] for face and landmark detection. They also used GoogleNet to extract deep features then fed into a mixture model of 10 age groups, each predictor being based on a combination of RF and SVR.

In relation to the apparent recognition methods, in the CVPR 2016 competition, the top ranked participants used a VGG-16 [28] pre-trained model. In [2] the authors first performed face detection, pose estimation and face alignment process. Then, a two-phase learning based on CNN models was used, one for age estimation and the second one for children age prediction. The authors of [15] used an ensemble of four fine-tuned CNN models, that were employed to extract the last full connected features, which were used by an ensemble method to generate the final result. Finally, [40] used [25] for face detection, and then an ensemble of 8 SO-SVM classifiers learned on the features from the last layer of VGG-16 network for age prediction.

Most very recent top methods described above were introduced for apparent age estimation (ICCV 2015 and CVPR 2016 competitions), however, since both apparent and real age estimation start from the same face images and are intimately related, it is rather straightforward to extend a method developed for one task to the other. This has been verified for the DEX method in [34]. While initially introduced for apparent age estimation, DEX shows state-of-the-art results also on group age estimation (OUI-Adience database) and real age estimation (FG-NET and MORPH2 databases) with minimal changes involving training data and adaptation of the range of age labels.

Different application scenarios can benefit from learning systems that predict the apparent age, such as medical diagnosis (premature aging due to environment, sickness, depression, stress, fatigue, etc.), effect of anti-aging treatment (hormone replacement therapy, topical treatments), or effect of cosmetics, haircuts, accessories and plastic surgery, just to mention a few. Some of the reasons age estimation is still a challenging problem are the uncontrollable nature of the aging process, the strong specificity to the personal traits of each individual, high variance of observations within the same age range, and the fact that it is very hard to gather complete and sufficient data to train accurate models.

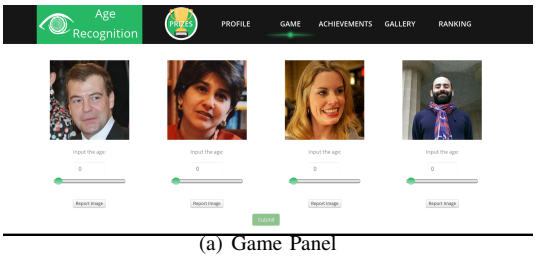
TABLE I
AGE-BASED DATABASES AND THEIR CHARACTERISTICS.

| Database | #Faces | #Subj. | Range | Age type | Controlled Environment |
|---|---|---|---|---|---|
| **FG-NET [20], [19]** | 1,002 | 82 | 0 - 69 | Real Age | No |
| **GROUPS [12]** | 28,231 | 28,231 | 0 - 66+ | Age group | No |
| **PAL [26]** | 580 | 580 | 19 - 93 | Age group | No |
| **FRGC [30]** | 44,278 | 568 | 18 - 70 | Real Age | Partially |
| **MORPH2 [32]** | 55,134 | 13,618 | 16 - 77 | Real Age | Yes |
| **YGA [11]** | 8,000 | 1,600 | 0 - 93 | Real Age | No |
| **FERET[29]** | 14,126 | 1,199 | - | Real Age | Partially |
| **Iranian face [3]** | 3,600 | 616 | 2 - 85 | Real Age | No |
| **PIE [35]** | 41,638 | 68 | - | Real Age | Yes |
| **WIT-BD [39]** | 26,222 | 5,500 | 3 - 85 | Age group | No |
| **Caucasian Face Database [4]** | 147 | - | 20 - 62 | Real Age | Yes |
| **LHI [1]** | 8,000 | 8,000 | 9 - 89 | Real Age | Yes |
| **HOIP [37]** | 306,600 | 300 | 15 - 64 | Age Group | Yes |
| **Ni's Web-Collected Database [27]** | 219,892 | - | 1 - 80 | Real Age | No |
| **OUI-Adience [7]** | 26,580 | 2,284 | 0 - 60+ | Age Group | No |
| **IMDBWIKI [34]** | 523,051 | 20,284+ | 0 - 100 | Real Age | No |
| **APPA-REAL (ours)** | **7,591** | **7,000+** | **0 - 95** | **Real and Apparent Age** | **No** |

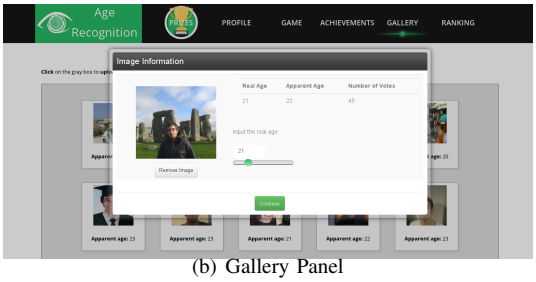
In this paper, to the best of our knowledge we (i) contribute with the first state of the art database with faces in the wild containing both real and apparent age annotations (Section II); (ii) analyze the relationship between real and apparent age (whose distribution is shown in Figure 3); (iii) develop a residual age estimator method (described in Section III-B) to further improve the performance on age estimation of the state-of-the-art DEX method [34] that won ICCV 2015 apparent age competition (Section III-A); (iv) we show for the first time that the real age estimation can be interpreted and successfully tackled as an apparent age estimation followed by an apparent to real age residual correction. By doing so, we can achieve superior performance to a standard (baseline) method using only the real age annotations. In Section IV we discuss the experimental setup and the achieved results and also provide a visualization tool of the sensitivity of the prediction model on a couple of images when trained for apparent, real, or real-apparent age estimation. Finally, Section V concludes the paper.

## II. APPA-REAL DATABASE

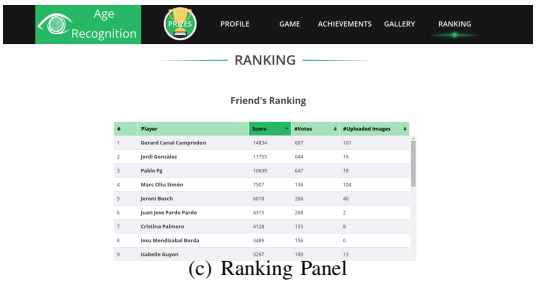
Due to the nature of the age estimation problem, there is a restricted number of publicly available databases providing a substantial number of face images labeled with accurate age information. Table I shows the summary of the existing databases with main reference, number of samples,




(a) Game Panel




(b) Gallery Panel




(c) Ranking Panel

Fig. 1. Age Recognition Application. (a) User can see the images of the rest of participants and vote for apparent age. (b) User can upload images and see their uploads and the opinion of the users regarding the apparent age of people in their images. (c) User can see the points he/she achieves by uploading and voting photos and the ranking among his/her friends and all the participants of the application.

number of subjects, age range, type of age and additional information. The large MORPH-Album 2 [32] database has extensively been used in recent works. However, all existing databases are based on real age estimation. In this work we present APPA-REAL, the first state-of-the-art database containing both real and apparent age labels (Last row in Table I).[1]

We collected the data to recognize the apparent age of people based on the opinion of many subjects using a new crowd-sourcing data collection and labeling application, data from the AgeGuess platform[2], as well as with the support of Amazon Mechanical Turk (AMT) workers. We developed a web application in order to collect and label an age estimation database online by the community. The application uses the Facebook API to facilitate access, hence reach more people with a broader background. We show some panels of the application in the Figure 1(a), 1(b) and 1(c).

The web application was developed so that the users get points for uploading and labeling images. The closer the age guess was to the apparent age the more points the player

[1]Database available at http://chalearnlap.cvc.uab.es/

[2]http://www.ageguess.org/

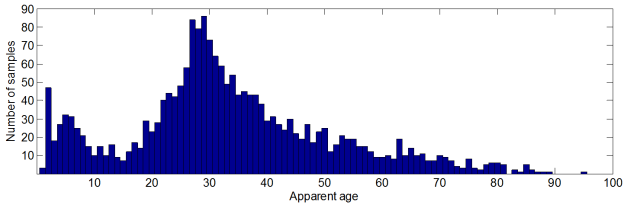



Fig. 2. Number of samples of the APPA-REAL database per apparent age. The age distribution is biased towards young adults, since the dataset is collected from public Internet repositories.

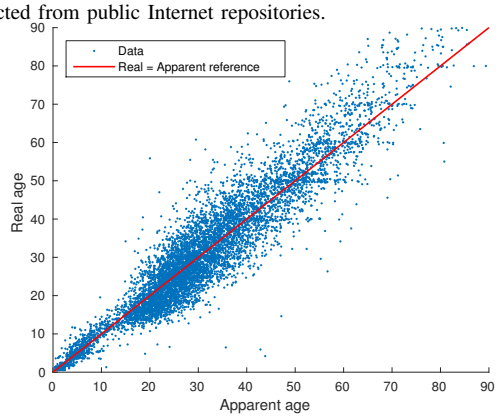



Fig. 3. Relationship between Apparent and Real age. The line Real = Apparent is shown for reference.

obtained. With the purpose of increasing the engagement of the players, we included two leaderboards: global and friends, where the users can check their position in the ranking in relation to their rest of users. Users were asked to upload images of a single person and we gave them tools to crop the image.

Images and their real and apparent votes collected from the designed application were combined with the ones donated by the AgeGuess platform. Furthermore, in order to increase the size of the database, additional images from Internet were uploaded to AMT and were labeled by many users (workers), assuring a minimum of 30 votes per image. In total, the new APPA-REAL database contains 7,591 images with associated real and apparent age labels. The total number of apparent votes is nearly 300,000. On average we have around 38 votes per each image and this makes our average apparent age very stable (0.3 standard error of the mean). For the apparent age, the data contains not only the mean apparent age but also the raw votes given by the raters after outliers removal. Last row in Table I shows some characteristics of the proposed database. The distribution of samples per each apparent age in our database is shown in Fig. 2. The images of our database have been taken under very different conditions, which makes it more challenging for recognition purposes.

In Figure 3 we show a scatter plot of the real and apparent age annotations of the images in our proposed database. As expected, there is a strong correlation between the two variables. However, the individual differences can be even larger than 20 years. This is no surprise since it is commonplace that some people “show their age”, while others “hide their age well”, some “age well” and others “age badly”, indicating that people perceive age not necessarily in agreement with the biological age. It is also interesting to

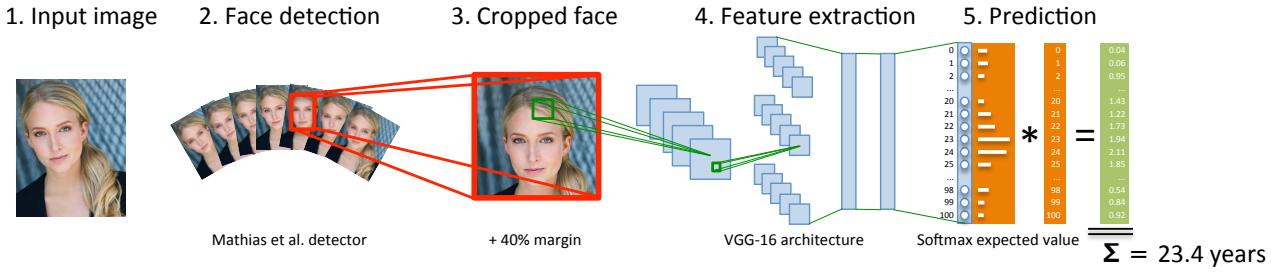



Fig. 4. Pipeline of DEX method for age estimation, figure taken from [34].

note that the apparent age is on average larger than real age for young adults but lower for the elderly. This is consistent with the effort made by young adults to appear more mature while the elderly attempt to look younger.

## III. Method

In this section we briefly review the DEX (*D*eep *EX*pectation) regression model of Rothe *et al.* [33] which provides state-of-the-art results on both apparent and real age estimation [34] on a number of standard benchmarks. Then, we propose our Residual DEX method that is able to further improve the performance of DEX on age estimation tasks.

*Notations:* We denote the real and apparent age of the $i$-th image as $a_i^{(R)}$ and $a_i^{(A)}$ respectively. We omit the superscript and simply write $a_i$ to refer to either real or apparent age.

### A. DEX Regression

As the baseline method of our study we use the DEX method or Rothe *et al.* [33], [34]. We are motivated by its state-of-the-art results, that achieved the first prize at ICCV ChaLearn LAP 2015 competition, and availability of the source codes. The processing pipeline of DEX is outlined in Figure 4. For each input image, first a face detector is deployed to obtain a robust face detection, then the face is aligned to a frontal face pose and the image is cropped with a $40\%$ margin around the detected face. The cropped face $I$ is the input image for the subsequent operations. DEX uses the VGG-16 architecture of Simonyan and Zisserman [36] for deep learning. VGG-16 is a Convolutional Neural Network (CNN) validated first on the ImageNet benchmark for image classification and then broadly adopted by the research community. DEX modifies the last layer (the CNN outputs) of the VGG-16 architecture to correspond to $Y$ age ranges, where each range $j$ covers $(y_j - \delta_j/2, y_j + \delta_j/2)$, with center $y_j$ and width $\delta_j$ such that the ranges touch $(y_{j-1} + \delta_{j-1}/2 = y_j - \delta_j/2\ )$.

In the training phase, the network of DEX is trained for classification, where for a cropped face image $I_i$, the age $a_i$ is assigned to the class $c(i)$ corresponding to the closest center:

$$c(i) = \arg\min_j |a_i - y_j|. \tag{1}$$

Therefore the regression problem is mapped to a classification problem.

In the prediction phase, the *expectation* is taken, using the output class probabilities $p_j(I)$ over the $Y$ age ranges,

$$\mathrm{D}(I) = \sum_{j=1}^{Y} p_j(I) y_j, \tag{2}$$

in order to obtain the predicted age $D(I)$ for image $I$.

For more details on DEX we refer to the original papers of Rothe *et al.* [33], [34].

### B. Residual DEX

Our original contribution is to propose the Residual DEX to further improve on DEX. The (original) DEX regressor is a rough estimator of the age, which extracts robust features from the input face image. The idea is that residuals (or errors) between the rough DEX estimation and the ground truth labels can be tackled with a specialized model. These residuals span a smaller range of values than those of the ground truth labels and are usually centered on 0 (most DEX errors are within 20 years). A good estimation of the residuals can allow us to correct and boost the performance of DEX. For this we learn a new regressor (using the same DEX architecture for CNN and the same expectation) to predict DEX residuals and we call it Residual DEX. The intuition is that most of the age estimation job is done by the rough DEX regressor while the Residual DEX models specialized facial features from the same cropped image to further correct the age estimation.

Given a trained DEX regressor $D_1$ for either real or apparent age estimation, for an cropped face image $I_i$, we denote the residual as:

$$r_i = a_i - D_1(I_i), \tag{3}$$

where $a_i$ is the ground truth (real or apparent) age and $D_1(I_i)$ is the predicted age. Hence, we improve the model by training a *second* regressor to estimate the residual. First, on the same training set, we learn a DEX model $D_2$ in order to predict $r_i$. Then, for an image $I_t$ in the test set, the combined prediction is formed as:

$$D_1(I_i) + D_2(I_i).$$

In the same way, we can repeat the previous procedure and learn a new regressor $D_3$ for the residual of $D_1 + D_2$, and so on and so forth. Within the framework of residual DEX, we can also combine regressors for real and apparent age: e.g. learn a regressor $D_1$ for apparent age, and a regressor $D_2$ for the residual $r_i = a_i^R - D_1(I_i)$.

TABLE II

MEAN ABSOLUTE ERROR (MAE) BETWEEN THE APPARENT AGE AND THE PREDICTED AGE FOR THE EVALUATED METHODS ON TEST SPLIT.

| Method | MAE Apparent |
|---|---|
| Apparent GT | 0 |
| Real GT | 4.573 |
| **Apparent DEX** | **4.082** |
| Real DEX | 4.513 |
| Real + Residual DEX | 4.450 |

TABLE III

MEAN ABSOLUTE ERROR (MAE) BETWEEN THE REAL AGE AND THE PREDICTED AGE FOR THE EVALUATED METHODS ON THE TEST SPLIT.

| Method | MAE Real |
|---|---|
| Real GT | 0 |
| Apparent GT ("wisdom of the crowd") | 4.573 |
| Real DEX | 5.468 |
| Real + Residual DEX | 5.352 |
| Apparent DEX | 5.729 |
| Apparent DEX + SVR | 5.426 |
| **Apparent + Residual DEX** | **5.296** |

## IV. EXPERIMENTS

### A. Experimental Setup

APPA-REAL database has a default split into train, test and validation images representing 4113, 1500 and 1978 images, respectively. This was obtained via a stratified random split evening out the age distribution.

The quantitative results are reported in terms of Mean Absolute Error (MAE), as commonly used in the literature. For the apparent age estimation another metric called $\epsilon$-error was proposed in [8], taking into account the ground truth standard deviation. However, because MAE can be used both for apparent and real age estimation this is the metric we chose in this paper.

When using DEX, we start from a pre-trained DEX model on the IMDB-WIKI dataset for real age estimation [34], and fine-tune it on the proposed database. We use the same training parameters as the original DEX, with $y_j = j$ and $\delta_j = 1$ for $j = 1, \cdots, 100$, and stop the training when the model starts to overfit on the validation split. For Residual DEX we set $y_j = j$ and $\delta_j = 1$ for $j = -50, \cdots, 50$ but otherwise use the same training parameters as DEX.

### B. Method Settings

**Real GT** and **Apparent GT** are the ground truth labels for real and apparent age which when available can be used as predictors for the other age labels (*i.e.* Apparent GT used to predict the real age). **Real DEX** is the model obtained by finetuning DEX for real age prediction on the proposed database, whereas **Apparent DEX** is finetuned for apparent age prediction.

**Apparent + Residual DEX** and **Real + Residual DEX** denote the models obtained by learning the residuals as detailed in Section III-B,*i.e.* **Apparent + Residual DEX** employs Apparent DEX estimation combined with its Residual DEX trained to predict the residuals of Apparent DEX.

In our experiments the application of more than one Residual DEX led to no significant performance improvements over just one Residual DEX. Therefore, we report results with just one level of Residual DEX, sparing useless computational burden. In our experiments, the execution of the deep CNN model takes $\sim 0.1$s on a NVidia TitanX GPU.

**SVR** denotes Support Vector Regression [5] using a RBF kernel, which will be used to map apparent to real age.

### C. Quantitative results

*Apparent Age estimation*

Table II shows the performance (MAE) of the DEX model for Apparent age estimation on the proposed APPA-REAL database. DEX achieves a MAE of 4.08 for apparent age estimation when trained with apparent age labels, significantly lower than the 4.51 MAE when DEX is trained for real age estimation and the 4.57 MAE when using directly the ground truth real age as apparent age predictor. Interestingly, the Real DEX achieves a better MAE than the Real GT at *apparent age estimation*. This shows that Real DEX picks up face features from the image, which are favorable also to apparent age estimation. This is not so surprizing since the DEX predictor, even trained on Real GT, bases itself on features of the image (much like humans when they attempt to predict the apparent age), hence its predictions can correlate better with apparent age than actual real age.

*Real Age estimation*

Table III shows the MAE results for Real age estimation. As mentioned before the apparent age correlates better with (and is a function of) the face image than the real (biological) age. This is validated by the results. Real age estimation is harder than the apparent age estimation, for the deployed models. Real DEX achieves 5.47 MAE on real age estimation, while Apparent DEX gets 4.08 MAE on apparent age estimation.

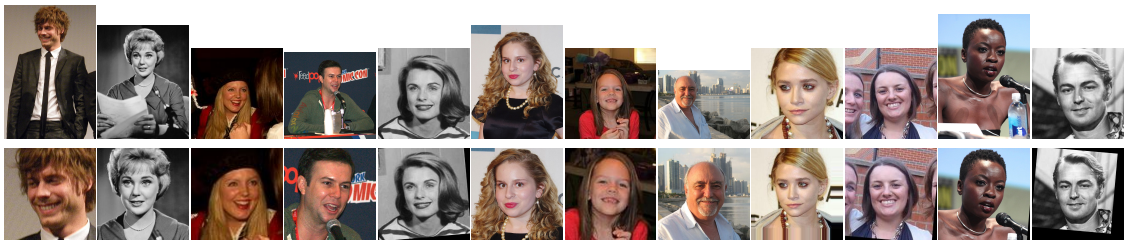
Surprisingly, the by far best real age estimation is provided by the apparent age (*i.e.* the "wisdom of the crowd") with a 4.57 MAE, while Real DEX gets 5.47. This suggest that there is a large room for improvement in real age estimation since the human crowd reference is 0.9 year better than the Real DEX.

Our proposed Residual DEX trained on top of Real DEX significantly improves the performance lowering the MAE from 5.468 to 5.352, or 0.8 year close to the "wisdom of the crowd" reference of 4.573.

*From apparent to real age estimation*

Since apparent age has been shown to be discriminative for real age estimation, we further analyze how a model trained for apparent age performs can be used for real age estimation.

In Table III we see that Apparent DEX gives a slightly higher MAE of 5.729 ($+0.26$) compared to Real DEX when used for real age estimation. However, as shown in



| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Input | | | | | | | | | | | | |
| Cropped Face | | | | | | | | | | | | |
| GT Apparent | 28.84 | 34.30 | 30.11 | 33.05 | 34.84 | 26.16 | 6.92 | 61.14 | 23.53 | 26.44 | 31.29 | 36.18 |
| Apparent DEX | 26.04 | 29.28 | 28.69 | 30.33 | 32.76 | 23.57 | 4.98 | 59.32 | 20.42 | 24.78 | 29.26 | 40.03 |
| GT Real | 24.00 | 30.00 | 25.00 | 31.00 | 29.00 | 18.00 | 8.00 | 68.00 | 25.00 | 30.00 | 37.00 | 29.00 |
| Real DEX | 22.81 | 25.90 | 21.18 | 28.96 | 28.51 | 18.12 | 4.29 | 63.03 | 17.12 | 22.89 | 29.29 | 38.10 |
| **Apparent + Residual DEX** | 24.00 | 29.40 | 26.21 | 29.29 | 31.38 | 21.17 | 4.03 | 63.12 | 19.15 | 22.94 | 28.22 | 40.72 |

Fig. 5. Representative examples of apparent and real age estimations when using the Apparent DEX, Real DEX and Apparent+Residual DEX. The examples are sorted from left to right and sampled according to the Apparent+Residual DEX error.

Figure 3 there is a slight distribution mismatch between the two apparent and real age labels. Correcting for this, by training a simple one dimensional SVR, mapping the predicted apparent age to real age (the 'Apparent DEX + SVR' setting), gives a MAE of 5.426 which is slightly lower than the MAE of Real DEX (5.468). This shows that a model trained for apparent age can be converted into a model for real age estimation with minimal effort, even outperforming the state-of-the-art DEX model trained for real age estimation. This is not so surprising, since apparent age is a function of the image, and thus 'easier' to learn from image than the real age, while still being a very strong predictor for real age. However, apparent age is obtained through the ensemble of human opinions and as such it is likely that not *all* the relevant information for real age estimation is captured. Using our proposed Residual DEX, we can go back to the image and pick-up these remaining features for real age estimation. Our results in III show that training a Residual DEX on top of the Apparent DEX model for real age prediction gives the lowest MAE of 5.296.

If we reverse the scenario, we find that this relationship is not symmetric. This is because the real age is actually a worse predictor of the apparent age (4.573 vs 4.082, see Table II) than DEX. Therefore, in Table II we only marginally improve the Real DEX prediction from from 4.513 to 4.450 when training for the apparent age with Real + Residual DEX, significantly worse than the simple Apparent DEX model.

### D. Visual assessment

In Figure VI we show the performance of DEX and Apparent + Residual DEX on 12 images selected from the test set side by side with the ground truth (GT) apparent and real age labels. To get a representative set, we sorted the images according to the MAE of the Apparent + Residual DEX and show images uniformly spaced from the list.

We see that for most images, the Residual DEX adjusts the age of Apparent DEX in the right direction. In the failure cases (*e.g.* the last 3 columns), the adjustment is either in the wrong direction, or too small compared to the large difference between apparent and real age ground truth labels for the image.

### E. Model Visualization

To visualize the DEX regressors we compute the sensitivity of each pixel with respect to the predicted age. The sensitivity is defined as the gradient of the predicted age with respect to the input image. We map the RGB-gradient to grayscale, normalize and smooth with a Gaussian of $\sigma = 2.5$, to get a heatmap with values in the range 0 to 1. We overlay the heatmap on top of the input image, encoding the value with the color (0 blue, 1 yellow) and the transparency (0 transparent, 1 solid).

In Figure 6 we show this sensitivity map overlaid over various images of the test set, for the Apparent and Real DEX models, as well as and the residual component of Apparent + Residual DEX (column "Residual DEX"). To visualize the difference between the Apparent and Real models, we also show the (absolute) difference between the heatmaps overlaid over the images (column "Real DEX - App. DEX").

As expected, mainly the face triggers the regressors, but the regions of high sensitivity (yellow) vary between the models. In particular, depending on the image, the models respond differently to the forehead, nasal and neck regions. For example, in the first row image we see that Apparent and Real DEX are mainly sensitive to the forehead, while the Residual DEX responds to the nose and the upper lip. In rows 2 and 3, the Apparent model gives a higher focus on the neck, while the chin is more strongly emphasized for Real DEX in rows 3 and 6.

Interestingly, for almost all the images, the models show a very low sensitivity to the hair, ears and mouth. These are the regions with high variance in the training images. Both ears are not always visible, the hair can be occluded, has various styles and (artificial) colors, while the mouth is a very expressive region which varies greatly defining the facial expression but not necessarily the age.

In the fourth row we can also see that the models are not sensitive to the second (partial) face in the image, focusing on the main central face.

The image examples are row-wise sorted by age and we can easily note that for both Apparent DEX and Real DEX the sensitive regions shift from the fore-head and between the eyes for young people to a relatively uniform spread over the face for middle age people and, finally, to chin and

neck regions for the older people. As expected, the Residual DEX combines the sensitivity zones of the Apparent DEX and the Real DEX as it learns to map the Apparent DEX estimation to the real age. At the same time Residual DEX is relatively more sensitive to information from outside the face in regions such as neck, hair, and background.

## V. Conclusion

In this paper we studied the relationship between real and apparent age estimation based on a unique face database with both real and apparent age annotations, introduced with this work. We proposed a residual age estimator and show further improvements in age estimation. For the first time we show that real age estimation can be decomposed into an apparent age estimation and an apparent to real age residual estimation, leading to improved accuracies over a standard real age estimation approach. Our database and this study can foster advances in both real and apparent age estimation research.

## References


[1] LHI image database, 2010.

[2] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay. Apparent age estimation from face images combining general and children-specialized deep learning models. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.

[3] A. Bastanfard, M. Nik, and M. Dehshibi. Iranian face database with age, pose and expression. In *Int. Conf. Machine Vision, 2007*, pages 50–55, Dec 2007.

[4] D. M. Burt and D. I. Perrett. Perception of age in adult caucasian male faces: Computer graphic manipulation of shape and colour information. *Royal Society of London. Series B: Biological Sciences*, 259(1355):137–143, 1995.

[5] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[6] K. Chen, S. Gong, T. Xiang, and C. Change Loy. Cumulative attribute space for age and crowd density estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2467–2474, 2013.

[7] E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimation of unfiltered faces. *Information Forensics and Security, IEEE Transactions on*, 9(12):2170–2179, Dec 2014.

[8] S. Escalera, J. Fabian, P. Pardo, X. Bar, J. Gonzlez, H. J. Escalante, D. Misevic, U. Steiner, and I. Guyon. Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 243–251, Dec 2015.

[9] S. Escalera, M. Torres Torres, B. Martinez, X. Baro, H. Jair Escalante, I. Guyon, G. Tzimiropoulos, C. Corneou, M. Oliu, M. Ali Bagheri, and M. Valstar. Chalearn looking at people and faces of the world: Face analysis workshop and challenge 2016. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2016.

[10] Y. Fu, G. Guo, and T. Huang. Age synthesis and estimation via faces: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(11):1955–1976, Nov 2010.

[11] Y. Fu and T. Huang. Human age estimation with regression on discriminative aging manifold. *Multimedia, IEEE Transactions on*, 10(4):578–584, June 2008.

[12] A. Gallagher and T. Chen. Understanding images of groups of people. In *Proc. CVPR*, 2009.

[13] X. Geng, C. Yin, and Z.-H. Zhou. Facial age estimation by learning from label distributions. *IEEE transactions on pattern analysis and machine intelligence*, 35(10):2401–2412, 2013.

[14] H. Han, C. Otto, and A. K. Jain. Age estimation from face images: Human vs. machine performance. In *ICB'13*, pages 1–8, 2013.

[15] Z. Huo, X. Yang, C. Xing, Y. Zhou, P. Hou, J. Lv, and X. Geng. Deep age distribution learning for apparent age estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.

[16] M. Inc. Face++ research toolkit. www.faceplusplus.com, Dec. 2013.

[17] K. Kim, S. Kang, S. Chi, and J. Kim. Human age estimation using multi-class svm. In *Ubiquitous Robots and Ambient Intelligence (URAI), 2015 12th International Conference on*, pages 370–372. IEEE, 2015.

[18] Z. Kuang, C. Huang, and W. Zhang. Deeply learned rich coding for cross-dataset facial age estimation. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2015.

[19] A. Lanitis. FG-NET Aging Data Base, November 2002.

[20] A. Lanitis, C. Taylor, and T. Cootes. Toward automatic simulation of aging effects on face images. volume 24, pages 442–455, 2002.

[21] G. Levi and T. Hassner. Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015.

[22] C. Li, Q. Liu, J. Liu, and H. Lu. Learning ordinal discriminative features for age estimation. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2570–2577. IEEE, 2012.

[23] T. Liu, Z. Lei, J. Wan, and S. Z. Li. Dfdnet: Discriminant face descriptor network for facial age estimation. In *Chinese Conference on Biometric Recognition*, pages 649–658. Springer, 2015.

[24] X. Liu, S. Li, M. Kan, J. Zhang, S. Wu, W. Liu, H. Han, S. Shan, and X. Chen. Agenet: Deeply learned regressor and classifier for robust apparent age estimation. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2015.

[25] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool. Face detection without bells and whistles. In *Computer Vision–ECCV 2014*, pages 720–735. Springer, 2014.

[26] M. Minear and D. C. Park. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4):630–633, 2004.

[27] B. Ni, Z. Song, and S. Yan. Web image mining towards universal age estimator. In *Proceedings of the 17th ACM International Conference on Multimedia*, MM '09, pages 85–94, New York, NY, USA, 2009. ACM.

[28] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *BMVC*, 2015.

[29] P. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The {FERET} database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295 – 306, 1998.

[30] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *CVPR*, pages 947–954. IEEE, 2005.

[31] N. Ramanathan, R. Chellappa, and S. Biswas. Computational methods for modeling facial aging: A survey. *Journal of Visual Languages and Computing*, 20(3):131 – 144, 2009.

[32] K. Ricanek and T. Tesafaye. MORPH: a longitudinal image database of normal adult age-progression. In *Int. Conf. FG*, pages 341–345, 2006.

[33] R. Rothe, R. Timofte, and L. Van Gool. Dex: Deep expectation of apparent age from a single image. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2015.

[34] R. Rothe, R. Timofte, and L. Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 2016.

[35] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *Int. Conf. FG*, pages 46–51, May 2002.

[36] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[37] . Softopia Japan Foundation. Human and Object Interaction Processing (HOIP) Face Database.

[38] P. Thukral, K. Mitra, and R. Chellappa. A hierarchical approach for human age estimation. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1529–1532. IEEE, 2012.

[39] K. Ueki, T. Hayashida, and T. Kobayashi. Subspace-based age-group classification using facial images under various lighting conditions. In *Int. Conf. FG*, pages 43–48, 2006.

[40] M. Uricar, R. Timofte, R. Rothe, J. Matas, and L. Van Gool. Structured output svm prediction of apparent age, gender and smile from deep features. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.

[41] J. Zhang, S. Shan, M. Kan, and X. Chen. *Coarse-to-Fine Auto-Encoder Networks (CFAN) for Real-Time Face Alignment*, pages 1–16. Springer International Publishing, Cham, 2014.

[42] Y. Zhu, Y. Li, G. Mu, and G. Guo. A study on apparent age estimation. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2015.

| 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** |

## VI. ASDF

| Method | MAE Apparent |
|---|---|
| Apparent GT | 0 |
| Real GT | 4.573 |
| **Apparent DEX** | **4.082** |
| Real DEX | 4.513 |
| Real + Residual DEX | 4.450 |

| 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** |

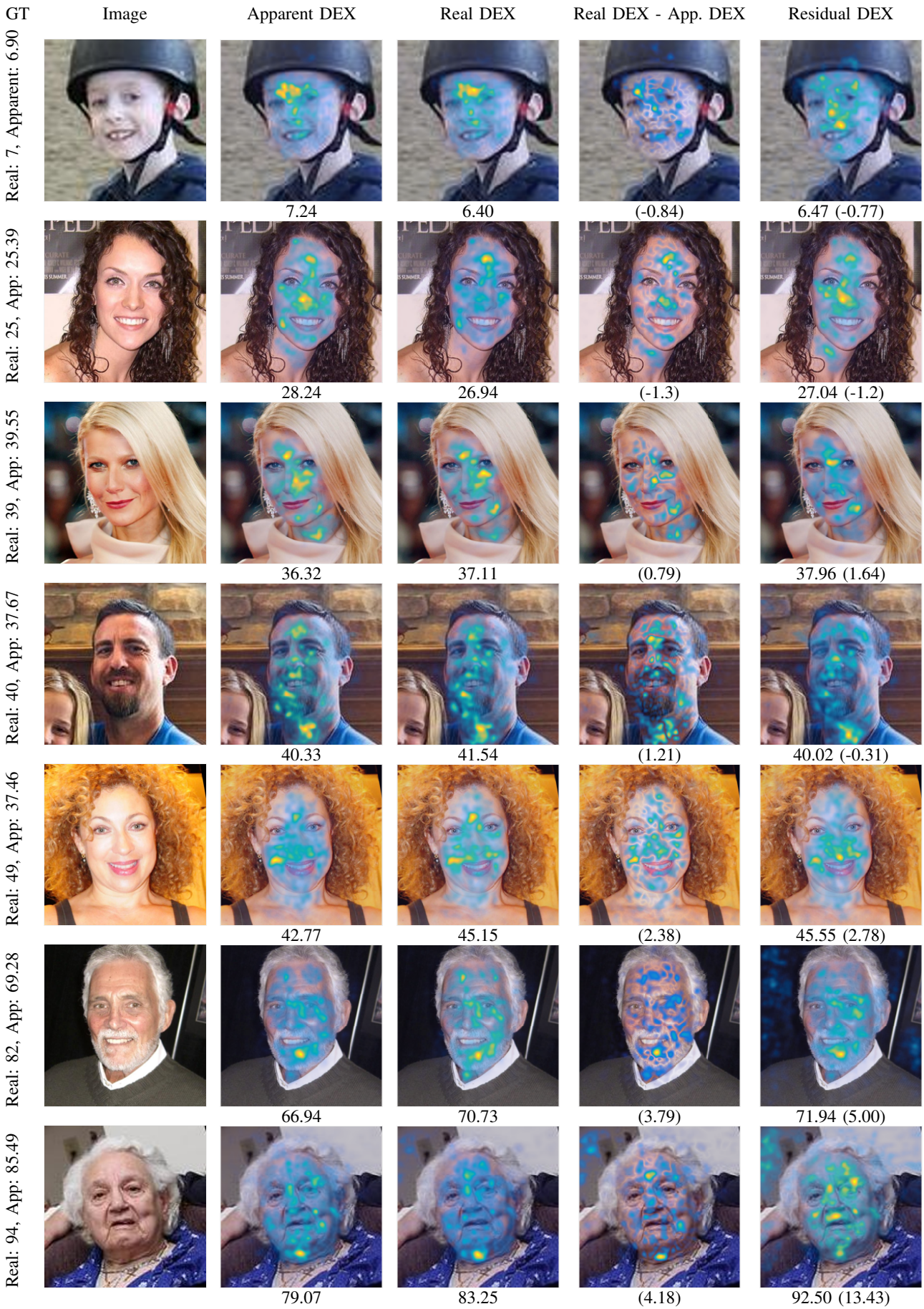
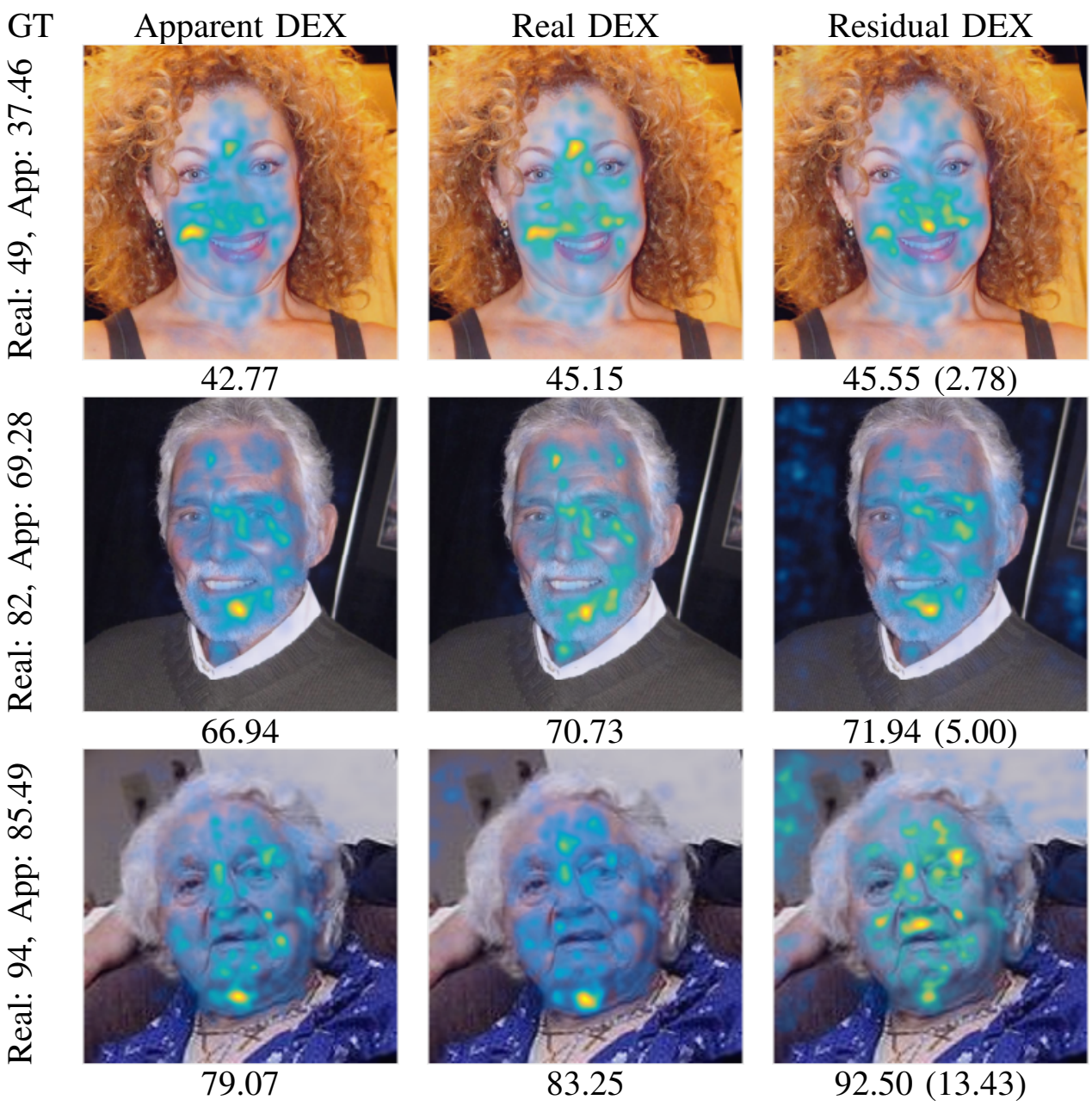
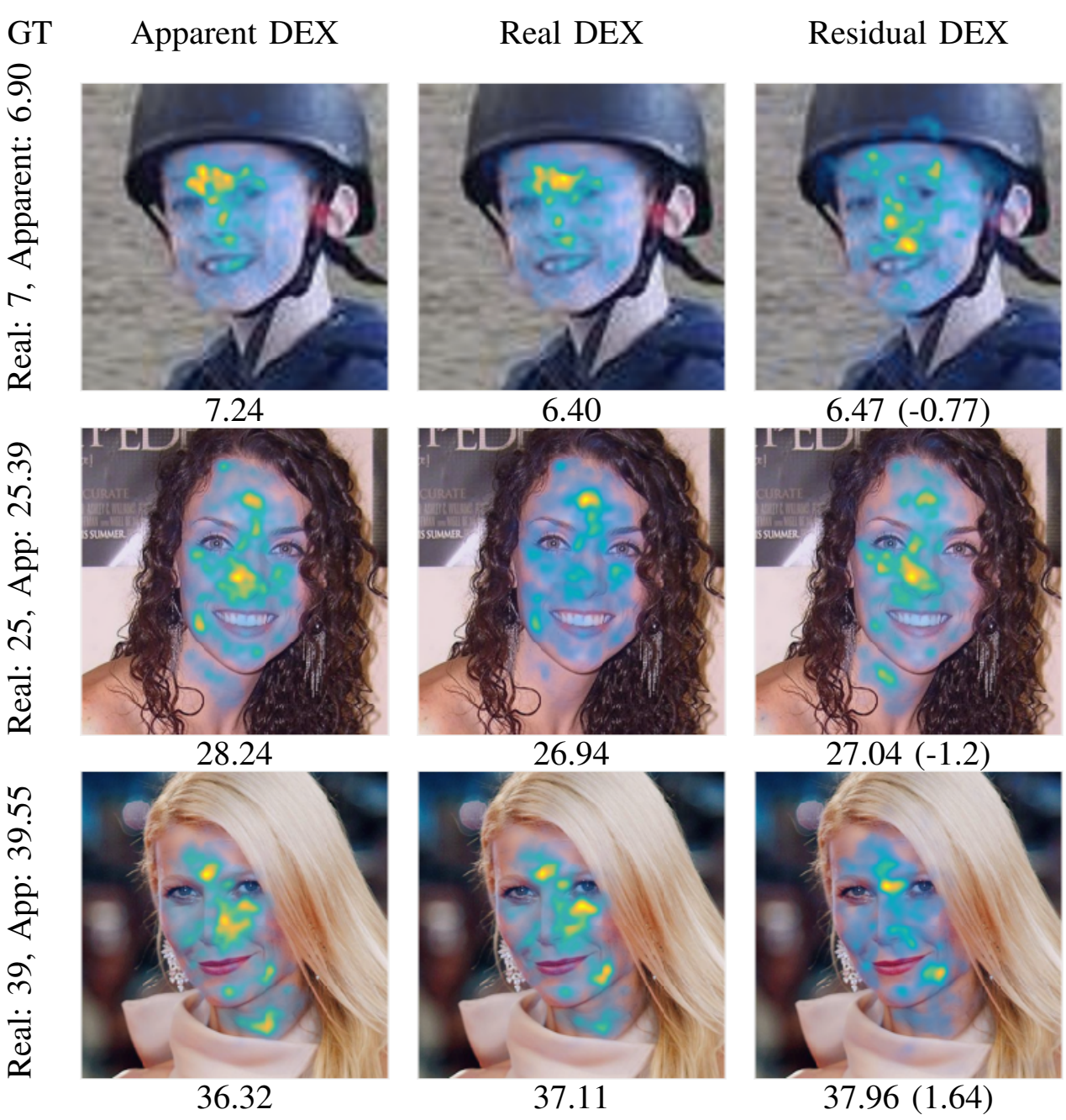
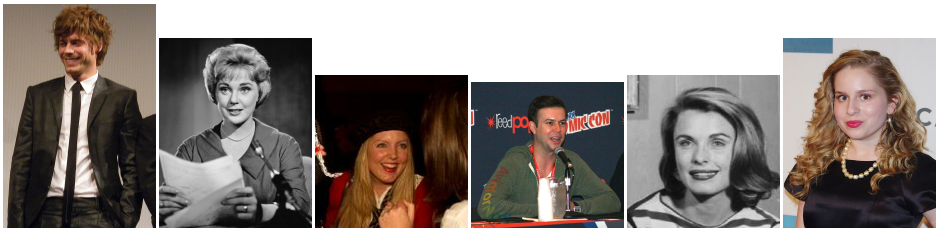
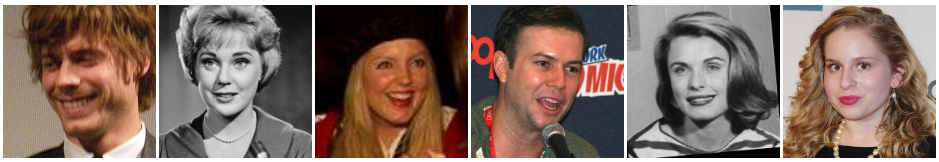


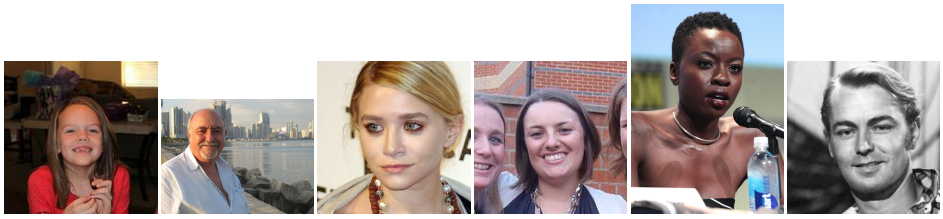

Fig. 6. Sensitivity map for apparent, real, and residual age estimation. The predicted age of each model is shown below the images and differences shown where applicable. Best zoomed on screen.


GT
Apparent DEX
Real DEX
Residual DEX
Real: 49, App: 37.46
42.77
45.15
45.55 (2.78)
Real: 82, App: 69.28
66.94
70.73
71.94 (5.00)
Real: 94, App: 85.49
79.07
83.25
92.50 (13.43)




| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Input |  | | | | | | | | | | | |
| Cropped Face |  | | | | | | | | | | | |
| GT Real | 24.00 | 30.00 | 25.00 | 31.00 | 29.00 | 18.00 | 8.00 | 68.00 | 25.00 | 30.00 | 37.00 | 29.00 |
| GT Apparent | 28.84 | 34.30 | 30.11 | 33.05 | 34.84 | 26.16 | 6.92 | 61.14 | 23.53 | 26.44 | 31.29 | 36.18 |
| Apparent DEX | 26.04 | 29.28 | 28.69 | 30.33 | 32.76 | 23.57 | 4.98 | 59.32 | 20.42 | 24.78 | 29.26 | 40.03 |
| Residual DEX | -2.04 | 0.12 | -2.48 | -1.04 | -1.38 | -2.40 | -0.95 | 3.80 | -1.27 | -1.84 | -1.04 | 0.69 |
| '' **Apparent + Residual DEX** | 24.00 | 29.40 | 26.21 | 29.29 | 31.38 | 21.17 | 4.03 | 63.12 | 19.15 | 22.94 | 28.22 | 40.72 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 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 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 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 |