EGroupNet: A Feature-enhanced Network for Age Estimation with Novel Age Group Schemes

MINGXING DUAN and KENLI LI, Hunan University, China
AIJIA OUYANG, Zunyi Normal University, China
KHIN NANDAR WIN, Hunan University, China
KEQIN LI, State University of New York, China
QI TIAN, University of Texas at San Antonio, USA

Although age estimation is easily affected by smiling, race, gender, and other age-related attributes, most of the researchers did not pay attention to the correlations among these attributes. Moreover, many researchers perform age estimation from a wide range of age; however, conducting an age prediction over a narrow age range may achieve better results. This article proposes a hierarchic approach referred to as EGroupNet for age prediction. The method includes two main stages, i.e., feature enhancement via excavating the correlations among age-related attributes and age estimation based on different age group schemes. First, we apply the multi-task learning model to learn multiple face attributes simultaneously to obtain discriminative features of different attributes. Second, we project the outputs of fully connected layers of several subnetworks into a highly correlated matrix space via the correlation learning process. Third, we classify these enhanced features into narrow age groups using two Extreme Learning Machine models. Finally, we make predictions based on the results of the age groups mergence. We conduct a large number of experiments on MORPH-II, LAP-2016 dataset, and Adience benchmark. The mean absolute errors of the two different settings on MORPH-II are 2.48 and 2.13 years, respectively; the normal score (ϵ) on the LAP-2016 dataset is 0.3578; and the accuracy of age prediction on Adience benchmark is 0.6978.

CCS Concepts: • **General and reference → Reference works**; • **Computing methodologies → Neural networks**;

Additional Key Words and Phrases: Age estimation, age groups, correlations, enhancement, multi-task learning

This work was supported in part by the National Outstanding Youth Science Program of National Natural Science Foundation of China under Grant No. 61625202, in part by the International (Regional) Cooperation and Exchange Program of National Natural Science Foundation of China under Grant No. 61661146006, in part by the National Key R&D Program of China under Grant No. 2016YT80201900, in part by the National Youth Science Program of National Natural Science Foundation of China under Grant 61902119, and in part by the Project funded by China Postdoctoral Science Foundation under Grants 2019M652758 and 2019TQ0087.

Authors' addresses: M. Duan and Kenli Li, Hunan University, College of Information Science and Engineering, Changsha, Hunan, 410000, China; emails: {duanmingxing, lkl}@hnu.edu.cn; A. Ouyang, Zunyi Normal University, College of Information Engineering, Zunyi, Guizhou, 563006, China; email: ouyangaijia@163.com; K. N. Win, Hunan University, School of Information Science and Engineering, Changsha, Hunan, 410000, China; email: knandarwin@hnu.edu.cn; Keqin Li, Department of Computer Science, State University of New York, New Paltz, New York 12561, USA; email: lik@newpaltz.edu; Q. Tian, Computer Science, University of Texas at San Antonio, San Antonio, Texas, USA; email: qitian@cs.utsa.edu.

1551-6857/2020/05-ART42 $15.00
https://doi.org/10.1145/3379449

ACM Reference format:
Mingxing Duan, Kenli Li, Aijia Ouyang, Khin Nandar Win, Keqin Li, and Qi Tian. 2020. EGroupNet: A Feature-enhanced Network for Age Estimation with Novel Age Group Schemes. *ACM Trans. Multimedia Comput. Commun. Appl.* 16, 2, Article 42 (May 2020), 23 pages.
https://doi.org/10.1145/3379449

1 INTRODUCTION

1.1 Motivation

In the past few years, a large number of researchers have paid a lot of attention to age prediction because of its crucial applications in our daily lives [12, 30]. For example, electronic customer relationship management (ECRM) helps a company to find the interests of different age groups, which can give much profit to the company. The system of age prediction generally includes two classes: age feature extraction and age prediction. The quality of the extracted features greatly affect the performance of age prediction [45], and the system can produce adequate results from these features. A detailed survey on age prediction can be found in Reference [17]. Especially, implementing using convolutional neural network (CNN) produces better results and good performance for age prediction [10, 30, 33]. In our article, we fully utilize the different levels features of the CNN model for age prediction as one of our contributions.

Concurrently, human age estimation can be influenced by many aspects, such as their lifestyles, their environments, body posture, gender and race, and so on. Therefore, the more factors we considered in prediction, the better performance we will obtain. Much research has been done to reveal the correlations between these factors and age attributes, such as age estimation by combining smile gestures [9], age estimation by fully considering gender and race attributes [10, 25], and totally deploying the influence of facial expression to predict age [48]. Although these methods have achieved better performance, they have two weaknesses: (1) just utilizing one or two related attributes to enhance the discriminative age feature representation but neglecting other related factors, such as, hair, eyebrows, and so on; (2) adopting a unified correlation mechanism to explore the useful information but ignoring the different influence degree of these attributes on age estimation. In other words, when CNNs are used to extract different attribute features of the same image, the lower layer features are mainly related to the spatial information of the image, involving less semantic information, such as the outputs of the first two convolution layers. After the third layer, there is more semantic information about the image, and the semantic information of different attributes is different. A conventional deep-learning algorithm needs multiple CNN models to extract different attribute features of the same image and each of the features is completely independent. These features have both correlation and negative relationships, and negative relationships are difficult to eliminate, since this module is not an end-to-end learning. Therefore, to solve that limit, our EGroupNet shares the previous two convolutional layers and splits these features from the third layer to the last layer. Then a relation matrix fully exploits the positive relationships among the features of different attributes.

Many researchers have tried for age prediction, most of them carried out their research with a wide range of age groups. Typically, we can achieve better performance and more valuable results with a specific range of age group; for example, the estimation can be done more efficiently if we predict an age from the range of 20–25 rather than that of 0 to 100. Therefore, before the final prediction, we first attempt to classify face image into a small age range. After that, the final age group can be predicted by combining the two age groups. As shown in Figure 1, we classify an image into two types of age groups and these groups have overlapping parts. As an example, first, the human image is classified into two groups, and if the groups have common area such as (0–10)

group and (5–15) group, then we merge this range as an another small group (5–10); otherwise, misclassification will occur as in Figure 1. Tan et al. [60] successfully utilized age group strategy to predict age. The method first applies Age Group-n Encoding method to classify images into n age groups, and then employs Local Age Decoding method to achieve accurate age. This method makes full use of relationships between age group labels, which results in better predictive performance. It can be seen that the age group strategy proposed in this article is completely different from that proposed in Reference [60]. As a result, one of the contributions of our article is getting better performance using age group schemes.

1.2 Our Contributions

In this article, a hierarchical age prediction structure named EGroupNet is proposed to estimate facial images. EGroupNet includes three main processes: extracting discriminative features, excavating the correlations and differences among these age-related attributes, and obtaining a narrow age range of an image to make a final prediction. This article just concerns the age-related facial attributes, such as smiling, gender, race, and so on. The first two processes realize the end-to-end learning, and cross-loss functions are used to update the network and the output weights of concat features. We pretrain the network on ImageNet [37] and IMDB-WIKI [56]. As shown in Figure 1, during each iteration, each subnetwork updates its own parameters according to corresponding loss. At the same time, the feature vectors of the second fully connected layer of each subnetwork is merged into a feature matrix and the output weight matrix is updated with age-related cross-loss function. The fine-tuned network can make a prediction and this prediction can be used as a baseline, which is utilized to determine whether the results of age grouping for the same image is within the appropriate age range. After that, the features extracted by fine-tuned neural network are used to train the three Extreme Learning Machine (ELM) models. Finally, the testing image is classified into several age groups. Simultaneously, based on the baseline, the system will determine whether these groups are within reasonable ranges or not and the two selected age groups with the common range are merged into new one with a specific range. After that, final age prediction is determined. The major contributions of this article are summarized as follows:

- We propose a hierarchical EGroupNet to perform age prediction. The system includes two stages: feature correlations and differences excavation, and performing prediction via different age group schemes, which are adequate to explore the advantages of CNN and ELM.
- End-to-end learning method is used to extract distinct features and excavate the correlations among age-related face attributes in the first stage. At the same time, we map the outputs of the fully connected layers of several subnetworks into a highly correlated matrix space via correlation matrix. The correlation matrix is learned by the stochastic gradient method and in this process, we just update the parameters of correlation matrix.
- To achieve a narrow interval age group, we design different kinds of age group schemes for Morph-II, LAP-2016, and Adience. The image is first classified into two different narrow age groups, and then a narrower age group is obtained by merging the first two age groups. To allow the model to identify these narrow age groups, we tag the three datasets with these age groups based on the original labels and then train the model with the new labels. The detailed labeling and training processes are presented in Section 3.
- Finally, we conduct many experiments on the MORPH-II, Adience benchmark, and LAP-2016 to verify the performance of our EGroupNet.

The rest of this article is organized as follows. Section 2 illustrated the related work. Section 3 presents our proposed method. The experimental results are analyzed in Section 4. Finally, we make conclusions in Section 5.

2 RELATED WORK

2.1 Deep Neural Network for Age Prediction

Recently, CNN has shown preferable performance in image recognition and achieved better performance by hybrid neural networks in a large number of applications. Lawrence et al. [40] used a deep neural-network method that combined two different deep networks to recognize faces. That method presented a better performance than other methods on the same applications. By fully utilizing the advantages of CNN and support vector machine (SVM), Niu et al. [51] employed a novel structure for age prediction and a more accurate result was achieved. Liu et al. [44] combined the CNN with the Conditional Random Field (CRF) to handle face recognition. Through many experiments on different datasets, the hybrid structure achieved the best results than other methods on the same datasets. Xie et al. [67] proposed a hybrid approach to handle scene recognition and domain adaption. The network fully utilized the local discriminative information from the input images. Next, a SVM structure was applied to classifying the features, which obtained a better performance. Tang et al. [62] combined deep neural network (DNN) and ELM to detect ships of images. DNN was used to extract the features, while ELM made the final prediction. Extensive experiments have been conducted to demonstrate that the system cost least detection time to make a prediction and obtained a better results. Malli et al. [49] utilized an ensemble CNN structure to make an apparent age prediction on ChaLearn LAP 2016 dataset [15], which finally achieved a 0.3668 error. Rothe et al. [55] introduced a Deep EXpectation (DEX) for apparent age prediction and it won 1st place in the ChaLearn LAP 2015 challenge. Liu et al. [46] proposed ensemble CNNs for age estimation and it won 2st place in the ChaLearn LAP 2015 challenge place. A grouping estimation fusion (GEF) system was proposed by Liu et al. [45] to predict human age. GEF achieved better results via fusing several decisions. Liu et al. [47] proposed a hybrid structure to estimate an apparent age and the prediction results proved its effectiveness. Gl̇źrpinar et al. [28] utilized kernel ELM with CNN to predict human age and a better result is achieved. Wan et al. [65] proposed five cascaded structure frameworks for age estimation and achieved superior performance on the Morph-II and CACD datasets. By learning global, local, and global-local features of facial images, Tan et al. [61] adopted a Deep Hybrid-Aligned Architecture to predict facial age and achieved a better performance on different datasets. Li et al. [43] utilized a BridgeNet to estimate age via excavating the continuous relation between age labels and extensive experiments demonstrates its efficiency. Zhang et al. [68] proposed an extremely Compact yet efficient Cascade Context-based Age Estimation model for small-scale images and obtained competitive performance. In this article, we attempt to combine CNN with ELM models to enhance the performance of age prediction.

2.2 Age Grouping

Recently, facial age grouping has proved its advantages in age prediction and a large number of approaches have presented their advantages. Kwon et al. [39] first utilized the age grouping approach for age prediction, which categorized samples into infants, youth, and seniors. Horng et al. [32] proposed a hierarchical system for age prediction, and they categorized images into four age groups. It obtained 81.58% accuracy on 230 facial images. Thukral et al. [63] achieved 70.04% accuracy on different age groups (i.e.., 0–15, 15–30, and 30+). Gunay et al. [23] used local binary patterns (LBP) method to estimate age prediction automatically [3]. The images were split into several sectors, which are categorized into six age groups via spatial LBP histograms: 10 ± 5, 20 ± 5, 30 ± 5, 40 ± 5, 50 ± 5, 60 ± 5. Finally, k-nearest neighbor made the final classification and the accuracy was 80%. Hajizadeh et al. [29] adopted a probabilistic neural network (PNN) to categorize samples into several age groups, and the accuracy was 87.25%. Gl̇źrpinar et al. [28] used a CNN and ELM models for age prediction, and ELM classified the images into 8 age groups. ELM

regressor made the final prediction, and it has been reached to 0.3740 normal score on LAP-2016 dataset. Iqbal et al. [36] used a local face descriptor, Directional Age-Primitive Pattern (DAPP) to predict age through characterizing discernible facial aging sign (e.g., craniofacial growth and skin aging) from a detailed and more finer point of view and the proposed approach achieved the latest methods by an acceptable margin.

2.3 Age Estimation

Age classification has got huge amount of attention, because it supports an efficient approach for achieving implicit social information [16]. Fu et al. [17] conducted a detailed survey of age prediction, and much useful information can be obtained in Reference [41]. Age prediction was first proposed by Kwon et al. [38], who proved that detecting the appearance of wrinkles and calculating ratios could categorize images into several age classes. And the approach was utilized to model craniofacial growth using both psychophysical evidences and anthropometric evidence [53]. The method needed detailed localization of facial features.

Geng et al. [22] used an AGing pattErn Subspace to conduct an automatic age prediction, while an age manifold learning mechanism was introduced in Reference [24] to extract features and a locally adjusted robust regressor was used to make final prediction. Although the approaches have many advantages, the weakness is that the samples need to be near-frontal and well-aligned. Chang et al. [4] proposed a cost-sensitive ordinal hyperplanes ranking approach to make predictions from facial images, while a hierarchical system called the "grouping estimation fusion" (DEF) was used to predict age. Li et al. [42] predicted age adopting feature choice approach and proved the effectiveness of the method through experiments. Because estimating an accurate age or age group of a facial image needed extensive face dataset attached with corresponding age labels, Hu et al. [33] presented a novel learning method to make full use of these inadequate labeled data through CNN and the method achieved the state-of-the-art performance. All of these approaches mentioned above have exhibited the advantage in age prediction while they do not fully excavate the correlations among age-related factors, which may benefit age prediction. Considering these limitations, our method focused on exploring the correlation and differences among the related factors.

3 PROPOSED METHOD

3.1 Multi-task Learning Model

Because the different convolutional layers of a CNN model provide different level representations of image, we attempt to make full use of these representations to achieve more robust and discriminative features. It is well known that features in the earlier layers of CNN retain more spatial resolution while higher semantic information is kept in high-level layers. Therefore, during the process of multi-task learning, sharing the information in the low-level layer and splitting these information in high-level layer can not only benefit the system learning the correlation and difference among age-related attributes but also reduce the whole parameters. At the same time, achieving a good result for age prediction needs more discriminative features and these features are more about semantic information. Consequently, we merge high-level features of different attributes into new features, after which, the final prediction is made.

3.1.1 Feature Extraction Models. Based on the analysis above, our EGroupNet shares the information from the first convolutional layer till the third one and then the network is split into several subnetworks. With different targets for the same image, the subnetworks extract the different feature representations. As shown in Figure 1, this process realizes an end-to-end learning.

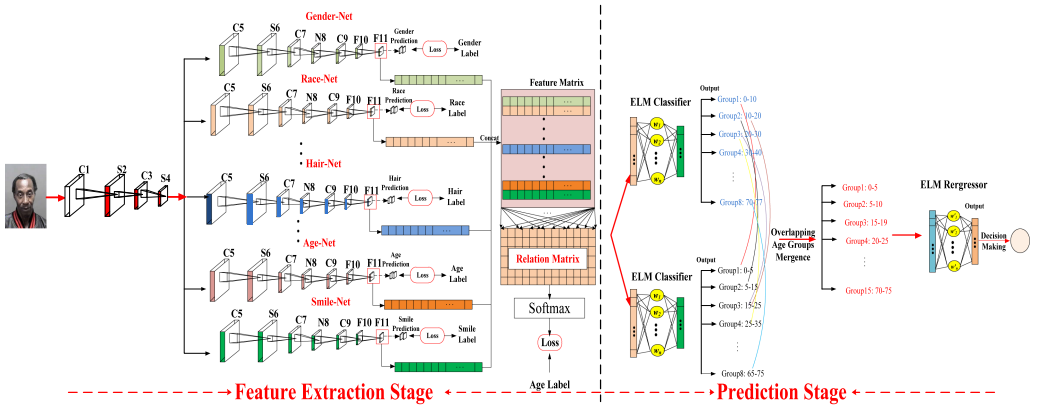


Fig. 1. Full schematic diagram of EGroupNet. The system includes two stages: feature extraction and prediction. During the first stage, the subenetworks are trained with corresponding attribute labels, and the outputs of the FC layers are used to learn the relation matrix (mapping matrix) via stochastic gradient algorithm. Through the relation matrix, the EGroupNet maps the outputs into a highly correlated matrix. This matrix is used to train the three ELM models. In this stage, the system can make a prediction, which will be used as a baseline to judge whether the classifications of age groups of the same image are right or not. In the second stage, two ELM models classify the image into two kinds of age groups, and we merge these age groups into a narrower one, based on which, the final prediction is made.

There are two convolution operations in the shared layer and three of those are existed in each subnetwork. Specifically, (C1, C3, ..., C9) denote the corresponding convolutional layers, (S2, S4, S6) represent pooling and normalization operations, N8 signifies only the normalization operation, and (F10 and F11) express the fully connected layers. Local response normalization (LRN) is used to enhance the generalization of the system following the ReLU operation and it can be expressed as follows:

$$c_{norm} = \frac{c}{(1 + (\beta/n) \sum_i c_i^2)^{\gamma}}, \tag{1}$$

where c is a square local region and n is set to 5, β is 0.0001 and it expresses the scaling parameter, γ is 0.75, and c_{norm} denotes the normalized region.

We use the cross-entropy loss function as the classifiers of these subnetworks, and the loss is

$$L = -\frac{1}{N} \sum_{i=1}^{N} (y_i \ln p_i + (1 - y_i)(\ln (1 - p_i))), \tag{2}$$

where p_i expresses the probability of an attribute, y_i denotes the ground-truth of the attribute, and N signifies the number of training examples.

Our goal is to excavate the correlation between age-related facial attributes and age attribute to enhance the valuable information of age features. Therefore, our system tries its best to learn a relation matrix during the training process. The features in the F11 layer mainly retain more semantic information and our approach attempts to explore correlation among those features. As shown in Figure 2, during each iteration, the feature vectors of the F11 layer of the whole subnetworks are merged into a feature matrix. Then a relation matrix is used to explore the correlation between age-related facial attributes and age, and map these outputs into a highly correlated space. At last, the output features with more robust and discriminative information are used to make final prediction.

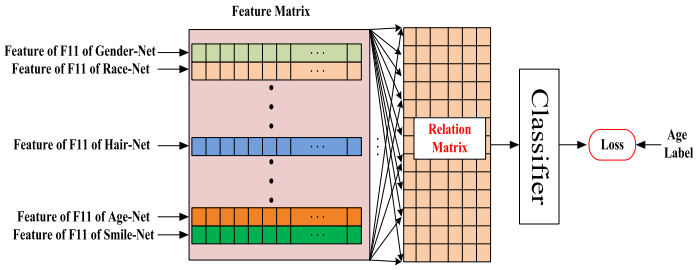


Fig. 2. The learning process of relation matrix.

The Back Propagation method is utilized to update the relation matrix. Note that the loss is just used to update the relation matrix and it is not propagated to the subnetworks. According to our best knowledge, whether classification tasks or regression tasks, a large amount of features may lead the tasks to overfitting. Reducing or punishing the unimportant weights of features are two common approaches to alleviate the problems, while we do not know which weights of the features are unimportant. Regularization is an effective way to solve the problem and L_2 regularization is used to calculate the loss. The total loss of the network can be denoted as

$$L(\theta) = -\frac{1}{N}\sum_{i=1}^{N}(y_i \ln p_i + (1-y_i)(\ln(1-p_i))) + Total_{weight\_loss}, \tag{3}$$

where θ denotes the parameters, p_i expresses the probability of an attribute, y_i denotes the ground-truth of the attribute, and N represents the number of training examples. $Total_{weight\_loss}$ expresses the whole loss of relation matrix in each iteration.

M_i expresses the feature matrix of ith sample and W_i expresses the corresponding relation matrix. Therefore, we update the relation matrix as follows:

$$W_{i+1} = W_i - \eta \Delta W_i, \tag{4}$$

where W_{i+1} denotes the weights of $(i+1)$th sample and η is the learning rate. Due to that nonlinear function is used during this process, $\frac{\partial L_i}{\partial W_i}$ is a constant and then the relation matrix can be updated as follows:

$$W_{i+1} = W_i - \eta \Delta \frac{\partial L_i}{\partial W_i}. \tag{5}$$

3.1.2 Back Propagation. Back Propagation is a critical process in the process of multi-task learning and each subnetwork is updated with the corresponding loss. Transferring gradients of each subnetwork to the shared part of our EGroupNet is an important part of the process of Back Propagation. We adopt a joint gradient transfer approach to compute the gradients. We use W^i_{layer} and b^i_{layer} to designate the weights and biases of some convolutional layer, for example, W^i_{C3} and b^i_{C3} denote the weights and biases of the C3 layer of ith sample. $L^i_{gender}, L^i_{race}, \ldots, L^i_{age}$ express the corresponding loss of each subnetwork. We use $X^i_{(subnetwork,layer)}(I)$ to denote the input of some layer of the subnetwork, for example, $X^i_{(age,C5)}(I)$ expresses the input of the C5 layer of the Age-Net. The joint gradient transferred approach is like the following:

$$\Delta W^i_{c3} = \frac{\partial L^i_{gender}}{\partial W^i_{c3}} + \frac{\partial L^i_{race}}{\partial W^i_{c3}} + \cdots + \frac{\partial L^i_{age}}{\partial W^i_{c3}}, \tag{6}$$

$$\Delta b^i_{c3} = \frac{\partial L^i_{gender}}{\partial b^i_{c3}} + \frac{\partial L^i_{race}}{\partial b^i_{c3}} + \cdots + \frac{\partial L^i_{age}}{\partial b^i_{c3}}. \tag{7}$$

We apply the chain rule to compute the partial derivative as follows:

$$\begin{cases} \frac{\partial L^i_{gender}}{\partial W^i_{c3}} = \frac{\partial L^i_{gender}}{\partial X^i_{(gender,\ C9)}(I)} \frac{\partial X^i_{(gender,\ C9)}(I)}{\partial X^i_{(gender\ C7)}(I)} \frac{\partial X^i_{(gender,\ C7)}(I)}{\partial X^i_{(gender\ C5)}(I)} \frac{\partial X^i_{(gender,\ C5)}(I)}{\partial W^i_{c3}} \\ \frac{\partial L^i_{race}}{\partial W^i_{c3}} = \frac{\partial L^i_{race}}{\partial X^i_{(race,\ C9)}(I)} \frac{\partial X^i_{(race,\ C9)}(I)}{\partial X^i_{(race,\ C7)}(I)} \frac{\partial X^i_{(race,\ C7)}(I)}{\partial X^i_{(race,\ C5)}(I)} \frac{\partial X^i_{(race,\ C5)}(I)}{\partial W^i_{c3}} \\ \quad\vdots \\ \frac{\partial L^i_{age}}{\partial W^i_{c3}} = \frac{\partial L^i_{age}}{\partial X^i_{(age,\ C9)}(I)} \frac{\partial X^i_{(age,\ C9)}(I)}{\partial X^i_{(age,\ C7)}(I)} \frac{\partial X^i_{(age,\ C7)}(I)}{\partial X^i_{(age,\ C5)}(I)} \frac{\partial X^i_{(age,\ C5)}(I)}{\partial W^i_{c3}} \end{cases}, \tag{8}$$

$$\begin{cases} \frac{\partial L^i_{gender}}{\partial b^i_{c3}} = \frac{\partial L^i_{gender}}{\partial X^i_{(gender,\ C9)}(I)} \frac{\partial X^i_{(gender,\ C9)}(I)}{\partial X^i_{(gender\ C7)}(I)} \frac{\partial X^i_{(gender,\ C7)}(I)}{\partial X^i_{(gender\ C5)}(I)} \frac{\partial X^i_{(gender,\ C5)}(I)}{\partial b^i_{c3}} \\ \frac{\partial L^i_{race}}{\partial b^i_{c3}} = \frac{\partial L^i_{race}}{\partial X^i_{(race,\ C9)}(I)} \frac{\partial X^i_{(race,\ C9)}(I)}{\partial X^i_{(race,\ C7)}(I)} \frac{\partial X^i_{(race,\ C7)}(I)}{\partial X^i_{(race,\ C5)}(I)} \frac{\partial X^i_{(race,\ C5)}(I)}{\partial b^i_{c3}} \\ \quad\vdots \\ \frac{\partial L^i_{age}}{\partial b^i_{c3}} = \frac{\partial L^i_{age}}{\partial X^i_{(age,\ C9)}(I)} \frac{\partial X^i_{(age,\ C9)}(I)}{\partial X^i_{(age,\ C7)}(I)} \frac{\partial X^i_{(age,\ C7)}(I)}{\partial X^i_{(age,\ C5)}(I)} \frac{\partial X^i_{(age,\ C5)}(I)}{\partial b^i_{c3}} \end{cases}. \tag{9}$$

We use the definition of ReLU function, $max(0, x)$, and then $f'(x) = \begin{cases} 1 & x > 0 \\ 0 & x \le 0 \end{cases}$, where $x = \sum_i WX^i_C(I) + b$. According to Equations (8) and (9), we can learn that the solution of partial derivative of each subnetwork is nearly the same, and we just present the detailed process for Gender-Net. Therefore, we obtained as

$$\frac{\partial L^i_{gender}}{\partial X^i_{(gender,\ C9)}(I)} = \begin{cases} \frac{1}{N} \sum_{i=1}^{N} W^i_{C9} \frac{p_i - y}{(1-p_i)p_i} & x > 0 \\ 0 & x \le 0 \end{cases}, \tag{10}$$

and

$$p_i = f(W^i_{C9},\ X^i_{(gender,\ C9)}(I),\ b^i_{C9}), \tag{11}$$

then

$$\frac{\partial X^i_{(gender,\ C9)}(I)}{\partial X^i_{(gender,\ C7)}(I)} = \begin{cases} W^i_{C7} & x > 0 \\ 0 & x \le 0 \end{cases}, \tag{12}$$

$$\frac{\partial X^i_{(gender,\ C7)}(I)}{\partial X^i_{(gender,\ C5)}(I)} = \begin{cases} W^i_{C5} & x > 0 \\ 0 & x \le 0 \end{cases}, \tag{13}$$

$$\frac{\partial X^i_{(gender,\ C5)}(I)}{\partial W^i_{C3}} = \begin{cases} X^i_{(gender,\ C3)}(I) & x > 0 \\ 0 & x \le 0 \end{cases}, \tag{14}$$

$$\frac{\partial X^i_{(gender,\ C5)}(I)}{\partial b^i_{C3}} = \begin{cases} 1 & x > 0 \\ 0 & x \le 0 \end{cases}. \tag{15}$$

For each iteration, the results of Equations (10), (11), (12), (13), (14), and (15) can be calculated according to Back Propagation method. Based on those results, we can quickly achieve the results of Equations (8) and (9). Finally, we can update the weights and biases of the C3 layer as follows:

$$W_{C3}^{i+1} = W_{C3}^{i} - \eta \Delta W_{C3}^{i}, \tag{16}$$

and

$$b_{C3}^{i+1} = b_{C3}^{i} - \eta \Delta b_{C3}^{i}, \tag{17}$$

where W_{C3}^{i+1} and b_{C3}^{i+1} denote the weights and biases of $(i+1)$th sample and η is the learning rate.

3.2 Age Grouping and Prediction

3.2.1 Interval Partition Strategy for MORPH-II and LAP-2016. As we have analysed above, age prediction from a small age range can achieve a better result. Therefore, according to the age distribution characteristics of different datasets, we first adopt corresponding interval partitioning strategy to partition output intervals of ELM models. For a single ELM classifier, the face images are divided into different non-overlapping ranges, while the age ranges in these two ELM models have several overlapping ranges. MORPH-II and LAP-2016 datasets are used to verify the performance of our EGroupNet. Because those datasets provide the accurate ages, we add two kinds of age groups (Table 1) into the labels of each dataset, which means that each dataset has two kinds of age groups. Because the processes of adding two types of age groups into the labels of dataset are almost the same, we present only one kind of age grouping method. The detailed process is presented in Algorithm 3. After we obtain the new labels with the age groups for Morph-II and LAP-2016. From Table 1, we learn that group (a) and group (b) have overlapping ranges, and we present the detailed merging process of age group in the next subsection.

ALGORITHM 1: Interval Partition Strategy

Require:
The number of images in dataset: N;
The accurate ages of images: Age_i, $i = \{1, 2, \ldots, N\}$;
Age groups: (a_k, b_k), $k = \{1, 2, \ldots, n\}$, where $b_{(k-1)} = a_k$ and $b_k = b_{(k+1)}$;
n denotes the number of the age groups;
Ensure:
The new labels of each dataset with age groups.

```
for i in range (N)
    if (a_k < Age_i < b_k)
        add (a_k, b_k) into the labels of images with Age_i;
    else if (Age_i = a_k)
        add (a_k, b_k) or (a_(k-1), b_(k-1)) into the labels of images with Age_i;
    else if (Age_i = b_k)
        add (a_k, b_k) or (a_(k+1), b_(k+1)) into the labels of images with Age_i;
end if
Return the new label lists.
```

3.2.2 Interval Partition Strategy for Adience Benchmark. Although Adience Benchmark provides the age groups rather than the accurate age labels, the ranges of these groups, (such as 0–2, 4–6, ..., 48–53), are so small that the results of state-of-the-art methods are less than 70%. We learn that if we train only a neural network with those age groups, it is hard to obtain the better performance because of the small dataset and its real-world characteristic that is without prior manual filtering. Therefore, we attempt to design a hierarchical age prediction mechanism on our EGroupNet. First, the part of neural network in our system is still used to extract the features and

Table 1. The Detailed Age Groups for Morph-II and LAP-2016

Datasets	Age Groups		Datasets	Age Groups	
	Group (a)	Group (b)		Group (a)	Group (b)
Morph-II	10–19	15–24	LAP-2016	0–9	5–14
	20–29	25–34		10–19	15–24
	30–39	35–44		20–29	25–34
	40–49	45–54		30–39	35–44
	50–59	55–64		40–49	45–54
	60–69	65–74		50–∞	55–∞
	70–77	-		-	-

explore the correlations. Then the two ELM models classify the feature matrix into two types of age groups and the groups are presented in Table 2. Finally, the two age groups are merged into a new age group according to Algorithm 3 and through comparing the new age group with the origin age groups, we can learn that whether the prediction is right or not. We added the two kinds of age groups into the labels. We present only the process of adding group (a), because the processes of adding group (a) and group (b) are almost the same. The detailed process is shown in Algorithm 2.

ALGORITHM 2: Interval Partition Strategy for Adience

Require:

The number of images in dataset: N;

The original age groups: $m_\kappa(i)$-$n_\kappa(i)$), $\kappa = \{1, 2, \ldots, 8\}$, $i = \{1, 2, \ldots, N\}$;

Age groups: (a_k, b_k), $k = \{1, 2, \ldots, n\}$, where $b_{(k-1)} = a_k$ and $b_k = b_{(k+1)}$ n denotes the number of the age groups;

Ensure:

The new labels of each dataset with age groups.

1: **for** i in range (N)
2: **if** $(a_k \leq m_\kappa(i)$ and $n_\kappa(i) \leq b_k)$
3: add (a_k, b_k) into the labels of images with Age_i;
4: **end if**
5: Return the new label lists.

3.2.3 ELM Models Training. In the first stage, training datasets are applied to train the model and then the fine-tuned network is applied to extract the features from the original for constructing the feature matrices, which are used to train the ELM models. Note that one-third of training datasets are used to train ELM models. As we analyzed above, ELM is an efficient and fast-learning algorithm and the weights of the input layer, and biases are randomly initialized to achieve the output of the hidden (second) layer. By a simple generalized inverse operation of the second layer output matrix, the weights of hidden layer and bias are computed. It has been proved in Reference [11] that the performance of radial basis function (RBF) is superior compared to its alternatives such as liner and polynomial kernels; therefore, RBF is used to calculate $\mathbf{HH}^T$ from the input feature vectors. ELM with RBF are used to train a structure and the faces are classified into different age groups. The process is shown in Figure 3.

The detailed training stages of ELM classifier models are as follows:

Stage 1. For each image of training dataset, the features extracted by the fine-tuned neural network are merged into a feature matrix and then a final correlation matrix is achieved via relation matrix. We store the corresponding correlation matrix as one sample of a new dataset;

Table 2. The Detailed Age Groups for Adience Benchmark

Datasets	Age Groups		
	Origin Age Groups	Group (a)	Group (b)
Adience	0–2	0–4	0–6
	4–6	4–8	6–13
	8–13	8–15	13–20
	15–20	15–25	20–32
	25–32	25–38	32–43
	38–43	38–48	43–53
	48–53	48–60	53–
	60–	60–	-

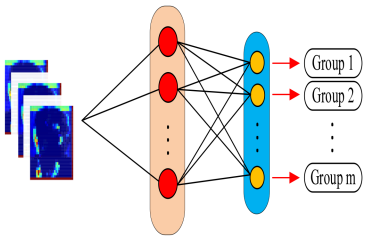


Fig. 3. The process of ELM classifier training.

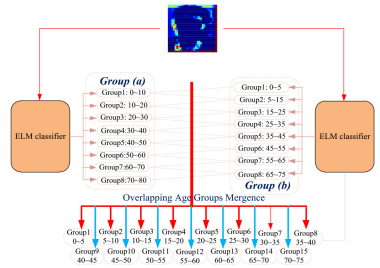


Fig. 4. The process of overlapping age groups mergence.

Stage 2. Each of the samples in the new training datset with the age groups is utilized to train the two ELM models, and then we obtain the weights and biases of the hidden layer.

Stage 3. We achieve the average weights and biases of the hidden layer for ELM classier models during the training process.

During the process of the training of ELM models, its outputs are used to train the ELM regressor. The detailed training stages of ELM regressor model are as follows:

Stage 1. For the outputs of the two ELM classifiers for each sample, we merge the groups into a smaller age group according to Algorithm 3;

Stage 2. We train the ELM regressor using the smaller age group with the accurate age label;

Stage 3. We achieve the average weights and biases of hidden layer for ELM regressor model during the training process.

3.2.4 The Process of Mergence on the Overlapping Age Groups. The goal of designing the two types of age groups is to obtain a narrow age range for a face image. The adjacent age groups of two ELM classifiers have overlapping intervals. The image is first classified into two age groups via ELM models and then a narrow age range is obtained via Algorithm 3. Figure 4 is a detailed example for the process mergence of age groups on Morph-II.

3.2.5 Age Estimation with Age Groups. During the validation and testing stages, each image of validation and testing datasets is first transformed into a feature correlation matrix via the fine-tuned neural network and then classified into two age groups. After that, a narrow age group is achieved via Algorithm 3. Finally, a final age prediction is made through the fine-tuned ELM regressor model.

ALGORITHM 3: Overlapping Age Groups Mergence Method

Require:
Given two age groups results for a face image (a, b) and (m, n) , $a < b$, $m < n$. (a, b) denotes the prediction result using one ELM classifier and (m, n) is the other result using the second classifier for the same image;

Ensure:
The final age group.

Start with the empty final age group set $\chi = \{\emptyset\}$;
if $(a < m < b$ and $b < n)$
  $\chi = \{m, b\}$;
else if $(m < a$ and $a < n < b)$
  $\chi = \{a, n\}$;
else if $(n < a$ or $b < m)$
  $\chi = \{\emptyset\}$.
end if
Return χ.

4 EXPERIMENTS

We verify the performance of our EGroupNet on MORPH-II, LAP-2016, and Adience Benchmark and compare the results with the latest methods. Our EGroupNet are trained on NVIDIA Tesla P100 using the $Tensorflow$ 1.0.1 framework [1]. First, the image is resized into 256×256 pixels. After that, 224×224 crop is chosen from the center or the four corners of the whole image. We adopt several dropout methods to reduce the risk of overfitting. For our ELM classifier and regressor models, L is set as {1,800, 2,000, ..., 4,000}. We conduct each experiment ten times, and we achieve the corresponding average results. The following shows the details of results.

4.1 Dataset Preparation

4.1.1 MORPH-II Dataset. **MORPH-II** [54] includes more than 55,000 facial images and two kinds of experimental settings on Morph-II are used to verify the performance of our EGroupNet. The first setting (S(1)) used in References [2, 6, 7, 24, 56–58, 60, 66] selects 5,492 images of Caucasian Descent people from Morph-II to lower the effects of cross-ethnicity, 80% of which are randomly selected as training dataset and the rest as testing dataset. Final results are achieved via 10-fold cross-validation. The second setting (S(2)) introduced in References [18–20, 58, 59] uses 80% of the whole images as training datset and the rest are as testing datasets, and 10-fold cross-validation is utilized in this setting.

4.1.2 LAP-2016 Dataset. The ChaLearn Looking at People 2016—Apparent Age Estimation challenge dataset [15] has 7,591 images that were recorded by human. Each label sample is noted with the mean μ and the standard deviation σ. For the reason of providing age label by the dataset, we pretrain the network on ImageNet and CelebA dataset [47] containing smiling, gender, age, and skin attributes. Then, we just fine-tune the Age-Net not the shared part with LAP-2016. Images of 4,113, 1,500, and 1,978 are randomly chosen as the training samples, validation samples, and testing samples, respectively.

The images are preprocessed as follows. First, a DPM detector [50] detect the input facial images. Because the images in this dataset are in unconstrained poses, the input images are rotated in the interval of $[-60°, 60°]$ by 5° as in Reference [64], and by $-90°$, 90°, and 180°. The highest rotation angle and detection score are obtained by the face box. Second, we increase the face box size by 40% in both width and height, and the face image is cropped. Finally, the image is resized into

Table 3. The Age Interval Distributions of the Three Categories on Adience

	0–2	4–6	8–13	15–20	25–32	38–43	48–53	60–	Total
M	745	928	934	734	2308	1294	392	442	8192
F	682	1234	1360	919	2589	1056	433	427	9411
B	1427	2162	2294	1653	4897	2350	825	869	19487

Table 4. Subnetwork Parameters

Layers	Parameters	Layers	Parameters	Layers	Parameters
Conv1	Num_output: 96 Kernel_size: 5 Stride: 2	Pool1	Num_output: 96 Kernel_size: 3 Stride: 2	Norm1	Local_size: 5 alpha: 1e-1 beta: 0.75
Conv2	Num_output: 256 Kernel_size: 3 Stride: 1	Pool2	Num_output: 256 Kernel_size: 3 Stride: 2	Norm2	Local_size: 5 alpha: 1e-1 beta: 0.75
Conv3	Num_output: 384 Kernel_size: 3 pad: 1	Pool3	Num_output: 384 Kernel_size: 3 Stride: 2	Norm3	Local_size: 5 alpha: 1e-1 beta: 0.75
Conv4	Num_output: 384 Kernel_size: 3 Stride: 1	Norm4	Local_size: 5 alpha: 0.01 beta: 0.75	Conv5	Num_output: 256 Kernel_size: 3 Stride: 1

256×256 pixels, and then a 224×224 cropped image is chosen from the center or the four corners of the whole image.

4.2 Adience Benchmark

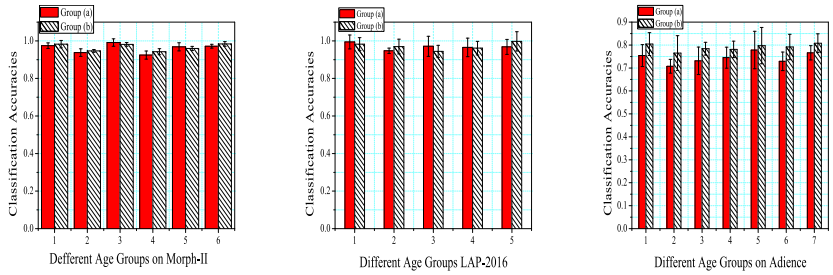
We use the recently released unconstrained Adience benchmark [14, 41] to verify the performance of EGroupNet. Because these images were uploaded to Flickr and they are without prior manual filtering, they are highly unconstrained, which means that they are closed to the real-world applications. Table 3 gives the detailed age interval distributions.

4.2.1 Network Structure. The part of neural network in EGroupNet consists of two parts, the shared network and several subnetworks. These subnetworks are with the same network layers, such as convolutional layers, contrast normalization layer, pooling layer, ReLU nonlinear function, and with identical network parameters. The detailed subnetwork configurations are shown in Table 4. Convolutional layer if followed by ReLU, a max pooling and a local response normalization layer. Every F10 layer has 4,098 units and is followed by a ReLU and 50% dropout to avoid overfitting. Each F11 layer is fully connected to a corresponding F10 layer, also with 4,098 units. and it also followed by ReLU and a 50% dropout. The finial fully connected layer fully connects F11 with 1,000 units.

4.3 Evaluation Criteria

4.3.1 Mean Absolute Error (MAE). We use ELM with RBF kernel to estimate human face images. MAE is adopted to estimate the accuracy of predicted age and it is calculated as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i|, \tag{18}$$



(a) (b) (c)

Fig. 5. Classification accuracies of different age groups on Morph-II, LAP-2016, and Adience. (a) 1 → (10–19) and (15–24), 2 → (20–29) and (25–34), 3 → (30–39) and (35–44), 4 → (40–49) and (45–54), 5 → (50–59) and (55–64), 6 → (60–69) and (65–74). (b) 1 → (0–9) and (5–14), 2 → (10–19) and (15–24), 3 → (20–29) and (25–34), 4 → (30–39) and (35–44), 5 → (40–49) and (45–54), 6 → (50–∞) and (55–∞). (c) 1 → (0–4) and (0–6), 2 → (4–8) and (6–13), 3 → (8–15) and (13–20), 4 → (15–25) and (20–32), 5 → (25–38) and (32–43), 6 → (38–48) and (43–53), 7 → (48–60) and (53–).

where x_i is the true label, y_i denotes the predicted value, and N represents the size of testing images. The cumulative score (CS) is described as

$$CS(L) = (n_{e \leq L}/N) \times 100\%, \tag{19}$$

where $n_{e \leq L}$ denotes the number of test images whose absolute error e of the age prediction is lower than L years.

4.3.2 Normal Score (ϵ). Because the LAP-2016 is labelled by different human, the results of an age prediction structure might be more conclusive by considering the variance of the annotations for each sample. Hence, by fitting a normal distribution with mean μ and standard deviation σ of the annotations for each sample, the ϵ-score is

$$\epsilon = 1 - e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{20}$$

Hence, $\epsilon \in (0, 1)$, and 0 denotes best case while 1 is worst case.

4.4 Results of Age Grouping

In the age grouping stage, we evaluate the classification accuracies of our ELM classifiers using MORPH-II, LAP-2016, and Adience Benchmark databases. Figure 5 presents the detailed classification accuracies and corresponding standard deviations of different age groups on Morph-II, LAP-2016, and Adience. Figure 6 shows the detailed classification accuracies and corresponding standard deviations under different hidden nodes on Morph-II, LAP-2016, and Adience.

As shown in Figures 5(a), 5(b), and 5(c), most of the classification results of different age groups on Morph-II and LAP-2016 have reached high accuracies. The size of the whole training dataset for corresponding age groups affects the classification accuracies, which means that large size of face images in some age range can fine-tune the model and a better performance can be achieved. According to the statistics of different age groups on these datasets, some groups have small size of face images but obtain better performance, for example, the age groups (48–53) and (53–) on Adience just have 825 and 869 images, respectively. But, the performance can not be more decreased than these of other groups that have more than 1,000 face images. More importantly, the trend of the entire classification accuracies fluctuates a little. These are mainly attributed to the novel

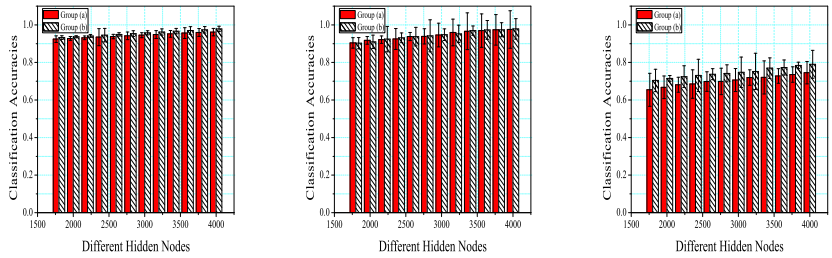


Fig. 6. Average classification accuracies under different hidden nodes on Morph-II, LAP-2016, and Adience.

Table 5. MAEs of Different Age Estimation Methods on MORPH-II

Method (S(1))	MAE	Method (S(1))	MAE	Method (S(2))	MAE	Method (S(2))	MAE
Human workers [31]	6.30	AGES [22]	8.83	IIS-LDL [20]	5.67	CPNN [21]	4.87
MTWGP [69]	6.28	CA-SVR [7]	5.88	Huerta et al. [35]	4.25	BFGS-LDL [19]	3.94
SVR [24]	5.77	OHRank [6]	6.07	OHRank [6]	3.82	LSVR [27]	4.31
DLA [66]	4.77	Rank [5]	6.49	CCA [26]	4.73	OR-CNN [52]	3.27
Rothe et al. [57]	3.45	DEX [56]	3.25	DLDL [18]	2.42	SMMR [34]	3.24
ARN [2]	3.00	DRFs [58]	2.91	Ranking-CNN [8]	2.96	dLDLF [59]	2.24
MO-CNN [60]	2.52	**EGroupNet (Ours)**	2.48	DRFs [58]	2.17	**EGroupNet (Ours)**	2.13

designed neural network, which not only extracts the discriminative features but also excavates the correlations between face attributes and age to enhance the age features quality.

At the same time, we present the entire average accuracies of groups (a) and (b) under different hidden nodes in Figure 6. We can learn that the entire average accuracies are improved with the increasement of hidden nodes. Furthermore, when the hidden nodes exceed 4,000, the improvement of performance are not obvious but increase the computational complexity of the system. Therefore, we set the hidden nodes as 4,000.

4.5 Age Estimation

4.5.1 Age Estimation on the MORPH-II Database Under Different Conditions. MORPH-II is utilized to evaluate the performance of EGroupNet, and we compare the results with the latest methods. The robustness and effectiveness of EGroupNet are analyzed in terms of the MAEs and cumulative scores (CS). We compare the performance of EGroupNet with the state-of-the-art methods. Table 5 presents the detailed results, which listed the training datasets and corresponding networks.

We can learn that our EGroupNet obtains the best results. Although DEX is an effective approach for age prediction, it may be attributed to the structure pretrained on IMDB-WIKI. The approach in Reference [2] obtains competitive results, the main reason is that the system fully utilizes the smoothed relaxation of a piecewise linear regressor. dLDLF [59] adopts a novel label distribution learning algorithm for age estimation and DRFs [58] utilizes a Deep Regression Forests (DRFs) to predict age by making full use of homogeneous data. However, these methods do not exploit the influence of other attributes of the face on the age prediction, which is one of the main reasons why our EGroupNet has achieved best prediction results among all comparison algorithms.

Three main reasons are attributed to the exciting results of our EGroupNet. First, EGroupNet is pretrained on ImageNet and IMDB-WIKI datasets. Second, our approach tries its best to excavate

Fig. 7. Examples of (a) good and (b) poor estimation on MORPH-II. "m/n/q/l" expresses group (a)/group (b)/mergence group/prediction results.

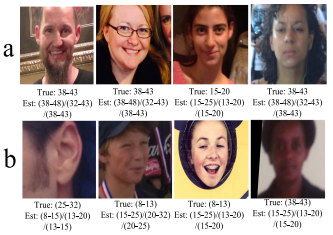


Fig. 8. Examples of (a) good and (b) poor estimation on Adience Benchmark. "m/n/q/l" expresses group (a)/group (b)/prediction results.

Table 6. ChaLearn Looking at People 2016 Apparent Age Estimation Challenge Final Results

Position	Team	Test error	Position	Team	Test error
1	OrangeLabs	0.2411	6	ITU_SiMiT	0.3668
2	palm_seu	0.3214	7	Bogazici	0.3740
3	cmp+ETH	0.3361	8	MIPAL_SNU	0.4569
4	WYU_CVL	0.3405	9	DeepAge	0.4573
5	EGroupNet (Ours)	0.3578	-	-	-

the correlations between facial attributes and age attribute, and then projects the high-level features for an image into a closely related feature matrix, which is used to make a prediction. Third, our EGroupNet makes prediction from a narrow age range instead of wide age range, which leads to better results. Although our EGroupNet has obtained satisfactory performance, the predictions of some images are not very good. Examples of good and poor results by EGroupNet are presented in Figure 7.

4.5.2 Age Estimation on the LAP-2016 Database Under Different Conditions. We also list the results of our EGroupNet on LAP-2016 dataset. Table 6 presents the final results of the challenge. Our EGroupNet model achieves 0.3419 ϵ-score in the development phase and 0.3578 ϵ-score in the testing phase. We pretrain the system on ImageNet dataset and then fine-tuned on CelebA dataset with smiling, gender, age, and skin attributes. Finally, we fine-tune the network on LAP-2016 dataset. The results verify the effectiveness of excavating the correlations between facial attributes and age attribute, and age group schemes.

Some prediction results of the validation images are presented in Figures 9 and 10. Figure 9 illustrates the invariance of CNN features to common problems, such as blur, pose, and occlusions. Figure 10 displays the failed results of age predictions, because of many possible reasons, such as face misdetection, insufficient number of images to our EGroupNet, and alignment errors.

4.5.3 Age Estimation on the Adience Benchmark Under Different Conditions. We also use an unconstrained dataset to test the performance of our proposed system. It is difficult to distinguish the race categories of the dataset, because of its absent of prior manual filtering. Dropout structure is used to limit the risk of overfitting and the dropout ratio is set to 0.5 (50% probability to set the output value of a neural as 0). Table 7 shows the age estimation results on the Adience benchmark.

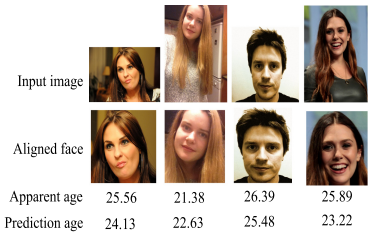


Fig. 9. Examples of satisfactory prediction.

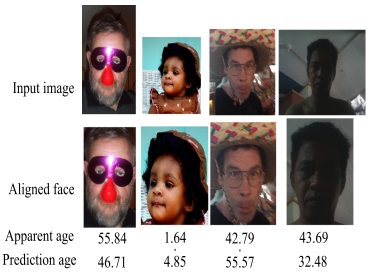


Fig. 10. Examples of poor prediction.

Table 7. Age Estimation Results on the Adience Benchmark

Method	Average Accuracy
Best from Reference [41]	0.5071 ± 0.051
Best from Reference [13]	0.523 ± 0.057
DEX [56]	0.64 ± 0.042
CNN2ELM [10]	0.6649 ± 0.0508
EGroupNet	0.6978 ± 0.0713

Listed are the mean accuracy ± standard error over all age categories.

Obviously, the results of our EGroupNet outperforms these of the reported state-of-the-art methods with considerable intervals. One obvious evident is the contribution of exploring the correlations between gender and age attribute, which enhance the useful information of age features. Furthermore, our EGroupNet utilizes the age group scheme to make the final age predictions from a narrow age range, which leads to a better results.

Some example of results are provided in Figure 8. Figure 8(a) shows the fine predictions by our EGroupNet, which proves the reliability of our method again. Figure 8(b) presents the examples of misclassification made by our system. Most obvious errors are caused by low resolution or blur and occlusions (particularly from the heavy makeup).

4.6 Ablation Study

4.6.1 Investigation on SubNetworks. The features extracted via neural network are closely related to the target. When EGroupNet only extracts age-related attribute features, the neural network pays more attention to the feature information about the age and will lose many other attribute information of the face that may be closely related to age. Our EGroupNet adopts multiple sub-networks to extract the different attribute features of the same face image, and then enhances the age feature information through correlation learning to achieve better prediction results. Adding sub-networks effectively improves the performance of the system, but at the same time makes the model more complex, which makes the training process to take more time. Therefore, this section analyzes the impact of subnetworks on the efficiency of the system in detail and Table 8 presents the detailed results.

As the numbers of sub-networks continue to increase, the parameters of the system almost double. However, since the sub-networks are trained in parallel, the training time does not increase double, but it is not the same as the training time of a single network, mainly because the convergence speed between different networks is not the same. Simultaneously, we can see that

Table 8. Ablation Study of the Impact of Subnetworks on the Efficiency of the System
Parameter refers to the relevant parameters of the CNN.

Datasets	SubNetworks	FLOPs	Parameters	Final Accuracies
Morph-II (S(1))	Age-Net	2,293M	8,473,850	2.78
	Age-Net+Race-Net	4,543M	16,723,764	2.61
	Age-Net+Race-Net+Gender-Net	6,120M	24,973,678	2.48
LAP-2016	Age-Net	2,293M	8,473,850	0.3694
	Age-Net+Race-Net	4,543M	16,723,764	0.3651
	Age-Net+Race-Net+Gender-Net	6,120M	24,973,678	0.3593
	Age-Net+Race-Net+Gender-Net+Smile-Net	7,697M	33,223,592	0.3578
Adience	Age-Net	2,293M	8,473,850	0.6741
	Age-Net+Gender-Net	4,543M	16,723,764	0.6978

Table 9. Ablation Study of Feature Correlation Schemes on Different Datasets

Datasets	Networks	Middle Results	Training Datasets
Morph-II (S(1))	Age-Net	3.74	Morph-II
	Predictions with Correlations	3.19	
LAP-2016	Age-Net	0.3842	CelebA and LAP-2016
	Predictions with Correlations	0.3792	
Adience	Age-Net	0.524	CelebA and Adience
	Predictions with Correlations	0.596	

the system performance is greatly improved in accordance with the increasing of the numbers of sub-networks. However, the different sub-networks have different effects on the performance of the system, for example, Gender-Net has the greatest impact on the age attribute, followed by race, and finally smile.

4.6.2 Investigation on Feature Correlation Schemes. As discussed earlier, exploring the correlation between face attributes and age is a crucial procedure that achieves robust and discriminative features and obtains a better performance. As shown in Figure 1, our fine-tuned EGroupNet can make three times age prediction for an facial image. The first one is the Age-Net, and then, by mapping these different kinds of features into a closely related feature matrix, the age prediction for the same image can be obtained. Our EGroupNet can make the final prediction. Table 10 shows the detailed results on the three datasets.

From Table 9, we can find that through exploring the correlation operations, these predictions achieve better performance than that of just using a single CNN without seeking these correlations. On Morph-II, we just consider the correlation among age, gender, and race, and the MAE decreases 0.55 points compared with that of Age-Net. On Adience, our approach fully explores the correlation among age and gender features, and the accuracy rate increases 0.072 points. On LAP-2016, we consider the influences of smiling, gender, and race for age attribute and the ϵ decreases 0.005 points. From the analysis above, we can learn that both exploring the correlations between age-related attributes and age attribute and fully utilizing the correlations can improve the performance of age prediction.

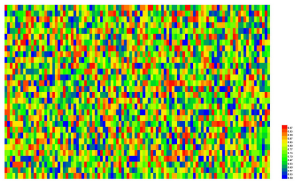


Fig. 11. Feature matrix of a testing image for Morph-II.

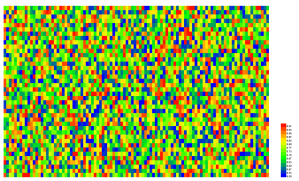


Fig. 12. Feature matrix of a testing image for LAP-2016.

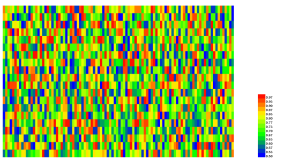


Fig. 13. Feature matrix of a testing image for Adience.

Table 10. Ablation Study of Age Group Schemes on Different Datasets

Datasets	Groups	Accuracy Rates	Networks	Results
Morph-II (S(1))	Group (a)	0.9717 ± 0.0103	Predictions with Correlations	3.19
	Group (b)	0.9837 ± 0.0124	EGroupNet	2.48
LAP-2016	Group (a)	0.9683 ± 0.0398	Predictions with Correlations	0.3792
	Group (b)	0.9975 ± 0.0511	EGroupNet	0.3578
Adience	Group (a)	0.7662 ± 0.0312	Predictions with Correlations	0.596
	Group (b)	0.8079 ± 0.041	EGroupNet	0.6978

Which relationships have the proposed model learned? The best way to do this is to visualize the age attribute feature information that are enhanced by correlation learning. The relation matrix projects the features of the fully connected layer into a space that is only closely related to age attribute and the mapped feature sizes are $3 \times 1{,}000$, $4 \times 1{,}000$, and $2 \times 1{,}000$, respectively. At this time, the space can fully reflect the relationships of age. A random testing image is transformed via relation matrix, and we visualize the feature information as Figures 11, 12, and 13. To better visualize the mapped features, we resize the feature matrices into 30×100, 40×100, and 20×100, which keep the size of these matrices uncharge. Through the analysis of the above figures, we can find that the model will fully exploit and utilize the feature information closely related to the age attribute from correlation learning. After that, it generates adequate feature information about age attribute while filtering most redundant and useless feature information unrelated to age attribute, which enhances the original age feature information.

4.6.3 Investigation on Age Group Schemes. Age group schemes also help to increase the performance of the final prediction, and in this section, we analyze the importance of age group for age prediction in details. Table 10 presents the results with different age group schemes. For accuracy rates, we present the results of two ELM classifiers on corresponding testing datasets. At the same time, we list the final prediction results and the intermediate results with correlations on Morph-II, LAP-2016, and Adience datasets in detail.

It can be seen that the accuracy rates of group (a) and (b) on Morph-II and LAP-2016 exceed 96%. Adience benchmark just provides the age groups and our method is to make a final age group prediction from a a narrow age range. We can find that making final age predictions based on narrow age ranges can improve the final performance. On Morph-II, the final average accuracy of age prediction decreases 0.61 points and the final ϵ on LAP-2016 decreases 0.0214 points. The final prediction on Adience benchmark is 0.6978, which increases 0.1018 points compared with results without using age group scheme.

4.6.4 Investigation on ELM Models. From the previous experimental analysis, it can be seen that ELM classifiers and regressor greatly improve the age prediction performance. However, we still

Table 11. Accuracies in Different Conditions on Morph-II, LAP-2016, and Adience

	Hidden Nodes		2,000	2,200	2,400	2,600	2,800	3,000	3,200	3,400	3,600	3,800	4,000
M	ELM_S	ACC_G	0.9322	0.9407	0.9448	0.9501	0.9521	0.9582	0.9619	0.9681	0.9715	0.9739	0.9762
		$MAEs_P$	2.97	2.94	2.91	2.88	2.85	2.81	2.77	2.71	2.67	2.64	2.62
	ELM_T	ACC_G	0.9276	0.9313	0.9347	0.9383	0.9417	0.9452	0.9486	0.9521	0.9564	0.9597	0.9613
			0.9368	0.9413	0.9453	0.9495	0.9537	0.9579	0.9623	0.9667	0.9703	0.9741	0.9789
		$MAEs_P$	2.79	2.78	2.76	2.75	2.73	2.71	2.69	2.66	2.64	2.61	2.48
L	ELM_S	ACC_G	0.9174	0.9233	0.9289	0.9367	0.9408	0.9462	0.9504	0.9673	0.9726	0.9741	0.9788
		$Error_P$	0.3739	0.3734	0.3727	0.3722	0.3715	0.3708	0.3703	0.3698	0.3692	0.3687	0.3682
	ELM_T	ACC_G	0.9141	0.9219	0.9251	0.9361	0.939	0.9447	0.9596	0.9661	0.9674	0.9736	0.9758
			0.9198	0.9241	0.9311	0.9379	0.9415	0.9481	0.9519	0.9699	0.9728	0.9744	0.9791
		$Error_P$	0.3655	0.365	0.3645	0.3643	0.3636	0.3627	0.3619	0.3611	0.3603	0.3593	0.3578
A	ELM_S	ACC_G	0.7148	0.7241	0.7316	0.7373	0.7409	0.7491	0.7525	0.7694	0.7731	0.7841	0.7901
		ACC_P	0.5634	0.5726	0.5837	0.5914	0.6009	0.6127	0.6228	0.6327	0.6411	0.6519	0.665
	ELM_T	ACC_G	0.6679	0.6817	0.6853	0.6978	0.697	0.7063	0.7197	0.7207	0.7281	0.7359	0.7459
			0.7153	0.7246	0.7319	0.7372	0.7413	0.7482	0.7528	0.7693	0.7734	0.7848	0.7906
		ACC_P	0.6579	0.6618	0.6657	0.6696	0.6734	0.6772	0.6813	0.6854	0.6891	0.6935	0.6978

M → the first setting (S(1)) of Morph-II, L → LAP-2016, A → Adience, ELM_S → a single ELM classifier, ELM_T → two ELM classifiers, ACC_G → classification accuracies of different age groups, $MAEs_P$ → MAEs of final prediction, $Error_P$ → final test errors, ACC_P → final prediction accuracies.

don't know what is the responsibilities and main functions of ELM in EGroupNet, so we analyze the performance of EGroupNet on Morph-II (S(1), LAP-2016, and Adience datasets under different ELM classifiers and different hidden nodes. When only a single ELM classifier is selected, the age interval is divided according to the Group (a) method, and the other settings remain the same as those of the original model. Table 11 shows the final prediction results.

From Table 11, we can conclude that the performance of EGroupNet adopting two ELM regressors is better than that using one ELM regressor. The main reason is that EGroupNet divides an image into a narrower and more precise age range, for example, input image → (10–20), (15–25) → (15–20) → final prediction, while the final age division interval with one ELM classifier is not accurate enough, for example, input image → (10–20) → final prediction. However, as the hidden nodes continue to increase, the accuracy of age grouping is continuously improved. When the node reaches 4,000, EGroupNet achieves the best accuracy, and after that, the accuracy of the system remains stable. In the case of the same network structure, if we utilize a softmax instead of using the ELM regressor and ELM classifier, then the system performance will be greatly reduced. Through the above analysis, it can be learnt that the ELM classifier and the regenerator play an important role in the whole prediction process.

5 CONCLUSION AND FUTURE WORK

We propose a learning framework called EGroupNet for age estimation. This system includes two processes: correlation excavation and age prediction. Extensive experiments conducted on MORPH-II, LAP-2016, and Adience prove the effectiveness of our EGroupNet. The experimental results verify that fully exploring the correlations between age-related attributes and age attribute can improve the final performance of prediction, and classifying the image into narrow age groups before reaching an age decision can achieve a better result. In the future, we will optimize the hierarchical structure and apply them to processing multi-facial attributes, such as smiling, hair, chin, and so on.

REFERENCES

[1] MartÍln Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, and Michael Isard. 2016. TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16)*. 265–283.

[2] Eirikur Agustsson, Radu Timofte, and Luc Van Gool. 2017. Anchored regression networks applied to age estimation and super resolution. In *Proceedings of the International Conference on Computer Vision (ICCV'17)*. 1652–1661.

[3] T. Ahonen, A. Hadid, and M. Pietikainen. 2006. Face description with local binary patterns: Application to face recognition. *IEEE TPAMI* 28, 12 (Dec. 2006), 2037–2041.

[4] K. Y. Chang and C. S. Chen. 2015. A learning framework for age rank estimation based on face images with scattering transform. *IEEE TIP* 24, 3 (Mar. 2015), 785–798.

[5] Kuang Yu Chang, Chu Song Chen, and Yi Ping Hung. 2010. A ranking approach for human ages estimation based on face images. In *Proceedings of the International Conference on Pattern Recognition (ICPR'10)*. 3396–3399.

[6] Kuang Yu Chang, Chu Song Chen, and Yi Ping Hung. 2011. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'11)*. 585–592.

[7] Ke Chen, Shaogang Gong, Tao Xiang, and Change Loy Chen. 2013. Cumulative attribute space for age and crowd density estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'13)*. 2467–2474.

[8] S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao. 2017. Using ranking-CNN for age estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'17)*. 742–751.

[9] H. Dibeklioğlu, F. Alnajar, A. Ali Salah, and T. Gevers. 2015. Combining facial dynamics with appearance for age estimation. *IEEE TIP* 24, 6 (June 2015), 1928–1943.

[10] Mingxing Duan, Kenli Li, and Keqin Li. 2018. An ensemble CNN2ELM for age estimation. *IEEE TIFS* 13, 3 (Mar. 2018), 758–772.

[11] Mingxing Duan, Kenli Li, Xiangke Liao, and Keqin Li. 2017. A parallel multiclassification algorithm for big data using an extreme learning machine. *IEEE TNNLS* 29, 6 (2017), 2337–2351.

[12] Mingxing Duan, Kenli Li, Xiangke Liao, Keqin Li, and Qi Tian. 2019. Features-enhanced multi-attribute estimation with convolutional tensor correlation fusion network. *ACM TOMM* 15, 3s (2019), 91.

[13] Mingxing Duan, Kenli Li, Canqun Yang, and Keqin Li. 2018. A hybrid deep-learning CNN-ELM for age and gender classification. *Neurocomputing* 275 (2018), 448–461.

[14] E. Eidinger, R. Enbar, and T. Hassner. 2014. Age and gender estimation of unfiltered faces. *IEEE TIFS* 9, 12 (Dec. 2014), 2170–2179.

[15] Sergio Escalera, Mercedes Torres Torres, Brais Martinez, Xavier Baró, Hugo Jair Escalante, Isabelle Guyon, Georgios Tzimiropoulos, Ciprian Corneou, Marc Oliu, Mohammad Ali Bagheri, et al. 2016. ChaLearn looking at people and faces of the world: Face analysisworkshop and challenge 2016. In *Proceedings of the Computer Vision and Pattern Recognition Workshop (CVPRW'16)*. 706–713.

[16] S. Fu, H. He, and Z. G. Hou. 2014. Learning race from face: A survey. *IEEE TPAMI* 36, 12 (Dec. 2014), 2483–2509.

[17] Y. Fu, G. Guo, and T. S. Huang. 2010. Age synthesis and estimation via faces: A survey. *IEEE TPAMI* 32, 11 (Nov. 2010), 1955–1976.

[18] B. B. Gao, C. Xing, C. W. Xie, J. Wu, and X. Geng. 2016. Deep label distribution learning with label ambiguity. *IEEE TIP* 26, 99 (2016), 1–1.

[19] X. Geng. 2016. Label distribution learning. *IEEE TKDE* 28, 7 (July 2016), 1734–1748.

[20] Xin Geng, Kate Smith-Miles, and Zhi-Hua Zhou. 2010. Facial age estimation by learning from label distributions. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI'10)*. AAAI Press, 451–456.

[21] X. Geng, C. Yin, and Z. H. Zhou. 2013. Facial age estimation by learning from label distributions. *IEEE TPAMI* 35, 10 (Oct. 2013), 2401–2412.

[22] X. Geng, Z. H. Zhou, and K. Smith-Miles. 2007. Automatic age estimation based on facial aging patterns. *IEEE TPAMI* 29, 12 (Dec. 2007), 2234–2240.

[23] Asuman Gunay and Vasif V. Nabiyev. 2008. Automatic age classification with LBP. In *Proceedings of the International Symposium on Computer and Information Sciences*. 1–4.

[24] Guodong Guo, Yun Fu, Charles R. Dyer, and Thomas S. Huang. 2008. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE TIP* 17, 7 (2008), 1178–1188.

[25] G. Guo and G. Mu. 2010. Human age estimation: What is the influence across race and gender? In *Proceedings of the Computer Vision and Pattern Recognition Workshop (CVPRW'10)*. 71–78.

[26] G. Guo and G. Mu. 2013. Joint estimation of age, gender and ethnicity: CCA vs. PLS. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR'13)*. 1–6.

[27] Guodong Guo, Guowang Mu, Yun Fu, and T. S. Huang. 2009. Human age estimation using bio-inspired features. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'09)*. 112–119.

[28] F. Gİźrpinar, H. Kaya, H. Dibeklioglu, and A. A. Salah. 2016. Kernel ELM and CNN based facial age estimation. In *Proceedings of the Computer Vision and Pattern Recognition Workshop (CVPRW'16)*. 785–791.
[29] M. A. Hajizadeh and H. Ebrahimnezhad. 2011. Classification of age groups from facial image using histograms of oriented gradients. In *Proceedings of the Conference on Machine Vision and Image Processing (MVIP'11)*. 1–5.
[30] H. Han, K Jain A, S. Shan, and X. Chen. 2017. Heterogeneous face attribute estimation: A deep multi-task learning approach. *IEEE TPAMI* 40, 99 (2017), 1–1.
[31] Hu Han, Charles Otto, Xiaoming Liu, and Anil K. Jain. 2015. Demographic estimation from face images: Human vs. machine performance. *IEEE TPAMI* 37, 6 (2015), 1148–1161.
[32] Wen Bing Horng, Cheng Ping Lee, and Chun Wen Chen. 2001. Classification of age groups based on facial features. *Tamkang J. Sci. Eng.* 4, 4 (2001), 183–192.
[33] Z. Hu, Y. Wen, J. Wang, M. Wang, R. Hong, and S. Yan. 2017. Facial age estimation with age difference. *IEEE TIP* 26, 7 (July 2017), 3087–3097.
[34] Dong Huang, Longfei Han, and Fernando De La Torre. 2017. Soft-margin mixture of regressions. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'17)*. 4058–4066.
[35] Ivan Huerta, Carles Fernández, and Andrea Prati. 2014. Facial age estimation through the fusion of texture and local appearance descriptors. In *Proceedings of the European Conference on Computer Vision (ECCV'14)*. Springer, 667–681.
[36] M. T. B. Iqbal, M. Shoyaib, B. Ryu, M. A. A. Wadud, and O. Chae. 2017. Directional age-primitive pattern (DAPP) for human age group recognition and age estimation. *IEEE TIFS* 12, 99 (2017), 1–1.
[37] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. *NIPs* 25, 2 (2012), 2012.
[38] Young Ho Kwon and N. da Vitoria Lobo. 1994. Age classification from facial images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'94)*. 762–767.
[39] Young H. Kwon and Niels Da Vitoria Lobo. 1999. Age classification from facial images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'99)*. 762–767.
[40] S. Lawrence, C. L. Giles, Ah Chung Tsoi, and A. D. Back. 1997. Face recognition: A convolutional neural-network approach. *IEEE TNN* 8, 1 (Jan. 1997), 98–113.
[41] Gil Levi and Tal Hassncer. 2015. Age and gender classification using convolutional neural networks. In *Proceedings of the Computer Vision and Pattern Recognition Workshop (CVPRW'15)*. 34–42.
[42] Changsheng Li, Qingshan Liu, Weishan Dong, Xiaobin Zhu, Jing Liu, and Hanqing Lu. 2015. Human age estimation based on locality and ordinal information. *IEEE TCYB* 45, 11 (Nov. 2015), 2522–2534.
[43] Wanhua Li, Jiwen Lu, Jianjiang Feng, Chunjing Xu, Jie Zhou, and Qi Tian. 2019. BridgeNet: A continuity-aware probabilistic network for age estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'19)*. 1145–1154.
[44] Fayao Liu, Guosheng Lin, and Chunhua Shen. 2015. {CRF} learning with {CNN} features for image segmentation. *Pattern Recogn.* 48, 10 (2015), 2983–2992.
[45] Kuan Hsien Liu, Shuicheng Yan, and C. C. Jay Kuo. 2015. Age estimation via grouping and decision fusion. *IEEE TIFS* 10, 11 (2015), 2408–2423.
[46] Xin Liu, Shaoxin Li, Meina Kan, Jie Zhang, Shuzhe Wu, Wenxian Liu, Hu Han, Shiguang Shan, and Xilin Chen. 2015. AgeNet: Deeply learned regressor and classifier for robust apparent age estimation. In *Proceedings of the Computer Vision and Pattern Recognition Workshop (CVPRW'15)*. 258–266.
[47] Z. Liu, P. Luo, X. Wang, and X. Tang. 2015. Deep learning face attributes in the wild. In *Proceedings of the International Conference on Computer Vision (ICCV'15)*. 3730–3738.
[48] Z. Lou, F. Alnajar, J. M. Alvarez, N. Hu, and T. Gevers. 2018. Expression-invariant age estimation using structured learning. *IEEE TPAMI* 40, 2 (Feb. 2018), 365–375.
[49] R. C. Malli, M. Aygİźn, and H. K. Ekenel. 2016. Apparent age estimation using ensemble of deep-learning models. In *Proceedings of the Computer Vision and Pattern Recognition Workshop (CVPRW'16)*. 714–721.
[50] Markus Mathias, Rodrigo Benenson, Marco Pedersoli, and Luc Van Gool. 2014. Face detection without bells and whistles. In *Proceedings of the European Conference on Computer Vision (ECCV'14)*. 720–735.
[51] Xiao Xiao Niu and Ching Y. Suen. 2012. A novel hybrid CNN-SVM classifier for recognizing handwritten digits. *Pattern Recogn.* 45, 4 (2012), 1318–1325.
[52] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. 2016. Ordinal regression with multiple output CNN for age estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'16)*. 4920–4928.
[53] N. Ramanathan and R. Chellappa. 2006. Modeling age progression in young faces. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 1. 387–394.
[54] K. Ricanek and T. Tesafaye. 2006. MORPH: A longitudinal image database of normal adult age-progression. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR'06)*. 341–345.

[55] R. Rothe, R. Timofte, and L. V. Gool. 2015. DEX: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE International Conference on Computer Vision Workshop (ICCVW'15)*. 252–257.
[56] Rasmus Rothe, Radu Timofte, and Luc Van Gool. 2018. Deep expectation of real and apparent age from a single image without facial landmarks. *Int J Comput Vis* 126 (2018), 144–157. https://doi.org/10.1007/s11263-016-0940-3
[57] Rasmus Rothe, Radu Timofte, and Luc Van Gool. 2016. Some like it hot visual guidance for preference prediction. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'16)*. 5553–5561.
[58] Wei Shen, Yilu Guo, Yan Wang, Kai Zhao, Bo Wang, and Alan L Yuille. 2018. Deep regression forests for age estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'18)*. 2304–2313.
[59] Wei Shen, Kai Zhao, Yilu Guo, and Alan L Yuille. 2017. Label distribution learning forests. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS'17)*. 834–843.
[60] Zichang Tan, Jun Wan, Zhen Lei, Ruicong Zhi, Guodong Guo, and Stan Z Li. 2017. Efficient group-n encoding and decoding for facial age estimation. *IEEE TPAMI* 40, 11 (2017), 2610–2623.
[61] Zichang Tan, Yang Yang, Jun Wan, Guodong Guo, and Stan Z. Li. 2019. Deeply-learned hybrid representations for facial age estimation. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI'19)*. 3548–3554.
[62] J. Tang, C. Deng, G. B. Huang, and B. Zhao. 2015. Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine. *IEEE TGRS* 53, 3 (Mar. 2015), 1174–1185.
[63] P. Thukral, K. Mitra, and R. Chellappa. 2012. A hierarchical approach for human age estimation. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'12)*. 1529–1532.
[64] M. Uricĺcr, R. Timofte, R. Rothe, J. Matas, and L. V. Gool. 2016. Structured output SVM prediction of apparent age, gender and smile from deep features. In *Proceedings of the Computer Vision and Pattern Recognition Workshop (CVPRW'16)*. 730–738.
[65] J. Wan, Z. Tan, Z. Lei, G. Guo, and S. Z. Li. 2018. Auxiliary demographic information assisted age estimation with cascaded structure. *IEEE TCYB* 48, 9 (Sep. 2018), 2531–2541.
[66] Xiaolong Wang, Rui Guo, and Chandra Kambhamettu. 2015. Deeply-learned feature for age estimation. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. 534–541.
[67] G. S. Xie, X. Y. Zhang, S. Yan, and C. L. Liu. 2015. Hybrid CNN and dictionary-based models for scene recognition and domain adaptation. *IEEE TCSVT* 27, 99 (2015), 1–1.
[68] Chao Zhang, Shuaicheng Liu, Xun Xu, and Ce Zhu. 2019. C3AE: Exploring the limits of compact model for age estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'19)*. 12587–12596.
[69] Yu Zhang and Dit Yan Yeung. 2010. Multi-task warped Gaussian process for personalized age estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'10)*. 2622–2629.

Received June 2019; revised January 2020; accepted January 2020