A Cascaded Convolutional Neural Network for Age Estimation of Unconstrained Faces

Jun-Cheng Chen[1]*, Amit Kumar[1]*, Rajeev Ranjan[1]*,
Vishal M. Patel[2], Azadeh Alavi[1] and Rama Chellappa[1]
1. University of Maryland, College Park
2. Rutgers, The State University of New Jersey
pullpull@cs.umd.edu,{akumar14,rranjan1}@umd.edu
vishal.m.patel@rutgers.edu,{azadeh,rama}@umiacs.umd.edu

Abstract

We propose a coarse-to-fine approach for estimating the apparent age from unconstrained face images using deep convolutional neural networks (DCNNs). The proposed method consists of three modules. The first one is a DCNN-based age group classifier which classifies a given face image into age groups. The second module is a collection of DCNN-based regressors which compute the fine-grained age estimate corresponding in each age class. Finally, any erroneous age prediction is corrected using an error-correcting mechanism. Experimental evaluations on three publicly available datasets for age estimation show that the proposed approach is able to reliably estimate the age; in addition, the coarse-to-fine strategy and the error correction module significantly improve the performance.

1. Introduction

Face analysis is an active research topic in computer vision with applications in surveillance, human-computer interaction, access control, and security. In this work, we focus on apparent age estimation. Traditionally, the problem is tackled through pure classification or regression approaches. In this paper, we present a cascaded approach which incorporates the advantages of both classification and regression approaches. Given an input image, we first apply the age group classification algorithm to obtain a rough estimate and then perform age group specific regression to obtain an accurate age estimate.

Like other facial analysis techniques, age estimation is affected by many intrinsic and extrinsic challenges, such as illumination variation, race, attributes, etc. One may define the age estimation task as a process of automatically labeling face images with the exact age, or the age group (age range) for each individual. It was suggested in [7] to differentiate the problem of age estimation along four concepts:

Figure 1: Estimated age on sample images from [6]. Our method is able to predict the age in unconstrained images with variations in pose, illumination, age groups, and expressions.

- Actual age: real age of an individual.
- Appearance age: age information shown on the visual appearance.
- Apparent age: suggested age by human subjects from the visual appearance.
- Estimated age: recognized age by an algorithm from the visual appearance.

The proposed cascaded classification and regression approach for apparent age estimation is based on a deep convolutional neural network. Our method consists of three main stages: (1) a single coarse age classifier, (2) multiple age regressors, and (3) an error correcting stage to correct the mistakes made by the age group classifer. Since the number of samples for apparent age estimation is limited, we exploit a DCNN model pretrained for large-scale

*The first three authors equally contribute to this work.

face identification task and finetune the model for age group classification and age regression tasks. This strategy is effective since the face recognition model trained on the CASIA-WebFace dataset [29] (*i.e.* it consists of 10,575 subjects and 494,414 images.) encodes rich information reflecting large variations in facial appearances due to aging and variations in pose, expression and illumination.

The main contribution of this work is to propose the age error correction module which mitigates the common disadvantage of coarse-to-fine approaches. Typically, the errors made at the initial classification stage cannot be recovered by the regressors at the following stage. In this work, we set up the baseline algorithm which is based on the proposed regression algorithm in Section 3.6 and study how the coarse-to-fine strategy and the error correction module improve the prediction performance. Figure 2 presents an overview of the proposed age estimation method.

The rest of the paper is organized as follows: Section 2 provides a brief overview of the related works. The proposed approach is presented in Section 3 with a concrete example. Experimental results are provided in Section 4, and Section 5 concludes the paper with a brief summary and discussion.

2. Related Work

Most of the earlier age estimation methods have focused on using shape or textural features. These features are then fed to a regression method or a classifier to estimate the apparent age [20, 26, 19, 28].

Holistic approaches usually adopt subspace-based methods, while feature-based approaches typically extract different facial regions and compute anthropometric distances. Geometry-based methods [26, 19] are inspired by studies in neuroscience, which suggest that facial geometry strongly influences age perception [19]. As such, these methods address the age estimation problem by capturing the face geometry, which refers to the location of 2D facial landmarks on images. Recently, Wu *et al.*[28] proposed an age estimation method that presents the facial geometry as points on a Grassmann manifold. To solve the regression problem on the Grassmann manifold, [28] then used the differential geometry of the manifold. However, the Grassmannian manifold-based geometry method suffers from a number of drawbacks. First, it heavily relies on the accuracy of landmark detection step, which might be difficult to obtain in practice. For instance, if an image is taken from a bearded person, then detecting landmarks would become a very challenging task. In addition, different ethnic-groups usually have slightly different face geometry, and to appropriately learn the age model, a large number of samples from different ethnic groups is required.

Unlike the traditional methods discussed, the proposed method is based on DCNN to encode the age information from a given image. Recent advances in deep learning methods have shown that compact and discriminative image representation can be learned using DCNN from very large datasets [2]. There are various neural-network-based methods, which have been developed for facial age estimation [9, 23, 16] . However, as the number of samples for estimating the apparent age task is limited, (i.e. not enough to properly learn discriminative features, unless a large number of external data is added), the traditional neural network methods often fail to learn an appropriate model.

Thukral *et. al.* [25] proposed a cascaded approach for apparent age estimation based on classifiers using the naive-Bayes approach and a support vector machine (SVM) and regressors using the relevance vector machine (RVM). However, the difference between [25] and the proposed approach is that we leverage the rich information contained in the DCNN model pretrained using a large-scale face dataset for age estimation. Also, the proposed error correction module mitigates the influences of the errors made at initial classification stage.

3. Proposed Method

Figure 2 shows an overview of our CNN-based cascaded age estimation method. Our approach consists of three main components: (1) age group classifier, (2) age regressor to predict the relative age with respect to each age group mean, and (3) apparent age error correction. Given a face image, we first apply the age group classifier to get a rough estimate of the age range from the image. Then, we choose the corresponding age regressor based on the classification results to predict the relative age with respect to the predicted group mean and combine them to get the apparent age estimate. Then, we utilize the characteristic of the classification plus regression framework to design an age error correction scheme to correct age classification and regression errors. Finally, the algorithm outputs the final age estimate for the given input image. In what follows next, we will describe each of these component in detail.

3.1. Face Preprocessing

In our work, all the face detection and facial landmark detection are handled using the open source library dlib [27][15]. Three landmark points (the center of the left eye, the center of the right eye, and the nose base) are used to align the detected faces into the canonical coordinate system using the similarity transform.

3.2. Deep Face Feature Representation

We use the DCNN model with the architecture similar to the one proposed in [29] which is pretrained for the face-identification task with softmax loss using the CASIA-WebFace dataset [29]. The CASIA-WebFace dataset consists of 10,575 subjects and 494,414 images. The architec-

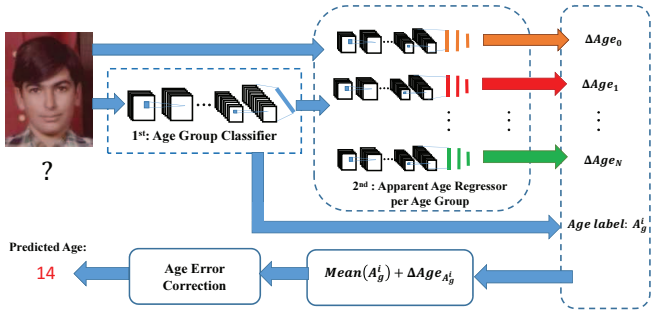

Figure 2: An overview of the proposed age cascade apparent age estimator.

ture is composed of 10 convolutional layers, 5 pooling layers and 1 fully connected layer. In our work, we use PReLU [12] instead of ReLU as the nonlinear activation function and data augmentation to train the network. The input is a color image of aligned faces of dimension $100 \times 100 \times 3$. The details of this architecture are given in Table 1. We do net surgery on this network (*i.e.*, we cut off the part after pool5 layer.) and use its pretrained weights on the CASIA-WebFace dataset to finetune on the age group dataset and apparent age estimation dataset to perform age group classification and relative age regression with respect to each age group.

3.3. Age Group Classifier

Inspired by the Viola and Jones face detection algorithm [27], we quantize the human age into several age groups (*e.g.* 0-7, 8-14, 15-23, etc.) which is an easier problem than directly performing classification or regression for the whole age range which requires a large amount of training data. To train the age group classifier, we remove the original fully connected layer, add the PReLU units and the fully connected layer with 512 outputs and finetune it on the the Images of Groups [8], Adience [4] and FGNet [11] datasets to obtain the DCNN-based age group classifier.

3.4. Apparent Age Regressor Per Age Group

To train the age regressor for each age group, we prepare the training data by splitting each training sample into the corresponding age group based on its ground truth age, and then subtract the mean of that group. The regressors are trained in two ways. The first one is to extract the pool5 features and use them to train the regressors with a large batch size. The other is to train the regressor through end-to-end network finetuning but with a smaller batch size. (*i.e.*, Similarly, we keep the part before pool5 layer and add fully connected layers.) Since the pool5 feature in the face identification task is followed by the fully connected layer with 10,575 output corresponding to the number of subject in the CASIA-WebFace dataset, the pool5 features should contain strong discriminative information from all the face images to classify a large number of subjects in the training data. In addition, we also adopt a novel loss function called, the Gaussian Loss, which takes the a rough age (*i.e.* the age is represented as a mean and a standard derivation instead of the exact age) as input and is robust for apparent age estimation. The role of the new loss function in learning the nonlinear regression method is discussed in Section 3.6.

For the pre-training of DCNN face representation model, we use the standard batch size 128 for the training phase. The initial negative slope for PReLU is set to 0.25 as suggested in [12]. The weight decay rates of all the convolutional layers are set to 0, and the weight decay of the final fully connected layer to 5e-4. In addition, the learning rate is set to 1e-2 initially and reduced by half every 100,000 iterations. The momentum is set to 0.9. Finally, we use the snapshot of 1,000,000th iteration as our pretrained model. For the finetuning of the age group classifier, we use the learning rate, 1e-4, for the convolutional layers and 1e-3 for the fully connected layers with 100,000 iterations. For training each age regressor, we first extract all the 320-d feature vectors for each age group and feed them at once into the age regressor network. We train it with 30,000 iterations using the learning rate, 1e-2, and momentum, 0.9. For the end-to-end finetuning of the regressors, we use batch size, 128, with the learning rate, 1e-4, for the convolutional layers and 1e-3 for the fully connected layers. The 120,000th models are used for each age regressor. Data augmentation is performed by randomly cropping 100×100 regions from a 128×128 box and horizontally face flipping.

3.5. Age Error Correction

In practice, the age group classifier will make errors and these errors significantly affect the final age estimation results for the second stage regressors. To handle these errors, we employ an error correcting approach. When we train the regressor for each age group, we also include the training examples from the neighboring age group. For example, given 3 age groups, (1) 8-14, (2) 15-21, and (3) 22-

28, if we want to train the age regressor for the first age group, besides the training samples with ages ranging from 8 to 14 years old, we also add the training samples from its neighboring group (*i.e.*, we added the samples from ± 2 groups for the experiments.), that is the second age group. Thus, when the classifier mistakenly assigns the subject to the neighboring age group, the regressor is able to predict a large enough value and correct the error caused by the age group classifier. Furthermore, to take the classifier error into consideration, we also add the misclassified samples to augment the training samples of all the regressors in between the true and wrong groups to increase the chance of correcting the imprecise age estimate so that it is close to the ground truth through our error correction scheme. The detailed step-by-step illustration for the age error correction scheme and other components will be presented in the following subsection. The pseudo code for our age correction approach is given in Algorithm 1.

Algorithm 1 AGE ESTIAMTION ALGORITHM

Input: (a) Input face image, I, (b) maxIter iterations, (c) age group classifier, G_0, and age regressor per age group, A_0, A_1, ..., A_{N-1} where N is the number of age groups and both age group classifier and age regressors are all DCNN-based models.

Output: Predicted apparent age, $\hat{a}$.

$g_\ell = G_0(I)$, where g_ℓ is the predicted age group label.
For i = 0 to N-1
    $\Delta a_i = A_i(I)$.
End For
$\hat{a}$ = mean(g_ℓ)+Δa_{g_ℓ}.
// Age estimation error correction
For i = 0 to maxIter - 1
    $\hat{g_\ell} = L(\hat{a})$, where $L(\cdot)$ returns the age group label of $\hat{a}$.
    IF $\hat{g_\ell} = g_\ell$
        Return $\hat{a}$
    ELSE
        $\hat{a}$ = mean($\hat{g_\ell}$)+$\Delta a_{\hat{g}_\ell}$
    End IF
    $g_\ell = \hat{g_\ell}$
End For
Return $\hat{a}$

3.6. Non-linear Regression

We use a 3-layer neural network to learn the age regressor for each age group. The number of layers is determined experimentally to be 3. The regression is learned by optimizing the Gaussian loss function as follows [6]. The Gaussian loss function is useful since the apparent age labels are usually not exact.

$$L = \frac{1}{N}\sum_{i=1}^{i=N} 1 - e^{-\frac{(\Delta x_i - \mu_i)^2}{2\sigma_i^2}}, \qquad (1)$$

Name	Type	Filter Size/Stride	#Params
Conv11	convolution	3×3×1 / 1	0.28K
Conv12	convolution	3×3×32 / 1	18K
Pool1	max pooling	2×2 / 2	
Conv21	convolution	3×3×64 / 1	36K
Conv22	convolution	3×3×64 / 1	72K
Pool2	max pooling	2×2 / 2	
Conv31	convolution	3×3×128 / 1	108K
Conv32	convolution	3×3×96 / 1	162K
Pool3	max pooling	2×2 / 2	
Conv41	convolution	3×3×192 / 1	216K
Conv42	convolution	3×3×128 / 1	288K
Pool4	max pooling	2×2 / 2	
Conv51	convolution	3×3×256 / 1	360K
Conv52	convolution	3×3×160 / 1	450K
Pool5	avg pooling	7×7 / 1	
Dropout	dropout (40%)		
Fc6	fully connection	10575	3305K
Cost	softmax		
total			5015K

Table 1: The base architecture of DCNN model used in this paper [29] to finetune on the age group classification and Δage regression for each age group.

where L is the average loss for all the training samples, Δx_i is the predicted shift in age from the mean of the corresponding age group. μ_i is the ground truth shift in age and σ_i is the standard deviation in age increment for the i^{th} training sample. The network parameters are trained using the back-propagation algorithm [21] with batch gradient descent. The gradient obtained for the loss function is given by (2). This gradient is used for updating the network weights during training using back-propagation.

$$\frac{\partial L}{\partial \Delta x_i} = \frac{1}{N\sigma^2}(\Delta x_i - \mu_i)e^{-\frac{(\Delta x_i - \mu_i)^2}{2\sigma_i^2}}. \qquad (2)$$

We apply dropout [24] after each fully connected layers to reduce the over-fitting due to the limited number of training data. The amount of dropout applied is 0.4, 0.3 and 0.2 for the input, first and second layers of the network respectively. The dropout ratio is applied in a decreasing manner to cope up with the decrease in the number of parameters for the deeper layers. Each layer is followed by the (PReLU) [12] activation function except the last one which predicts the age. The first layer is the input layer which takes the 320 dimensional feature vector obtained from the face-identification task. The output of this layer, after the dropout and PReLU operation, is fed to the fist hidden layer containing 320 hidden units. Subsequently, the output propagates to the second hidden layer containing 160 hidden units. The output from this layer is used to generate a scalar value that would describe the apparent age. Figure 3 depicts the 3-layer neural network used.

3.7. A Toy Example

To illustrate the end-to-end pipeline of the proposed age estimation algorithm, we present a toy example below. In

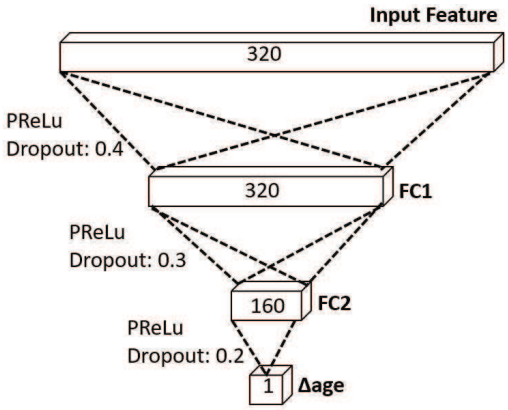

Figure 3: The 3-layer neural network used for estimating the increment in age for each age group.

this example, we use the 3 age group setting for the age group classifier where (1) the first age group is from 8 to 14 years, (2) the second 15 to 21, and (3) the third 22 to 28. The age regressor will predict Δage with respect to the mean age of its corresponding group. For example, the regressor for the first age group takes charge of predicting the real value ranging from -3 (*i.e.* 8 - 11 = -3, where 11 is the mean age of the first group) to +3 (*i.e.* 14 - 11 = 3). Now, given a face image with ground truth age 27 years old, ideally the predicted age group label should be 3 after passing the image into the age group classifier. Then, we will use the third age regressor to predict its Δage which should ideally predict the value as +2 and then we can estimate the apparent age as 25 + 2 = 27 by combining the results of the age group classifier and its corresponding age regressor where 25 is the group mean for the third age group. However, as mentioned in Section 3.5, in practice, if the age group classifier makes mistakes, the age estimation results will be wrong. To handle this error, we do the age error correction as described in Section 3.5. Now, given another face image with ground truth age 14, incorrectly being classified into third age group, we augment the misclassified samples when we train the regressor. Thus, it can be expected that the Δage should be negative enough, say -5, and as a result, the age estimation will be 25 - 5 = 20 which is still wrong but falls in the range of the second group. Then, we can pass the image again to the second group regressor to get a new estimate, say 18 - 4 = 14. We stop correcting the error when the predicted age and the previous predicted age falls in the same group or reach the maximum number of iterations. That is, we will pass the image to the first regressor again and it will predict 11 + 3 = 14 and then we stop. Otherwise, we continue to perform the correction.

The proposed age estimation algorithm is summarized in Algorithm 1. The execution orders for both the classification and regression parts are written in parallel, and thus it runs in one age group classification plus N Δage regression simultaneously in total. The maximum number of iterations is preset to avoid looping.

4. Experimental Results

We evaluate the proposed method on two publicly available datasets: Adience [4] and FG-Net [11]. Both datasets include unconstrained images of individuals which are labeled by their actual biological ages. In addition to these two datasets, we present results on the ICCV 2015 Chalearn 'Looking at people-Age Estimation' challenge dataset [6]. The main difference between this dataset and Adience and FG-Net datasets is that Chalearn includes unconstrained images of individuals labeled by their apparent ages.

4.1. Datasets

Adience dataset [4] consists of $26,580$ unconstrained images of $2,284$ subjects in 8 age groups (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60+). The standard five-fold, subject-exclusive cross-validation protocol is used for testing (*i.e.*, we merge 0-2 and 4-6 into one for the experiments of Challenge and FG-Net datasets.)

FG-Net aging dataset [11] contains a collection of $1,002$ images of 82 subjects, where each image is annotated with true age.

Images of groups [8] consists of $28,231$ faces in $5,080$ images. Each face is annotated with a label corresponding to one of the seven age groups; 0-2, 3-7, 8-12, 13-19, 20-36, 37-65, 66+ .

Chalearn Workshop Challenge dataset is the first dataset on apparent age estimation containing annotations. The dataset consists of $2,476$ training images, $1,136$ validation images, and $1,087$ test images, which were taken from individuals aged between 0 to 100. The images are captured in the wild, with variations in pose, illumination and quality. Figure 4 shows the distribution of the 'Chalearn Looking at People' Challenge dataset across the different age groups. It is evident from this figure that most of the data are distributed around the age group of 20-50, while there are very few samples in the range of 0-15 and above 55. The remaining data consists of the test set which has not been released publicly.

4.2. Experimental Details

For the first stage of age classification, we augmented the training set with the training splits of Adience[4], FG-Net[11] and Images of groups [8] datasets. To evaluate on the FG-Net, we train the seven regressor networks and then pass them through our proposed error correcting mechanism to predict the final age. Although the recently released IMDB-WIKI dataset [22] contains a large collection of images with ages, the number of the images for the young and

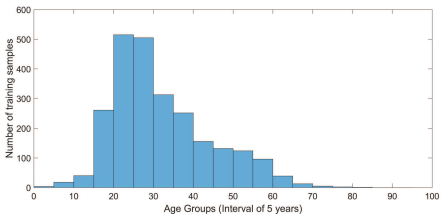

Figure 4: Training data distribution of ICCV-2015 Chalearn Looking at People Apparent Age Estimation Challenge, with regard to age groups.

old age groups is much smaller than other groups and some of the annotations for the dataset are noisy. Due to these factors, we confine the age group ranges to the ones defined by Adience[4] and focus on those previosly well-labelled datasets for this paper. The study of the influences by different ranges of age group intervals is left for future work. All the models were trained using Caffe [14]. We also compare the performance of our proposed method with a recently proposed geometry-based method [28], which is referred to as Grassmann-Regression (G-LR).

4.3. Results

To evaluate the performance of age classification algorithm, we conduct experiments on the Adience dataset [4], by following the 5 fold cross validation protocol described in [17]. From Table 2, it can be seen that our approach achieve better performance than the previous state-of-the-art methods. One thing worth noticing is that the accuracy for exact age group classification is around 53%, but the 1-off accuracy is 88.45% (*i.e.*, 1-off means the predicted label is within the neighboring groups of the true one, and 2-off means $\pm$ 2 groups). The results demonstrate the need of our error correction module to make the coarse-to-fine strategy to work better.

Method	Exact	1-off
Best from [4]	45.1 ± 2.6	79.5 ± 1.4
Best from [17]	50.7 ± 5.1	84.7 ± 2.2
Ours	$\mathbf{52.88 \pm 6}$	$\mathbf{88.45 \pm 2.2}$

Table 2: Age estimation results on the Adience benchmark. Listed are the mean accuracy $\pm$ standard error over all age categories. Best results are marked in bold.

After age group classification, we evaluated the performance of the proposed method following the protocol provided by the Chalearn 'Looking at People' challenge dataset to further investigate how the coarse-to-fine strategy and error correction mechanism help the age estimation. The error is computed as follows:

$$\varepsilon = 1 - e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{3}$$

where x is the estimated age, μ is the provided apparent age label for a given face image, average of at least 10 different user opinions, and σ is the standard deviation of all (at least 10) gauged ages for the given image. We evaluate our method on the validation set of the challenge [6], as the test set annotations are not available for performing analysis. Our baseline approach is to perform age estimation by a single deep regressor (as described in Section 3.6) on top of all the DCNN features. From Table 3, it shows that the coarse-to-fine strategy improves the prediction results of the baseline approach, and the error correction module further significantly boosts the performance which also demonstrates that the error correction module effectively fixes the errors made by the age classification step. In addition, we also show that the results of end-to-end finetuning on the training data of the challenge data for both baseline and our approach outperform the ones which are trained separately. (*i.e.*, For the results of baseline with end-to-end finetuning, we use the 500,000th model which are trained with the same batch size and learning rate for the proposed approach.) Some prediction sample results from this dataset are shown in Figure 5.

Method	Gaussian Error
G-LR [28]	0.62
Baseline	0.39
Our method without error correction	0.382
Our method with error correction	0.355
Baseline with end-to-end finetuning	0.312
Our method with end-to-end finetuning and error correction	0.297

Table 3: Performance comparison on the Chalearn Challenge dataset.

By looking at the images, we can infer that our method is robust to pose and resolution changes to a certain extent. It fails mostly for extreme illumination and extreme pose scenarios. On further inspection of the Chalearn challenge dataset, we observe the the first stage classification fails to classify correctly when the images have attributes such as hats, glasses, microphone, etc. However, the proposed error correcting mechanism makes it robust to such artifacts. The performance of our method can be improved considerably if we train using age labeled data.

Finally, we further evaluate the proposed method with end-to-end finetuning on the FG-Net dataset (*i.e.*, For

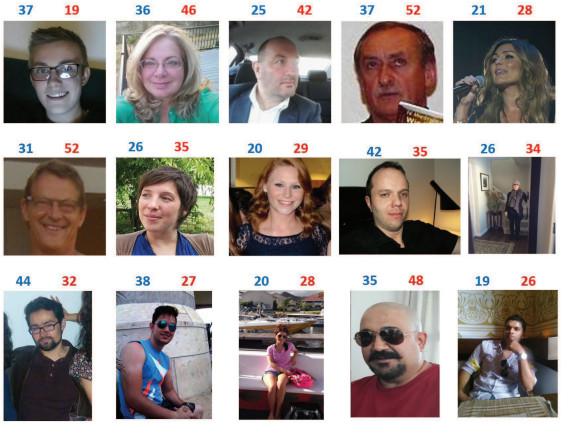

Figure 5: Age estimates on the Chalearn Validation set. The incorrect age obtained without using the self correcting module is shown in blue, while the corrected age is given in red.

FGNet, we set $\sigma = 2$ for Gaussian loss.). Since the training of DCNN is computationally intensive, a fair amount of time is needed to complete the full leave-one-person out (LOPO) evaluations. Thus, we chose to compromise and show a result that demonstrates the performance level as compared to other methods. We randomly chose 73 subjects and used their images as the training data and the rest for testing. Table 4 shows the comparison of our method with several other methods proposed in recent years. From this table, it can be seen that our method performs comparable to other state-of-the-art age estimation methods. The approach with error correction module performs much better than the one without considering neighboring samples for error correction during training.

4.4. Runtime

All the experiments were performed using NVIDIA GTX TITAN-X GPU and the CUDNN library on a 2.3Ghz computer. The first stage training for the classification task took approximately 8 hours whereas training for the second stage took approximately 8 hours per regressor. The system is fully automated with minimal human intervention. The end-to-end system takes about 2.5 seconds per image for age estimation, with only 0.8 seconds being spent in age estimation given the aligned face while the remaining time being spent on face detection and alignment.

5. Conclusions

In this work, we proposed a cascaded classification-regression framework to perform unconstrained facial apparent age estimation. The proposed approach estimates the apparent age in a coarse-to-fine manner. The age group classifier gives the rough age estimate, the regressor per age group gives the fine-grained age estimate, and the age error correcting module fixes incorrect prediction. Our experimental results demonstrate the effectiveness of the proposed approach, especially when only a limited number of training data available in the target domain.

Although our age classifiers and regressors are all based on DCNN, our framework is generic and can be extended to other non-DCNN models. In addition, the same classification-regression framework can be also applied to other vision problems, such as head pose estimation.

6. Acknowledgments

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2014-14071600012. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

[1] W.-L. Chao, J.-Z. Liu, and J.-J. Ding. Facial age estimation based on label-sensitive learning and age-oriented regression. *Pattern Recognition*, 46(3):628 – 641, 2013.

[2] J. Chen, V. M. Patel, and R. Chellappa. Unconstrained face verification using deep CNN features. *CoRR*, abs/1508.01722, 2015.

[3] K. Chen, S. Gong, T. Xiang, and C. Loy. Cumulative attribute space for age and crowd density estimation. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 2467–2474, June 2013.

[4] 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.

[5] M. El Dib and M. El-Saban. Human age estimation using enhanced bio-inspired features (ebif). In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1589–1592, Sept 2010.

[6] S. Escalera, J. Fabian, P. Pardo, X. Bar, J. Gonz?lez, H. Escalante, and I. Guyon. Chalearn 2015 apparent age and cultural event recognition: datasets and results.

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

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

[9] X. Geng, C. Yin, and Z. Zhou. Facial age estimation by learning from label distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(10):2401–2412, 2013.

Reference	Method	Training/Testing method	Result (MAE)
Luu2009 [18]	2 stage SVR in AAM subspace	802 training 200 test images	4.37
Ylioinas2013 [30]	LBP Kernel Density Estimate	LOPO	5.09
Geng2013 [10]	Label Distribution (CPNN)	LOPO	4.76
Chen2013 [3]	Cumulative Attribute SVR	LOPO	4.67
El Dib2010 [5]	Enhanced Biologically -Inspired features	LOPO	3.17
Han2013 [11]	Component and holistic BIF	LOPO	4.6
Hong2013 [13]	Biologically InspiredAAM	LOPO	4.18
Chao2013 [1]	Label-sensitive learning	LOPO	4.38
Ours proposed method	**Classification+Regression**	**890 train , 112 test**	**4.8**
Ours proposed method	**Classification+Regression+Error Correction**	**890 train , 112 test**	**3.49**

Table 4: Performance comparison of different age estimation algorithms on the FG-Net aging database using mean absolute error(MAE).

[10] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(12):2234–2240, Dec 2007.

[11] H. Han, C. Otto, and A. Jain. Age estimation from face images: Human vs. machine performance. In *Biometrics (ICB), 2013 International Conference on*, pages 1–8, June 2013.

[12] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *arXiv preprint arXiv:1502.01852*, 2015.

[13] L. Hong, D. Wen, C. Fang, and X. Ding. A new biologically inspired active appearance model for face age estimation by using local ordinal ranking. In *Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service*, ICIMCS '13, pages 327–330, New York, NY, USA, 2013. ACM.

[14] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *ACM International Conference on Multimedia*, pages 675–678, 2014.

[15] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1867–1874, 2014.

[16] S. N. Kohail. Using artificial neural network for human age estimation based on facial images. In *International Conference on Innovations in Information Technology*, pages 215–219. IEEE, 2012.

[17] G. Levi and T. Hassner. Age and gender classification using convolutional neural networks. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops*, June 2015.

[18] K. Luu, K. Ricanek, T. Bui, and C. Suen. Age estimation using active appearance models and support vector machine regression. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*, pages 1–5, Sept 2009.

[19] A. J. O'Toole, T. Price, T. Vetter, J. C. Bartlett, and V. Blanz. 3d shape and 2d surface textures of human faces: The role of "averages" in attractiveness and age. *Image and Vision Computing*, 18(1):9–19, 1999.

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

[21] M. Riedmiller and H. Braun. A direct adaptive method for faster backpropagation learning: The rprop algorithm. In *IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS*, pages 586–591, 1993.

[22] R. Rothe, R. Timofte, and L. V. Gool. Dex: Deep expectation of apparent age from a single image. In *ICCV, ChaLearn Looking at People workshop*, December 2015.

[23] A. Saxena, S. Sharma, and V. K. Chaurasiya. Neural network based human age-group estimation in curvelet domain. *Procedia Computer Science*, 54:781–789, 2015.

[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

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

[26] P. Turaga, S. Biswas, and R. Chellappa. The role of geometry in age estimation. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 946–949. IEEE, 2010.

[27] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[28] T. Wu, P. Turaga, and R. Chellappa. Age estimation and face verification across aging using landmarks. *IEEE Transactions on Information Forensics and Security*, 7(6):1780–1788, 2012.

[29] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[30] J. Ylioinas, A. Hadid, X. Hong, and M. Pietikinen. Age estimation using local binary pattern kernel density estimate. In A. Petrosino, editor, *Image Analysis and Processing ICIAP 2013*, volume 8156 of *Lecture Notes in Computer Science*, pages 141–150. Springer Berlin Heidelberg, 2013.