Unconstrained Age Estimation with Deep Convolutional Neural Networks

Rajeev Ranjan[1] Sabrina Zhou[2] Jun Cheng Chen[1] Amit Kumar[1] Azadeh Alavi[1]
Vishal M. Patel[3] Rama Chellappa[1]
[1]University of Maryland [2]Montgomery Blair High School [3]Rutgers University
rranjan1@.umiacs.umd.edu, sabrina.zhou.m@gmail.com, {pullpull,akumar14,azadeh}@umiacs.umd.edu,
vishal.m.patel@rutgers.edu, Rama@umiacs.umd.edu

Abstract

We propose an approach for age estimation from unconstrained images based on deep convolutional neural networks (DCNN). Our method consists of four steps: face detection, face alignment, DCNN-based feature extraction and neural network regression for age estimation. The proposed approach exploits two insights: (1) Features obtained from DCNN trained for face-identification task can be used for age estimation. (2) The three-layer neural network regression method trained on Gaussian loss performs better than traditional regression methods for apparent age estimation. Our method is evaluated on the apparent age estimation challenge developed for the ICCV 2015 ChaLearn Looking at People Challenge for which it achieves the error of 0.373.

1. Introduction

In this paper, we present a method for estimating the age of a person from unconstrained images. This method was specifically developed for the ICCV 2015 ChaLearn Looking at People Apparent Age Estimation Challenge[4]. The Age Estimation Challenge dataset is one of the first datasets on age estimation that contains annotations on apparent age and it contains nearly 5000 age-labeled RGB face images [4]. Samples images from this dataset are shown in Figure 1.

Face image-based age estimation has recently become a topic of interest in both industry and academia [6, 20]. 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. To further understand the task, it was suggested to differentiate age estimation into four concepts [6]:

- Actual age: real age of an individual.
- Appearance age: age information shown on the visual appearance.

Figure 1. Sample images from the ICCV apparent age estimation challenge dataset [4].

- Apparent age: suggested age by human subjects from the visual appearance.
- Estimated age: recognized age by machine from the visual appearance.

We developed a method for automatic age estimation based on DCNNs. Our method consists of four main stages: face detection, face alignment, features extraction, and a 3-layer neural network regression. We employ a *deep pyramid deformable part* model for face detection [21], and *ensemble of regression trees* method for face alignment [14]. We then encode the age information using a *deep convolutional neural network (DCNN)*. However, as the number of samples is limited, we propose to obtain features from the $pool_5$ layer of a pre-trained DCNN model for the task of face-identification. As such, we show that without re-tuning the pre-trained DCNN network for face-identification task on age estimation data, we are able to obtain a reasonably good performance. For age estimation, we adopted a

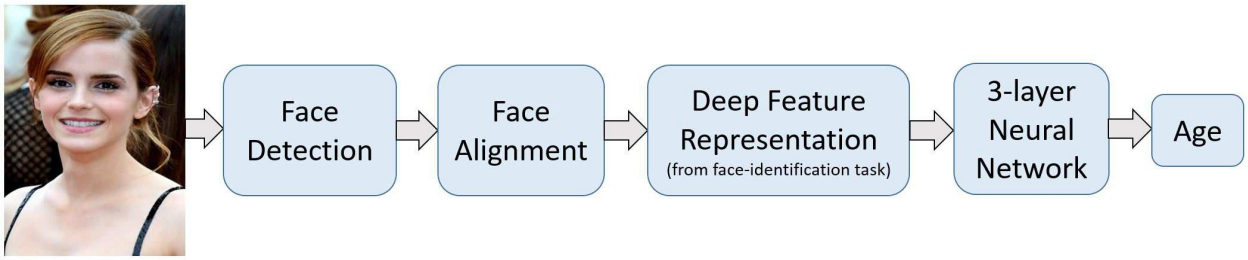

Figure 2. An overview of the proposed DCNN-based system for apparent age estimation.

3-layer neural network regression model (3NNR) with the Gaussian loss function. Finally, to further enhance the regression method, we adopt a hierarchical learning approach. Figure 2 presents an overview of our age estimation method.

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

2. Related Works

Age estimation techniques are often based on shape-based cues and texture-based cues from faces, which are then followed by traditional classification or regression methods. To estimate age, holistic approaches usually adopt subspace-based methods, while feature-based approaches typically extract different facial regions and compute anthropometric distances [20, 27, 19, 28].

Inspired by studies in neuroscience, which suggest that facial geometry is a strong element that influences age perception [19], geometry-based methods [27, 19] address the age estimation problem by capturing the face geometry. Facial geometry, in this context, 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. A Grassmann manifold is a quotient space of the special orthogonal group[1] $SO(n)$ and is defined as a set of p-dimensional linear subspaces of $\mathbb{R}^n$. In practice an element $\mathbf{X}$ of $\mathcal{G}n, p$ is represented by an orthonormal basis as a $n \times p$ matrix, *i.e.*, $\mathbf{X}^T\mathbf{X} = \mathbf{I}_p$. In the current scenario, $p = 2$ and n represents the number of facial landmarks. To solve the regression problem on the Grassmann manifold, [28] then used the differential geometry of the manifold. As such, all the points on the manifold are embedded onto the tangent space at a mean-point. To this end, standard vector-space regression method can be directly applied on the tangent plane, which is an Euclidean vector space. The experimental results conducted on various datasets showed that this geometry-based method can outperform several state-of-the-art feature-based approaches [18, 9, 17] as well as other geometry-based methods [10]. 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 ethnicity groups usually have separate face geometry, and to appropriately learn the age model, a large number of samples from different ethnic groups is required.

In contrast, our 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 are developed for facial age estimation [8, 24, 15] . 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.

Contribution: In this paper, we propose to

1. Obtain features from the $pool_5$ layer of the DCNN model pre-trained for the face-identification task, as described in [2]. We show that using features from such a network, without fine-tuning on age estimation data, we can still obtain good estimates of apparent ages. In section 4, we show that compared to the neural network-based methods which are trained using large number of age estimation data, the proposed method performs reasonably well.

2. Then we propose a 3-layer neural network based regression model learned using Gaussian loss function, and show that this model can outperform traditional regression methods.

[1] Special orthogonal group $SO(n)$ is the space of all $n \times n$ orthogonal matrices with the determinant $+1$. It is not a vector space but a differentiable manifold, i.e., it can be locally approximated by subsets of an Euclidean space.

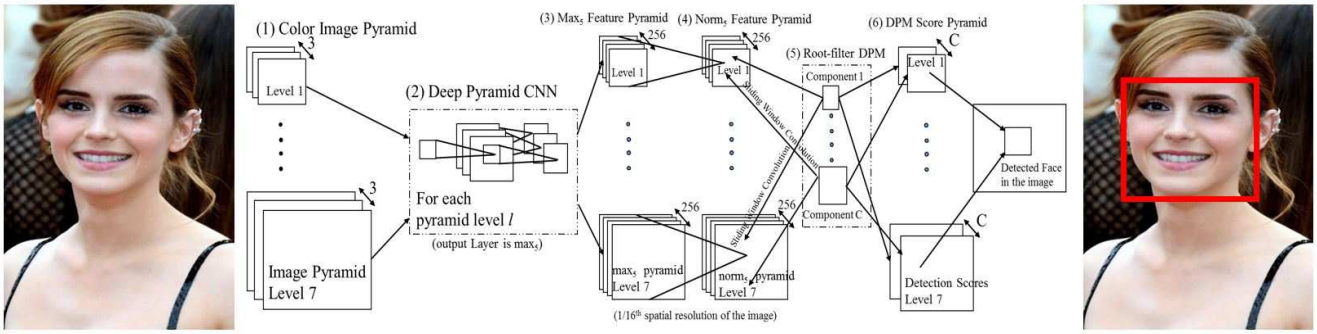

Figure 3. An overview of the face detection system using Deep Pyramid Deformable Parts Model[21].

3. Finally, as the number of samples for young and old age groups is very limited, we adopt a hierarchical learning method and further enhance the performance of the proposed regression model.

3. Proposed Approach

The proposed system consists of the complete pipeline for face detection, face alignment and age estimation. Given an image, we first detect the location of the face using the face detection algorithm[21]. We then crop the face and use the face alignment method to detect crucial fiducial points. These points are used to properly align the detected faces. Each aligned face is then passed through a deep CNN face-identification network to compute the features required for age estimation. Finally, a 3-layer neural network-based regression is performed on these features to estimate the apparent facial age. The system architecture is illustrated in Figure 2. The details of each component are presented in the following sections.

3.1. Face Detection

We use the Deep Pyramid Deformable Parts Model [21] for face detection. It consists of mainly two modules. The first one takes an input image of variable size and constructs an image pyramid with seven levels. Using this image pyramid, the network generates a pyramid of 256 feature maps at the fifth convolution layer ($conv_5$). A 3×3 max filter is applied to the feature pyramid at a stride of one to obtain the max_5 layer which essentially incorporates the $conv_5$ "parts" information. The activations at each level are normalized in the ($norm_5$) layer to remove the bias from face size.

The second module is a root-only DPM[5] trained on the ($norm_5$) feature pyramid using a linear SVM. The root filter of a certain width and height is convolved with the feature pyramid at each resolution level in a sliding window manner. Locations having scores above a certain threshold are mapped to their corresponding regions in the image to generate the bounding boxes. These boxes undergo a greedy non-maximum suppression to prune low scoring detection regions with Intersection-Over-Union (IOU) overlap above 0.3. In order to localize the face as accurately as possible, the selected boxes undergo bounding box regression.

The face detection method described above has the advantage of detecting faces at multiple resolutions and multiple pose angles. It can detect faces as low as (20×20) pixels in resolution, which makes it capable of detecting and localizing faces in unconstrained scenarios. Since the images provided for the age estimation challenge[4] contained only one face, we select the bounding box with highest score as the face location. The architecture overview of this face detection system is illustrated in Figure 3.

3.2. Face Alignment

For face alignment, we use the dlib C++ library of the method described in [14]. It uses an ensemble of regression trees to estimate the face landmark positions directly from a sparse subset of pixel intensities, achieving super-real-time performance with high quality predictions. It optimizes an appropriate loss function and performs feature selection in a data-driven manner. In particular, it learns each regressor using gradient boosting with a squared error loss function. The sparse pixel set, used as the regressors input, is selected via a combination of the gradient boosting algorithm and a prior probability on the distance between pairs of input pixels. The prior distribution allows the boosting algorithm to efficiently explore a large number of relevant features. The result is a cascade of regressors that can localize the facial landmarks when initialized with the mean face.

Given the input image with the detected face bounding box, the method first crops the face and assigns an initial mean shape S^0. The shape estimate $S^{(t+1)}$ at $(t+1)^{th}$ stage can be written in terms of the shape estimate at the previous stage $S^{(t)}$ and a shape regressor r_t which determines the increment in the shape estimate using the current shape estimate and the difference in pixel intensity values as the

features. Equation (1) summarizes the cascade regression process.

$$S^{(t+1)} = S^{(t)} + r_t(I, S^{(t)}). \quad (1)$$

Once the facial landmarks are detected, the face is aligned into the canonical coordinate system with the similarity transform using the 3 landmark points (i.e. left eye center, right eye center, and nose base). After alignment, the face is converted to gray-scale image with the resolution of 100×100 pixels. Figure 4 shows landmark detections and the aligned face obtained by this method.

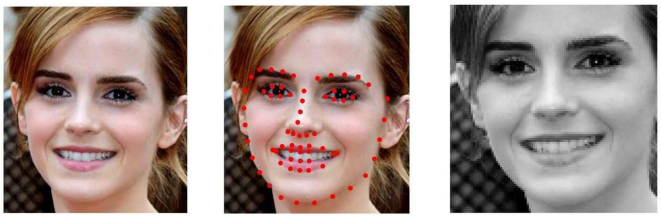

Figure 4. (a) The cropped face from the image using face detection. (b) Landmarks detected on the face. (c) Aligned gray-scale face image using 3 landmark points (eye centers and nose base).

3.3. Feature Representation

We compute the features using the CNN learned for the face-identification task as described in [2]. The feature vector representation obtained from this network is compact and discriminative for faces. We use the same network architecture proposed in [2] for age estimation.

The network is trained for a classification task on the CASIA-WebFace dataset [29] with $10,575$ subjects as the classes. It requires grey-scale aligned faces of dimensions 100×100 as input, which is obtained from the face alignment step. The network includes 10 convolutional layers, 5 pooling layers and 1 fully connected layer. Each convolutional layer is followed by a parametric rectified linear unit (PReLU) [11] except the last one, $Conv_{52}$. Moreover, two local normalization layers are added after $Conv_{12}$ and $Conv_{22}$, respectively to mitigate the effect of illumination variations. The kernel size of all filters are 3×3. The first four pooling layers use the max operator, and the average pooling for the last layer, $pool_5$. The feature dimension of $pool_5$ is thus equal to the number of channels of $Conv_{52}$ which is 320. To classify a large number of subjects in the training data, this low-dimensional feature should contain strong discriminative information from all the face images. Consequently, the $pool_5$ feature is used as our feature representation for age estimation.

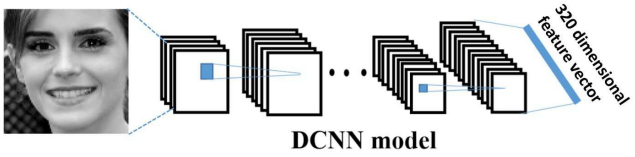

Figure 5. Feature Representation obtained for the aligned gray-scale face image using the CNN trained for face-identification task.

3.4. Age Estimation

Non-linear Regression: We use a 3-layer neural network to learn the age regression function. The number of layers is determined experimentally to be 3. The regression is learned by optimizing the loss function provided in [4]

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

where L is the average loss for all the training samples, x_i is the predicted age, μ_i is the ground truth label for age and σ_i is the standard deviation in age for the i^{th} training sample. The network parameters are trained using the back-propagation algorithm[23] with batch gradient descent. The gradient obtained for the loss function is given by 3. This gradient is used for updating the network weights during training with back-propagation

$$\frac{\partial L}{\partial x_i} = \frac{1}{N\sigma^2}(x_i - \mu_i) e^{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}}. \quad (3)$$

Since, the feature vector is dedicated for a more complex task of face-identification, not all the dimension would be necessary for age information. This calls for a dimensionality reduction technique to be used. However, we do not know the exact number of reduced dimension to retain. So, instead of reducing the dimension manually, we let the network decide for itself the best dimensions to be used for age estimation. This is done by adding Dropout [25] after each layer which reduces the over fitting due to less number of training data and high dimension of feature. The amount of dropout applied is 0.4, 0.3 and 0.2 for the input, first and second layer 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 (PReLU) [11] 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 160 number of hidden units. Subsequently, the output propagates to the second hidden layer containing 32 hidden units. The output

from this layer is used to generate a scalar value that would describe the apparent age. Figure 6 depicts the 3-layer neural network used along with the parametric details for each layer.

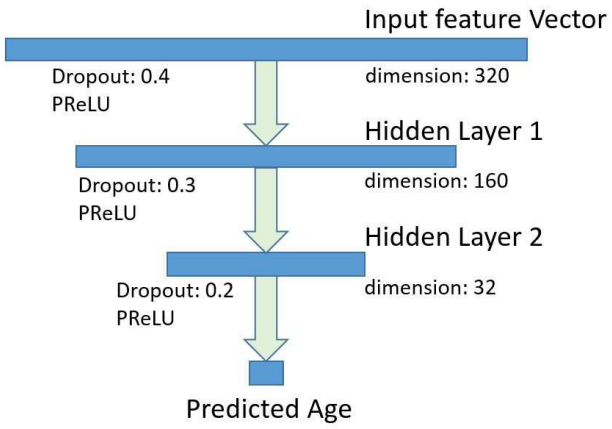

Figure 6. The 3-layer neural network used for estimating the apparent age.

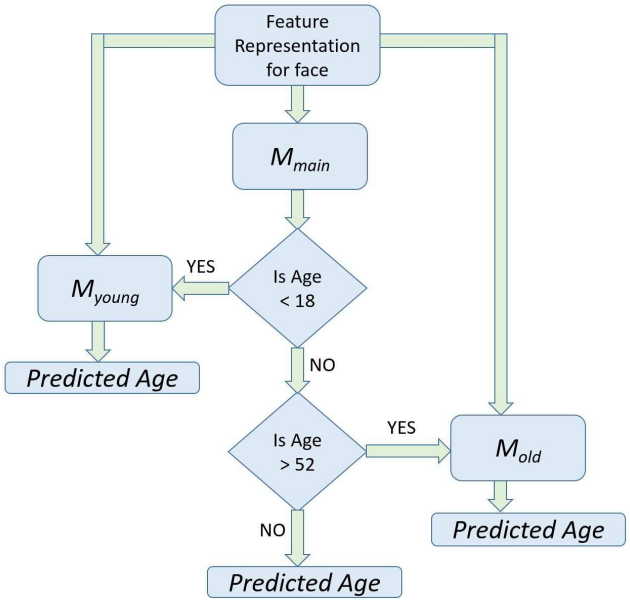

Figure 7. The flowchart illustrating the algorithm for choosing between the young model, old model and the original model for correct predicted age.

Hierarchical Learning: The number of training samples from the challenge dataset [4], with young age labels (less than 20) and old age labels (greater than 50) are fewer as compared to the mid age labels (between 20 and 50). Due to this uneven distribution in data, the regressor doesn't fit well for young and old age groups. To overcome this issue, we adopt the hierarchical approach to estimate the age as described in [26].

Let the regression model trained using training samples from [4] be denoted by M_{main}. We additionally train two separate models, M_{young} for the young age group (0 to 20) using data from Adience OUI dataset [3] and M_{old} for the old age group (50 and above) using the MORPH dataset [22]. The new models are trained using the same 3 layer neural network architecture. We estimate the final age by selecting from the predicted ages from all the 3 models. If M_{main} estimates the age between 18 and 52, we select it as the final age. If M_{main} estimates age to be less than 18, we choose the age predicted by M_{young}. In scenarios where M_{main} estimates age to be greater than 52, we choose the age predicted by M_{old}. The flowchart in Figure 7 illustrates the selection process for the age estimated by the three models.

4. Experimental Results

We evaluate the proposed method using the ICCV-2015-Challenge validation and test data sets [4]. We first compare the performance of our proposed method with the geometry based method [28], which from this point we will refer to as Grassmann-Regression (G-LR). We follow the same method as described in section 3.2 to detect landmarks for the G-LR method. We also compare the performance of our proposed 3-layaer NN regression method with the standard linear regression on the provided validation dataset. We then show the effect of adopting a hierarchical learning step for improving the proposed 3-layer NN regression method. Finally, we report our results on the test dataset.

4.1. Data Description

For performing automatic apparent age estimation from face-images, 4699 images were provided which were collectively labelled (each label is the average of at least 10 different users opinions) [4]. This is the first dataset on apparent age estimation containing annotations. The dataset consists of 2476 training images, 1136 validation images, and 1087 test images, which were taken from individuals aged between 0 to 100.

The images are captured in unconstrained environment in the wild, with variation in position, illumination and quality. Figure 8 shows the distribution of the ICCV-2015-Challenge dataset across the different age groups. It is evident from the 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.

In our experiments, we train the M_{main} model using only the training samples from [4]. We augment our data by sampling 1000 images for the age group of 0-20 from Adience [3] and 1000 images for the age group of above

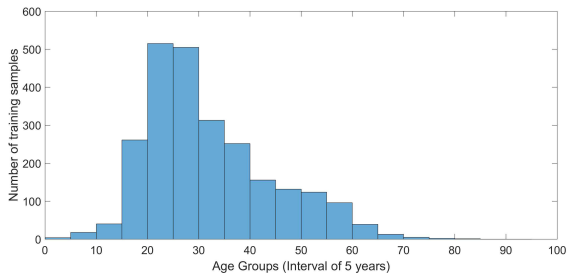

Figure 8. Training data distribution of ICCV-2015 Apparent Age Estimation Challenge, with regard to age groups

50 years old from MORPH [22]. We then train the models M_{young} and M_{old} using samples from the augmented data with age groups 0-20 years and 50 years or more, respectively. We also use the features for flipped faces to train our models. All the models were trained using Caffe [13]. We would like to emphasize that our Deep CNN network was not fine-tuned for age estimation. Even after using features obtained from a CNN trained for the face-identification task, and applying 3-layer NN based regression on them with less amount of training data, we still can achieve close to human level performance of 0.37 error for apparent age estimation. We conjecture that the features for face-identification are effective for age estimation as apparent age is one of the contributing factors for face identification. For evaluating on the test set, we also include the validation set samples in training our regression models. Table 3 shows the comparison of our method with other methods submitted for the competition, on the test set. It clearly shows that our method performs comparably even though we use very few age labeled external data in comparison to the other methods.

4.2. Results

To evaluate the performance of the proposed method, we follow the protocol provided by the ICCV-2015-Challenge [4]. We compute the error rate as follows:

$$\varepsilon = 1 - e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4)$$

where x is the estimated age, μ is the provided apparent age label for a given face-image, average of at least 10 different users opinions, and σ is the standard deviation of all (at least 10) gauged ages for the given image. We evaluate our methods on the validation set of the challenge[4], as the test set annotations are not available yet for performing analysis.

Table 1 shows that the proposed feature extraction method followed by linear regression (DCNN-LR) notably outperforms the Geodesic-based method G-LR by 12%. Also, replacing LR with the proposed 3NNR leads to significant improvement of 17% for G-LR and 13% for DCNN-LR. In addition, including the augmented data followed by hierarchical method leads to additional 1.8% performance improvement. This also shows that even without using the augmented data, the proposed method can perform well (37.7 error rate) .

	Error
G-LR	62
DCNN-LR	50
G-3NNR	45
DCNN-3NNR	37.7
DCNN-H-3NNR	**35.9**

Table 1. Error rate (in %) for the apparent age estimation task on validation set. G refers to Geometry base method using Grassmannian Manifold [4], LR refers to Linear Regression, DCNN is the proposed feature extraction method that uses the output of 5conv layer of pre-trained DCNN model for face-identification task [2], 3NNR is our proposed 3-layer Neural Network Regression model, and H refers to hierarchical learning method.

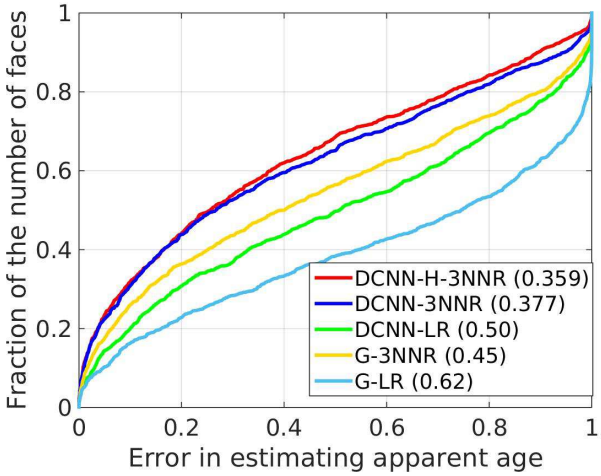

Figure 9. Cumulative error distribution curves for apparent age estimation on validation set. Numbers in the legend are mean errors for different methods.

Figure 9 shows the cumulative error distribution with respect to the number of faces in the validation set. The plots show that the estimated age for more than 60% of faces lie within the standard deviation of their apparent age (which corresponds to having error$<$0.4), when using DCNN-H-3NNR or DCNN-3NNR.

Table 2 provides a deeper insight on how our top two methods (DCNN-H-3NNR and DCNN-3NNR) perform for five different age groups on validation set. It can be clearly seen that the error is lowest for $15-30$ years age group and increases for younger and older age groups, for both these methods. On further analysis, we find that the error from our method is inversely proportional to the number of train-

Age Groups	DCNN-H-3NNR	DCNN-3NNR
$0-15$	0.62	0.82
$15-30$	0.265	0.267
$30-45$	0.445	0.421
$45-60$	0.506	0.561
$60+$	0.453	0.824

Table 2. Error rate for the DCNN-H-3NNR and DCNN-3NNR across different age groups in the challenge validation set

ing data present for that age group as shown in Figure 8. Thus, the performance is mainly limited by the uneven distribution in the training data. Also, the hierarchical learning method(DCNN-H-3NNR) significantly improves the error for $0-15$ years and $60+$ years age groups as compared to DCNN-3NNR. The overall improvement is just 1.8% because the number of samples for these age groups is less in the validation set as well.

	Error rate	External data usage
$Team-1$	26.5	D_1, D_2, D_3
$Team-2$	27.1	D_4, D_5, D_6, D_7
$Team-3$	29.5	D_1 and O_1
$Team-4$	30.6	D_6 , and O_2
DCNN-H-3NNR	**37.3**	D_4, D_6, D_7
$Team-6$	37.5	D_5, D_7, D_8,
$Team-7$	42	$D_1, D_5, D_7, D_9, D_{10}$
$Team-8$	49.9	D_7
$Team-9$	52.4	-
$Team-10$	59.4	-

Table 3. Error rate (in %) for the apparent age estimation task. More information about other contributors can be found in [4]. In addition, D and O numbering refers to D_1 : ImageNet[16], D_2: IMDB, D_3: Wikipedia, D_4: CASIA-WebFace[29], D_5: Cross-Age Celebrity (CACD)[1], D_6: MORPH[22], D_7: Adience[3], D_8: FG-NET Aging, D_9: Images of Groups[7], and D_{10}: LFW[12]. O_1 : 240,000 face images with age labels, and O_2 : own collected data.

Tabel 3 compares the performance of the participant teams in the competition on the test set. The top teams have proposed neural network-based approaches, where the top four teams have used large number of external images to boost their performances.

Figures 10 and 11 show sample images from validation set with successful and failed prediction respectively, using our DCNN-H-3NNR method for different age groups. By looking at the images, we can infer that our method is robust to pose and resolution to a certain extent. It fails mostly for extreme illumination and extreme pose scenarios. It is robust to certain attributes like glasses, but inefficient with respect to face expression or occlusion. Also, the method seldom provides correct estimation for age>70 because of the lack of training data for this age group. The performance of our method can be improved considerably if we train using more number of age labeled data.

4.3. Runtime

All the experiments were performed using nvidia GTX TITAN-X GPU without using CUDNN library on a 2.3Ghz computer. Training on the augmented dataset took approx. 2 hours. The system is fully automated with minimal human intervention. The end to end system takes $3-4$ seconds per image for age estimation, with only 1 second being spent in age estimation given the aligned face while the remaining time being spent in face detection and alignment.

5. Conclusion

In this paper, we study the performance of a DCNN method on the newly released unconstrained ICCV-2015-challenge dataset for face-image-based apparent age estimation. It was shown that obtaining features from the $pool_5$ layer of a pre-trained DCNN model for the task of face-identification, without further fine-tuning the network for age estimation task, can lead to reasonably good performance. It was also demonstrated that the proposed 3-layer neural network regression model (3NNR) with the Gaussian loss function outperforms the traditional linear regression model for age-estimation. Finally, including augmented data followed by adding the hierarchical method leads to additional 1.8% performance improvement. This also shows that even without using augmented data, the proposed method can perform well.

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.

References

[1] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014.

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

[3] 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.

Figure 10. Samples from the Validation set [4] with successful age estimation (error <0.2) using our algorithm for different age groups: (a) 0-15 (b) 15-30 (c) 30-45 (d) 45-60 (e) 60+. The estimated age is shown on the top of each sample image.

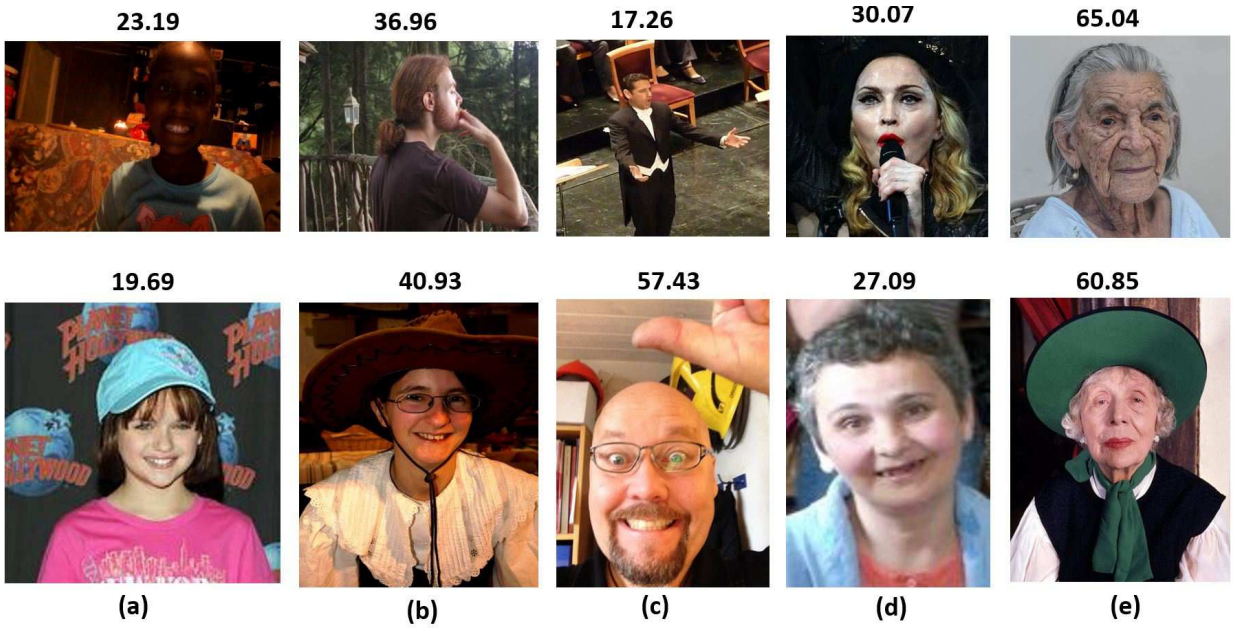

Figure 11. Samples from the Validation set [4] with incorrect age estimation (error >0.8) using our algorithm for different age groups: (a) 0-15 (b) 15-30 (c) 30-45 (d) 45-60 (e) 60+. The estimated age is shown on the top of each sample image.

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

[5] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, Sept 2010.

[6] 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.

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

[8] 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.

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

[10] G. Guo, Y. Fu, C. Dyer, and T. S. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing*, 17(7):1178–1188, 2008.

[11] 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.

[12] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[13] 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.

[14] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *CVPR*, 2014.

[15] 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.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[17] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B Cybernetics*, 34(1):621–628, 2004.

[18] A. Lanitis, C. Taylor, J. C. J, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):442–455, 2002.

[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] R. Ranjan, V. M. Patel, and R. Chellappa. A deep pyramid deformable part model for face detection. *CoRR*, abs/1508.04389, 2015.

[22] K. Ricanek and T. Tesafaye. Morph: a longitudinal image database of normal adult age-progression. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 341–345, April 2006.

[23] 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.

[24] 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.

[25] 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.

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

[27] 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.

[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.