AgeNet: Deeply Learned Regressor and Classifier for Robust Apparent Age Estimation

Xin Liu*[1], Shaoxin Li*[2], Meina Kan[1], Jie Zhang[1], Shuzhe Wu[1], Wenxian Liu[1], Hu Han[1], Shiguang Shan[1], Xilin Chen[1]
[1]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, 100190, China
[2]Tencent BestImage Team, Shanghai, 100080, China
{xin.liu, meina.kan, jie.zhang, shuzhe.wu, wenxian.liu, hu.han}@vipl.ict.ac.cn
{darwinli}@tencent.com, {sgshan, xlchen}@ict.ac.cn

Abstract

Apparent age estimation from face image has attracted more and more attentions as it is favorable in some real-world applications. In this work, we propose an end-to-end learning approach for robust apparent age estimation, named by us AgeNet. Specifically, we address the apparent age estimation problem by fusing two kinds of models, i.e., real-value based regression models and Gaussian label distribution based classification models. For both kind of models, large-scale deep convolutional neural network is adopted to learn informative age representations. Another key feature of the proposed AgeNet is that, to avoid the problem of over-fitting on small apparent age training set, we exploit a general-to-specific transfer learning scheme. Technically, the AgeNet is first pre-trained on a large-scale web-collected face dataset with identity label, and then it is fine-tuned on a large-scale real age dataset with noisy age label. Finally, it is fine-tuned on a small training set with apparent age label. The experimental results on the ChaLearn 2015 Apparent Age Competition demonstrate that our AgeNet achieves the state-of-the-art performance in apparent age estimation.

1. Introduction

Facial age estimation has drawn increasing attention in computer vision with its potential applications on video surveillance, access control, and demography [4, 10]. However, it is very hard to annotate the real age of a person in a given photo unless the data of birth of the person and the photo acquisition date are both known. Morph-II [22] and FG-NET [11] are two prevalent benchmarks for age estimation, howbeit Morph-II contains only Mugshot photos and FG-NET has only 1,002 images. Thus, lacking of large-scale real-world benchmark has become a big problem in studying robust age estimation algorithms.

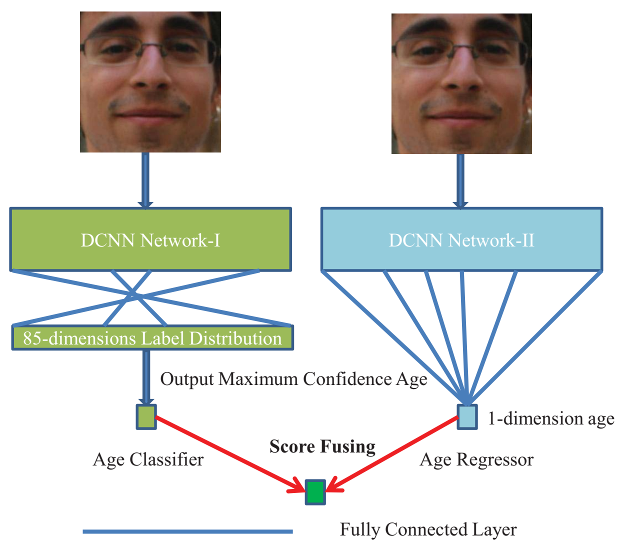


Figure 1. Overview of the proposed AgeNet: deeply learned regressor and classifier approach for robust apparent age estimation.

Recently, apparent age becomes a new measurement towards real age. As defined, apparent age is labeled by different volunteers given only the images containing the single individuals. Compared with real age, the annotated apparent age could be mutable, but the mean of the labels from different annotators are highly stable and thus can be defined as the apparent age. In this work, we adopt the first state-of-the-art apparent age dataset provided by the ICCV2015 Looking at People Challenge to study the apparent age estimation problem [3]. This dataset contains 4,699 images in total, each with a mean apparent age an-

*The first two authors contributed equally to this work.

978-1-4673-9711-7/15 $31.00 © 2015 IEEE
DOI 10.1109/ICCVW.2015.42

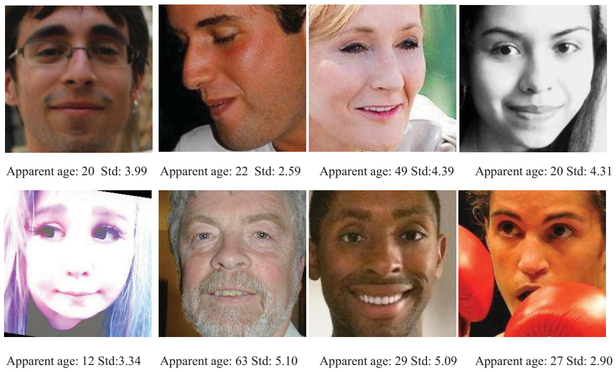


Figure 2. Some example images in the ICCV2015 Looking for People Challenge apparent age dataset, where apparent age and standard deviation are given under the images.

notation and the standard deviation of labeling. Figure 2 shows some examples in this apparent age dataset. Facial images in this dataset is collected in real-world environment and have variations in pose, occlusion, lighting, ethnicity and color mode.

This challenge is the first study of apparent age estimation from facial images. So we mainly give a brief literature review of real age estimation from facial image. Briefly speaking, face age estimation consists of two crucial stages: age pattern representation and age estimator learning.

Age pattern representation: Like many other computer vision tasks, feature representation is a basic problem in age estimation. Many state-of-the-art methods employ hand-crafted features to represent the age pattern of a face. For instance, Guo *et al.* proposed the Biologically-Inspired Features (BIF) [9] for aging pattern representation. Furthermore, Liu *et al.* proposed to fuse multiple hand-crafted local descriptors such as LBP [21], HOG [2] and BIF [9] for more robust representation [19].

Age estimator learning: Given age pattern representation, age estimator aims to predict the age of the face in the image. Generally, age estimator can be modeled as classification model [15, 6], regression model [8, 32, 5, 20], a combination of both regression and classification [7, 19] or ranking model [28, 17]. In methods of modeling the age estimator as classification, usually a class label is assigned for each age. Geng *et al.* proposed an age estimation method named Aging Pattern Subspace (AGS), which models the long-term aging process of a person and estimates the person's age by minimizing the reconstruction error in corresponded AGS [6]. On the contrary, some other methods model the age estimation as regression problem, i.e., directly regress the age value by using the typical regression methods such as support vector regression (SVR) or partial least squares (PLS). Among these representative works, Guo *et al.* adopted the kernel partial least squares (KPLS) regression for age estimation, which learns dimension-reduced feature and the aging estimator simultaneously in a joint learning framework [8]. Zhang *et al.* formulated age estimation as a multi-task regression problem and learn person-specific age estimator via a multi-task warped Gaussian process (MTWGP) model [32]. Geng *et al.* proposed a label distribution method to encode age where one face image can contribute to not only the learning of its real age, but also the learning of its adjacent ages [5]. Ni *et al.* proposed a multi-instance kernel regression algorithm to learn universal age estimator from noisy web face images [20]. For the combination of classification and regression approaches, Guo *et al.* adopted a locally adjusted robust regressor to divide age into several age groups and used classifier to determine the age within a group [7]. [19] proposed a group based age estimation framework named GEF which consists of three stages, i.e., age grouping, age estimation within age groups and decision fusion for final age estimation. In ranking based method, the order between ages is modeled. [28] proposed a rank-boosting based method and [17] proposed a relative attribute tree based method to model the relative aging relationship between different ages.

Most of the above methods model the age pattern representation and age estimator learning separately, and thus the benefits of both parts cannot be explored completely. The latest deep learning technology makes it possible to learn feature representation and age estimator together [14]. Pioneer works of employing deep learning technology for age estimation can be found in [16, 25, 29]. Among these works, [16] applied deep convolution network to classify age into age groups. [25] used a 7-layers deep convolutional network to learn deep age patterns, followed by both SVR and CCA for final age estimation. [29] used a multi-scale deep convolutional neural network for fully end-to-end age regressor learning. These methods have achieved better performance benefited from the deep learning technology.

In light of the existing works in age estimation and explosive progresses in deep learning, this paper proposes to estimate age by combining classification models and regression models, both exploiting very large-scale deep convolutional neural network. The two age estimators are complementary to each other and are further fused for robust age estimation. Figure 1 demonstrates the general idea of the proposed AgeNet. The contributions of this work are summarized as below:

1) We propose an end-to-end apparent age estimation approach, named by us AgeNet, in which the age estimation is formulated as deep regression and deep classification models, both exploiting large-scale Deep Convolutional Neural Network (DCNN). In our DCNNs, regression of age and classification of Gaussian label distribution are used as the loss function respectively;

2) To reduce the risk of over-fitting on the small apparent

age training set, a general-to-specific deep transfer learning scheme is developed;

3) Intensive experimental study of the proposed method for robust apparent age estimation is conducted and our approach wins the runner-up of the ICCV2015 Looking at People Challenge - Track 1 Apparent Age Estimation.

The rest of this paper is organized as follows. Section 2 reviews the related works about very large-scale deep convolutional neural network and the face aging encoding strategies. Section 3 details the proposed method and section 4 gives the experimental results. Section 5 concludes this work and discusses the further research.

2. Related Works

In this section, we firstly give a short introduction of the very large-scale deep convolutional neural network, and then we present three commonly used age encoding strategies.

In practice, the very large-scale deep convolutional neural network has yielded quite impressive performance in image recognition problem [24]. In this work, we deploy the GoogLeNet [24], which is a very large-scale deep convolutional neural network with 22 layers. Figure 3 shows the architecture of GoogLeNet. Considering the promising performance the GoogLeNet achieved, we adopt the similar architecture as it for the end-to-end learning and make some modification, which is discussed in section 3.1.

There are three commonly used strategies to encode age in state-of-the-art age estimation methods, namely, 1-dimension real-value encoding, 0/1 encoding and label distribution encoding.

1-dimension real-value encoding: It is straightforward to encode age as a 1-dimension real value as it is, e.g., 1, 3, 25, 82 and a regression model can be learned to directly predict the age.

0/1 encoding: The 0/1 encoding is widely used in neural network to encode one class with only single 1 and all the rest as 0, e.g., [0, 0, 0,...,1, 0, 0, 0,..., 0]. Age can be also encoded in this way if regarding each age as a class.

Label distribution encoding: Geng *et al.* [5] proposed a label distribution based age encoding strategy. The general idea of label distribution is representing the label of an instance by its description degree in each label. In this work, we adopt the Gaussian label distribution. For a given image I, if its age is y, then the corresponding age label is represented as a multi-dimension vector, with the j-th dimension as follow:

$$l_j = \exp\left(\frac{-(j-y)^2}{2*\sigma^2}\right)/\sigma, j = 1, ..., M \tag{1}$$

where j denotes the j-th chronological age, y is the ground-truth age, σ is the label standard deviation, and M is the dimension of the age label vector which is also the maximum age, e.g., 85 in this work. Figure 4 demonstrates an example of label distribution based age encoding.

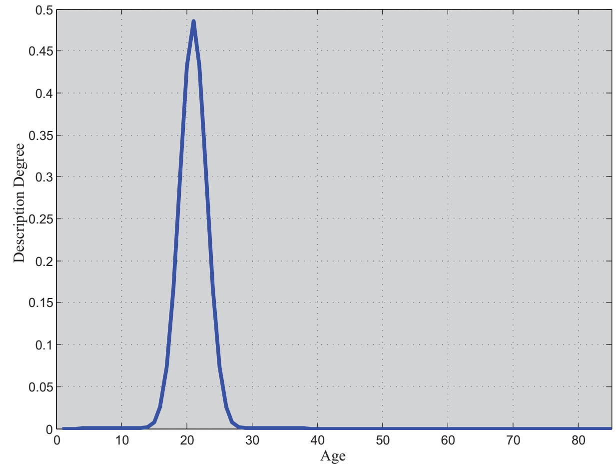


Figure 4. An example of label distribution based age vector with apparent age 21 and a labeling standard deviation of 4.0265.

3. Proposed Method

This section details the proposed deeply learned regressor and classifier for robust apparent age estimation.

3.1. Method Overview

Figure 1 presents an overview of our method. The basic idea of our method is that the age estimator is modeled as a classifier and a regressor respectively, and then the two models are complementarily fused for better performance. With the real-value based regression and Gaussian label distribution based classification as the loss functions, very large-scale deep neural network is used to learn layer-wise aging pattern representations and the final age estimator together, formulating an end-to-end deep architecture for apparent age estimation. In this work, we make two modifications of the GoogLeNet. First, we remove the two auxiliary loss layers. Second, we add batch normalization layer [12] before each ReLU operation and remove all the dropout operations to accelerate the convergence of this very large-scale deep network. Owe to the usage of batch normalization, we found it unnecessary to add two auxiliary loss layers for the purpose of avoiding vanish of gradient problem. By removing the two auxiliary loss layers, the performance is even improved slightly.

In the deep age regressor, the Euclidean loss is used to measure the 1-dimensional real-value encoding. To avoid the risk of scale unbalance in network, we add the sigmoid operation before the Euclidean loss. If the sigmoid operation is not included, the network optimization will crashes due to a gradient overflow problem. In the deep age classifier, age is encoded as an 85-dimension Gaussian label

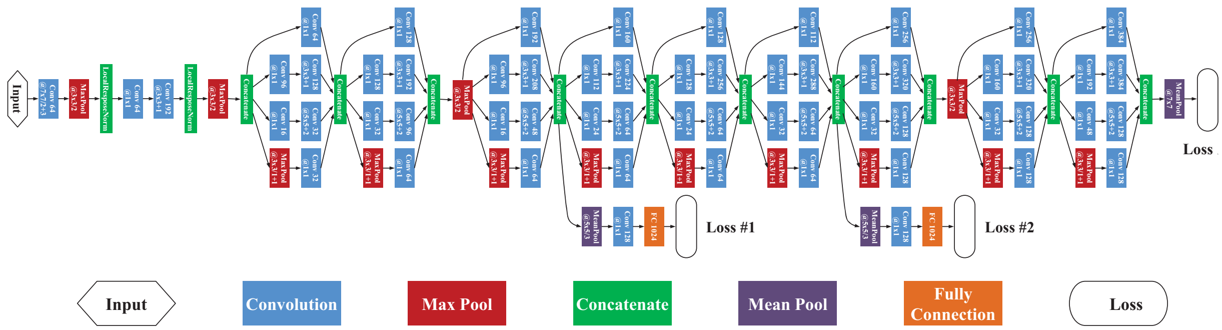


Figure 3. The architecture of the GoogLeNet deep convolutional neural network.

distribution vector and cross-entropy loss is adopted to accommodate the probability representation of the label distribution. Details for the deep regressor and age classifier will be presented in the following two subsections.

3.2. Deeply Learned Age Regressor

The deep age regressor models apparent age estimation as an end-to-end deep regression problem. A Sigmoid layer is added above the one-dimension output layer to normalize the output to $[0, 1]$, so the apparent age should be normalized to [0,1] in advance. In this work, we normalize the apparent age by dividing the age with 100. The Euclidean loss function for deep age regressor learning is formulated as below:

$$E(W) = \frac{1}{2N} \sum_{i=1}^{N} \parallel \hat{y}_n - y_n \parallel_2^2 \tag{2}$$

where W denotes the parameter of the deep convolutional network, y_n denotes the normalized ground-truth apparent age, $\hat{y}_n$ is the network output, and N is the batch size. The final age estimation will be:

$$R = f(\hat{y}_n * 100 + 0.5) \tag{3}$$

where R denotes the age regressor and $f(x)$ operation output the maximum integer less than x.

3.3. Deeply Learned Age Classifier

The deep age classifier models apparent age estimation as an end-to-end deep classification problem. A straightforward way is to take the 0/1 age encoding coupled with the Softmax loss. However, this kind of strategy encodes the distance between all ages equally and does not take the relationship between adjacent ages into consideration. It is usually harder to tell the difference between adjacent age (e.g., 40 and 41) than the non-adjacent age (e.g., 40 and 80). The label distribution based age encoding [5] models the adjacent age patterns and achieves more robust real age estimation. Thus, we employ the label distribution based age encoding in the proposed deep age classifier framework. For the label distribution based age encoding, the cross-entropy loss is used in deep age classifier as follow:

$$E(W) = \frac{-1}{N} \sum_{i=1}^{N} \sum_{n=1}^{L} [p_{in} log \hat{p}_{in} - (1 - p_{in}) log(1 - \hat{p}_{in})] \tag{4}$$

where $p_{in} \in [0, 1]$ denotes the n-th dimension of the ground-truth age label distribution vector for training image i as presented in Equation. 1, $\hat{p}_{in}$ is the corresponding network output, L denotes the length of age label distribution vector, and N is the batch size.

The final estimated age of image from deep age classifier is calculated as the age with maximum confidence in the output label vector:

$$C = \underset{n}{\operatorname{argmax}}\, \hat{p}_{in} \tag{5}$$

where C denotes the age classifier and $\hat{p}_{in}$ is the n-th dimension of the output age label vector of network.

3.4. General-to-Specific Deep Transfer Learning

To reduce the risk of over-fitting, we propose a general-to-specific deep transfer learning scheme for deep network training. Figure 5 shows the three stages of this scheme, i.e., pre-train with face identities, fine-tune with real age, and fine-tune with apparent age. We believe that adapting the deep network from general domain to specific task can reduce the un-certainty and diversity of representations as seen from the experimental evaluations in section 4.3.

Stage 1. Face identification is different from face age, however they are correlated to each other and there are quite large-scale images for face identities. Therefore, we firstly employ the large-scale face identities database CASIA-WebFace [30] to pre-train the network, which is much better than random initialization.

Stage 2. Real age is different from apparent age, but they are similar to each other in most cases. So, we further employ the face images with real age are deployed to fine-tune the deep network from stage 1, including CACD [1], Morph-II [22] and WebFaceAge [20].

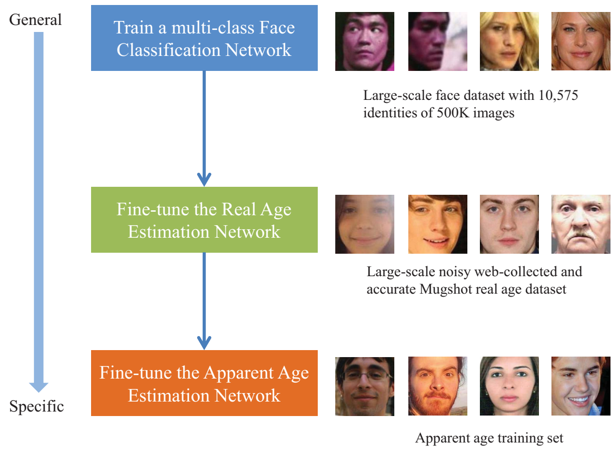
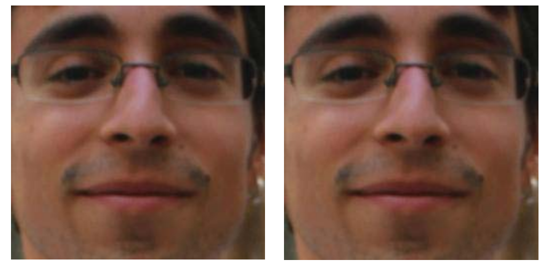


Figure 5. The general-to-specific deep transfer learning strategy for learning robust transparent age estimator.

Table 1. Outside training data for deep transfer learning.

Train Set	Short description
CASIA-WebFace[30]	500K images of 11,575 persons
CACD [1]	160K images of age 14 to 62
WebFaceAge [20]	600K images of age 1 to 82
Morph-II [22]	55,135 images of age 16 to 77



(1) Exterior (2) Interior

Figure 6. Demonstration of Exterior and Interior face normalization templates: (1) Exterior face template (2) Interior face template.

Table 2. Parameters of eight models fused in this work.

Face Template	Crop Size	Type of Age Estimator
Exterior	248	Regressor
Exterior	227	Regressor
Exterior	248	Classifier
Exterior	227	Classifier
Interior	248	Regressor
Interior	227	Regressor
Interior	248	Classifier
Interior	227	Classifier

Stage 3. Finally, we employ apparent age training set to fine-tune the deep network from stage 2, producing a robust apparent age deep network.

In stage 2 and 3, both the age classifier and age regressor are learned. Table 1 presents the four outside datasets used in stage 1 and 2. It should be noticed that, both CACD and WebFaceAge are web-collected age dataset by searching using keywords like "10 years old" or "Emma Watson 2004", so partial age labels are noisy and inaccurate. The Morph-II dataset is accurate, but it contains Mugshot photos only, which has very different distribution with the apparent age dataset. We believe some aging patterns should be shared between the real age dataset and the apparent age dataset.

3.5. Face Preprocessing

The face images for age estimation are preprocessed with three steps including face detection, facial landmark localization and face normalization.

Face Detection: In face detection stage, we adopt the face detection toolkit developed by VIPL lab of CAS [26]. One can refer to [27] for more details.

Facial Landmark Localization: We apply the Coarse-to-Fine Auto-Encoder Networks (CFAN) [31] to detect the five facial landmarks in the face the left and right center of the eyes, the nose tip, the left and right corner of mouth.

Face Normalization: Inspired by Li *et al*'s work [18], we take two different face normalization methods named as Exterior and Interior. Figure 6 displays the exemplars from the two face normalizations. The Exterior face template contains not only intrinsic but also holistic contextual information, while the Interior face template contains only the internal facial organs. In both Exterior and Interior face templates, the face is normalized into 256×256 pixel size.

3.6. Ensemble Learning

The final age estimation output is the fusion of eight deep neural networks, details of these eight models are shown in Table 2. The crop size is set for random data argumentation during training. In the test phase, only the center face patch with the same crop size is used. The ensemble strategy can be abstracted in two perspectives:

Model Ensemble: From the perspective of model ensemble, we combine deep age regressor and deep age classifier for more robust age estimation.

Face Template Ensemble: From the perspective of multi-patch face representation, we combine the models of different face normalization template and crop size.

4. Experiments

In this section, we present the experimental evaluations of the proposed method. First, we briefly review the ICCV2015 Looking for People Apparent Age Estimation Challenge. Then, we present the implementation details.

Table 4. Comprehensive model evaluations on the validation set.

Face Templates	Crop Size	Age Estimator	Mean Normalized Error	MAE
Exterior	248	0/1 Encoding with Softmax Loss	0.3900	4.4877
Exterior	248	Regressor	**0.3339**	**3.7711**
Exterior	227	Regressor	0.3498	3.9357
Exterior	248	Classifier	0.3360	3.9489
Exterior	227	Classifier	0.3413	3.9445
Interior	248	Regressor	0.3346	3.8539
Interior	227	Regressor	0.3548	4.0308
Interior	248	Classifier	0.3432	4.0317
Interior	227	Classifier	0.3513	4.0854
Fusion of Exterior	248 / 227	Classifier / Regressor	0.2932	3.3820
Fusion of Interior	248 / 227	Classifier / Regressor	0.3010	3.4921
Fusion of Classifiers	248 / 227	Classifier	0.3086	3.5704
Fusion of Regressors	248 / 227	Regressor	0.3036	3.4938
All fusion	248 / 227	Classifier / Regressor	**0.2872**	**3.3345**

Table 3. The three subsets in the apparent age dataset.

SubSet	Number of Images
Training Set	2,476
Validation Set	1,136
Test Set	1,087

Finally, we present a comprehensive study of the proposed method.

4.1. ICCV2015 Apparent Age Estimation Challenge

The ICCV2015 Apparent Age Estimation Challenge aims to investigate the performance of estimation methods on apparent age rather than real age. This is the first dataset on age estimation containing annotations of apparent age. This dataset is composed of 4,699 images, with each image collectively labeled at least 10 different users, and the average of them is used as the ground truth. The data is split into three sets as shown in Table 3.

The challenge is composed of development phase and final test phase. In the development phase, the model is trained on the training set and evaluation is conducted on the validation set. In the final test phase, both the training set and validation set can be deployed for model training. The performance is measured by mean normalized error calculated as:

$$\bar{\epsilon} = \frac{1}{N}\sum_{i=1}^{N}\left[1 - exp(\frac{-(y_i-\hat{y}_i)^2}{2\sigma_i^2})\right] \tag{6}$$

where $\hat{y}_i$ denotes the estimated apparent age, y_i denotes the ground-truth apparent age, and σ_i denotes the standard deviation for the test image i, and N is the total number of test images. Besides, we also introduce the widely used mean absolute error (MAE) measure:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} \mid y_i - \hat{y}_i \mid \tag{7}$$

4.2. Implementation Details

For all our deep networks, we set the base_lr as 0.01 and reduce the learning rate by polynomial with gamma value equals to 0.5. The momentum is set as 0.9 and the weight decay is set as 0.0005. All the experiments are conducted in Titan-X GPU with 12GB memory using the Caffe deep learning toolbox [13].

For the stage 1 (pre-train on face identity dataset), we set the batch size as 24 and the total iterations as 320K, it takes 1.5 days to train one model.

For the stage 2 (fine-tune on the real age dataset), we set the batch size as 50 and the total iterations as 100K, it takes 1 day to train one model.

For the stage 3 (fine-tune on the final apparent age train set), we set the batch size as 50 and the total iterations as 10K, it takes 2.5 hours to train one model.

4.3. Experimental Evaluations

In this section, we conduct a comprehensive evaluation of the proposed method.

Comparisons of age regressor and age classifier: In Table 4, we compare the performance of the age regressor and classifier using two face normalization templates and two different crop sizes. The experimental results demonstrate that:

1) The deeply learned regressor and classifier achieve comparable performance. The best performance of single model is achieved by deeply learned age regressor with Exterior face template and crop size as 248. Overall, the

Table 5. Comparisons of different transfer learning stages on the final performance.

Transfer Learning Stage	Mean Error	MAE
No Pre-train	0.5381	7.0924
Real Age Net Pre-train	0.3994	4.7526
Face Net Pre-train	0.3504	4.2095
Face Net and Real Age Net	**0.3360**	**3.9489**

deeply learned age regressor achieves slightly better performance than deeply learned age classifier.

2) The Gaussian label distribution is significantly better than 0/1 encoding for age classifier learning.

3) The ensemble of age classifier and regressor or face templates can both improve the performance of apparent age estimator. The best performance is achieved by fusing all the eight models. The MAE of our best result is 3.3345, which is a very impressive performance for apparent age estimation.

The efficiency of general-to-specific deep transfer learning: In Table 5, we study how the general-to-specific deep transfer learning improves the performance of apparent age estimation. We set the face normalization template and crop size as Exterior and 248 respectively, and train age classifier only. It can be seen that:

1) Random initialization without pre-train yields poor performance due to the the over-fitting of large-scale deep network on so small-scale training set.

2) The pre-train of multi-class face classifier network is much efficient than the pre-train of real age network, which means that pre-train with large-scale images from related tasks can help improving the generalization ability of the network.

3) Though the web-collected face age label is noisy, it can still help to improve the performance of apparent age performance, which demonstrates great robustness of the general-to-specific deep transfer learning.

Success and Failure cases: Figure 7 presents some good and bad cases of apparent age estimation results by the proposed approach. It can be seen that our approach is robust to variations in pose, lighting, ethnicity, occlusion and color mode. However, our approach does not work very well for face blur, mis-alignment or senior people.

Comparison with the other competitors on the final evaluation: Table 6 presents the result of our method and the other teams on final evaluation of ICCV2015 Looking for People Apparent Age Estimation Challenge. According to the final official report [3], we rank 1st in the development phase and 2nd in the final test phase. Our approach is slightly worse than the method of CVL_ETH by 0.5%. However, while the method of CVL_ETH employs 20 VGG deep neural networks [23], we only employ 8 GoogLeNets, leading to much less computation cost.

Table 6. Comparisons of the performance in the final evaluation.

Rank	Team	Development	Test
1	CVL_ETH	0.295116	**0.264975**
2	**ICT-VIPL**	**0.292297**	0.270685
3	AgeSeer	0.327321	0.287266
3	WVU_CVL	0.316289	0.294835
4	SEU-NJU	0.380615	0.305763
5	UMD	-	0.373352
6	Enjunto	0.370656	0.374390
7	Sungbin Choi	-	0.420554
8	Lab219A	0.477079	0.499181
9	Bogazici	0.483337	0.524055
10	Notts CVLab	-	0.594248

5. Conclusions and Future Works

In this paper, we propose a deep end-to-end learning approach, named as AgeNet, for robust apparent age estimation. We deploy a very large-scale 22-layers deep convolution neural network with regression and label distribution based classification as the objective to encode the apparent age respectively. To reduce the risk of over-fitting, we propose a general-to-specific deep transfer learning scheme, which consists of pre-training deep network with face images labeled with identity, followed by two successive finetuning steps with real age images and apparent age images. The ICCV2015 Look for People 2015 Apparent Age Estimation Challenge demonstrates the effectiveness of our approach.

For future work, we will study how to jointly learn the age regressor and classifier in a unified learning framework and conduct experiments in real age estimation benchmarks such as Morph-II and FG-Net.

Acknowledgements

This work is partially supported by 973 Program under contract No. 2015CB351802, Natural Science Foundation of China under contracts Nos. 61402443, 61390511, 61379083, 61222211. We thank Huajian Ni for fruitful discussions.

References

[1] B. Chen, C. Chen, and W. Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *European Conference on Computer Vision (ECCV)*, pages 768–783, 2014.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision*

GT: 43 Estimation: 43 GT: 19 Estimation: 19 GT: 24 Estimation: 24 GT: 56 Estimation: 56 GT: 41 Estimation: 42

(a) Good Cases of the proposed system

GT: 27 Estimation: 3 GT: 42 Estimation: 25 GT:23 Estimation: 9 GT:68 Estimation: 56 GT:69 Estimation: 59

(b) Bad Cases of the proposed system

Figure 7. Examples of estimation results of the proposed apparent age estimator.

and Pattern Recognition (CVPR), volume 1, pages 886–893, 2005.

[3] S. Escalera, J. Fabian, P. Pardo, X. Baró, G. Jordi, H. J. Escalante, and G. Isabelle. Chalearn 2015 apparent age and cultural event recognition: datasets and results. In *ICCV ChaLearn Looking at People workshop*, 2015.

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

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

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

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

[8] G. Guo and G. Mu. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 657–664, 2011.

[9] G. Guo, G. Mu, Y. Fu, and T. Huang. Human age estimation using bio-inspired features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 112–119, 2009.

[10] H. Han, C. Otto, X. Liu, and A. Jain. Demographic estimation from face images: Human vs. machine performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 37(6):1148–1161, 2015.

[11] http://www.fgnet.rsunit.com/. The fg-net aging database.

[12] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Internatioal Conference on Machine Learning (ICML)*, 2015.

[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 *Proceedings of the ACM International Conference on Multimedia (ACMM)*, pages 675–678, 2014.

[14] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems (NIPS)*, pages 1097–1105, 2012.

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

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

[17] S. Li, S. Shan, and X. Chen. Relative forest for attribute prediction. In *Asian Conference on Computer Vision (ACCV)*, pages 316–327, 2012.

[18] Y. Li, S. Shan, H. Zhang, S. Lao, and X. Chen. Fusing magnitude and phase features for robust face recognition. In *Asian Conference on Computer Vision (ACCV)*, pages 601–612. 2012.

[19] K. Liu, S. Yan, and C. J. Kuo. Age estimation via grouping and decision fusion. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(11):2408–2423, 2015.

[20] B. Ni, Z. Song, and S. Yan. Web image and video mining towards universal and robust age estimator. *IEEE Transactions on Multimedia (TMM)*, 13(6):1217–1229, 2011.

[21] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPMAI)*, 24(7):971–987, 2002.

[22] K. Ricanek Jr and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pages 341–345, 2006.

[23] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[25] X. Wang, R. Guo, and C. Kambhamettu. Deeply-learned feature for age estimation. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 534–541, 2015.

[26] www.vipl.ict.ac.cn. Visual information processing and learning (vipl) group in institute of computing technology (ict) chinese academy of sciences (cas).

[27] S. Yan, S. Shan, X. Chen, and W. Gao. Locally assembled binary (lab) feature with feature-centric cascade for fast and accurate face detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–7, 2008.

[28] P. Yang, L. Zhong, and D. Metaxas. Ranking model for facial age estimation. In *International Conference on Pattern Recognition (ICPR)*, pages 3404–3407, 2010.

[29] D. Yi, Z. Lei, and S. Li. Age estimation by multi-scale convolutional network. In *Asian Conference on Computer Vision (ACCV)*, pages 144–158. 2015.

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

[31] J. Zhang, S. Shan, M. Kan, and X. Chen. Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment. In *European Conference on Computer Vision (ECCV)*, pages 1–16, 2014.

[32] Y. Zhang and D. Yeung. Multi-task warped gaussian process for personalized age estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2622–2629, 2010.