Deep Age Distribution Learning for Apparent Age Estimation

Zengwei Huo, Xu Yang, Chao Xing, Ying Zhou, Peng Hou, Jiaqi Lv and Xin Geng*
Key Lab of Computer Network and Information Integration (Ministry of Education)
School of Computer Science and Engineering, Southeast University, Nanjing 211189, China.
{huozw, x.yang, xingchao, zhouying1, hpeng, lvjiaqi, xgeng}@seu.edu.cn

Abstract

Apparent age estimation has attracted more and more researchers since its potential applications in the real world. Apparent age estimation differs from chronological age estimation that in apparent age estimation each facial image is labelled by multiple individuals, the mean age is the ground truth age and the uncertainty is introduced by the standard deviation. In this paper, we propose a novel method called Deep Age Distribution Learning(DADL) to deal with such situation. According to the given mean age and the standard deviation, we generate a Gaussian age distribution for each facial image as the training target instead of the single age. DADL first detects the facial region in image and aligns the facial image. Then, it uses deep Convolutional Neural Network(CNN) pre-trained based on the VGGFace and fine-tuned on the age dataset to extract the predicted age distribution. Finally it uses ensemble method to get the result. Our DADL method got a good performance in ChaLearn Looking at People 2016-Track 1: Age Estimation and ranked the 2^{nd} place among 105 registered participants.

1. Introduction

Human faces contain a lot of important characteristics, such as identity, gender, age and expression. The information of these characteristics is widely used in the fields such as personal identification [10], human-computer interaction [9] and security control [7]. Among these tasks, apparent age estimation has become a challenging and attractive topic. For example, ChaLearn organized one challenge track on Apparent Age Estimation in ICCV2015 workshop [1]. Then they extended the previous challenge dataset and organized three parallel quantitative challenge tracks on RGB face analysis. Track 1 is the apparent age estimation [2].

Apparent age is different from chronological age, since the age of each image in the dataset is labelled by multiple individuals rather than its real age. Each facial image is assigned with a mean age μ and a standard deviation σ. Thus the uncertainty introduced by standard deviation makes this age estimation task different from the traditionally chronological one. In the past years, various types of algorithms have been proposed to solve this problem. Rasmus *et al*. [13] tackle the estimation of apparent age in still face images with Deep EXpectation (DEX). DEX uses 20 VGG-16 architectures of ImageNet [14] and these architectures are fine-tuned on the IMDB-WIKI dataset [13] by using a softmax expected value refinement. Finally they ensemble the predictions of 20 networks by fine-tuning them on the competition dataset of ICCV2015 workshop. Liu *et al*. [11] propose a very large-scale 22-layers deep convolution neural network named AgeNet for robust apparent age estimation. To reduce the risk of over-fitting, they also propose a general-to-specific deep transfer learning scheme, which consists of the pre-trained deep network. Zhu *et al*. [17] has utilized the deep representations that are trained in a cascaded way on deep neural networks. They can make full use of unlabelled data for pre-training and the data with the age labels are used to fine-tune the network.

The above methods are faced with the difficulties of apparent age estimation which can be concluded as follows. Firstly, it is hard to manually design a suitable feature learning method for apparent age estimation. Fortunately, lots of researches on deep learning have shown its natural advantages in feature learning. Thus, in this work we choose deep learning as our basic model and train this model on suitable datasets.

If deep architecture is chosen as the basic model, we should collect facial images as many as possible. This is the second problem since even the lots of facial images can be collected from the Internet, it is almost impossible to label these huge number of images with suitable apparent ages manually. Thus, we should exploit existent labelled dataset as sufficient as possible.

Finally, the traditional age estimation tasks usually give facial images with integral ages (such as 25), while in the

*X. Geng is the corresponding author.

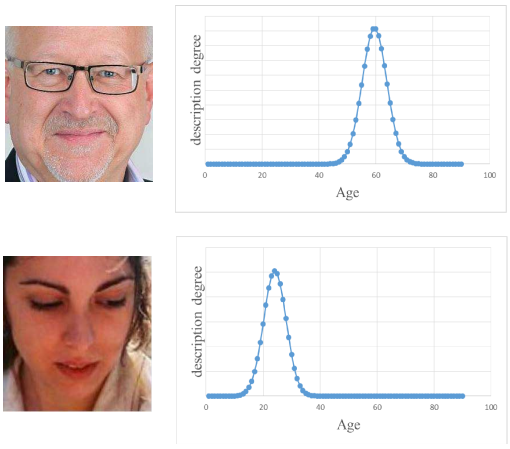

Figure 1: The Examples of the Age Distribution

ChaLearn Looking at People 2016 - Track 1: Age Estimation competition [2], real ages (such as 25.2) are assigned to each facial image. Another difference between apparent age estimation and chronological age estimation is the existent of uncertainty which is introduced by the standard deviation. Thus, apparent age estimation should not be simply treated as traditional classification problem.

In order to achieve better results, we must carefully deal with the above three issues. We can find that the growth of the age is a slow process, and the age which is closer to the mean age has a lager description degree. Therefore we can generate the Gaussian distribution by using the mean age μ and the standard deviation σ. Through this way, we can exploit the uncertainty information and solve the problem that the provided ages are real. As can be seen from Fig.1, the horizontal axis shows the apparent ages and the vertical axis represents the description degree d_x^y of each age y assigned with the facial image x. The description degree d_x^y can be called Age Distribution with the constraints $d_x^y \in [0, 1]$ and $\sum_y d_x^y = 1$. We can learn a mapping from the input facial image to its related age distribution in this way. On the one hand, we can exploit the prior knowledge more sufficient. For example, the description degree of the age near the real mean age is larger than the age far from the real mean age and the difference is controlled by the standard deviation. On the other hand, we can use the rank information of the age distribution to generate a back-propagation updating value for each age associated with the input image. This method can efficiently relax the requirement for large amount of training images. Although Geng *et al*. [5] proposed an adaptive distribution learning for a more complex distribution form, the Gaussian distribution used in this paper is simple and efficient for learning.

Then, we propose a Deep Age Distribution Learning (DADL) method which utilizes the idea of traditional age distribution learning [6] for apparent age estimation and enhances it by exploiting CNNs. This method can effectively use the correlation among neighboring labels based on the label distribution. In this method we develop different deep CNN models based on VGGFace model [12] by fine-tuning VGGFace model on different facial image datasets. Next, we extract the age distributions from the obtained CNN models which have been trained on the competition dataset. Finally, we design an ensemble method to generate the final predicted ages by using these extracted age distributions based on similarity with the training samples. These learned ages are our final predicted results.

This paper is organized as follows. In Section 2, we will introduce the proposed Deep Age Distribution Learning method. Implementation details and experimental results will be presented in Section 3. Finally, we conclude our method and discuss the future issues in Section 4.

2. Approach

In this section, we will first introduce the DADL method. Next, we show the whole process of our DADL method used for apparent age estimation.

2.1. Deep Age Distribution Learning method

Human facial images look quite similar if the ages of these images are close. For example, one's face looks same when he's 25 or 26. This prompts us when learning a particular age, we can use the facial images at neighboring ages. As Geng *et al*. described in [6], this idea can be implemented by assigning a label distribution rather than a single label of the age to each facial image, so it's called the Age Distribution. The age distribution contains a number of real values which represent the description degree of each age to the facial image. It also reflects the correlation information among neghboring ages, that is the close ages have the high correlation. A suitable age distribution will make a facial image contribute to not only learning the real age, but also learning the neighboring ages, and it can provide more flexibility in representing ambiguity [6].

We propose the DADL method. The goal is to use the deep CNN model to predict the age distribution. By define the loss function according to Kullback-Leibler (KL) divergence instead of considering each facial image as an example with one age, each facial image is treated as an example associated with an age distribution. Thus we combine the Deep Learning and Label Distribution Learning [4, 3] as whole and this is the Deep Age Distribution Learning. We want to minimize the the following KL divergence

$$\sum_j y_j \ln \frac{y_j}{p_j} \tag{1}$$

The y_j is the j^{th} value in the ground truth distribution and p_j is the j^{th} value in the predicted distribution. The loss

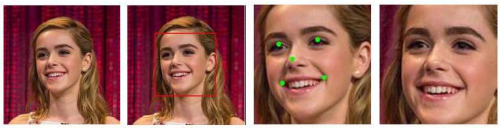

Figure 2: Three steps of the images pre-processing in our method.

function is defined as follows:

$$L = -\sum_j y_j \ln p_j \tag{2}$$

We use softmax function to calculate the predicted distribution according to the last fully connected layer, which is:

$$p_i = \frac{\exp(o_i)}{\sum_t \exp(o_t)} \tag{3}$$

where o_i is the i^{th} output in the last fully connected layer of one sample. Stochastic gradient descent is used to minimize the loss function Eq 2, according to the chain rule, for any fixed i, we have

$$\frac{\partial L}{\partial o_i} = \sum_j \frac{\partial T}{\partial p_j}\frac{\partial p_j}{\partial o_i} = p_i - y_i \tag{4}$$

Then we can use the Backpropagation Algorithm to update the parameters in the model.

For our proposed loss function, $y_j \in [0, 1]$, the age that close to the ground truth age has a higher value in the distribution and the ground truth age has the maximum value in the distribution, while for softmax loss function, y_j is 0 or 1. 1 means it's the ground truth age and 0 means other ages. So compared with the softmax loss function, our Deep Age Distribution Learning can effectively utilize the correlation among neighboring labels.

2.2. Deep Age Distribution Learning method for Apparent Age Estimation

In this section, we detail the proposed DADL method used in the apparent age estimation.

2.2.1 Facial images pre-procession

We first preprocess the facial images. Figure 2 shows three steps of the facial images pre-procession. The original images are fed into a public available face detector and facial point detector software [15] to detect the facial region in the image and five facial key points, including left/right eye centers, nose tip and left/right mouth corners. After this step, we align all the faces based on the detected five points and resize all the images into 256×256 pixels and then use the data augmentation method as Krizhevsky *et al*. described in [8]. After preprocessing, the final images which fed into the CNN model are 224×224 pixels.

2.2.2 Training single model using Deep Age Label Distribution Learning

In our method, we use the pre-trained VGGFace model [12] as our basic model for age estimation.The VGGFace model is used to solve the face recognition problem, although face recognition is different from the age estimation, they are correlated to each other. Figure 3 shows the architecture of this deep CNN model. So in this step we fine-tune the VGGFace model on different age datasets and get the new deep models.

2.2.3 Ensemble different results of different deep CNN models

After training on different datasets, we can get different deep models. Since different datasets are used to train these deep models, these deep CNN models will capture different kinds of interest from one facial image. And the distributions got by deep CNNs can provide different useful information for predicted ages. Besides this, the learned distribution can also be treated as one special type of representations for one facial image. Thus, we design a distance-based voting ensemble method to predict ages from these learned distributions. For each facial image in the test dataset, we can learn different 90 dimensional distributions (We assume the age range is 1 to 90). We sum these distributions to get one 90 dimensional distribution and also normalize this distribution vector to make sure the sum of 90 elements to be one. We denote this normalized distribution vector as $\boldsymbol{x}$. For the n^{th} facial image with its age which is denoted as t_n in the training set, we can get a corresponding distribution vector $\boldsymbol{x}_n$. When a new facial image comes, we first compute its distribution vector $\boldsymbol{x}^*$ from deep CNNs and then use the following approach to predict its age t^*:

$$t^* = \sum_{n=1}^{N} t_n K(\boldsymbol{x}^*, \boldsymbol{x}_n)\,. \tag{5}$$

The distance measure $K(\cdot,\cdot)$ is defined as follows:

$$K(\boldsymbol{x}^*, \boldsymbol{x}_n) = \exp(-\alpha \times ||\boldsymbol{x}^* - \boldsymbol{x}_n||_2^2)\,, \tag{6}$$

where $\boldsymbol{x}_n^{th}$ is the concatenated feature of the n^{th} image in the training dataset, t_n is the age of the n^{th} image and N is the number of images in the training dataset. α can be selected by using ten-fold cross validation.

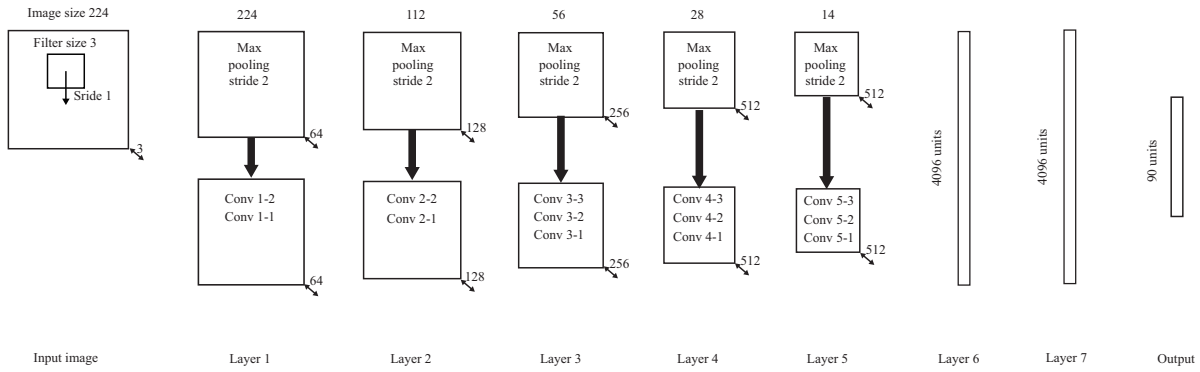

Figure 3: Architecture of the deep CNN model. In this deep CNN, a 224×224 pixels aligned facial image is presented as input. The convolutional filter is 3×3 with stride and pad are both 1, followed by a ReLU (not shown) layer and a 2×2 max-pooling layers with stride is 2. This deep CNN has 5 convolutional layers with the same parameters. The last 2 layers are fully connected layers, and a ReLU (not shown) layer follows each fully connected layers. The final layer is an 90-dimensional output layer with defined loss function.

3. Experiments

In this section we will show the DADL method used in the ChaLearn Looking at People 2016-Track 1:Age Estimation.

3.1. Datasets

We will first describe the details about the datasets used in our method. The number of each age in five datasets is shown in Figure 4. The ages in the CVPR 2016 ChaLearn Looking at People workshop dataset are real numbers, we round ages to the integers for the sake of calculating.

3.1.1 IMDB-WIKI dataset

The IMDB-WIKI dataset [13] is the largest publicly available dataset of facial images with gender and age labels for training. The facial images in this dataset are crawled from the IMDB and Wikipedia website. There are 461,871 facial images from the IMDB website and 62,359 from the Wikipedia website. Considering our limited computation ability, we randomly choose 240,000 images and generate three small datasets, each small dataset has 80,000 facial images. In order to use the Deep Age Distribution Learning, we assume that age distribution follows a discrete Gaussian distribution, the assigned age is the mean value μ and standard deviation σ is assumed to be 3.

3.1.2 ICCV 2015 ChaLearn Looking at People workshop dataset

The ICCV 2015 ChaLearn Looking at People workshop dataset [1] has 3,651 face images with their apparent ages and standard deviations in total. This is a small dataset which contains the apparent ages and standard deviations and this dataset is used to fine-tune the deep model. We turn over the images as the image augmentation and use all 7,302 images. The ages in this dataset are integers. In the training dataset, the minimum and maximum standard deviations are 1.0326 and 6.7192, while in the validation dataset, the minimum and maximum standard deviations are 0.6999 and 6.8980.

3.1.3 CVPR 2016 ChaLearn Looking at People workshop dataset

In the CVPR 2016 workshop competition, totally 5,613 facial images (4,113 in the training dataset and 1,500 in the validation dataset) with their apparent ages and standard deviations are given. Most ages in this dataset are not integers. And in the training dataset, the minimum and maximum standard deviations are 0 and 12.3769, while in the validation dataset, the minimum and maximum standard deviations are 0 and 14.1198. The standard deviation covers a larger range and we must predict the image with small standard deviation as accurate as possible. Thus this competition is more difficult than the ICCV 2015 ChaLearn Looking at People workshop. In our Deep Age Distribution Learning method, according to the scripts provided by the competition organizers, if the standard deviation is less than 0.162, we set it to 0.162.

3.2. Implementation details

The whole process of our DADL method is shown in Figures 5. On the one hand, we fine-tune three deep models on the different IMDB-WIKI datasets and then we use the obtained models to fine-tune on the final dataset which consists of ICCV 2015 ChaLearn Looking at People workshop

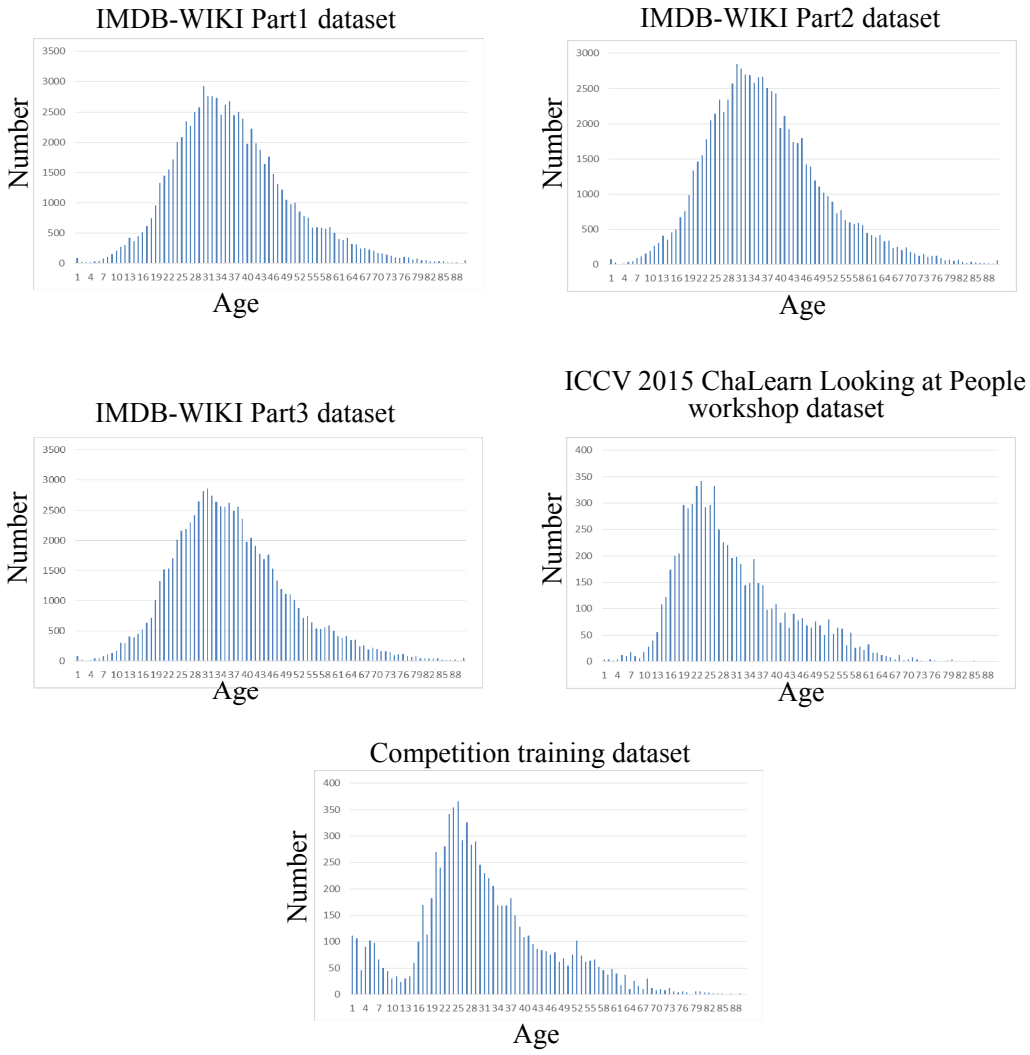

Figure 4: The number of each age in five datasets. The horizontal axis of each histogram represents age and the vertical axis of each histogram means the number of age in the corresponding dataset. The name of dataset is shown at the top of each figure.

dataset [1] and CVPR 2016 ChaLearn Looking at People workshop dataset. We mix two datasets as the final dataset because the facial images are obtained in the same method in these datasets. On the other hand, we also fine-tune a pre-trained VGGFace model directly on final dataset without fine-tuning on IMDB-WIKI dataset. Therefore we get four deep models in total at last. Then we use the ensemble method to get the final results.

The deep CNNs are trained on Nvidia Tesla K80 GPUs by using the MatConvNet framework [16]. Fine-tuning a single model on the small dataset of IMDB-WIKI dataset takes about 12h and fine-tuning a single model on the final dataset takes about 3h. The distribution extraction of each image takes 0.01s. The preprocessing of each facial image takes about 10s per image.

3.3. Evaluation Criteria

The performance of age estimation is evaluated by the ϵ-error provided by the competition, which is:

$$\epsilon = 1 - \exp(-\frac{(x-\mu)^2}{2\sigma^2}), \tag{7}$$

where x is the predicted age, μ is the mean apparent age and σ is the standard deviation. This formula gets an error value between 0 (correct) and 1 (far from age). Not predicted images are evaluated with 1.

We also use the standard mean absolute error (MAE) to evaluate the result. The MAE is the average of absolute

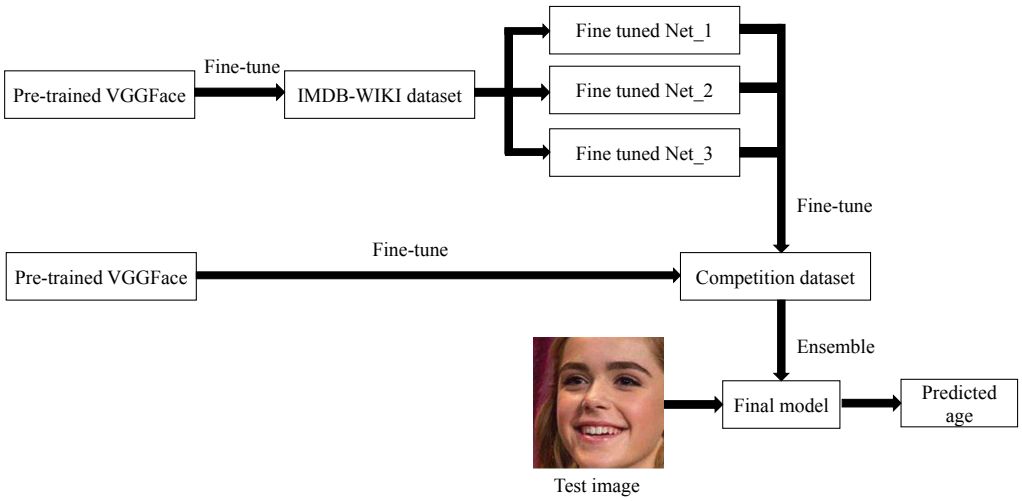

Figure 5: The framework of proposed Deep Age Label Distribution Learning method. In our method, the pre-trained VGGFace model are used to fine-tune on the IMDB-WIKI dataset, then the obtained deep models and original VGGFace model are fine-tuned on the competition dataset respectively. For test images, we use ensemble method to get the final results.

errors between the predicted age and the ground truth age.

3.4. Validation Result

On the validation set, the final ϵ-error of our model is 0.133. The results of four different deep CNNs on the validation set are shown in Table 1. In Figure 6, there are nine examples of facial images from the validation dataset with their generated distribution using provided mean value and standard deviation and predicted distribution. It can be seen that our method is not sensitive to the variations created by pose, lighting and image color mode. However, our method show poor performance in facial images which have occlusions, wrong face alignment or low-resolution. We show these bad examples in Figure 7.

From the Table 1, we find that the model trained on the competition dataset directly is better than fine-tuned on the IMDB-WIKI dataset. But the model fine-tuned on the IMDB-WIKI dataset and then fine-tuned on the competition dataset needs only 20 epoch, while fine-tuned on the competition dataset directly needs 50 epoch or more. So it accelerates the training process and makes the model find the optimal solution quickly.

To reflect the advantage of our DADL method, we use the softmax loss as loss function and train deep models on the competition dataset. The results of four deep models on the validation set are shown in Table 2. Compare Table 1 and Table 2 we can find that our method is better than the softmax loss method.

Model	Net_1	Net_2	Net_3	VGGFace
ϵ-error	0.1519	0.1413	0.1493	0.1341
MAE	1.8471	1.9223	1.8530	1.7569

Table 1: Performance of our method on validation dataset.

Model	Net_1	Net_2	Net_3	VGGFace
ϵ-error	0.2887	0.2514	0.2631	0.2594
MAE	3.5366	3.0112	3.1682	3.1182

Table 2: Performance of using softmax loss on validation dataset.

We also provide the results of the deep CNN which only trains on the small IMDB-WIKI datasets and test on the validation dataset directly. These results are shown in the Table 3. The results of the IMDB-WIKI datasets are not very good, this maybe caused by the unsuitable setting of the σ, which is 3, in IMDB-WIKI dataset.

3.5. Final Results

Table 4 shows the performance of our method on the test set. The compared results of the top 8 teams are also shown in Table 4.

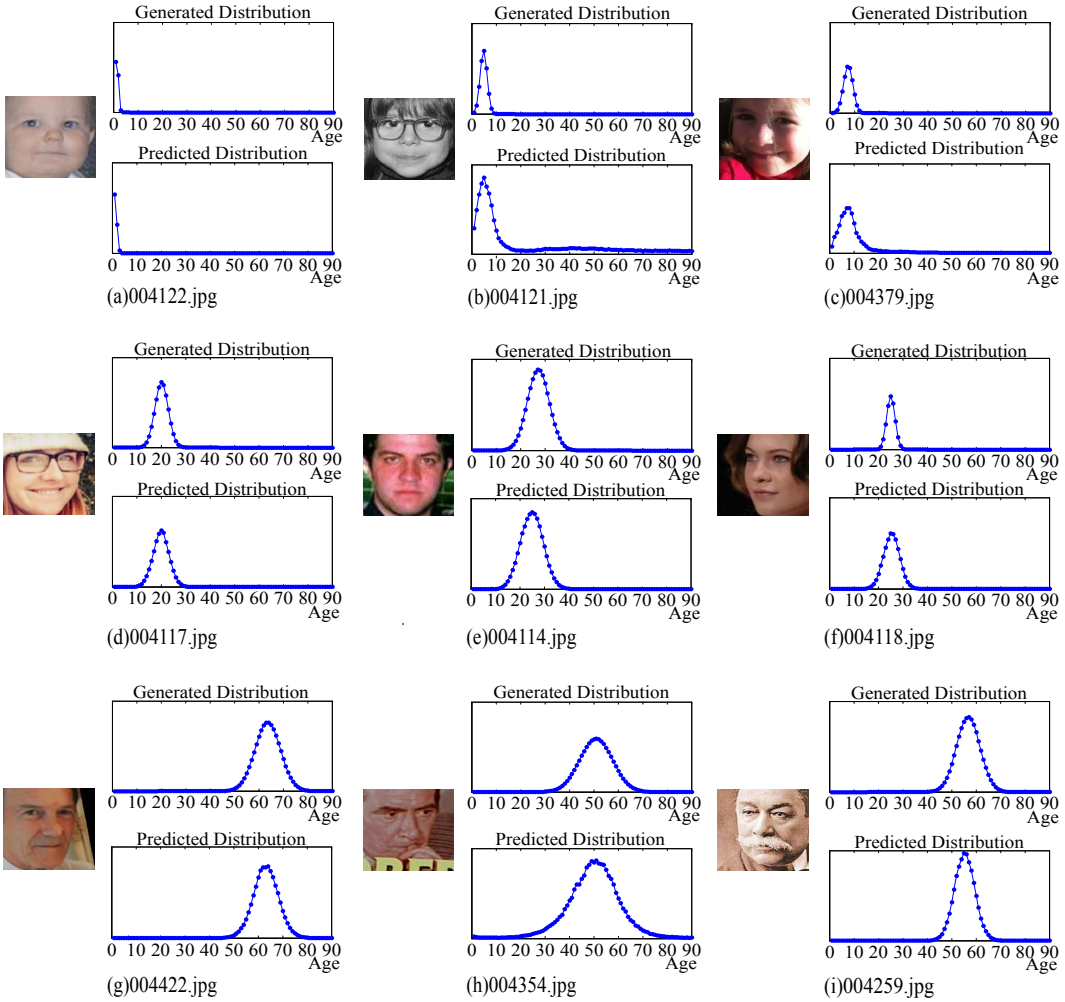

Figure 6: Facial images examples which predicted well by our method from the validation dataset . In each subfigure, the left is the facial image, the top right is the generated distribution using provided mean value and standard deviation, and the bottom right is the predicted distribution provided by our method. These three rows present the distribution of child, young people and old people respectively.

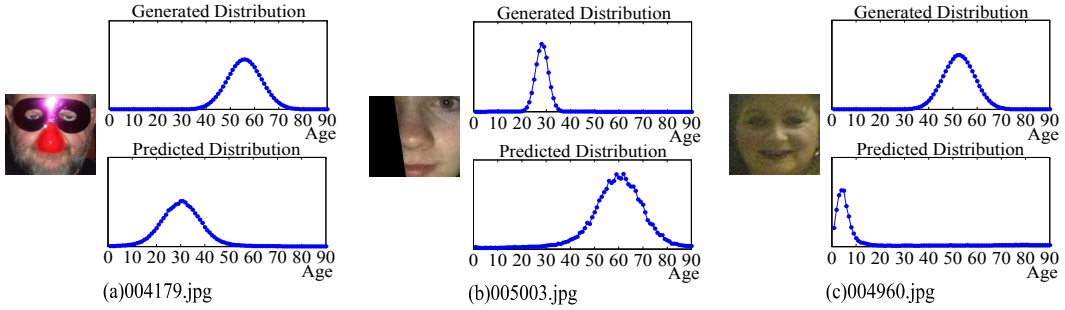

Figure 7: Bad examples of facial images from the validation dataset. The reason of poor performance in these images are occlusions, wrong face alignment and low-resolution.

Model	Net_1	Net_2	Net_3
ϵ-error	0.4815	0.5176	0.5240
MAE	5.6651	6.5221	6.9214

Table 3: Performance of our method without training on the competition train dataset and testing on the validation dataset directly.

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

Table 4: ChaLearn Looking at People 2016-Track 1: Age Estimation final ranking in the test set. There are 105 registered participants in total.

4. Conclusion

In this paper, we propose the Deep Age Distribution Learning method to solve the apparent age estimation problem. Our DADL method uses VGGFace model as the basic model and defines the loss function to take full advantage of the standard deviations. This method extracts the predicted age distribution from the fine-tuned models and uses ensemble method to get the result. Promising results are reported and we get a good performance in ChaLearn Looking at People 2016 - Track 1: Age Estimation and ranked the 2^{nd} place.

In the future, we will explore more advantages of our DADL method and combine the adaptive distribution learning with deep learning in the further work.

References

[1] S. Escalera, J. Fabian, P. Pardo, X. Baró, J. Gonzàlez, H. J. Escalante, and I. Guyon. ChaLearn 2015 apparent age and cultural event recognition: Datasets and results. *IEEE International Conference on Computer Vision, ChaLearn Looking at People workshop*, 2015.

[2] S. Escalera, M. Torres, P. Pardo, B. Martnez, X. Baró, H. J. Escalante, I. Guyon, G. Tzimiropoulos, C. Corneanu, M. Oliu, M. A. Bagheri, and M. Valstar. ChaLearn looking at people and faces of theworld: Face analysis workshop and challenge 2016. *IEEE Conference on Computer Vision and Pattern Recognition, ChaLearn Looking at People and Faces of the World workshop*, 2016.

[3] X. Geng. Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 2016, in press.

[4] X. Geng and R. Ji. Label distribution learning. In *IEEE Conference on Data Mining Workshops*, pages 377–383, 2013.

[5] X. Geng, Q. Wang, and Y. Xia. Facial age estimation by adaptive label distribution learning. In *Proceedings of the 22nd International Conference on Pattern Recognition Stockholm*, 2014.

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

[7] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai. Learning from facial aging patterns for automatic age estimation. In *Proceedings of the 14th ACM International Conference on Multimedia*, pages 307–316, 2006.

[8] 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.

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

[10] A. Lanitis, C. J. Taylor, 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.

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

[12] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *Proceedings of the British Machine Vision*, volume 1, page 6, 2015.

[13] R. Rothe, R. Timofte, and L. Gool. Dex: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 10–15, 2015.

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[15] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3476–3483, 2013.

[16] A. Vedaldi and K. Lenc. Matconvnet – convolutional neural networks for matlab. 2015.

[17] Y. Zhu, Y. Li, G. Mu, and G. Guo. A study on apparent age estimation. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, December 2015.