Multi-Region Ensemble Convolutional Neural Networks for High-Accuracy Age Estimation

Yiliang Chen[1]
elichan5168@gmail.com
Zichang Tan[23]
tanzichang2016@ia.ac.cn
Alex Po Leung[1]
pleung@must.edu.mo
Jun Wan[23]
jun.wan@ia.ac.cn
Jianguo Zhang[4]
jnzhang@dundee.ac.uk

[1] Faculty of Information Technology Macau University of Science and Technology, Macau SAR
[2] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[3] University of Chinese Academy of Sciences
[4] Computing, School of Science and Engineering, University of Dundee

Abstract

In real life, when telling a person's age from his/her face, we tend to look at his/her whole face first and then focus on certain important regions like eyes. After that we will focus on each particular facial feature individually like the nose or the mouth so that we can decide the age of the person. Similarly, in this paper, we propose a new framework for age estimation, which is based on human face sub-regions. Each sub-network in our framework takes the input of two images each from human facial region. One of them is the global face, and the other is a vital sub-region. Then, we combine the predictions from different sub-regions based on a majority voting method. We call our framework Multi-Region Network Prediction Ensemble (MRNPE) and evaluate our approach using two popular public datasets: MORPH Album II and Cross Age Celebrity Dataset (CACD). Experiments show that our method outperforms the existing state-of-the-art age estimation methods by a significant margin. The Mean Absolute Errors (MAE) of age estimation are dropped from 3.03 to 2.73 years on the MORPH Album II and 4.79 to 4.40 years on the CACD.

1 Introduction

Age estimation has been a very challenging problem in computer vision areas with its potential applications on access control, precision advertising and video surveillance. However, age estimation is still a very complicated problem for us, because there are so many elements affecting our judgment such as wrinkles, smooth degree and even various genes, thus creating great uncertainties in age estimation. The earliest age recognition work originated in 1994 by Kwon and Lobo [14], which simply classified ages into ages ranges instead of a single chronological year. After that, regression and classification methods played most significant role in predicting the age from the human face images, such as Support Vector Machines

(SVM), Support Vector Regression (SVR), Partial Least Squares (PLS), and Canonical Correlation Analysis (CCA). Among these traditional methods, the most representative work is BIF+CCA (KCCA) [10]. In recent years, deep learning methods have drawn increasingly attention in facial age estimation. For instance, in 2011, Yang *et al.* [24] proposed a convolutional neural network (CNN) for age estimation. However, they mainly paid attention to face tracking, and just adopted the original CNN without any modification for age estimation. There are three prevalent benchmarks for age estimation, which are FG-NET [15], MORPH Album II [19] and CACD [2]. Owing to these aging datasets, age estimation techniques are developing faster.

Despite so much effort, telling a person's age from a single image is still a very challenging task, because the facial features reflecting aging are of different types at different age ranges. For instance, facial aging process is conspicuous in the shape of face during childhood, while the feature associated with aging is distinctive in the skin of texture during adulthood. Hence, our method not only focuses on the global region, but also pays attention to the local regions, because local characteristics are also very important for age estimation, such as the corner of eyes, mouths, noses, etc. For example, Yi *et al.* [25] and Ting *et al.* [17] used multiple local regions and the whole face as input images, and every region was trained by its own subnet and all subnets were concatenated on the same fully connected layer. However, in fact the effect of aging of every single facial sub-region is different. For example, with aging one obvious observable feature can be wrinkles, other features may not be so distinctive like the nose area which ages fairly slowly. To this end, it might be less effective to directly train all sub-regions together in the same network, which makes it difficult for the neural network to learn the differences among the sub-regions.

Therefore, we propose an age estimation framework to consider the effect of aging on individual local region by constructing different sub-networks, each taking the input of a global region and one sub-region. The results of all the subnets are used together for prediction. In this way, the generalization of learned model can be enhanced. Moreover, each module of the corresponding sub-region can sufficiently learn together with the global information of the whole face so that we can take full advantage of all information from different components. The main contributions of our work are summarized as below:

- A novel MRNPE method is proposed for age estimation. It includes some sub-networks in a unified framework, and each sub-network takes full advantage of multiple sub-regions to capture the local and global features from face images.

- The experimental results show that the proposed Multi-Region Network Ensemble Prediction (MRNPE) framework is significantly better than a direct multi-region ensemble prediction method [17, 25].

- Our MRNPE framework outperforms state-of-the-art age estimation methods by a significant margin on both the MORPH Album II and the CACD datasets.

2 Related Work

Geometry features and the features of skin wrinkle are often adopted in the methods of age estimation in the early year of age prediction [14]. However, these methods can only tell the range of a person's age from a single facial image. Later on, Horng *et al.* [12] adopt the sobel edge operator and region labeling to locate eyes, mouths and noses in human facial images so

as to improve the accuracy of age estimation. For the single-year age estimation, many novel methods were proposed based on the AAM [4] tool, especially the AGing PattErn Subspace (AGES) [7] method which achieve the mean absolute error (MAE) on the FG-NET dataset to 6.22 years. However, the effect of AAM method is prone to various factors such as illumination and poses. Therefore, more recently, the methods that using local features have been increasingly drawing our attention and becoming the mainstream on age estimation such as Gabor [6], Local Binary Patterns [8], and Biologically Inspired Features (BIF) [11]. After extracting the features from the images, regression or classification methods are often adopted for age estimation, such as BIF+SVM [11], BIF+SVR [11], and BIF+CCA [10]. Therefore, the majority of traditional methods estimate the age from a face image by two steps: 1) local feature extraction 2) regression or classification.

In the last few years, the CNNs have made huge success on age estimation [18, 20, 21, 25]. A novel approach named Deep Expectation (DEX) model was proposed by Rothe *et al.* [20] based on the VGG-16 architecture network which is pretrained by an aging dataset named IMDB-WIKI [20] with 500K images, and such deeper CNN model won the 1st place at the ChaLearn LAP challenge 2015. After that, Rothe *et al.* [21] improved their method, which didn't involve facial landmarks techniques and reduced the MAE to 4.785 years on the CACD dataset in 2016. DEX model was further improved by the champion [1] at the ChaLearn Lap challenge 2016, using a separate model for the images of children. In the same year, Niu *et al.* [18] proposed an end-to-end deep learning method to solve ordinary regression problems, which achieves the result of 3.27 MAE on the MORPH Album II dataset.

There also exists some research adopting pre-partitioned facial regions to predict age[17, 25]. Yi *et al.* [25] proposed 46 parallel CNNs with different regions of the face images as inputs, which decreased MAE to 3.63 years on the MORPH Album II dataset. However, it does not consider the relationship between local regions and the global region, and its network structure is pretty complicated. Ting *et al.* [17] adopted a method similar to that of Yi *et al.* [25], but didn't get too much progress because it ignored the differences among sub-regions resulting in increasing the difficulty of training.

In this paper, our work is also inspired by using the pre-partitioned sub-regions as inputs. We consider the overall combination of the global region and local regions, and therefore take full advantage of the sub-regions by separating them and utilize an appropriate ensemble method to predict the age.

3 The Proposed Method

In this section, we introduce the pipeline of our proposed framework (MRNPE) shown in Fig. 1, and then describe in detail each stage of MRNPE.

3.1 Local Region Aligned and Cropped

Preprocessing to align and crop faces is needed because faces on images are mostly irregularly placed on the images. Through alignment, we can eliminate some irrelevant factors, such as the posture of the faces. As shown in Fig. 1, in our experiments, we use facial landmarks by ASM [3] to locate the 21 facial points, and then we align face images based on the midpoint between two pupils and the middle point of the upper lip. In addition, rich knowledge of ages can be found in different facial areas such as eyes, noses, mouths and eyebrows as well as from the whole faces. Therefore, in our experiments, we also use facial

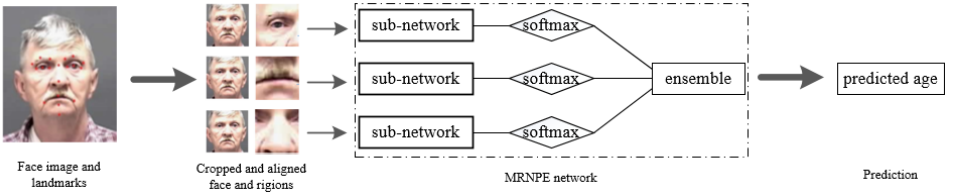

Figure 1: MRNPE framework structure

landmarks technique to crop the areas of the three most important local components which are the regions of the left eye, the regions of the nose and the regions of the mouth. All the images are resized and cropped into 224 x 224 pixels. Besides, according to Niu *et al.* [18], the facial color attribute is helpful for age estimation, so we also keep the color element in our formulations. We discard facial images which cannot be detected by the face detector. Note that among detected faces, there still exists occlusion (e.g, wearing sun glasses) or the viewing angle is not ideal. Although those images present challenges for age estimation, we keep them in our training and testing dataset [17, 21, 23, 25].

3.2 MRNPE Framework Structure

Our framework is illustrated in Fig. 1, which is comprised of three subnets because three important facial sub-regions are used in our experiments. In other words, the left eye, the nose and the mouth are utilized in our experiments. Each subnet takes a pair of facial region inputs which are a global face image and a sub-region image. All the images are in RGB with the size 224x224 pixels. Every input of the sub-region is a unique region and there are three age predictions from three subnets respectively, and subsequently each subnet adopts the softmax loss function which is defined as:

$$l(\theta) = -\frac{1}{n}\left[\sum_{i=1}^{n}\sum_{j=1}^{u} 1\left\{y^{(i)} = j\right\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{u} e^{\theta_l^T x^{(i)}}}\right] \tag{1}$$

where θ is the parameter matrix of the softmax function; $x^{(i)}$ indicates the features of the i-th sample, and $x^{(i)} \in R^d$ and $y^{(i)}$ is the age label of the i-th sample. n is the number of samples and u is the maximum label of age. Besides, $1\{\bullet\}$ is an indicator function, which means 1{a true statement} = 1.

For Mean Absolute Error (MAE), the widely adopted approach to evaluate MAE is the function of the metric Expected Value (EV) [20]. The predicted age P_k in Eq. 2 from each subnet's softmax is equal to $\sum_{i=0}^{100} p_i c_i$, where p_i is the predicting probability of the corresponding age c_i, and the subindex i ranges from 0 to 100 because our softmax is a hundred-and-one-dimensional vector [17, 21, 23, 25]. Finally, we combine the three predictions from three subnets using Eq. 2

$$P = \sum_{k=1}^{m} \alpha_k P_k \tag{2}$$

where α_k is the weight for each subnet k and m is defined as the number of the sub-regions and we just use three important regions in our following experiments, so the m is equal to 3.

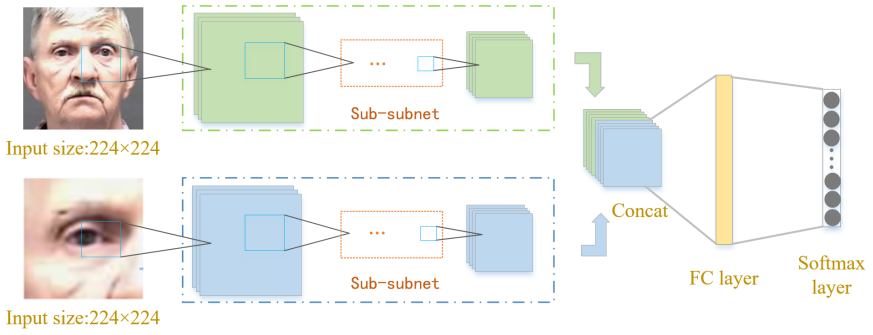

Figure 2: The detailed architecture of the sub-network in MRNPE.

Afterward we combine three subnets' predictions and get a final prediction P. In this paper, we do not comprehensively investigate the different ensemble methods in our framework like Adaboost, because we pay more attention to the framework itself and we briefly discuss different ensemble methods in Section 5.1. Therefore, in our following experiments, we adopt three weak learners and combine them through averaging to create a strong learner that can make accurate predictions. We only utilize three regions in our experiments, so all the weights are equal to 1/3.

3.3 The Detailed Architecture of the Sub-network in MRNPE

Our framework adopts two types of network structures and the detailed architecture is shown in Fig. 2. One of them is based on AlexNet [13] network architecture and the other is a deeper network based on VGG-16 [22] network structure. Actually, from the results in Section 4.3, both network structures work well on age estimation.

The input of each subnet is a pair of images including a global face and a sub-region, and two inputs go through their own sub-subnet. Each sub-subnet is a variant of either Alexnet or VGG net.

For AlexNet architecture, we used 5 convolutional layers and 3 max pooling layers with the pipeline structure "Conv+Pooling+Conv+Pooling+Conv+Conv+Conv+Pooling", and the sub-subnets are concatenated in a fully connected layer. Then, the output from the FC layer will go through a softmax layer and get a prediction of the subnet. Our MRNPE framework with AlexNet architecture do not use any pre-trained models.

For VGG-16 architecture, it is a very deep structure with smaller kernel size (3x3). Similarly, we adopted 13 convolutional layers and 5 max pooling layers with the structure "Conv+Conv+Pooling+Conv+Conv+Pooling+Conv+Conv+Conv+Pooling+Conv+Conv+Conv+Pooling+Conv+Conv+Conv+Pooling", and the outputs from the two sub-subnets are combined together and go through the fully connected layers before reaching the softmax loss layer. Besides, we utilized ImageNet [5] dataset to pretrain our MRNPE(VGG-16) network structure, and finetune our pre-trained network with MORPH Album II and CACD respectively in the following experiments.

3.4 Prediction Strategy

In our experiments, during the prediction, every global face image and every sub-regions image are tested with their mirror patches following the previous work [17, 25]. In other

words, we double the testing images by creating their mirrors. About mirroring the images, our method is slightly different from the method used by Yi *et al.* [25] and Ting *et al.* [17]. They mirrored all images except the left eyes, because they used the region of right eyes to replace the mirrored images of the left eyes. However, we find that the operation is not useful, so in our experiments, we construct mirrored images for the left eyes. We combine two results from mirrored and original images to arrive at the final decision by using ensemble.

4 Experiments

4.1 Datasets and Setup

In our experiments, our framework is based on the MORPH Album II [19] and the CACD [2] datasets which are two popular datasets for human facial age estimation.

MORPH Album II contains approximately 55,000 facial images and their ranges of the ages are from 16 to 77 years, and however some pictures have negative influence on predictions because of the uneven illumination. CACD is the biggest public cross-age dataset, and it is collected from the famous Internet Movie DataBase (IMDB). Besides, CACD includes more than 160K images of 2000 celebrities. However, compared with MORPH Album II, in some pictures, the face is partially occluded or the viewing angle, which is challenging for age estimation. .

For the MORPH Album II, to follow the previous approach [9, 17, 23, 25] with the same test protocols[1] provided by Yi *et al.* [25], the dataset is randomly partitioned into three non-overlapping subsets S1, S2 and S3. Therefore, there are two different combinations of training set and testing set: 1) Training set is S1, and testing sets are S2+S3; 2) Training set is S2, and testing sets are S1+ S3. For CACD dataset, the images of 200 celebrities were filtered to remove noises in the work of [16, 23]. We follow exactly the same protocol, and include these 200 celebrities for testing and others for training in our experiments.

Following the section 3.1, we crop the regions from the human faces detected, and images with no face detected are removed from the dataset. After such processing, MORPH Album II includes 55244 images, while CACD contains 162941 images [23]. The size of MORPH Album II dataset may not be sufficient for training a deep net, and therefore we augment the set of training images by flipping, rotating each with $\pm 8^o$ and $\pm 4^o$, and adding Gaussian white noises with variances of 0.001, 0.005, 0.01, 0.015 and 0.02.

4.2 Sub-Region Comparison

To illustrate the advantage of our method, we compare ours to the network structure of the method [17] named Multi-Region Convolution Neural Network (MRCNN). Each region in MRCNN is processed by a separated single CNN and the outputs from these subnets are concatenated on the same fully connected layer. And the Euclidean loss layer in MRCNN is replaced with Softmax loss layer in our following experiments. Therefore, in our experiments, the subnet structure of MRCNN is as same as our sub-subnet in Fig. 2 of our MRNPE framework. We can see the results from Table 1. The result of first row is based on the M-RCNN, pretty close to the Face+Mouth result which is one of the subnet in our MRNPE framework. However, the result is worse than that of Face+Nose. Furthermore, our MRNPE

[1] http://www.cbsr.ia.ac.cn/users/dyi/agr.html

Table 1: First row experiment is based on the Multi-Region Convolution Neural Network (MRCNN) method, and 2,3,4 rows are the results of the subnets in the MRNPE and the final row is our final result of MRNPE (AlexNet).

Architecture	Train Set	Test Set	MAE ↓	Avg. MAE with EV↓
Face+LeftEye+Nose+Mouth (MRCNN)	S1	S2 + S3	3.43	3.28
	S2	S1 + S3	3.13	
Face+LeftEye (Subnet in MRNPE)	S1	S2 + S3	3.25	3.12
	S2	S1 + S3	2.98	
Face+Nose (Subnet in MRNPE)	S1	S2 + S3	3.29	3.16
	S2	S1 + S3	3.03	
Face+Mouth (Subnet in MRNPE)	S1	S2 + S3	3.43	3.30
	S2	S1 + S3	3.16	
MRNPE (AlexNet)	S1	S2 + S3	2.98	2.86
	S2	S1 + S3	2.73	

Table 2: Comparisons with the state-of-the-art methods on MORPH Album II under the same testing protocol.

Architecture	Train Set	Test Set	MAE ↓	Avg. MAE with EV↓
BIF + KCCA [10]	S1	S2 + S3	4.00	3.98
	S2	S1 + S3	3.95	
Multi-scale CNN [25]	S1	S2 + S3	3.72	3.63
	S2	S1 + S3	3.54	
Single CNN Softmax	S1	S2 + S3	3.28	3.16
	S2	S1 + S3	3.03	
Soft softmax [23]	S1	S2 + S3	3.24	3.14
	S2	S1 + S3	3.03	
Pretrained Model Soft softmax [23]	S1	S2 + S3	3.14	3.03
	S2	S1 + S3	2.92	
MRNPE (AlexNet)	S1	S2 + S3	2.98	2.86
	S2	S1 + S3	2.73	
MRNPE (VGG16)	S1	S2 + S3	2.85	2.73
	S2	S1 + S3	2.60	

(AlexNet) method reduces the MAE to 2.86. It shows that if we train all the regions together, it is inevitable that each sub-region affects other's judgment so that we can not take full advantage of all regions. Thus, we separate the sub-regions and combine them with an appropriate ensemble method in our MRNPE framework.

4.3 Comparison with State-of-the-art Algorithms

In order to show the effectiveness of our method, we compare our method with other state-of-the-art algorithms, as summarized in Table 2 and Table 3 based on the MORPH Album II and the CACD, respectively. To further demonstrate our results, we also adopt Cumulate Score (CS) [11] to evaluate our performance, and the results are shown in Fig. 3 and Fig. 4.

Results on MORPH Album II We can see from the Table 2 and Fig. 3, for the MORPH Album II, our MRNPE (AlexNet) framework gets great results which are superior to other state-of-the-art methods without any pretrained model under MAE with EV, in all com-

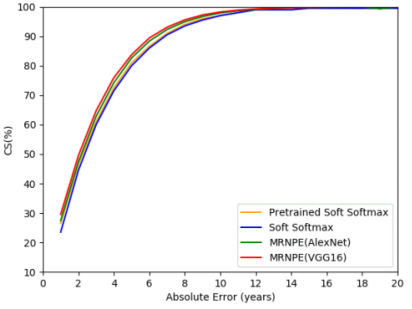

Figure 3: The CS curves based on MORPH Album II dataset.

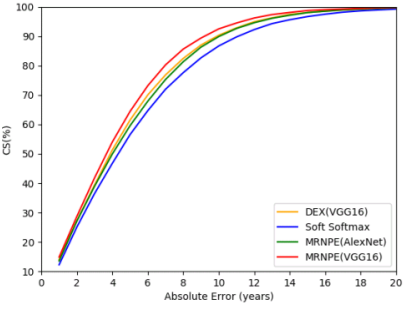

Figure 4: The CS curves based on CACD dataset.

bination of testing protocol. Compared with our MRNPE (AlexNet) method, the MAE of Pretrained Model Soft softmax method is decreased from 3.03 to 2.86. Besides, our MRNPE (VGG16) gets the most impressive result, which reduces the MAE to 2.73 on the MORPH Album II dataset. For the Cumulate Score, our methods also beat other the state-of-art methods with a considerable margin from CS1 to CS20.

Results on CACD For the CACD dataset, note the images in CACD are taken in the unconstrained environment, making age prediction more challenging. The result of our method in CACD can be seen from the Table 3 and Fig. 4. MRNPE (AlexNet) is better than the majority of the state-of-art methods, comparable to the DEX (VGG16) [21] method. DEX (VGG16) method adopt a very deep network structure, while our method(AlexNet) adopt a shallower network structure. Thus, it can be argued that our MRNPE (AlexNet) still works well on the CACD dataset. It is worth noting that our method with the VGG-16 structure surpass the result of DEX (VGG16) and it also outperforms all of the existing state-of-the-art age estimation methods by a significant margin. The MAE of age estimation is dropped from 4.79 to 4.40 years on CACD. Moreover, the Cumulate Score in CACD, our MRNPE (VGG-16) framework beats DEX (VGG16) in all years. These results also demonstrate that our method is suitable for the VGG16 structure on age estimation.

5 Discussion

5.1 Brief Comparison of the Ensemble Combination

Except the ensemble method with averaging for the final prediction, we try to adopt a simple ensemble method so as to briefly investigate the influence of the weights of different subnets. We assign 1/2, 1/3 and 1/6 to three subnets of their weight's parameter respectively according to the subnets' rank of the performance in Table 1. We further assign higher weights to the low performing sub-networks in order to investigate whether a slight change of the weights can have a huge influence on the final accuracy. We also evaluate the performance of our combination weights as setting the values of the weights of eye, nose, and mouth to 1 0 0, 0 1 0 and 0 0 1 respectively. Those cases correspond to the results in Table 1. The results are shown in the Table 4 and it shows that results are relatively stable w.r.t weights in a reasonable range (around 1/3), which indicates that our combination strategies are not sensitive to the combination parameters, but with one set of the best parameters being 1/3, 1/3 and 1/3 (averaging).

Table 3: Comparisons with state-of-the-art methods on CACD dataset under the same testing protocol.

Method	Train Set	Test Set	MAE with EV ↓
DFDNet [16]	1800 celebrities	200 celebrities	5.57
Single CNN Softmax	1800 celebrities	200 celebrities	5.28
Soft softmax [23]	1800 celebrities	200 celebrities	5.19
MRNPE (AlexNet)	1800 celebrities	200 celebrities	4.85
DEX (VGG16) [21]	1800 celebrities	200 celebrities	4.79
MRNPE (VGG16)	1800 celebrities	200 celebrities	4.40

Table 4: Different MRNPE (AlexNet) ensemble methods based on MORPH Album II (Training Set S1 and Testing Set S2+S3).

Left Eye Weight	Nose Weight	Mouth Weight	Avg. MAE with EV↓
1/3	1/3	1/3	2.98
1/2	1/3	1/6	2.98
1/2	1/6	1/3	2.99
1/3	1/2	1/6	2.99
1/3	1/6	1/2	3.02
1/6	1/2	1/3	3.01
1/6	1/3	1/2	3.04

5.2 Improvement Using Mirroring Prediction and Data Augmentation on MORPH Album II

We investigate the performances of using mirroring prediction or data augmentation strategies on the MORPH Album II dataset with training set S1 and testing set S2+S3, and the results are shown in the Table 5. From the table we can see that in our experiments, data augmentation improves the performance a lot and the MAE is dropped from 3.18 to 3.02. In contrast, using mirroring prediction strategy gets small but noticeable improvement, which slightly decreases the MAE from 3.18 to 3.13. Besides, it also shows that without these strategies our framework still can work well on age estimation.

6 Conclusion and Future Work

We propose a novel age estimation framework based on CNN in this paper, called MRNPE. We use different important regions from the human face, and take full advantage of these

Table 5: Comparisons with MRNPE(AlexNet) without data augmentation or using mirroring prediction on MORPH Album II (Training Set S1 and Testing Set S2+S3).

Method	Data augmentation	Using mirroring prediction	MAE with EV↓
MRNPE (AlexNet)	No	No	3.18
MRNPE (AlexNet)	No	Yes	3.13
MRNPE (AlexNet)	Yes	No	3.02
MRNPE (AlexNet)	Yes	Yes	2.98

features by combining them from separated networks. We utilize a proper ensemble method to combine these predictions of subnets. Moreover, our method can outperform the existing state-of-the-art age estimation methods by a significant margin on the MORPH Album II and the CACD datasets. Our method improves the results on the MORPH Album II dataset from 3.03 to 2.73, while the corresponding MAE on CACD dataset is dropped from 4.79 and 4.40 respectively. Future work could include 1) investigating the number of regions we used, such as the corner of eyes, eyebrows and so on; 2) other ensemble methods like Adaboost so as to find out the best weights for different subnets; and 3) pre-training our networks using a large dataset of age estimation. It will be also interesting to develop variants of our framework for other computer vision tasks such as person re-identification.

Acknowledgement

This work was supported by the Macau Science and Technology Development Fund of (No. 019/2014/A1, No. 112/2014/A3), the National Key Research and Development Plan (Grant No. 2016YFC0801002), the Chinese National Natural Science Foundation Projects ♯61502491, ♯61473291, ♯61572501, ♯61572536, NVIDIA GPU donation program and AuthenMetric *R&D* Funds.

References

[1] Grigory Antipov, Moez Baccouche, Sid-Ahmed Berrani, and Jean-Luc Dugelay. Apparent age estimation from face images combining general and children-specialized deep learning models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 96–104, 2016.

[2] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *European Conference on Computer Vision*, pages 768–783. Springer, 2014.

[3] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[4] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001.

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

[6] Feng Gao and Haizhou Ai. Face age classification on consumer images with gabor feature and fuzzy lda method. In *International Conference on Biometrics*, pages 132–141. Springer, 2009.

[7] Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 29(12):2234–2240, 2007.

[8] Asuman Gunay and Vasif V Nabiyev. Automatic age classification with lbp. In *Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on*, pages 1–4. IEEE, 2008.

[9] Guodong Guo and Guowang Mu. Human age estimation: What is the influence across race and gender? In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 71–78. IEEE, 2010.

[10] Guodong Guo and Guowang Mu. Joint estimation of age, gender and ethnicity: Cca vs. pls. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–6. IEEE, 2013.

[11] Guodong Guo, Guowang Mu, Yun Fu, and Thomas S Huang. Human age estimation using bio-inspired features. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 112–119. IEEE, 2009.

[12] Wen-Bing Horng, Cheng-Ping Lee, and Chun-Wen Chen. Classification of age groups based on facial features. *Tamkang Journal of Science and Engineering*, 4(3):183–192, 2001.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *International Conference on Neural Information Processing Systems*, pages 1097–1105, 2012.

[14] Young Ho Kwon et al. Age classification from facial images. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 762–767. IEEE, 1994.

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

[16] Ting Liu, Zhen Lei, Jun Wan, and Stan Z Li. Dfdnet: discriminant face descriptor network for facial age estimation. In *Chinese Conference on Biometric Recognition*, pages 649–658. Springer, 2015.

[17] Ting Liu, Jun Wan, Tingzhao Yu, Zhen Lei, and Stan Z Li. Age estimation based on multi-region convolutional neural network. In *Chinese Conference on Biometric Recognition*, pages 186–194. Springer, 2016.

[18] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. Ordinal regression with multiple output cnn for age estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4920–4928, 2016.

[19] Allen W Rawls and Karl Ricanek Jr. Morph: Development and optimization of a longitudinal age progression database. In *European Workshop on Biometrics and Identity Management*, pages 17–24. Springer, 2009.

[20] Rasmus Rothe, Radu Timofte, and Luc Van 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.

[21] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, pages 1–14, 2016.

[22] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *Computer Science*, 2014.

[23] Zichang Tan, Zhou Shuai, Wan Jun, Lei Zhen, and Stan Z Li. Age estimation based on a single network with soft softmax of aging modeling. In *Computer Vision–ACCV 2016*. 2016.

[24] Ming Yang, Shenghuo Zhu, Fengjun Lv, and Kai Yu. Correspondence driven adaptation for human profile recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 505–512. IEEE, 2011.

[25] Dong Yi, Zhen Lei, and Stan Z Li. Age estimation by multi-scale convolutional network. In *Asian Conference on Computer Vision*, pages 144–158. Springer, 2014.