Age Estimation based on Multi-Region Convolutional Neural Network

Ting Liu, Jun Wan, Tingzhao Yu, Zhen Lei, and Stan Z. Li*

[1]Center for Biometrics and Security Research & National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences
{ting.liu,jun.wan,zlei,szli}@nlpr.ia.ac.cn,yutingzhao2013@ia.ac.cn

Abstract. As one of the most important biologic features, age has tremendous application potential in various areas such as surveillance, human-computer interface and video detection. In this paper, a new convolutional neural network, namely MRCNN (Multi-Region Convolutional Neural Network), is proposed based on multiple face subregions. It joins multiple face subregions together to estimation age. Each targeted region is analyzed to explore the contribution degree to age estimation. According to the face geometrical property, we select 8 subregions, and construct 8 sub-network structures respectively, and then fuse at feature-level. The proposed MRCNN has two principle advantages: 8 sub-networks are able to learn the unique age characteristics of the corresponding subregion and the eight networks are packaged together to complement age-related information. Further, we analyze the estimation accuracy on all age groups. Experiments on MORPH illustrate the superior performance of the proposed MRCNN.

Keywords: Facial Age Estimation, MRCNN, Convolutional Neural Network, Age Group Analyzing

1 Introduction

Age estimation based upon facial images, as an emerging soft biometrics identification technology, has become a hot topic among computer vision areas. The task of age estimation is to compute the appearance age of a given facial image.

Most of the traditional age estimation methods published are reviewed in [1, 2]. Texture features such as Gabor, LBP, PCA, Haar, and BIF have been widely used to represent both the holistic and local face regions. Then, classification, regression methods or the combination of the two are adopted to predict the age of face image, such as SVM, SVR, PLS, CCA. Among existing traditional methods, BIF+KCCA[14] is almost the best method in terms of accuracy. However, traditional ways of age estimation are often broken down into several incoherent steps, including data preprocessing, hand-crafted feature extraction, feature selection or down-sampling, classification or regression, etc. As a consequence, it exists some gaps between above-mentioned steps. Recently, deep

* corresponding author

(a) 12 local face regions. (b) 8 groups local face regions.

Fig. 1. Facial component location.

learning has achieved state-of-the-art results in the field of computer vision. Convolutional Neural Network (CNN) has also been introduced into age estimation. Wang et.al. [3] applies CNN for age estimation for the first time. Instead of using the feature obtained at the top layer, they use feature maps obtained in different layers, and adopt manifold learning to learn aging patterns. Levi et.al. [4] improves AlexNet for sex and age prediction. They regard the age estimation problem as a 8-classification problems, because ages are divided into eight groups. Yi et.al. [5] introduces multi-scale analysis strategy, which improves the age estimation performance significantly. To make further improvement on the performance of age estimation based on CNN, similar to [5], we utilize individual facial landmarks to locate several face subregions which have abundant texture information, e.g., wrinkle and facial marks. Early researches have demonstrated that pose, illumination and expression (PIE) variation have less effects on local areas than the entire face image. Face representation based on local regions are more robust and hold greater potential. In this paper, we propose a novel framework, Multi-Region Convolutional Neural Network (MRCNN). The highlights and main contributions of the paper are summarized as follow.

1. A novel Convolutional Neural Network framework MRCNN for face age prediction is proposed. MRCNN makes full use of multiple subregions which contain rich age information.

2. MRCNN achieves state-of-the-art performance on MORPH database, validating its effectiveness and superiority.

3. Each facial component is analyzed to examine its contribution to age estimation. And we also reveal some prediction facts in and across all age groups.

2 Multi-Region Convolutional Neural Network

2.1 Facial Component Location

Intuitively, the locations where have the richest age information are eyes, nose, mouse, etc. Based on above cognition, given a face image, we first localize 21 facial landmarks with reasonable accuracy using ASM [6]. And then, we align the

Table 1. The detailed description of facial component location .

Regions	Rectangle Color	Size	Characteristics
head	-	300×300	-
face	white	240×240	no unrelated background
eyes	purple	100×100	vital component
nose	yellow	100×100	vital component
mouth	red	100×100	vital component
eyecorners	cyan	60×60	crows feet, wrinkle
eyebags	green	60×60	skin sag, dark circles and pigmentation
nosewings	blue	60×60	edict wrinkle

face images based on the two pupils and the middle of the mouth. The remaining landmarks follows the transformation. All images are resized and cropped into 300×300 (as shown in Fig. 1(a). The effect of age irrelevant features (*e.g.* posture) can be eliminated via face alignment, which is indeed helpful for age estimation. Beside, we convert color images into gray images, for that color information has poor stability and little effect on age estimation.

In this paper, we empirically select 12 local areas for face representation. The 12 local areas include the *head*, *face*, *eyes*, *nose*, *mouth*, *eyecorners*, *eyebags* and *nosewings*. In Fig. 1(a), they are described by face rectangles with different colors. And details can be found in TABLE. 1.

Considering the facial symmetry, we group the 12 local areas into 8 groups as follow, the head and its mirror, the face and its mirror, the left-eye and the right-eye' mirror, the nose and its mirror, the mouth and its mirror, the outer corners of left-eye and right-eye' mirror, the eyebags left-eye and right-eye' mirror, the left wing and the mirror of the right wing of nose. And the 8 groups pathes are normalized as 60×60, as shown in Fig. 1(b).

2.2 The Architecture of MRCNN

In this paper, a new convolutional neural network, namely MRCNN (Multi-Region Convolutional Neural Network), is proposed based on multiple face sub-regions. The whole architecture of our MRCNN model is described in Fig. 2, which is designed to predict the age of the given face image. The input is the 8 groups preprocessed local face regions, with 60×60 size, and the output is a age label. For the eight local face areas, we construct eight sub-networks respectively. The details of each sub-network is illustrated as Fig. 3. The sub-network of each group is composed of three convolution layers and two max-pooling layers. The input of sub-network is a certain group of local face areas. The output C3 layer is feature maps, and its dimension is $16 \times 15 \times 15 = 3600$. All feature maps of eight groups are concatenated in channel dimension. As a consequence, the input of F4 layer is $8 \times 3600 = 28800$ dimensions. Then, we fuse these responses in the two following full connected layers. $L2$ loss is chosen to be the loss function which can be regarded as a linear regression layer. Stochastic Gradient Descent

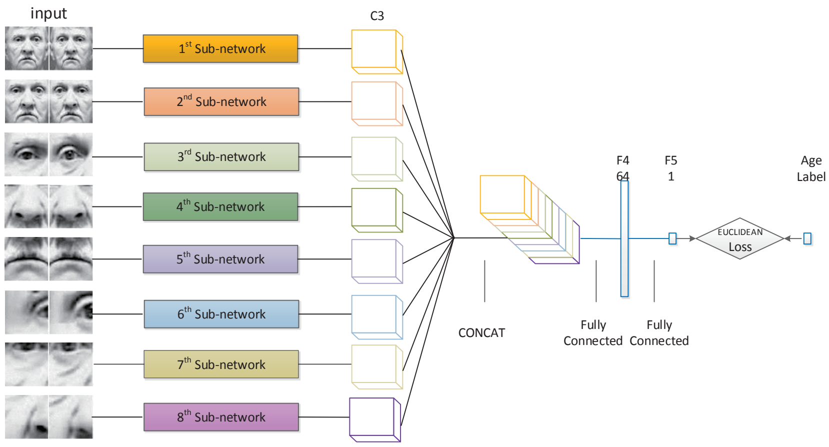

Fig. 2. The entire architecture of MRCNN.

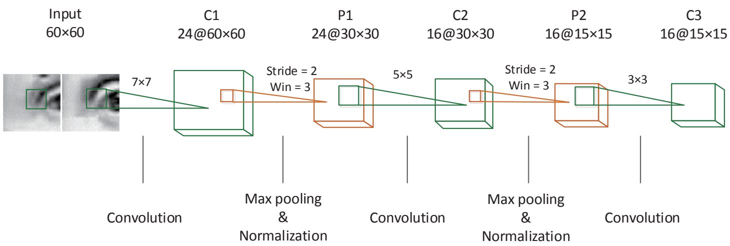

Fig. 3. The detailed architecture of the sub-network.

(SGD) is adopted to optimize the network. The detailed parameter settings are clearly laid out in Fig. 3.

Multiple face subregions join together to estimation age. MRCNN has several principle advantages: 8 sub-networks are able to learn the unique age characteristics of the corresponding sub-region and the final fully connected layer packages the eight together to complement age-related information. Compare with global-face representation, local-face representation is robust against rotation etc. image transformation. Unlike existing traditional methods, MRCNN automatically learns global optimal parameters instead of manual designed features and classifiers. Followed experiments can prove the effectiveness of MRCNN.

3 Experiments

We evaluate our proposed method on the MORPH database [7], which is the biggest available database for facial age estimation. We follow the standard evaluation protocol as [8]. Each targeted region is analyzed to explore the contribution degree to age estimation, and it also illustrates the superiority of multi-region method. Further, we analyze the estimation accuracy on all age groups.

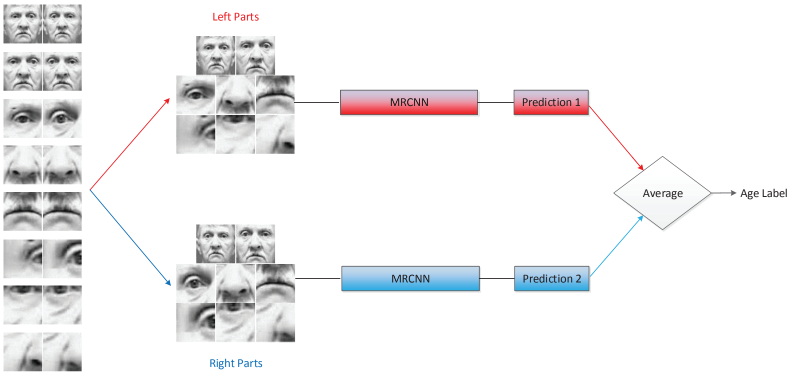

Fig. 4. The test phase of MRCNN.

3.1 Age Estimation

As referred in Section 2.1, each face image derives eight groups of local face patches. In the training phases, we regard the non-mirror patches as the left part, while the mirror patches as the right part. We merge these two parts together as the input of MRCNN. As a consequence, the number of the input of each sub-network is $10634 \times 2 = 21268$, which is equal to double the train data. In the testing phases, for a new given face image, two predicted results can be derived from the two training parts respectively. We fuse the two results at the decision level by calculating their average value. The specific operations can be found in Fig. 4.

3.2 Comparison of Age Estimation by Single and Multiple Regions

The MAE (Mean Absolute Error) and CS (Cumulate Score) [8] of our proposed method based on single region and multiple regions are shown in TABLE. 2 and Fig. 5. The MAE is the smaller the better, while the CS is the bigger the better. The facts are indicated with the red arrows in following TABLEs. MAE$_{left}$ is the error with left parts, while MAE$_{right}$ is the error with right parts. MAE$_{avg}$ is the error with average values. It is obviously that the performance of MRCNN is much superior to those methods based only upon global area and single region.

From the point of view of the horizontal comparison, we can separate the eight groups into four scales, namely, $head > face > eyes\&nose\&mouth > eyecorners\&eyebags\&nosewings$, thus MRCNN can be robust to distortion. Theoretically, the larger the scale is, the more information of age it will preserve. The MAE of head is much lower than nosewings'. Through the vertical comparison, it is obviously that *eyes* contains the most richest age information, and the MAE of *eyes* is lower than MAEs of *nose* and *mouth*, which are at the same scale with *eyes*. Even the methods based on the much smaller scale *eyecorners* and *eyebags* achieve better performances than *mouth* and *nose*. This emphasizes the importance of *eyes*. What's more, the contribution of *nose* is greater than *mouth*.

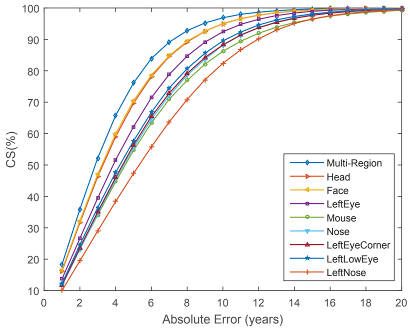

Fig. 5. The CS curves based on single and multiple regions.

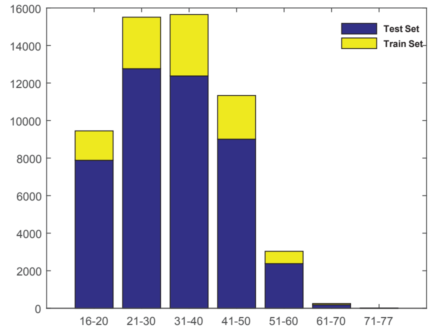

Fig. 6. The train and test set distribution of MORPH.

Though the different scales achieve different performances, they can complement each other via learning from all regions simultaneously. And the MAE of MRCNN is reduced to 3.48 years, which is superior to state-of-the-art algorithms.

Table 2. The Contribution Degree of Single and Multiple Regions.

Architecture		MAE↓			CS(5-year)↑
		MAE_{left}	MAE_{right}	$\mathbf{MAE}_{avg}$	
MultiRegion		3.60	3.63	**3.48**	**0.76**
SingleRegion	Head	4.08	4.15	3.97	0.70
	Face	4.06	4.05	3.92	0.70
	LeftEye	4.85	5.02	4.55	0.62
	Mouse	5.60	5.62	5.44	0.55
	Nose	5.31	5.39	5.21	0.56
	LeftEyeCorner	5.62	5.61	5.17	0.56
	LeftLowEye	5.35	5.43	4.99	0.58
	LeftNose	6.25	6.52	6.02	0.47

3.3 Comparison of Age Estimation in Different Age Groups

In order to demonstrate the effectiveness of our method on each age group, we implement our method on different age groups. The distribution of all age groups is shown in Fig. 6. The performances in different age groups are listed in TABLE. 3. Except the age groups of 61-70 and 71-77 whose numbers are too small, MRCNN get the best results in all age groups.

It is generally known that one's apparent age and biological age may be inconsistent. Apparent is affected by various intrinsic and extrinsic factors. With time going on, their gaps are increasing. Our experiments results reflected that well. Another finding is that MAE of mouse is lower than nose' in the youth, while contrary in the older. We consider the emerging edict wrinkle around nose can account for it.

Table 3. Age estimation in different age groups.

Indicators	Architecture		16-20	21-30	31-40	41-50	51-60	61-70	71-77
MAE$\downarrow$	MultiRegion		**3.49**	**2.88**	**3.32**	**3.93**	**5.32**	9.58	16.50
	SingleRegion	Head	4.27	3.31	3.64	4.46	5.89	9.57	17.14
		Face	4.02	3.25	3.68	4.44	5.95	10.00	15.04
		LeftEye	5.86	3.78	3.77	4.69	7.23	12.00	22.32
		Mouse	6.37	4.13	4.41	6.55	9.70	14.95	29.21
		Nose	6.60	4.16	4.25	5.92	8.03	12.44	22.05
		LeftEyeCorner	6.39	3.81	4.10	6.07	9.65	15.99	26.92
		LeftLowEye	6.33	3.91	4.23	5.55	7.64	12.45	21.87
		LeftNose	9.01	5.17	4.09	6.13	9.43	15.44	22.02
CS(5-year)$\uparrow$	MultiRegion		**0.78**	**0.84**	**0.77**	**0.70**	**0.54**	0.22	0
	SingleRegion	Head	0.67	0.78	0.73	0.63	0.50	0.25	0
		Face	0.70	0.79	0.72	0.63	0.49	0.18	0
		LeftEye	0.45	0.72	0.71	0.60	0.37	0.07	0
		Mouse	0.42	0.69	0.63	0.42	0.26	0.09	0
		Nose	0.37	0.67	0.65	0.49	0.33	0.10	0
		LeftEyeCorner	0.38	0.72	0.67	0.46	0.22	0.03	0
		LeftLowEye	0.41	0.71	0.65	0.51	0.36	0.13	0
		LeftNose	0.15	0.55	0.67	0.45	0.23	0.03	0

Table 4. Comparison with state-of-the-art algorithms.

Methods	MAE$\downarrow$	CS(5-year)$\uparrow$
BIF+CCA[9]	5.37	-
BIF+KCCA[9]	3.98	-
BIF+KPLS[10]	4.04	-
DFDnet[11]	4.65	0.60
LDL[12]	4.87	-
MSCNN[5]	3.63	-
DLA[3]	4.77	0.63
MRCNN(Ours)	**3.48**	**0.76**

3.4 Comparison with State-Of-The-Art Algorithms

We compare our method with other state-of-the-art algorithms, as summarized in TABLE. 4. Compared with traditional methods, MRCNN is a end-to-end self-learning system instead of using hand-crafted features. Besides, MRCNN introduces multi-regions fusion strategy to complement age-related information, which improves the age estimation performance significantly. On the MORPH database, the best traditional method achieves a MAE of 3.98 (BIF+KCCA), and the best method based on CNN achieves a MAE of 3.63 (MSCNN), while our method achieves a MAE of 3.63, reducing the error by 4.1 percents. All experiments show the effectiveness and superiority of MRCNN.

4 Conclusion

In this paper, we propose a novel method for age estimation, named MRCNN. We select eight face subregions, and construct eight sub-network structures respectively, and then fuse at feature-level. Eight sub-networks are able to learn the unique age characteristics of the corresponding subregion and the eight sub-networks are packaged together to complement age-related information. Each facial component is analyzed to examine its contribution to age estimation. We also analyze the estimation accuracy on all age groups. Experiments on MORPH illustrate the superior performance of the proposed MRCNN.

5 Acknowledge

This work was supported by the National Key Research and Development Plan (Grant No.2016YFC0801002), the Chinese National Natural Science Foundation Projects ♯61473291, ♯61572501, ♯61502491, ♯61572536, NVIDIA GPU donation program and AuthenMetric R&D Funds.

References

1. Fu, Y., Guo, G., Huang, T.S.: Age synthesis and estimation via faces: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(11) (2010) 1955–1976
2. Han, H., Otto, C., Liu, X., Jain, A.K.: Demographic estimation from face images: Human vs. machine performance. IEEE Transactions on Pattern Analysis and Machine Intelligence **37**(6) (2015) 1148–1161
3. Wang, X., Guo, R., Kambhamettu, C.: Deeply-learned feature for age estimation. In: IEEE Winter Conference on Applications of Computer Vision, WACV'15, IEEE (2015) 534–541
4. Levi, G., Hassner, T.: Age and gender classification using convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW'15, IEEE (2015) 34–42
5. Yi, D., Lei, Z., Li, S.Z.: Age estimation by multi-scale convolutional network. In: Asian Conference on Computer Vision, ACCV'14. Springer (2014) 144–158
6. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Computer Cision and Image Understanding **61**(1) (1995) 38–59
7. Ricanek, K., Tesafaye, T.: Morph: A longitudinal image database of normal adult age-progression. In: International Conference on Automatic Face and Gesture Recognition, FGR'06, IEEE (2006) 341–345
8. Guo, G., Mu, G., Fu, Y., Huang, T.S.: Human age estimation using bio-inspired features. In: IEEE Conference on Computer Vision and Pattern Recognitio, CVPR'09, IEEE (2009) 112–119
9. Guo, G., Mu, G.: Joint estimation of age, gender and ethnicity: Cca vs. pls. In: Automatic Face and Gesture Recognition, FG'13, IEEE (2013) 1–6
10. Guo, G., Mu, G.: Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR'11, IEEE (2011) 657–664

11. Liu, T., Lei, Z., Wan, J., Li, S.Z.: Dfdnet: Discriminant face descriptor network for facial age estimation. In: Chinese Conference on Biometric Recognition, CCBR'15. Springer (2015) 649–658
12. Geng, X., Yin, C., Zhou, Z.: Facial age estimation by learning from label distributions. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(10) (2013) 2401–2412