2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2017)

Deep Feature Selection and Projection for Cross-Age Face Retrieval

Kaihua Tang[1], Xiaonan Hou[1], Zhiwen Shao[1], Lizhuang Ma[1]
[1]Shanghai Jiao Tong University
tkhchipaomian@gmail.com, xnhou1989@gmail.com, shaozhiwen@sjtu.edu.cn, ma-lz@cs.sjtu.edu.cn

***Abstract*—While traditional PIE (pose, illumination and expression) face variations have been well settled by latest methods, a new kind of variation, cross-age variation, is drawing attention from researchers. Most of the existing methods fail to maintain the effectiveness in real world applications that contain significant gap of age. Cross-age variation is caused by the shape deformation and texture changing of human faces while people getting old. It will result in tremendous intra-personal changes of face feature that deteriorate the performance of algorithms. This paper proposed a deep feature based framework for face retrieval problem. Our framework uses deep CNNs feature descriptor and two well designed post-processing methods to achieve age-invariance. To the best of our knowledge, this is the first deep feature based method in cross-age face retrieval problem. The deep CNNs model we use is firstly trained on traditional PIE datasets and then fine-tuned by cross-age dataset. The feature selection and projection post-processing we propose is also proved to be very effective in eliminating cross-age variation of deep CNNs feature. The experiments conducted on Cross-Age Celebrity Dataset (CACD), which is the largest public dataset containing cross-age variation, show that our framework outperforms previous state-of-the-art methods.**

***Index Terms*—face retrieval; cross-age variation; deep feature**

I. Introduction

Face retrieval is one of the most important computer vision topics in application. It not only requires promising accuracy as face recognition or verification, but also needs convincing ordering of retrieved results. Like face recognition and verification, PIE (pose, illumination and expression) variations are three key factors that most of the previous works focus on. However, cross-age variation is crucial in real world applications. Take two common face retrieval applications into consideration, finding criminals and missing persons, the gallery dataset usually contains faces across different ages. When dealing with cases that contain over ten years gap of age, cross-age variation is extremely significant. Without an age-invariant feature, face retrieval systems will lose their effectiveness. The comparison between PIE variations and cross-age variation is illustrated in Fig. 1. Due to the tides of deep convolution neural networks (CNNs), latest deep feature based methods have been similar or even higher than human performance in various computer vision problems, such as face and object detection, face recognition, face verification, etc. However, as far as we know, there is only one CNNs based method in cross-age face recognition [1], which actually just uses a shallow CNNs model (5 convolution layers).

(a) Faces containing PIE variations

(b) Faces containing Cross-Age variation

Fig. 1. The above images give examples of different variations on faces. As we can see, cross-age variation is far more complicated than PIE variations.

Meanwhile, there is still no deep feature based methods are published in cross-age face retrieval research. This is probably because deep CNNs requires a mass of training data, but existing public cross-age face datasets can not satisfy this demand.

The existing age-related researches can be briefly divided into five categories: age estimation, aging simulation, cross-age face retrieval, cross-age face recognition and verification. Among all of these fields, age estimation and aging simulation don't directly deal with cross-age variation. Compared with face recognition and verification across ages, cross-age face retrieval has a higher demand in age-invariance of face feature. Because recognition and verification can further improve their performance by choosing classification methods, Besides, face retrieval also has a strict requirement for ordering. It makes the cross-age face retrieval the toughest task above them all.

To the best of our knowledge, the existing cross-age face retrieval methods [2]–[4] are all based on handcrafted local feature, which is referred as shallow feature in this paper. The most effective shallow feature in this field is high-dimensional LBP (HD-LBP) [6]. It extracts multi-scale patches on facial landmarks and concatenates their local LBP features to form a high-dimensional feature vector. After using PCA to eliminate the redundant information, the robustness and effectiveness of HD-LBP will dramatically increase. Most of the existing methods in this field [2]–[4] are all based on HD-LBP feature. In recent years, it gradually becomes a common sense that

978-1-5386-1937-7/17/$31.00 ©2017 IEEE

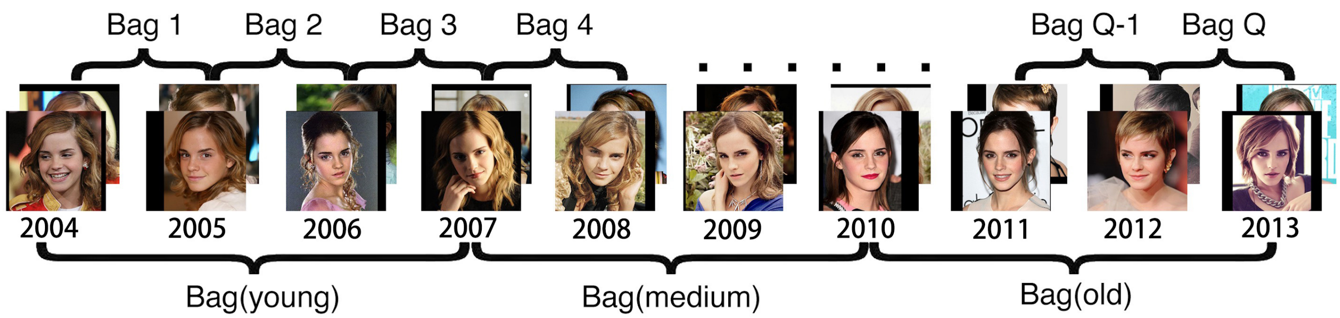

Fig. 2. The organization of the local (above) and global (bottom) age bags that used to ensure continuity and consistency of final feature.

deep feature is much more competitive than traditional shallow features. However, the existing cross-age face retrieval frameworks [2], [3] only work on shallow features. This is probably because these methods depend on the structural information of shallow features while deep feature has already lost its structural information in fully connected layer. To eliminate cross-age variation on deep feature, we propose a novel framework and two post-processing methods: variance based feature selection and age-bag based feature projection. The proposed framework successfully improves the performance of deep feature on cross-age face retrieval problem.

This paper implements VGG deep CNNs extractor as described in [7]. To tackle the data-deficiency problem, the model will be firstly trained on a traditional PIE dataset: VGG dataset [7] (2.6M images from 2,622 identities), and then fine-tuned on Cross-Age Celebrity Dataset (CACD). Since deep feature extracted from fully connected layer loses the structural information, we assume each feature dimension to be independent and thus design an age sensitive measure factor to eliminate those dimensions that contain significant cross-age variation. By using a threshold to select features from the output of fully connected layer, the total feature dimension can be reduced and the robustness of the system will increase. The other post-processing method is age-bag based feature projection. We organize the images of each train individual into two different types of age bag. The first type of age bag, which we call local age bag, contains images from neighboring years. It will be used to ensure the continuity of projected feature. The second type of age bag is global age bag. All of the training images from a specific person will be put into three global bags: young, medium and old, to guarantee the global consistency of final feature. Fig. 2 illustrates the organizing of age bags. The contributions of the this paper can be concluded into three:

- As far as we know, the proposed framework firstly applies deep CNNs feature into cross-age face retrieval research. Two-step training is used to tackle data-deficiency, and experiment results of our method outperform previous state-of-the-art methods.
- We propose a variance based feature selection for deep feature, which based on an age sensitive measure factor we design. Feature selection can reduce the feature dimension while improving the robustness of the system.
- We also take age specificities into consideration and propose an age-bag based feature projection. This projection ensures the local continuity and global consistency of final age-invariant face feature.

The rest of this paper will be organized as following: in section II, we describe the related work of our research; in section III, we introduce the details of proposed feature selection and feature projection; in section IV, we conduct several experiments on CACD, which is the largest public dataset containing cross-age variation; then the conclusion of this paper will be given in section V.

II. Related Work

A. Age-Related Researches

In early years, due to the absence of highly-qualified cross-age face datasets, most of the age-related researches only focus on aging simulation [10], [11] and age estimation [12], [13] problems. After several quality datasets [2], [16] are published and deep CNNs frameworks became popular, it further boosts the development of age estimation researches. Since age estimation can be simply considered as classification or regression problem, various CNNs based methods keep shattering the record for estimation mean absolute error (MAE). For example, a parallel multi-scale CNNs framework is proposed by Yi et al. [14]. It trains different sub-networks for each facial image patch and concatenates their output for estimation. Tan et al. [15] design a soft softmax regression function that considers age as intervals instead of discrete values. Although deep CNNs is not fit for aging simulation problem, an RNNs based method proposed by [17] also shows convincing results in aging simulation.

Because of the limited number of cross-age face images, most of the existing cross-age face verification and recognition methods [4], [5] apply shallow feature extractor in their systems. Although some commercial institutions, such as Baidu [9], claim to achieve highly age-invariant face recognition performance by deep CNNs, their voluminous datasets are not available to public. The commonly used public datasets in this field are MORPH [16] and CACD [2]. In 2016, Wen et al. [1] firstly combined CNNs feature with latent factor analysis and achieved the state-of-the-art result on both cross-age face verification and recognition. However, the networks they use only contain 5 convolutional layers, which can barely be considered as deep networks.

As far as we know, there is still no deep CNNs feature based method for cross-age face retrieval research. Cross-age reference coding (CARC) [2], which is published by Chen et al., uses HD-LBP [6] as original feature. It collects 600 specific individuals as reference to encode the original feature. Each individual can be considered as an aging model and each model will pool out an age-invariant value as part of final age-invariant feature. Hou et al. [4] propose a robust feature mapping (RFME) pre-processing to eliminate cross-age variation on original HD-LBP feature. In 2016, Tang et al. [3] and their eigen-aging reference coding (EARC) optimized the reference coding method by using eigen-faces instead of specific training individuals. The number of reference is reduced to 50 and the performance is also increased. Both RFME and EARC also use HD-LBP feature like CARC. Although their algorithms are proved to be effective in eliminating cross-age variation on shallow feature, the shallow feature itself limits their performance. This paper also tentatively combines EARC and CARC with deep feature (cf. Table. I). However, due to the distinction between shallow feature and deep feature, these methods fail to maintain their effectiveness. Therefore, EARC and CARC are not fit for deep CNNs feature. In fact, the performance of original VGG feature deteriorates heavily after applying these methods.

B. Deep Convolutional Neural Networks

CNNs have boosted the entire computer vision community by its overwhelming efficiency. Its application fields include but not limited to object detection and classification, face detection, face recognition and verification. However, the results of deep CNNs are highly dependent on the quality and quantity of training datasets. Benefit from various large-scale datasets that are available on the Internet, it is able to drive the development of recent scientific progress. There are two key usages of deep CNNs: (i) end-to-end classifier, or (ii) effective feature extractor. This paper implements a popular deep CNNs descriptor, VGG-Face descriptor [7], as original face feature extractor. The details of this model is described in section IV and Fig. 6.

III. The Proposed Method

A. Variance Based Feature Selection

Since it's inevitable for shallow feature descriptor to extract redundant information, dimensional reduction is essential to shallow feature based frameworks. The most commonly used dimensional reduction method is PCA. It will maintain the structural information and eliminate undesirable information. According to the previous research [6], this kind of pre-processing could dramatically improve the robustness and accuracy of the system.

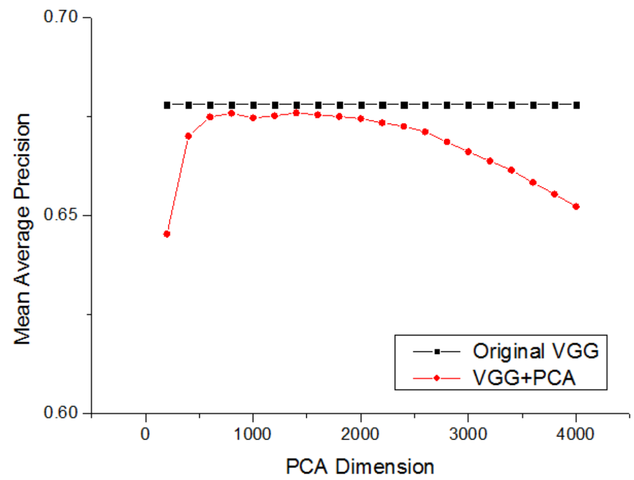

Fig. 3. The retrieval performance of combining VGG-Face feature and PCA.

In the meantime, due to the high representativeness of the deep feature, most of the researchers regard dimensional reduction to be unnecessary on deep feature based methods. It does be true to a certain extent. Actually, if a smaller size of deep feature is required, it's better to re-train a new fully connected layer with smaller number of output dimension. However, since cross-age variation has its own specificities, we construct an age sensitive measure factor by cross-age variance and cross-face variance. This variance based factor can be used to measure the age-sensitiveness of deep feature. Since each dimension of the output on fully connected layer is independent from each other, we apply a threshold of this factor to select features on fully connected layer. Those age-sensitive dimension, which contains significant cross-age variation or far worse than average performance, will be removed. As we can see from Fig. 3, if PCA is directly applied into deep CNNs feature, it comes with the cost of losing accuracy, which the last thing we want to see. At the same time, the proposed feature selection (FS) and feature projection (FP) can improve the performance while reduce the dimension about 20% from original deep feature (cf. Table. II).

B. Age Sensitive Measure Factor

The age sensitive measure factor is designed by comparing mean cross-age variance and mean cross-face variance of the feature.The cross-age variance is calculated on sets of features that comes from each identity at the different years (ages). It is formulated as Eq. 1, where i represents the identity from 1 to N, j is the year (age) of each training identity varying from 1 to M, X_i is the feature set of each identity, $x_{i,j}$ is the average feature of person i on year j, and μ_i is the average feature of person i across different ages. Cross-face variance is similarly computed in Eq. 2. X_j is the feature set of each year, which contains faces from different identities. μ_j is the

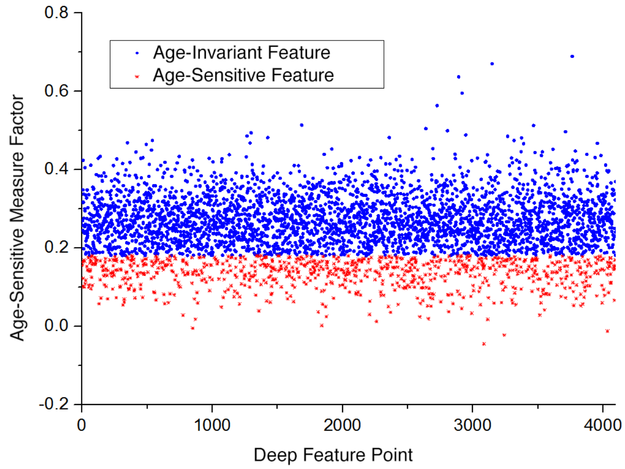

Fig. 4. Age-sensitive feature factors of 4096-dimensional VGG-Face feature. Red points that under threshold are considered as age-sensitive while the blue points are age-invariant.

average feature on a certain year across faces.

$$Var_{age}(X_i) = \frac{\sum_{j=1}^{M}(x_{i,j} - \mu_i)^2}{M}, \quad (1)$$

$$Var_{face}(X_j) = \frac{\sum_{i=1}^{N}(x_{i,j} - \mu_j)^2}{N}. \quad (2)$$

After these two types of variance are achieved, the mean $Var_{age}(X_i)$ and mean $Var_{face}(X_j)$ should be calculated across all training identities and years. It will further increase the robustness over single variance.

$$Var_{age} = E(Var_{age}(X_i)) = \frac{\sum_{i=1}^{N} Var_{age}(X_i)}{N}, \quad (3)$$

$$Var_{face} = E(Var_{face}(X_j)) = \frac{\sum_{j=1}^{M} Var_{face}(X_j)}{M}. \quad (4)$$

Eventually, we compute the difference between mean cross-face variance and mean cross-age variance of the feature and use maximum of them to normalize the factor as following:

$$Factor = \frac{Var_{face} - Var_{age}}{max(Var_{face}, Var_{age})}, \quad (5)$$

Two kinds of normalization methods are tested in our experiments: $max(Var_{face}, Var_{age})$ and $abs(Var_{face}) + abs(Var_{age})$. Since they barely have any difference, we only pick $max(Var_{face}, Var_{age})$ for all experiments in section IV.

By applying threshold, we can easily remove those outputs with low value of age-sensitive measure factor. Two potential reasons will cause the age-sensitiveness, either (i) Var_{age} is too larger that offsets its Var_{face} or (ii) both Var_{age} and Var_{face} are small, which means this feature output is redundant and undesirable in our system.

C. Age-Bag based Feature Projection

After deep CNNs feature is selected by the above method, we also design a projection method based on what we call age bags. An age bag is a set of face feature that contains features from a certain training individual on adjacent years. Based on common knowledge, the appearance of human face is continuously changing with aging process. Face features on neighboring years should be relatively more similar than those having a significant gap of age. This is the continuity of cross-age face feature. So we design local feature bag as we can see from Fig. 2. These local bags contains images from two adjacent years. Minimizing the difference between neighboring local bags can guarantee the continuity of projected feature. However, only taking continuity into consideration can not remove the cross-age variation of feature. So we further construct three global face bags, these bags only briefly categorize training images into: Bag_{young}, Bag_{medium} and Bag_{old}. They can further ensure the consistency of feature.

In CACD dataset, all the face images are capture in the years varying from 2004 to 2013. For each training individual, we separate the image set into $M - 1$ local face bags and 3 global ones. Local face bags are marked as $bag_q, q = 1, 2, ..., Q(M - 1)$ and average feature of each local bag is $Y_{i,q} = Mean(x_{i,j}), j \in bag_q$. Three global face bags respectively contain images of the same person on 2004-2007, 2007-2010 and 2010-2013. They are defined as bag_{young}, bag_{medium} and bag_{old}. Their corresponding feature vectors are $Z_{i,p} = Mean(x_{i,j}), j \in bag_p, p \in \{young, medium, old\}$. For the convenience of calculation, we gather above bag feature vectors into feature matrix: $Y_q = [Y_{1,q}, Y_{2,q}, ..., Y_{n,q}]$ and $Z_p = [Z_{1,p}, Z_{2,p}, ..., Z_{n,p}]$, where n is the number of training individuals. Then we come up with the optimization function: Eq. 6. I_0 is the identity matrix with size n; P is the projection matrix that we want to achieve.

$$\begin{aligned} min \sum_{q=1}^{Q-1} &||P(Y_q - Y_{q+1})||_F^2 \\ &+ \beta||P(2Z_{middle} - Z_{young} - Z_{old})||_F^2 \\ &+ \alpha||P - I_0||_F^2. \end{aligned} \quad (6)$$

By solving the above equation. we achieve the projection matrix P:

$$\begin{aligned} P = (&\sum_{q=1}^{Q-1}(Y_q - Y_{q+1})(Y_q - Y_{q+1})^T \\ &+ \beta(2Z_{middle} - Z_{young} - Z_{old})(2Z_{middle} - Z_{young} - Z_{old})^T \\ &+ \alpha I_0)^{-1}\alpha I_0. \end{aligned} \quad (7)$$

Two parameters $\{\beta, \alpha\}$ are used to adjust the weight among continuity, consistency and sparsity of projection matrix. This projection will map all the selected deep CNNs feature into a new space, which will minimize the cross-age variation.

TABLE I
THE PERFORMANCE OF EARC [3] AND CARC [2] ON DEEP CNNS FEATURE. VGG FEATURE IS USED ON THIS COMPARISON.

Method	2004-2006	2007-2009	2010-2012
VGG	0.636	0.668	0.731
VGG+CARC	0.617	0.651	0.716
VGG+EARC	0.621	0.656	0.716
VGG+FS+FP	0.662	0.696	0.753

TABLE II
THE COMPARISON OF DIFFERENT DIMENSIONAL REDUCTION METHODS APPLIED ON VGG FEATURE. FS AND FP ARE PROPOSED FEATURE SELECTION AND FEATURE PROJECTION METHODS.

Method	Dimension	2004-2006	2007-2009	2010-2012
VGG	4096	0.636	0.668	0.731
VGG+PCA	1500	0.634	0.664	0.728
VGG+FS	3253	0.643	0.675	0.737
VGG+FS+FP	3253	0.662	0.696	0.753

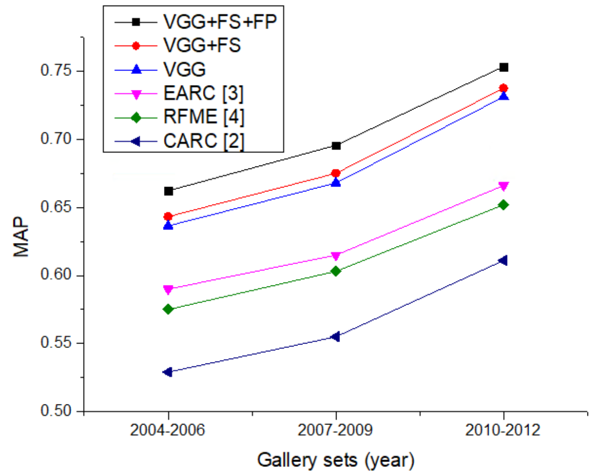

Fig. 5. Comparisons of the proposed method with previous state-of-the-art methods on CACD. FS means variance based feature selection. FP means the age-bag based feature projection.

IV. EXPERIMENTS

A. Cross-Age Celebrity Dataset

CACD [2] is the largest public dataset in cross-age fields. It contains about 160,000 images of 2,000 identities across 10 years (2004-2013). The age of identities varies from 16 to 62. Compared with other popular age-related face datasets: MORPH album II [16] and FG-NET, the total number of images on CACD is far more larger than these two. Besides, images on MORPH album II and FG-NET are all under highly constrained environments, which can not guarantee a robust model for real-world application.

To be fair, we use the same testing face set as the previous state-of-the-art methods [2]–[4]. CACD is devided into 4 parts in our approach: (i) 1000 out of 2000 identities are used as training set for VGG-Face model; (ii) 800 celebrities will be used to train projection matrix; (iii) 80 identities are used for parameter selection; and (iv) the rest 120 construct the testing set.

B. VGG-Face model and Pre-processing

This paper implement VGG-Face model [7] as the deep CNNs feature descriptor. VGG deep CNNs have long been proved to be one of the most effective face feature descriptors. The architecture of VGG networks in this paper is identical to [7]. 13 convolution layers and 5 maximum pooling layers are used as shown in Fig. 6. Stride and padding are set to be 1 and 1 for every convolution layers. Each convolution layer is connected with a nonlinear function ReLU, which isn't displayed on the figure. The batch size is 64 and learning rate is automatically changing from 10^{-2} to 10^{-4}.

Output of the FC_7 is the deep facial feature we need. To train this model, we add a classification layer FC_8 with softmax function. The training process is separated into two steps. The first step uses VGG dataset just like [7]. VGG dataset contains 2,622 identities and 2.6M images in total. Because VGG dataset doesn't contain cross-age information, it is just used to achieve basic performance. The second training step uses 1,000 identities from CACD dataset to fine-tune the model again. A new classification layer with 1,000 output is used to replace old one. To augment the size of training set, we apply rotating ($\pm 5^0, \pm 10^0 and \pm 15^0$), flipping and Gaussian white noise (variance: 0.005, 0.01, 0.015, 0.02).

We implement similar pre-processing as described in [1]. 5 facial landmarks (center of two eyes, tip of nose, corners of mouth) are detected from each face according to method [18]. Then each face image will be cropped into a 224×224 patch by similarity transformation.

C. Mean Average Precision (MAP)

In order to measure the performance of retrieval results. Mean Average Precision (MAP) is used in this paper, which has already been widely accepted in image or information retrieval researches. MAP firstly calculates all the Average Precisions (APs) from query image set Q, where the number of query image is Num_Q. For each Average Precision (AP), the positive results will be gathered in descending ranking. For each query image, the number of positive results is n_i. MAP can be explained by following equation:

$$MAP(Q) = \frac{1}{Num_Q} \sum_{i=1}^{Num_Q} \frac{1}{n_i} \sum_{j=1}^{n_i} Precision(Rank_{i,j}), \quad (8)$$

where $Precision(R)$ is the ratio of positive images from top to image rank R. The reason why all researchers prefer MAP is that it can not only show the accuracy of result but also reflect the quality of retrieval ordering.

D. Parameters Selection

To select parameters of our method, we use images (80 identities) captured in 2013 as query set and the rest years as database. In the proposed method, we have to decide a threshold of the FS, $\{\beta, \alpha\}$ of projection matrix, and the number of training individual we use in FP. For the threshold

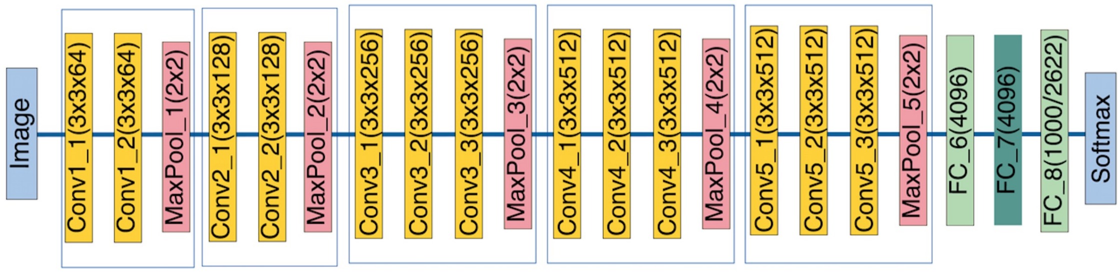

Fig. 6. The architecture of VGG deep CNNs model. The output of fully connected layer 7 is extracted feature. Fully connected layer 8 is used for classification based training. The size of fully connected layer 8 is either 2622 or 1000, which depend on the training datasets.

of FS, we binarily search and locate its value at 0.18 (see Fig.4). For $\{\beta, \alpha\}$, we greedily adjust their values from 10^{-5} to 10^{5} and then set $(\beta, \alpha) = (10^{-1}, 10^{4})$, where we can achieve the best performance. As to the number of training individuals for FP, we reserve 800 individuals at first place and change number from 100 to 800. The accuracy will barely increase after the number is over 500, so we fix it to 500 in our experiments.

E. Results and Analyses

To conduct retrieval experiments, Images from 120 testing celebrities are separated into one query set (2013) and three gallery databases (2004-2006, 2007-2009, 2010-2012). Cosine similarity and MAP measurement are chosen to evaluate the retrieval results.

This paper tentatively combines CARC [2] and EARC [3] with VGG deep feature. They are proved to be quite effective on HD-LBP shallow feature. However, as shown in Table. I, after applying CARC and EARC frameworks, MAP of VGG deep feature will decrease by $(1.7 \pm 0.2)\%$ and $(1.4 \pm 0.1)\%$ respectively.

As we mentioned above, traditional feature dimension reduction methods are not necessary and not fit for deep feature, because the dimension of deep feature can be manually manipulated by changing the size of fully connected layer. To measure the performance of traditional method PCA, we set PCA dimension from 100 to 4000 and draw all the MAP results on Fig. 3. The best PCA performance is compared to the proposed FS on Table. II. Although PCA is able to reduce over half of the dimension, MAP results decrease $(0.3{\pm}0.1)\%$ as well. Meanwhile, the proposed FS select about 80% dimensions from original feature and improve MAP by $(0.6 \pm 0.05)\%$.

The overall comparisons with previous methods are shown in Fig.5, VGG deep feature outperforms the existing shallow feature based methods over $(5.6 \pm 0.9)\%$ MAP. The proposed FS and FP further improve the performance of VGG deep feature by $(2.5 \pm 0.4)\%$. Compared with the former state-of-the-art method EARC [3], this paper improves $(7.9 \pm 0.8)\%$ MAP in total, which reflects much higher accuracy and better ordering among retrieved faces.

V. Conclusion

In this paper, we propose a deep feature based cross-age face retrieval framework. It implements two-step training on VGG model to solve the deficiency of cross-age face dataset. Then a variance based feature selection and a age-bag based feature projection are designed to remove the cross-age variation on deep feature. Variance based feature selection use cross-age variance and cross-face variance to measure the age-sensitiveness of deep feature and apply selection to remove undesirable outputs. Age-bag based feature projection collects images from adjacent years, and then use them to design a projection matrix, that can ensure the continuity and consistency of final age-invariant face feature.

VI. Acknowledgements

This research is partially supported by funds of Large-scale Applications and Recognition with Big Data of Human Faces Towards Internet Finance No.16511101300, Reverse-age Face Beautification Based on Hierarchical Saliency and Age Evolution Characteristics, NSFC No.61472245 and National High-tech R&D Program of China (863 Program) No.2015AA015904.

References

[1] Wen, Y., Li, Z. and Qiao, Y. "Latent factor guided convolutional neural networks for age-invariant face recognition." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4893-4901. 2016.

[2] Chen BC., Chen CS. and Hsu WH. "Cross-age reference coding for age-invariant face recognition and retrieval." In European Conference on Computer Vision, pp. 768-783. Springer International Publishing, 2014.

[3] Tang, K., Kamata, S. I., Hou, X., Ding, S. and Ma, L. "Eigen-Aging Reference Coding for Cross-Age Face Verification and Retrieval." In Asian Conference on Computer Vision, pp. 389-403. Springer, Cham. November 2016.

[4] Hou, X., Ding, S. and Ma, L. "Robust feature encoding for age-invariant face recognition." In Multimedia and Expo (ICME), 2016 IEEE International Conference on, pp. 1-6. IEEE, 2016.

[5] Gong, D., Li, Z., Lin, D., Liu, J. and Tang, X. "Hidden factor analysis for age invariant face recognition." In Proceedings of the IEEE International Conference on Computer Vision, pp. 2872-2879. 2013.

[6] Chen, D., Cao, X., Wen, F. and Sun, J. "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3025-3032. 2013.

[7] Parkhi, O. M., Vedaldi, A. and Zisserman, A. "Deep Face Recognition." In BMVC, vol. 1, no. 3, p. 6. 2015.

[8] Huang, G.B., Ramesh, M., Berg, T. and Learned-Miller, E. ”Labeled faces in the wild: A database for studying face recognition in unconstrained environments”. Vol. 1, no. 2. Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[9] Liu, J., Deng, Y., Bai, T., Wei, Z. and Huang, C. ”Targeting ultimate accuracy: Face recognition via deep embedding.” arXiv preprint arXiv:1506.07310 (2015).

[10] Suo, J., Chen, X., Shan, S. and Gao, W. ”Learning long term face aging patterns from partially dense aging databases.” In Computer Vision, 2009 IEEE 12th International Conference on, pp. 622-629. IEEE, 2009.

[11] Tsumura, N., Ojima, N., Sato, K., Shiraishi, M., Shimizu, H., Nabeshima, H., Akazaki, S., Hori, K. and Miyake, Y. ”Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin.” ACM Transactions on Graphics (TOG) 22, no. 3 (2003): 770-779.

[12] Fu, Y. and Huang, T. S. ”Human age estimation with regression on discriminative aging manifold.” IEEE Transactions on Multimedia 10, no. 4 (2008): 578-584.

[13] Geng, X., Zhou, Z. H. and Smith-Miles, K. ”Automatic age estimation based on facial aging patterns.” IEEE Transactions on pattern analysis and machine intelligence 29, no. 12 (2007): 2234-2240.

[14] Yi, D., Lei, Z. and Li, S. Z. ”Age estimation by multi-scale convolutional network.” In Asian Conference on Computer Vision, pp. 144-158. Springer, Cham, 2014.

[15] Tan, Z., Zhou, S., Wan, J., Lei, Z. and Li, S. Z. ”Age Estimation Based on a Single Network with Soft Softmax of Aging Modeling.” In Asian Conference on Computer Vision, pp. 203-216. Springer, Cham, 2016.

[16] Ricanek, K. and Tesafaye, T. ”Morph: A longitudinal image database of normal adult age-progression.” In Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, pp. 341-345. IEEE, 2006.

[17] Wang, W., Cui, Z., Yan, Y., Feng, J., Yan, S., Shu, X. and Sebe, N. ”Recurrent face aging.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2378-2386. 2016.

[18] Zhu, X. and Ramanan, D. ”Face detection, pose estimation, and landmark localization in the wild.” In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 2879-2886. IEEE, 2012.