Automatic Age Estimation from Face Images via Deep Ranking

Huei-Fang Yang[1]
hfyang@citi.sinica.edu.tw
Bo-Yao Lin[2]
boyaolin@iis.sinica.edu.tw
Kuang-Yu Chang[2]
kuangyu@iis.sinica.edu.tw
Chu-Song Chen[2]
song@iis.sinica.edu.tw

[1] Research Center for Information Technology Innovation
Academia Sinica
Taipei, Taiwan

[2] Institute of Information Science
Academia Sinica
Taipei, Taiwan

This paper focuses on automatic age estimation (AAE) from face images, which amounts to determining the exact age or age group of a face image according to features from faces. Although great effort has been devoted to AAE [1, 4, 6], it remains a challenging problem. The difficulties are due to large facial appearance variations resulting from a number of factors, *e.g.*, aging and facial expressions. AAE algorithms need to overcome heterogeneity in facial appearance changes to provide accurate age estimates.

To this end, we propose a generic, deep network model for AAE (see Figure 1). Given a face image, our network first extracts features from the face through a 3-layer scattering network (ScatNet) [2], then reduces the feature dimension by principal component analysis (PCA), and finally predicts the age via category-wise rankers constructed as a 3-layer fully-connected network. The contributions are: (1) Our ranking method is point-wised and thus is easily scaled up to large-scale datasets; (2) our deep ranking model is general and can be applied to age estimation from faces with large facial appearance variations as a result of aging or facial expression changes; and (3) we show that the high-level concepts learned from large-scale neutral faces can be transferred to estimating ages from faces under expression changes, leading to improved performance.

Our model is with the following characteristics:

(1) **The scattering features are invariant to translation and small deformations.** ScatNet is a deep convolutional network of specific characteristics. It uses predefined wavelets and computes scattering representations via a cascade of wavelet transforms and modulus pooling operators from shallow to deep layers. With the nonlinear modulus and averaging operators, ScatNet can produce representations that are discriminative as well as invariant to translation and small deformations. As ScatNet provides fundamentally invariant representations for discriminating feature extraction, only the weights of the fully-connected layers are learned in our network model, which considerably reduces the training time.

(2) **The rank labels encoded in the network exploit the ordering relation among labels.** Each category-wise ranker is an ordinal regression ranker. We encode the age rank based on the reduction framework [5]. Given a set of training samples $X = \{(x_i, y_i), i = 1 \cdots N\}$, let $x_i \in R^D$ be the input image and y_i be a rank label ($y_i \in \{1, \ldots, K\}$), respectively, where K is the number of age ranks. For rank k, we separate X into two subsets, X_k^+ and X_k^-, as follows:

$$\begin{aligned} X_k^+ &= \{(x_i, +1) | y_i > k\} \\ X_k^- &= \{(x_i, -1) | y_i \leq k\}. \end{aligned} \tag{1}$$

Next, we use the two subsets to learn a binary classifier from the network, which then conducts an answer to the query: "Is the image with a ranking score higher than k?" To each query, we simply make a binary decision between the positive and negative sides. Hence, each query reduces the age rank estimation task to a binary classification problem. A series of query results imply the ordinal relationship between the rank labels, where each query identifies the preferred classes.

Because the k-th binary classifier focuses on determining whether the age rank of an image is greater than k, $K-1$ such binary classifiers are required for K age ranks. Therefore, for a face image with true age k, the teaching vector with length $K-1$ is designed as $[1, \ldots, 1, -1, \ldots, -1]$, where the first $k-1$ values are 1 and the remaining -1.

(3) **The category-wise rankers perform age estimation within the same group.** Category-wise age estimation is commonly carried out in a hierarchical manner, first performing between-category classification and then within-category age estimation [3, 4]. Employing a hierarchical strategy in our network involves introducing more layers, which may result in more computational costs. We instead concatenate the encoding of multiple category-wise rankers altogether in the output layer. This design allows the category-wise rankers to learn age estimation according to the shared high-level representations.

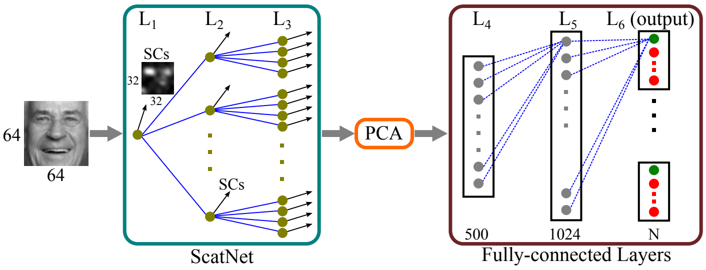

Figure 1: Our network model for human age prediction from face images. Our DeepRank+ comprises a 3-layer ScatNet, a dimensionality reduction component by principal component analysis (PCA) that reduces the feature dimension to 500, and a 3-layer fully-connected network with an architecture of 500-1024-N. The number of nodes N is application dependent. We use ReLUs in layer L_5 and sigmoids in layer L_6. The output layer is constructed with the category labels (the green nodes) and ranking ones (the red). When the category information is not used in training, the output layer can be constructed simply with the rank labels; such a network is coined as DeepRank.

Assume there are C category-wise rankers. The encoding for each ranker consists of two constituents: the category label(s) and the age rank. That is, for category j, its encoding is given as $e_j = [g_j, r_j]$, where g_j denotes the desired output of the category and r_j denotes the teaching vector of the rank. Concatenating the C encoding sets forms a final encoding: $E = [g_0, r_0, \ldots, g_j, r_j, \ldots, g_C, r_C]$.

We conduct two sets of sexperiments on three datasets. One is the experiments on age estimation from faces obtained from different gender and races on a large-scale MORPH dataset. The other is the experiments on age estimation from faces under expression changes from the Lifespan and FACES datasets. Our approach yields mean absolute errors (MAEs) of 3.49, 5.19 and 7.04 years on MORPH, Lifespan, and FACES, respectively, performing favorably against state-of-the-arts.

[1] Fares Alnajar, Zhongyu Lou, José Manuel Álvarez, and Theo Gevers. Expression-invariant age estimation. In *Proc. BMVC*, 2014.

[2] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):1872–1886, 2013.

[3] Guodong Guo and Guowang Mu. Human age estimation: What is the influence across race and gender? In *Proc. CVPRW*, 2010.

[4] Hu Han, Charles Otto, Xiaoming Liu, and Anil K. Jain. Demographic estimation from face images: Human vs. machine performance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2014.

[5] Hsuan-Tien Lin and Ling Li. Reduction from cost-sensitive ordinal ranking to weighted binary classification. *Neural Computation*, 24 (5):1329–1367, 2012.

[6] Dong Yi, Zhen Lei, and Stan Z. Li. Age estimation by multi-scale convolutional network. In *Proc. ACCV*, 2014.