Information Fusion 48 (2019) 84–94

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Learning deep compact similarity metric for kinship verification from face images

Xiuzhuang Zhou[a], Kai Jin[b], Min Xu[*,b], Guodong Guo[c]

[a] School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China
[b] College of Information Engineering, Capital Normal University, Beijing 100048, China
[c] Department of Computer Science and Electrical Engineering, West Virginia University, WV 26506, USA

ARTICLE INFO

Keywords:
Face recognition
Kinship verification
Metric learning
Deep neural network
Hierarchical compactness

ABSTRACT

Recent advances in kinship verification have shown that learning an appropriate kinship similarity metric on human faces plays a critical role in this problem. However, most of existing distance metric learning (DML) based solutions rely on linearity assumption of the kinship metric model, and the domain knowledge of large cross-generation discrepancy (*e.g.*, large age span and gender difference between parent and child images) has not been considered in metric learning, leading to degraded performance for genetic similarity measure on human faces. To address these limitations, we propose in this work a new kinship metric learning (KML) method with a coupled deep neural network (DNN) model. KML explicitly models the cross-generation discrepancy inherent on parent-child pairs, and learns a coupled deep similarity metric such that the image pairs with kinship relation are pulled close, while those without kinship relation (but with high appearance similarity) are pushed as far away as possible. Moreover, by imposing the intra-connection diversity and inter-connection consistency over the coupled DNN, we introduce the property of hierarchical compactness into the coupled network to facilitate deep metric learning with limited amount of kinship training data. Empirically, we evaluate our algorithm on several kinship benchmarks against the state-of-the-art DML alternatives, and the results demonstrate the superiority of our method.

1. Introduction

Recent evidence in psychology has indicated that face appearance is a reliable and critical cue for measure of the genetic similarity between the parent and their children [1–3]. Motivated by this, researchers from biometrics and computer vision societies have developed some computational models for kinship verification via face images [4–7]. The objective of this verification problem is to determine whether there exists a kin relationship between a given pair of face images. Potential applications based on such verification technique ranges from social media mining to children adoptions and missing children searching.

While encouraging results have been demonstrated over the past a few years, kinship verification using face images still remains open. On one hand, face images are often captured in wild conditions, and varying illumination, poses and expressions in such scenarios make the verification problem quite challenging. On the other hand, kinship verification aims to investigate the kin relationship between two different visual entities (*e.g.*, father and daughter), and thus the inherent appearance gap of intra-class in kinship verification is generally far larger than that in traditional face recognition [8–22].

Recent advances in kinship verification have indicated that learning an appropriate similarity metric on human faces plays a critical role in kinship verification. Distance metric learning (DML) methods [23–26] have been investigated in kinship verification [7,27,28] for the purpose of achieving an optimal distance metric rather than a pre-specified one for more robust kin-faces matching. Despite the success of DML-based approaches, existing solutions to kinship verification still suffer from two critical limitations:

(1) They are often proposed to learn a linear distance metric for input space, which is less powerful to capture the nonlinear manifold where the genetic traits inherent on human face lie. Moreover, in existing DML-based solutions the parent-child face images share a common linear transformation for visual matching, and hence the domain prior of large distribution gap between the parent and child has not been taken into account, leading to inaccurate measure of inherent kin similarity on human faces.

(2) While learning a nonlinear distance metric based on the deep neural networks (DNNs) [20,29] is a straightforward solution to this

* Corresponding author.
E-mail address: xumin@cnu.edu.cn (M. Xu).

https://doi.org/10.1016/j.inffus.2018.07.011
Received 24 November 2017; Received in revised form 5 June 2018; Accepted 31 July 2018
Available online 01 August 2018

problem, supervised metric learning with DNN typically requires a large number of labeled training samples, which is extremely expensive to collected in practical kinship verification due to the privacy concerns and involved time and human costs.

To address these issues, we propose in this paper a new kinship metric learning (KML) method for kinship verification from face images with a well-designed DNN architecture. The main contributions of this work are summarized as follows:

(1) We design a coupled DNN, named KinNet, for kinship verification from face images. KinNet explicitly models the cross-generation discrepancy inherent on parent-child pairs, and facilitates deep metric learning with limited amount of labeled kinship data. Particularly, by imposing the diversity regularization and cross-generation consistency regularization on the coupled connections, we introduce the property of hierarchical compactness into the coupled network to improve generalization performance of the kinship metric model.

(2) We develop a new deep metric learning algorithm with the proposed KinNet architecture to learn a deep compact cross-generation similarity metric. The learned similarity metric possesses some desirable properties that help address the limitations of most existing DML-based solutions to kinship verification.

From the information fusion point of view, the parent-child faces input to KML can be regarded as the two-view kin data for kinship verification, and hence KML can be considered as a *multi-view* metric learning in the deep learning framework. Essentially, our KML implicitly learns to fuse a pair of deep embeddings for robust similarity measure of the parent-child pairs.

On the other hand, by latent variable modeling, an ensemble of latent factors in weight matrices of the KinNet are enforced to be as diverse from one another as possible, such that the learned deep embeddings are compact enough to reduce information redundancy in metric learning. From the ensemble learning point of view, our KML implicitly learns to fuse a set of *diverse* latent factors in deep metric learning.

(3) We empirically evaluate our method on several benchmark datasets, and the results show that our proposed KML significantly boosts the current state-of-the-art level of kinship verification.

The remainder of this paper is organized as follows. We first briefly review the related work in Section 2 , and Section 3 details the kinship metric learning method with the proposed KinNet architecture. Experimental settings, results and discussions are presented in Section 4 , and Section 5 concludes the paper.

2. Related work

In this section, some related topics are briefly reviewed: (1) kinship verification, and (2) deep metric learning.

Roughly speaking, existing methods for kinship verification are either feature-based [4,6,30–34] or distance metric-based [5,7,27,28,35–38]. Feature-based methods extract discriminative feature from face images by hand-crafted image descriptors [4,6,30] or feature learning [32–34] to represent genetic traits on human face. Fang et al. [4] proposed to extract various local features (*e.g.*, skin color and histogram of gradient) from facial components to recover genetic traits on human face. It has been identified as one of the earliest attempts at tackling the kinship verification problem. After that, Zhou et al. [6] introduced a learning-based descriptor for representation and recognition of the kin faces. Guo and Wang [30] proposed to use the DAISY descriptors for matching of facial components with spatial Gaussian kernels. Yan et al. [33] proposed to construct a set of unlabeled face samples from the labeled face in the wild (LFW) dataset [39] as the reference set, and then introduced a prototype-based feature learning algorithm for the verification problem. Most recently, deep learning methods, such as the gated autoencoder [32] and representation learning based on convolutional neural network (CNN) [40] or deep belief network (DBN) [34], have been introduced to characterize the resemblance of parent-child pairs, and demonstrated powerful representation ability. Other feature representations or feature learning methods for kinship verification include facial dynamics [41], self similarity representation [42], and feature selection by spatially voting [43].

Distance metric-based methods aim to learn distance similarity metric (or embedding) for parent-child pairs based on some statistical learning methods. Recently, a variety of metric learning methods have been proposed over the past decade by following different settings in machine learning [23–26], and they have been extensively investigated in face recognition [8–10,12–14,17,18,20,21] and kinship verification [7,27,28,44,45]. The basic objective of these DML methods is to learn an appropriate distance metric, under which the distance between positive face pairs is reduced and that of negative pairs is enlarged as much as possible. In [7], Lu et al. introduced a neighborhood repulsed metric learning (NRML) method for kinship verification. They approach this by learning a Mahalanobis distance metric, based on which the face examples with kin relation are pulled close and those without kin relation are pushed far away. DML-based solutions to kinship verification can be further improved in multiview learning setting [28] or ensemble learning setting [44]. To reduce the large appearance gap of kin faces, Xia et al. [5,35] proposed to use intermediate young parent face images, and developed a transfer subspace learning (TSL) based algorithm to bridge the divergence between the children and old parent images.

In recent years, deep learning has drawn increasing attention in computer vision and machine learning. Most existing deep learning methods can be categorized in three different settings: unsupervised, semi-supervised, and supervised, and they have been successfully applied in various visual applications [11,40,46–49]. While many efforts have been made on deep learning [50,51], metric learning with a DNN has not been well studied. Most recently, Hu et al. [20] and Lu et al. [37] proposed a discriminative deep metric learning (DDML) method for face verification by imposing the large margin criterion on the output of the DNN. For cross-domain visual recognition, they further proposed to develop a deep transfer metric learning (DTML) method [21].

Similar to DDML methods [20,37,52,52] and DTML [21], our proposed KML also seeks to learn a deep metric for similarity measurement. However, the network architecture built for our KML is quite different from those of the DDML and DTML. In DDML and DTML, a set of connections of the DNN are shared in common for a pair of input images, while in our KML a couple of deep transformations are jointly learned for parent-child pair to take into account the domain knowledge of large distribution gap between the parent and child faces. Our method is essentially different from the heterogeneous metric learning with hierarchical couplings [53], which is focused on both the low-level value-to-attribute and the high-level attribute-to-class hierarchical couplings. Moreover, diversity regularization and cross-generation consistency regularization are imposed on the connections of our proposed coupled DNN, which facilitates deep metric learning with limited amount of labeled kinship data and improves generalization performance of the kinship metric model. This regularization has not yet been well investigated by previous deep metric learning methods. Conceptually, the hierarchical compactness on the connections of the coupled DNN is motivated by the work [54], where mutual angular regularization is enforced on latent factors of the linear transform to encourage its diversity. We extend the compactness on linear transformation to the hierarchical compactness on our coupled deep transformation in supervised metric learning setting. We empirically show that this hierarchical compactness imposed on DNN exactly benefits the deep metric learning for kinship verification in terms of the verification accuracy and computational efficiency.

There are some other related works [36,55–59] that address the view-specific (or cross-modal) metric learning problem. However, they have not considered the compactness of the view-specific (or model-

Table 1
Review of existing metric learning-based kinship verification methods. The columns Deep, Coupled, and Compact represent if the learned similarity metric is a deep, coupled, and compact model, respectively.

Year	Authors	Method	Deep	Coupled	Compact
2011	Xia et al. [37]	Transfer subspace learning	No	No	No
2014	Lu et al. [7]	Neighborhood repulsed metric learning	No	No	No
2014	Yan et al. [30]	Discriminative multimetric learning	No	No	No
2017	Hu et al. [29]	Large-margin multi-metric learning	No	No	No
2016	Zhou et al. [46]	Ensemble similarity learning	No	No	No
2017	Lu et al. [39]	Discriminative deep metric learning	Yes	No	No
2017	Liu et al. [62]	Status-aware projection learning	No	Yes	No
2018	Mahpod et al. [38]	Hybrid distance learning	No	Yes	No
2018	**Proposed**	Kinship metric learning	Yes	Yes	Yes

specific) transformation, and are not proposed in the setting of kinship verification. While the asymmetric metric learning methods reported in [36,45,59] also aim to learn a coupled distance metric for kinship verification, the learned linear transformation is fundamentally different from our proposed in that, our KML considers the cross-generation consistency regularization and diversity regularization in deep metric learning setting, and thus the learned kinship metric can be more discriminative and robust in measuring the genetic similarity on human faces.

A review of existing metric learning-based kinship verification methods is summarized in Table 1. The columns Deep, Coupled, and Compact represent if the learned similarity metric is a deep, coupled, and compact model, respectively.

3. Our approach

In this section, we first introduce the proposed KinNet architecture, and then elaborate our KML method with KinNet for kinship verification. Finally, we present the optimization algorithm to solve the KML problem.

The motivation figure of our proposed KML method is shown in Fig. 1. Suppose there is a quadruplet $(x_p, x_c, \hat{x}_p, \hat{x}_c)$ in the original metric space, where (x_p, x_c) are a pair of parent-child faces with kin relationship, and $\hat{x}_c$ and $\hat{x}_p$ are their nearest samples in the child and parent faces set, respectively. As illustrated in Fig. 1, the members of a quadruplet come from three different families, denoted by squares, circle, and triangle, respectively. The blue and green colors denote the parent and child faces, respectively. In the original metric space, there is large difference between the parent-child pair (x_p, x_c) in the square class due to large cross-generation discrepancy such as aging and sex difference. Often, there are some other parent and child faces lying in the neighborhoods of them in the square class, as shown in Fig. 1. As a result, there is a high chance to misclassify the faces in the neighborhoods under a single (shared) linear distance metric f. To address this issue, our KML aims to learn a coupled deep metric (f_p, f_c), under which facial images with kin relations are pulled as close as possible and those without kin relations (but with high appearance similarity) are pulled as far as possible.

3.1. KinNet

In recent years, CNN based deep learning methods have drawn increasing attention in computer vision, and they have been successfully applied in various visual applications [11,40,46–49]. For kinship verification, directly learning millions of parameters of a CNN from only a few hundreds labeled kin-face images is problematic. One feasible solution is that the internal layers of the CNN (*e.g.*, VGG [15]) can act as an initial image transformation for the kin-faces input. As a rich mid-level representation for face images, CNN can be pre-trained on a large scale face dataset (the source task) and then reused on other target task (*e.g.*, kinship verification from kin-face images). However, this is non-trivial as the relationship and distribution of the images between the source and target datasets can be quite different. The target task aims to investigate the kin relationship between two different visual entities (*e.g.*, father and daughter), and thus the inherent appearance gap of intra-class in kinship verification is significantly larger than that in face recognition. To address this issue, we design an adaptation network, named KinNet, to learn a robust kinship metric from kin-face dataset with limited amount of labeled data in supervised metric learning framework, such that the face images with kinship relation are pulled close and those without kin relation are pushed away.

As illustrated in Fig. 2, our KinNet consists of a pair of neural networks (Parent, Child) that are respectively designed for parent and children with different model parameters and outputs. This coupled architecture has explicitly considered the domain knowledge that the parent-child pair should not share the same image transformation due to the large age span and sex difference between them.

KinNet receives the quadruplets $(\hat{x}_c, x_p, x_c, \hat{x}_p)$ as input, which are generated from a mini-batch of the kin-face pairs (X_p, X_c). Here, (x_p, x_c) denotes a parent-child pair with kin relation, $\hat{x}_c \in X_c$ and $\hat{x}_p \in X_p$ denote the nearest neighbor of x_p and x_c, respectively. The output of the coupled network (Parent, Child) is fed into the *kinship metric* layer for loss evaluation. The kinship metric layer on the top is to evaluate the structured empirical loss of the quadruplets input. In our KinNet, it evaluates the violation of the genetic similarity constraint (GRC) for each quadruplet input, and then propagates discriminative (gradient) information back to the lower layers for update of the network parameters, so that the empirical loss is minimized under some regularization conditions. More specifically, by imposing intra-connection diversity and inter-connections consistency over the KinNet, a deep compact cross-generation metric (f_p, f_c) can be learned with limited amount of labeled kin-face data, under which the kin pairs (x_p, x_c) are pulled close, and the non-kin pairs $(x_p, \hat{x}_c)$ and $(\hat{x}_p, x_c)$ (especially with

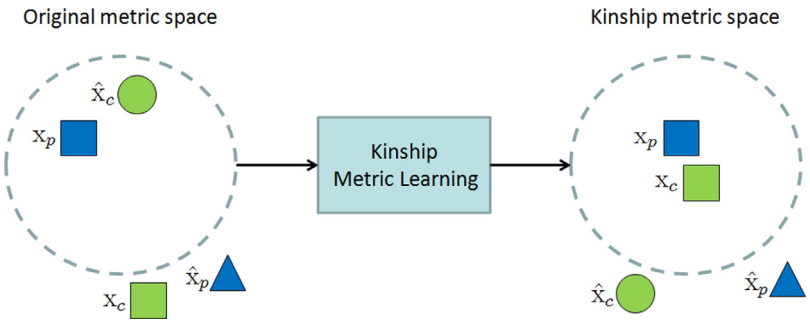

Fig. 1. The motivation figure of our proposed kinship metric learning (KML) method.

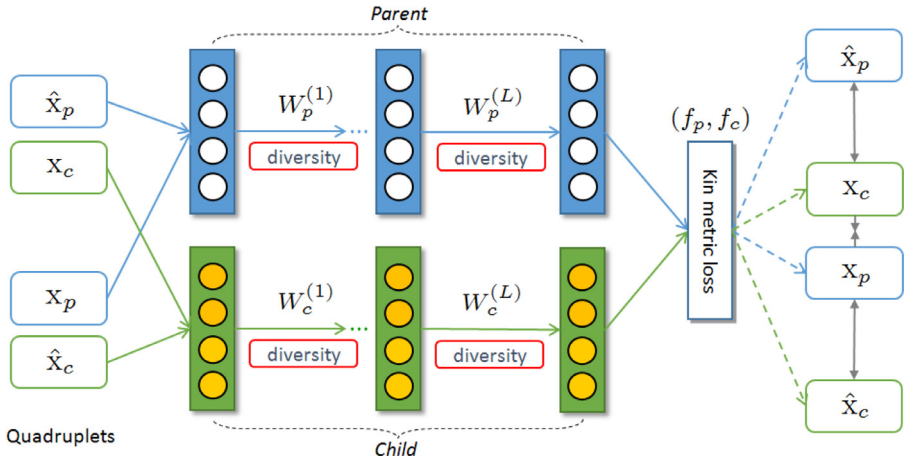

Fig. 2. The meta-view of our proposed KML method with a carefully designed deep architecture KinNet.

high appearance similarity in original metric space) are pushed away as far as possible.

3.2. Kinship metric learning

Existing DML-based kinship verification methods [7,27,28,44] seek to learn a Mahalanobis distance metric, such that the intra-class samples (with kin relation) are pulled close, and the inter-class samples (without kin relation) are pushed away as far as possible. Under the distance metric, the squared distance between a parent-child pair of face examples $(\mathrm{x}_p, \mathrm{x}_c)$ can be computed by

$$d_{\mathrm{M}}^2(\mathrm{x}_p, \mathrm{x}_c) = (\mathrm{x}_p - \mathrm{x}_c)^T \mathrm{M}(\mathrm{x}_p - \mathrm{x}_c) \tag{1}$$

where $\mathrm{x}_p, \mathrm{x}_c \in \mathbf{R}^D$, and $\mathrm{M} \succeq 0$ is a positive semidefinite (PSD) matrix. By factorizing M into $\mathrm{M} = \mathrm{W}^T\mathrm{W}$, the Mahalanobis distance can be written by $\|\mathrm{W}\mathrm{x}_p - \mathrm{W}\mathrm{x}_c\|^2$. This can be interpreted by first projecting the face example from the original feature space to a latent space $\mathbf{R}^d$ by using the linear transformation $\mathrm{W} \in \mathbf{R}^{d \times D}$, then measuring the squared Euclidean distance in the latent space. Different from most existing DML-based solutions that seek to linear transformations for feature mapping, we seek to learn a coupled nonlinear transformation to capture the nonlinear manifold where kin-face images may lie.

Let $\mathcal{T} = \{(\mathrm{x}_p^i, \mathrm{x}_c^i), i = 1, 2, ...N\}$ denote the training set of N pairs of images with kin relation, where x_p^i and x_c^i are the images from ith parent and child, respectively. Typically, most metric learning methods receive either pairwise or triplet input [20,60] for model training. For kinship verification problem, however, generating all possible pairs (or triplets) would result in a large number pairs (or triplets). In practice, most of them are too easy to distinguish and would not make any contribution to the loss convergence in training. To address this issue, our network takes quadruplets as input for kinship transformation learning. In particular, for a batch of N_b face pairs $\{(\mathrm{x}_p^i, \mathrm{x}_c^i), i = 1, 2, ...N_b\}$ sampled from $\mathcal{T}$, KN_b quadruplets $\{(\mathrm{x}_p^i, \mathrm{x}_c^i, \mathrm{x}_p^{ik}, \mathrm{x}_c^{ik}), i = 1, 2, ...N_b, k = 1, 2, ...K\}$ can be generated for stochastic gradient descent (SGD) based DNN training, where x_p^{ik} and x_c^{ik} respectively denote the k-nearest neighbor (kNN) of x_c^i and x_p^i, and K is the neighbor size. Under this setting, KML is formulated as the following optimization problem

$$\min_f J = \frac{1}{N_b}\sum_{i=1}^{N_b}\left(\ell_i + \frac{1}{K}\sum_{k=1}^{K}\ell_{ik} + \frac{1}{K}\sum_{k=1}^{K}\ell_{ki}\right) + \lambda r(f) \tag{2}$$

where $r(f)$ is the regularization term, λ is a trade-off parameter, and

$$\ell_i = \frac{1}{2}(\mathcal{S}_f(\mathrm{x}_p^i, \mathrm{x}_c^i) - 1)^2 \tag{3}$$

$$\ell_{ik} = \frac{1}{2}(\mathcal{S}_f(\mathrm{x}_p^i, \mathrm{x}_c^{ik}) + 1)^2 \tag{4}$$

$$\ell_{ki} = \frac{1}{2}(\mathcal{S}_f(\mathrm{x}_p^{ik}, \mathrm{x}_c^i) + 1)^2 \tag{5}$$

where the similarity score $\mathcal{S}_f$ for a parent-child pair $(\mathrm{x}_p, \mathrm{x}_c)$ is defined as the cosine of the angle between the transformed images

$$\mathcal{S}_f(\mathrm{x}_p, \mathrm{x}_c) = \frac{\langle f_p(\mathrm{x}_p), f_c(\mathrm{x}_c)\rangle}{\|f_p(\mathrm{x}_p)\|\|f_c(\mathrm{x}_c)\|} \tag{6}$$

where $f = (f_p, f_c)$ is the coupled deep transformation to be learned with KinNet. The objective in Eq. (2) enforces the GRC of kinship verification with a structured loss, which takes the quadruplets as input generated from a batch of face pairs. Under the coupled nonlinear transformation learned with KinNet, a positive pair of face images (with kin relation) are pulled as close as possible and those of the negative ones (without kin relation but with high appearance similarity) are pushed as far as possible, as illustrated in Fig. 2.

As illustrated in Fig. 2, the KinNet is a m-layers neural network with d_ℓ units at the ℓth layer, $\ell = 0, 1, \ldots, m$. It takes the output of any feature extractor (*e.g.*, LBP, HOG and VGG) as the initial transformation f_0 for the quadruplets input. For a image input x, the deep transformation $f^{(\ell)}(\mathrm{x})$ is parameterized by a set of weights and biases $(\mathrm{W}^{(j)}, \mathrm{b}^{(j)}), j = 1, 2, \ldots \ell$, where $\mathrm{W}^{(j)} \in \mathbf{R}^{d_j \times d_{j-1}}$. In particular, let $\mathrm{h}_p^{(\ell)}$ and $\mathrm{h}_c^{(\ell)}$ denote the output of the ℓth layer of our coupled network, respectively, *i.e.*,

$$\mathrm{h}_{pi}^{(\ell)} = f_p^{(\ell)}(\mathrm{x}_p^i) = \phi(\mathrm{W}_p^{(\ell)}\mathrm{h}_{pi}^{(\ell-1)} + \mathrm{b}_p^{(\ell)}) \tag{7}$$

$$\mathrm{h}_{ci}^{(\ell)} = f_c^{(\ell)}(\mathrm{x}_c^i) = \phi(\mathrm{W}_c^{(\ell)}\mathrm{h}_{ci}^{(\ell-1)} + \mathrm{b}_c^{(\ell)}) \tag{8}$$

$$\mathrm{h}_{pik}^{(\ell)} = f_p^{(\ell)}(\mathrm{x}_p^{ik}) = \phi(\mathrm{W}_p^{(\ell)}\mathrm{h}_{pik}^{(\ell-1)} + \mathrm{b}_p^{(\ell)}) \tag{9}$$

$$\mathrm{h}_{cik}^{(\ell)} = f_c^{(\ell)}(\mathrm{x}_c^{ik}) = \phi(\mathrm{W}_c^{(\ell)}\mathrm{h}_{cik}^{(\ell-1)} + \mathrm{b}_c^{(\ell)}) \tag{10}$$

where $\phi(\cdot)$ is an activation function. The goal of KML is then to learn the parameters $(\mathrm{W}_p^{(\ell)}, \mathrm{W}_c^{(\ell)}, \mathrm{b}_p^{(\ell)}, \mathrm{b}_c^{(\ell)})_{\ell=1}^m$ according to Eq. (2).

3.3. Consistency-diversity regularization

We now discuss the regularization term in Eq. (2). A commonly used solution to limit the model complexity is to use the ℓ_2-norm regularizer on $(\mathrm{W}_p^{(\ell)}, \mathrm{W}_c^{(\ell)})_{\ell=1}^m$ [20]. Due to the fact that the kinship dataset for model training is availability of only limited amount of data, we hope that the learned kinship metric is as *compact* as possible, so as to: (i) encourage the cross-generation correlation between W_p and W_c, and (ii) reduce overfitting in kinship metric learning. Motivated by the mutual

angular regularization on the latent variable models [54], we introduce the hierarchical compactness in our coupled network by enforcing the cross-generation consistency and hierarchical diversity on the coupled deep transformation (f_p, f_c):

$$r(f) = \frac{1}{m}\sum_{\ell=1}^{m}\left[\mathcal{D}_{\mathcal{B}}(\mathrm{W}_p^{(\ell)}, \mathrm{W}_c^{(\ell)}) - \Phi(\mathrm{W}_p^{(\ell)}) - \Phi(\mathrm{W}_c^{(\ell)})\right] \tag{11}$$

where $\mathcal{D}_{\mathcal{B}}(\mathrm{W}_p, \mathrm{W}_c)$ measures the cross-generation discrepancy over the coupled transformation (W_p, W_c), and $\Phi(\mathrm{W})$ measures the diversity of the latent factors of the weight matrix W. In this work, we adopt the Bregman divergence [61] to measure the cross-generation discrepancy:

$$\mathcal{D}_{\mathcal{B}}(\mathrm{W}_p, \mathrm{W}_c) = \mathcal{B}(\mathrm{W}_p) - \mathcal{B}(\mathrm{W}_c) - \langle \nabla\mathcal{B}(\mathrm{W}_c), \mathrm{W}_p - \mathrm{W}_c\rangle \tag{12}$$

where $\mathcal{B}(\cdot)$ is a continuously-differentiable and strictly convex function. In particular, let $\mathcal{B}(\mathrm{W}) = \mathrm{W}^T\mathrm{W}$, we have $\mathcal{D}_{\mathcal{B}}(\mathrm{W}_p, \mathrm{W}_c) = \|\mathrm{W}_p - \mathrm{W}_c\|_F^2$. The cross-generation consistency regularization penalizes the large discrepancy of the coupled transformation in a hierarchical manner. This is reasonable in kinship verification, as a coupled transformation with large difference would not capture the inherent correlation between the image pairs (with kinship relation), leading to degradation of the measurement ability in kinship verification.

By latent variable modeling [62], each row of the matrix $\mathrm{W}^{(\ell)}$ corresponds to a latent factor. The non-obtuse angle $\theta\left(\xi_i^{(\ell)}, \xi_j^{(\ell)}\right)$ between every pair of latent facotrs $\left(\xi_i^{(\ell)}, \xi_j^{(\ell)}\right)$ of $\mathrm{W}^{(\ell)}$ computed by

$$\theta\left(\xi_i^{(\ell)}, \xi_j^{(\ell)}\right) = \arccos\left(\frac{\left|\left\langle \xi_i^{(\ell)}, \xi_j^{(\ell)}\right\rangle\right|}{\|\xi_i^{(\ell)}\|\left\|\xi_j^{(\ell)}\right\|}\right) \tag{13}$$

can be employed for modeling the diversity of latent factors of $\mathrm{W}^{(\ell)}$, with the goal of reducing the redundancy of latent factors and improving the coverage of infrequent latent features and structures. The diversity (or compactness) of the weight matrix $\mathrm{W}^{(\ell)}$ is then defined as

$$\Phi(\mathrm{W}^{(\ell)}) = \bar{\theta}(\mathrm{W}^{(\ell)}) - \hat{\theta}(\mathrm{W}^{(\ell)}) \tag{14}$$

where $\bar{\theta}(\mathrm{W}^{(\ell)})$ and $\hat{\theta}(\mathrm{W}^{(\ell)})$ denote the mean and variance for all pairwise non-obtuse angles of the latent factors of $\mathrm{W}^{(\ell)}$, respectively, *i.e.*,

$$\bar{\theta}(\mathrm{W}^{(\ell)}) = \frac{2}{d_\ell(d_\ell-1)}\sum_{i=1}^{d_\ell-1}\sum_{j=i+1}^{d_\ell}\theta\left(\xi_i^{(\ell)}, \xi_j^{(\ell)}\right) \tag{15}$$

$$\hat{\theta}(\mathrm{W}^{(\ell)}) = \frac{2}{d_\ell(d_\ell-1)}\sum_{i=1}^{d_\ell-1}\sum_{j=i+1}^{d_\ell}\left[\theta\left(\xi_i^{(\ell)}, \xi_j^{(\ell)}\right) - \bar{\theta}(\mathrm{W}^{(\ell)})\right]^2 \tag{16}$$

We can observe that such a diversity regularization facilitates the latent factors to be evenly different from each other in latent spaces of each network layer, as illustrated in Fig. 3. Our KML *propagates* this compactness from bottom to top in a hierarchical manner, so as to reduce overfitting associated with the DNN while decreasing the number of hidden units required, which in effect decreases the computational cost and leads to faster network training and usage.

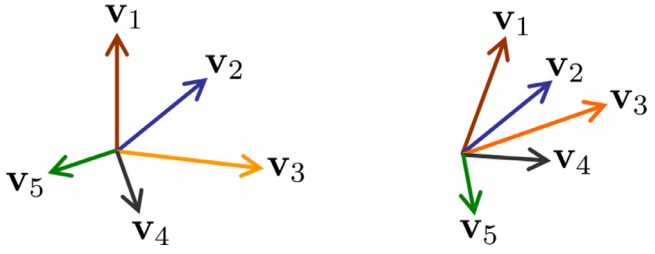

Fig. 3. An Illustration of the diversity of $\mathrm{W}^{(\ell)}$ at the ℓth network layer, $\ell = 1, 2, \ldots, m$. Compared to the latent factors (for example, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_5$) of the weight matrix (right), those of the weight matrix (left) are evenly different from each other in latent space, and hence they may possess higher diversity.

3.4. Optimization for KML

As the compactness regularizer $r(f)$ is non-convex and non-smooth, it is difficult to optimize the objective (2) directly. We develop an iterative algorithm to solve this problem. For each network layer ℓ, $\mathrm{W}^{(\ell)}$ is first factorized into $\mathtt{diag}(\mathrm{a}^{(\ell)})\widetilde{\mathrm{W}}^{(\ell)}$, where $\mathrm{a}^{(\ell)} \in \mathbf{R}^{d_\ell}$, and the ith element of $\mathrm{a}^{(\ell)}$ is the ℓ_2-norm of the ith row of $\mathrm{W}^{(\ell)}$. It is clear that the ℓ_2-norm of each column vector $\mathrm{w}_k^{(\ell)}$ of $\mathrm{W}^{(\ell)}$ is 1, for $k = 1, 2, \ldots, d_{\ell-1}$. Hence, the coupled transformation (f_p, f_c) can be rewritten by

$$f_p^{(\ell)}(\mathrm{x}_p) = \phi\left(\mathrm{diag}(\mathrm{a}_p^{(\ell)})\widetilde{\mathrm{W}}_p^{(\ell)}\mathrm{h}_p^{(\ell-1)} + \mathrm{b}_p^{(\ell)}\right) \tag{17}$$

$$f_c^{(\ell)}(\mathrm{x}_c) = \phi\left(\mathrm{diag}(\mathrm{a}_c^{(\ell)})\widetilde{\mathrm{W}}_c^{(\ell)}\mathrm{h}_c^{(\ell-1)} + \mathrm{b}_c^{(\ell)}\right) \tag{18}$$

for $\ell = 1, 2, \ldots, m$.

According to the compactness measure defined in (14), we have $\Phi(\mathrm{W}^{(\ell)}) = \Phi(\widetilde{\mathrm{W}}^{(\ell)})$, and moreover, as justified in [54], the task of maximizing $\Phi(\widetilde{\mathrm{W}}^{(\ell)})$ can be relaxed by maximizing a smooth lower bound $\Psi(\widetilde{\mathrm{W}}^{(\ell)})$

$$\Psi(\widetilde{\mathrm{W}}^{(\ell)}) = \arcsin(\sqrt{\widehat{\mathrm{W}}}) - \left(\frac{\pi}{2} - \arcsin\left(\sqrt{\widehat{\mathrm{W}}}\right)\right)^2 \tag{19}$$

of $\Phi(\widetilde{\mathrm{W}}^{(\ell)})$, where $\widehat{\mathrm{W}} = \det\left(\widetilde{\mathrm{W}}^{(\ell)T}\widetilde{\mathrm{W}}^{(\ell)}\right)$. Hence, the regularization term can be rewritten as

$$r(f) = \frac{1}{m}\sum_{\ell=1}^{m}\left[\left\|\mathrm{diag}(\mathrm{a}_p^{(\ell)})\widetilde{\mathrm{W}}_p^{(\ell)} - \mathrm{diag}(\mathrm{a}_c^{(\ell)})\widetilde{\mathrm{W}}_c^{(\ell)}\right\|_F^2 - \Psi\left(\widetilde{\mathrm{W}}_p^{(\ell)}\right) - \Psi\left(\widetilde{\mathrm{W}}_c^{(\ell)}\right)\right] \tag{20}$$

which is now smooth and convex with respect to $\widetilde{\mathrm{W}}_p^{(\ell)}$ (or $\widetilde{\mathrm{W}}_c^{(\ell)}$).

We use SGD to solve the optimization problem (2). During each iteration of SGD, a mini-batch of N_b positive pairs are sampled from the kinship training set, and then the quadruplets are generated from the mini-batch for model training. Forward propagation is performed to compute the hierarchical outputs $(\mathrm{h}_p^{(\ell)}, \mathrm{h}_c^{(\ell)})_{\ell=1}^m$. Then, the gradients are computed using back propagation. In back propagation, the model parameters are optimized by alternating between $\left(\widetilde{\mathrm{W}}_p^{(\ell)}, \widetilde{\mathrm{W}}_c^{(\ell)}, \mathrm{b}_p^{(\ell)}, \mathrm{b}_c^{(\ell)}\right)_{\ell=1}^m$ and $(\mathrm{a}_p^{(\ell)}, \mathrm{a}_c^{(\ell)})_{\ell=1}^m$, *i.e.*,

(1) With $\left(\widetilde{\mathrm{W}}_p^{(\ell)}, \widetilde{\mathrm{W}}_c^{(\ell)}, \mathrm{b}_p^{(\ell)}, \mathrm{b}_c^{(\ell)}\right)_{\ell=1}^m$ fixed, optimizing $\mathrm{a}_p^{(\ell)}$ and $\mathrm{a}_c^{(\ell)}$, for $\ell = 1, 2, \ldots, m$.

$$\frac{\partial J}{\partial \mathrm{a}_o^{(m)}} = \sum_i\left(\frac{\partial J}{\partial \mathrm{h}_{oi}^{(m)}}\frac{\partial \mathrm{h}_{oi}^{(m)}}{\partial \mathrm{a}_o^{(m)}} + \sum_k\frac{\partial J}{\partial \mathrm{h}_{oik}^{(m)}}\frac{\partial \mathrm{h}_{oik}^{(m)}}{\partial \mathrm{a}_o^{(m)}}\right) + \lambda\frac{\partial r}{\partial \mathrm{a}_o^{(m)}} \tag{21}$$

$$\frac{\partial J}{\partial \mathrm{a}_o^{(\ell)}} = \sum_i\left(\frac{\partial J}{\partial \mathrm{h}_{oi}^{(\ell+1)}}\frac{\partial \mathrm{h}_{oi}^{(\ell+1)}}{\partial \mathrm{h}_{oi}^{(\ell)}}\frac{\partial \mathrm{h}_{oi}^{(\ell)}}{\partial \mathrm{a}_o^{(\ell)}} + \sum_k\frac{\partial J}{\partial \mathrm{h}_{oik}^{(\ell+1)}}\frac{\partial \mathrm{h}_{oik}^{(\ell+1)}}{\partial \mathrm{h}_{oik}^{(\ell)}}\frac{\partial \mathrm{h}_{oik}^{(\ell)}}{\partial \mathrm{a}_o^{(\ell)}}\right) + \lambda\frac{\partial r}{\partial \mathrm{a}_o^{(\ell)}},\ \forall\ \ell < m \tag{22}$$

$$\mathrm{a}_o^{(\ell)} = \mathrm{a}_o^{(\ell)} - \gamma\frac{\partial J}{\partial \mathrm{a}_o^{(\ell)}},\ \forall\ \ell = 1, 2, \ldots, m \tag{23}$$

where $o \in \{p, c\}$, and γ is the learning rate.

(2) With $(\mathrm{a}_p^{(\ell)}, \mathrm{a}_c^{(\ell)})_{\ell=1}^m$ fixed, optimizing $\widetilde{\mathrm{W}}_p^{(\ell)}, \widetilde{\mathrm{W}}_c^{(\ell)}, \mathrm{b}_p^{(\ell)}, \mathrm{b}_c^{(\ell)}$, for $\ell = 1, 2, \ldots, m$.

$$\frac{\partial J}{\partial \widetilde{\mathrm{W}}_o^{(m)}} = \sum_i \left(\frac{\partial J}{\partial \mathrm{h}_{oi}^{(m)}} \frac{\partial \mathrm{h}_{oi}^{(m)}}{\partial \widetilde{\mathrm{W}}_o^{(m)}} + \sum_k \frac{\partial J}{\partial \mathrm{h}_{oik}^{(m)}} \frac{\partial \mathrm{h}_{oik}^{(m)}}{\partial \widetilde{\mathrm{W}}_o^{(m)}} \right) + \lambda \frac{\partial r}{\partial \widetilde{\mathrm{W}}_o^{(m)}} \tag{24}$$

$$\frac{\partial J}{\partial \mathrm{b}_o^{(m)}} = \sum_i \left(\frac{\partial J}{\partial \mathrm{h}_{oi}^{(m)}} \frac{\partial \mathrm{h}_{oi}^{(m)}}{\partial \mathrm{b}_o^{(m)}} + \sum_k \frac{\partial J}{\partial \mathrm{h}_{oik}^{(m)}} \frac{\partial \mathrm{h}_{oik}^{(m)}}{\partial \mathrm{b}_o^{(m)}} \right) \tag{25}$$

$$\frac{\partial J}{\partial \widetilde{\mathrm{W}}_o^{(\ell)}} = \sum_i \left(\frac{\partial J}{\partial \mathrm{h}_{oi}^{(\ell+1)}} \frac{\partial \mathrm{h}_{oi}^{(\ell+1)}}{\partial \mathrm{h}_{oi}^{(\ell)}} \frac{\partial \mathrm{h}_{oi}^{(\ell)}}{\partial \widetilde{\mathrm{W}}_o^{(\ell)}} + \sum_k \frac{\partial J}{\partial \mathrm{h}_{oik}^{(\ell+1)}} \frac{\partial \mathrm{h}_{oik}^{(\ell+1)}}{\partial \mathrm{h}_{oik}^{(\ell)}} \frac{\partial \mathrm{h}_{oik}^{(\ell)}}{\partial \widetilde{\mathrm{W}}_o^{(\ell)}} \right) + \lambda \frac{\partial r}{\partial \widetilde{\mathrm{W}}_o^{(\ell)}}, \forall\, \ell < m \tag{26}$$

$$\frac{\partial J}{\partial \mathrm{b}_o^{(\ell)}} = \sum_i \left(\frac{\partial J}{\partial \mathrm{h}_{oi}^{(\ell+1)}} \frac{\partial \mathrm{h}_{oi}^{(\ell+1)}}{\partial \mathrm{h}_{oi}^{(\ell)}} \frac{\partial \mathrm{h}_{oi}^{(\ell)}}{\partial \mathrm{b}_o^{(\ell)}} + \sum_k \frac{\partial J}{\partial \mathrm{h}_{oik}^{(\ell+1)}} \frac{\partial \mathrm{h}_{oik}^{(\ell+1)}}{\partial \mathrm{h}_{oik}^{(\ell)}} \frac{\partial \mathrm{h}_{oik}^{(\ell)}}{\partial \mathrm{b}_o^{(\ell)}} \right), \forall\, \ell < m \tag{27}$$

$$\widetilde{\mathrm{W}}_o^{(\ell)} = \widetilde{\mathrm{W}}_o^{(\ell)} - \gamma \frac{\partial J}{\partial \widetilde{\mathrm{W}}_o^{(\ell)}},\ \forall\, \ell = 1, 2, \ldots, m \tag{28}$$

$$\mathrm{b}_o^{(\ell)} = \mathrm{b}_o^{(\ell)} - \gamma \frac{\partial J}{\partial \mathrm{b}_o^{(\ell)}},\ \forall\, \ell = 1, 2, \ldots, m \tag{29}$$

where $o \in \{p, c\}$.

As the ℓ_2-norm of each column vector of $\widetilde{\mathrm{W}}_p^{(\ell)}$ (and $\widetilde{\mathrm{W}}_c^{(\ell)}$) is required to be 1 in optimization, a projection operation is needed after each gradient descent of $\widetilde{\mathrm{W}}_p^{(\ell)}$ (and $\widetilde{\mathrm{W}}_c^{(\ell)}$)

$$\widetilde{\mathrm{W}}_o^{(\ell)} \xrightarrow{\text{decomposition}} \mathrm{diag}(\mathrm{a}_o)\mathrm{W}_o \tag{30}$$

$$\widetilde{\mathrm{W}}_o^{(\ell)} \leftarrow \mathrm{W}_o \tag{31}$$

where $o \in \{p, c\}$, and $\ell = 1, 2, \ldots, m$.

In KML, network weights $(\mathrm{W}_p^{(\ell)}, \mathrm{W}_c^{(\ell)})_{\ell=1}^m$ of the KinNet are initialized by randomly sampling from a Gaussian distribution $\mathcal{N}(0, \sigma_w^2)$ with zero mean and standard deviation $\sigma_w = 10^{-2}$. The biases $\mathrm{b}_p^{(\ell)}$ and $\mathrm{b}_c^{(\ell)}$ are initialized to zero, for $\ell = 1, 2, \ldots, m$. The proposed KML algorithm is summarized in Algorithm 1.

Input: Training set: $\mathcal{T}$; parameters: $N_b, m, K, \lambda, \gamma, \sigma_w$, and T.
Output: Deep kinship transformation (f_p, f_c).
/* *Initialization* */.
$(f_p^{(0)}, f_c^{(0)}) \leftarrow (f_0, f_0)$;
for $\ell = 1$ to m **do**
  $(\mathrm{W}_p^{(\ell)}, \mathrm{W}_c^{(\ell)}) \sim \mathcal{N}(0, \sigma_w^2)$;
  $\mathrm{b}_p^{(\ell)} \leftarrow \mathbf{0}$, $\mathrm{b}_c^{(\ell)} \leftarrow \mathbf{0}$;
end
/* *Optimization by SGD* */.
for $t = 1$ to T **do**
  Random sampling: $\mathcal{D}_t = (\mathrm{x}_p^i, \mathrm{x}_c^i)_{i=1}^{N_b} \sim \mathcal{T}$;
  Quadruplets generating: $\mathcal{Q}_t = (\mathrm{x}_p^i, \mathrm{x}_c^i, \mathrm{x}_p^{ik}, \mathrm{x}_c^{ik}) \sim \mathcal{D}_t$;
  /* *Forward propagation* */.
  for $\ell = 1$ to m **do**
    Compute $(h_p^{(\ell)}, h_c^{(\ell)})$ according to Eq.(7)~(10);
  end
  /* *Backward propagation* */.
  while (not convergence) **do**
    /* *With* $(\widetilde{\mathrm{W}}_p^{(\ell)}, \widetilde{\mathrm{W}}_c^{(\ell)}, \mathrm{b}_p^{(\ell)}, \mathrm{b}_c^{(\ell)})_{\ell=1}^m$ *fixed* */
    for $\ell = m$ downto 1 **do**
      Compute the gradients according to Eq.(21)~(22);
    end
    Update $(\mathrm{a}_p^{(\ell)}, \mathrm{a}_c^{(\ell)})_{\ell=1}^m$ by Eq.(23);
    /* *With* $(\mathrm{a}_p^{(\ell)}, \mathrm{a}_c^{(\ell)})_{\ell=1}^m$ *fixed* */
    for $\ell = m$ downto 1 **do**
      Compute the gradients according to Eq.(24)~(27);
    end
    Update $(\widetilde{\mathrm{W}}_p^{(\ell)}, \widetilde{\mathrm{W}}_c^{(\ell)}, \mathrm{b}_p^{(\ell)}, \mathrm{b}_c^{(\ell)})_{\ell=1}^m$ by Eq.(28)~(29);
    /* *Projection* */
    Projection for $(\widetilde{\mathrm{W}}_p^{(\ell)}, \widetilde{\mathrm{W}}_c^{(\ell)})_{\ell=1}^m$ by Eq.(30)~(31);
  end
end

Algorithm 1. Kinship Metric Learning (KML).

4. Experiments

To evaluate the effectiveness of our proposed kinship verification method, we conduct experiments on four widely used datasets: KinFaceW-I[1], KinFaceW-II[2], Cornell KinFace[3], and UB KinFace[4]. Fig. 4 presents some sample kin pairs from the KinFaceW-II dataset. We elaborate the datasets, experimental settings, results and analysis in what follows.

4.1. Datasets and experimental settings

Four different kinship relations: mother–son (M–S), mother–daughter (M–D), father–son (F–S) and father–daughter (F–D), are included in each of the four datasets. More specifically, there are 127, 116, 134, and 156 pairs of parent-child face images for these four relations in KinFaceW-I, respectively. There are 250 pairs of parent-child images for each kin relation in KinFaceW-II. There are 150 pairs of parent-child images in the Cornell KinFace dataset, where 13%, 25%, 40%, and 22% of them are with the M–S, M–D, F–S, and F–D kin relations, respectively. There are 600 face images of 200 groups in the UB KinFace dataset, and each group contains three face images corresponding to the child, old parent, and young parent, respectively. Two subsets of the UB KinFace, UB-1 and UB-2, are created in our experiments: 200 child and 200 young parent for UB-1, and 200 child and 200 old parent for UB-2. Since near 80% of the face images are with F-S relation in the UB KinFace, we have not further divided the subsets into different kinship relations. Each face image in the four datasets is

Fig. 4. Sample kin pairs (aligned and cropped) from the KinFaceW-II dataset [7]. From top to bottom are father–son (F–S), father–daughter (F–D), mother–son (M–S), and mother–daughter (M–D) kinship relations, respectively.

[1] http://www.kinfacew.com.
[2] http://www.kinfacew.com.
[3] http://chenlab.ece.cornell.edu/projects/kinshipverification.
[4] http://www.ece.neu.edu/~yunfu/research/kinface/kinface.htm.

aligned and cropped of size 64 × 64 according to the eye positions. We then extract following features for different kinship verification methods:

LBP [63]: Each face image is first divided into 8 × 8 non-overlapping blocks of size 8 × 8, and we then extract a 59-dimensional uniform LBP feature for each block. These features are finally concatenated into a 3776-dimensional vector.

HOG [64]: We first divide each face image into 16 × 16 non-overlapping blocks of size 4 × 4, and then divide it into non-overlapping blocks of size 8 × 8. We obtain a 9-dimensional HOG feature for each block and they are then concatenated into a 2880-dimensional vector.

SIFT [65]: We first densely sampled and computed one 128-dimensional feature for each 16 × 16 patch of the image, where the spacing of the two neighboring patches is 8 pixels. Then, these SIFT features are concatenated into a 6272-dimensional vector.

VGG [15]: We employ the CNN architecture of VGG by removing the FC8 layer for the mid-level representation of kin faces. This network consists of five successive convolutional layers followed by two fully connected layers. As an initial feature transformation for our KML, the network takes as input a 224 × 224 face image and outputs a 4096-dimensional vector. Please refer to Parkhi et al. [15] for detailed description of the geometry of VGG Face.

There are two training settings for supervised learning on these kinship datasets: image restricted and image unrestricted. For fair comparison with other DML-based solutions to kinship verification [7,20,27,28], we follow the image unrestricted setting in our experiments, where the identity information of the person is available to potentially form additional negative pairs in the training set. We perform five-fold cross-validation in the experiments, and the parameters of our KML are carefully tuned on a subset of the KinFaceW-II dataset, since this is the largest one among the four datasets. For KML, total number of the layers of KinNet is empirically set to 3 (*i.e.*, $m = 2$) to prevent overfitting with limited number of training data, and the two sub-networks of KinNet are both of size (300, 100). The parameters N_b, K, λ, γ, σ_w are empirically set to 20, 3, 1.0, 0.01, and 0.01, respectively.

4.2. Results and analysis

4.2.1. Comparisons with the state-of-the-art kinship verification methods

To evaluate the verification performance of our method, we first compare our KML with several state-of-the-art kinship verification methods (8 DML-based and 2 feature learning-based) presented in the past a few years: TSL [35], NRML [7], DMML [28], LM^{3}L [27], ESL [44], DDML [37], SaPL [59], HDL [36], PDFL [33] and KVRL-fcDBN [34]. Note that most of them are implemented based on different feature descriptors, and we report their best results in the experiments. Table 2 tabulates the mean verification rate of the different methods on four kinship datasets. We notice that a hierarchical representation learning (KVRL) method proposed by Kohli et al. [34] achieved the best accuracy performance so far on several benchmark datasets. As shown in Table 2, our method achieves the second best accuracy performance and outperforms other metric learning based solutions on the benchmark datasets. More specifically, on KinFaceW-I, KML outperforms the other DML-based methods with the lowest gains in mean verification rate of 3.1%. On KinFaceW-II, KML achieves the lowest gains of 2.0%. As for the Cornell KinFace, KML achieves the lowest gains of 4.7%. On the UB KinFace, KML achieves the lowest gains of 3.2%. The ROC curves for our KML method are shown in Fig. 5.

Table 2
Mean verification rate (%) of the state-of-the-art kinship verification methods (including 9 DML-based and 2 feature learning-based methods) on the four kinship datasets.

Method	KinFaceW-I	KinFaceW-II	Cornell KinFace	UB KinFace
TSL [35]	N.A.	N.A.	N.A.	56.5
NRML [7]	77.5	74.7	71.6	67.1
DMML [28]	72.3	78.3	73.7	72.3
LM^{3}L [27]	73.5	78.7	N.A.	N.A.
ESL [44]	78.6	75.7	73.0	72.1
DDML [37]	78.8	80.4	76.7	71.6
SaPL [59]	78.3	79.0	N.A.	N.A.
HDL [36]	79.7	83.7	N.A.	N.A.
KML (Proposed)	**82.8**	**85.7**	**81.4**	**75.5**
PDFL [33]	70.1	77.0	71.9	67.3
KVRL-fcDBN [34]	**96.1**	**96.2**	**89.5**	**91.8**

Compared with other eight DML-based solutions to kinship verification, the superior verification performance of our KML mainly comes from two aspects: (*i*) Rich mid-level face representation learned with large-scale face datasets (*e.g.*, VGG) is transferred to kinship verification task with limited amount of training data, and (*ii*) KML takes into account the domain knowledge encoded in kinship data and explicitly learns cross-generation deep transformation in a compact and hierarchical manner, and thus it can learn more robust kinship similarity metric on human faces.

Finally, note that our method is focused on the kinship metric learning, and we believe it is complementary to most existing feature learning-based methods (e.g., KVRL in Ref. [34]) for kinship verification.

4.2.2. Comparisons with different feature descriptors

To investigate the effectiveness of the proposed parameters (feature) transferring in our KML, different feature descriptors (LBP, HOG, SIFT, and VGG) mentioned above are employed as the initial image transformation f_0 for KML in the experiments. For the hand-crafted descriptors LBP, HOG, and SIFT, the whitened PCA is employed to project it to a 500-dimensional vector for redundancy removal. Tables 3–6 summarize the verification rate of KML using different feature descriptors (as f_0) on the KinFaceW-I, KinFaceW-II, Cornell KinFace, and UB KinFace kinship datasets, respectively. From Tables 2–6 we make the following observations:

(*a*) The parameters transferring (from VGG Face) in KML outperforms the best local descriptors with the lowest gain in mean verification rate of 3.9%, 4.7%, 6.7% and 2.3% on the KinFaceW-I, KinFaceW-II, Cornell KinFace, and UB KinFace datasets, respectively. The results indicate that, despite the differences in image statistics and tasks between face recognition and kinship verification, the transferred mid-level representation from large-scale face dataset combined with our KML leads to significantly improved accuracy for kinship verification.

(*b*) Our KML can achieve significantly better verification rate than the state-of-the-art DML-based kinship verification methods. The superiority of KML can be attributed to the deep coupled and compact property of the learned similarity metric.

4.2.3. Comparisons with different metric learning strategies

The core component of KML is a metric learning algorithm. To further validate the effectiveness of our KML method, we conduct experimental comparisons with the following metric (transformation) learning strategies in kinship verification:

- Shallow kinship metric learning (S-KML), a variant of our KML with a single network layer.
- Common kinship metric learning (C-KML), a variant of KML with a common transformation (instead of the coupled one) shared by the parent and child. This indicates that the networks P and C share a set of common parameters in C-KML.
- Kinship metric learning without the hierarchical compactness regularization (KML-C). KML-C instead imposes the Frobenius-norm regularization on network parameters: $r(f) = \frac{1}{m}\sum_{\ell=1}^{m}\left(\|\mathrm{W}_p^{(\ell)}\|_F^2 + \|\mathrm{W}_c^{(\ell)}\|_F^2\right)$.
- NRML [7], LMNN [25] and ITML [26] with deep face

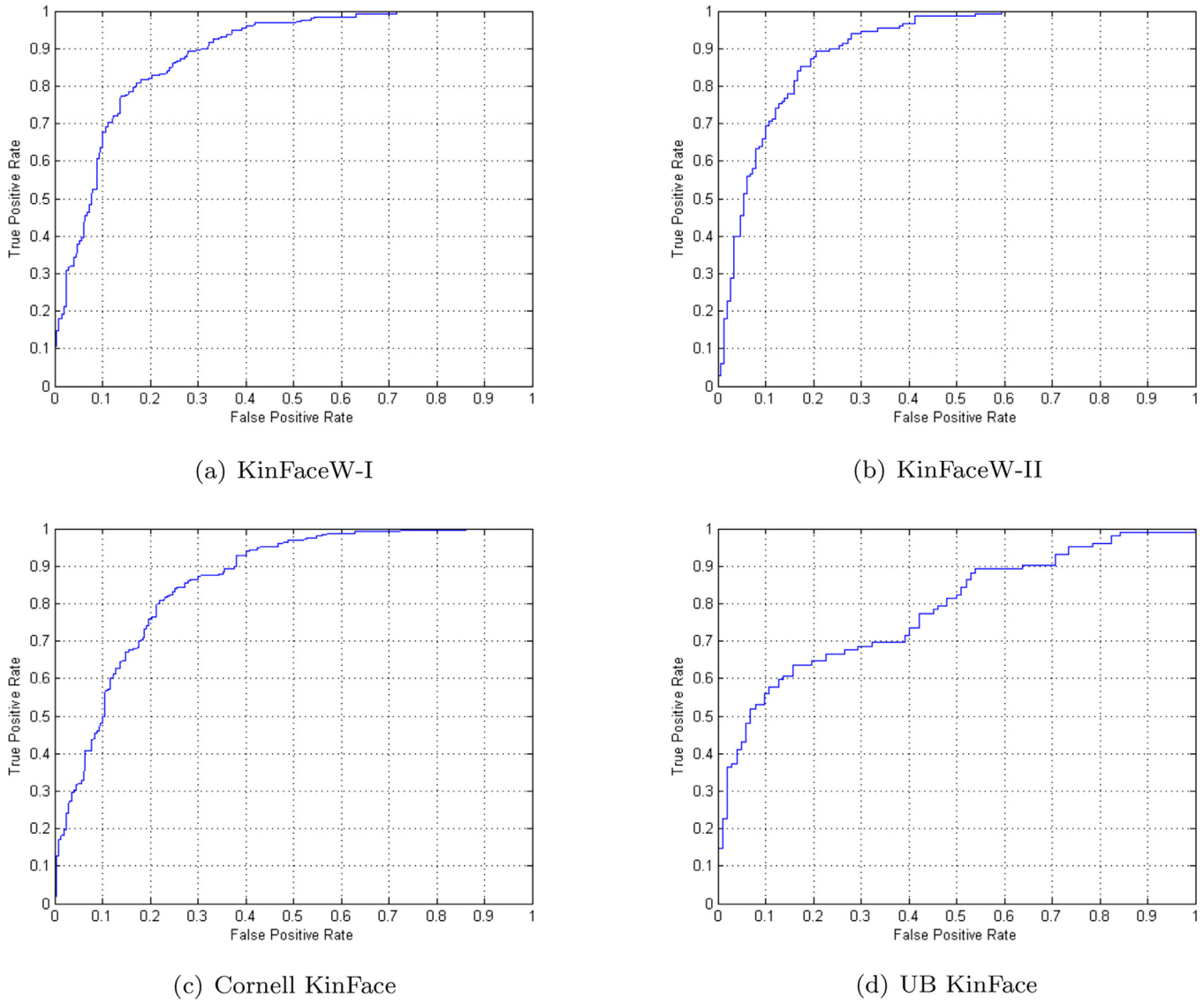

Fig. 5. The ROC curves for the proposed KML method on the four datasets.

Table 3
Mean verification rate (%) of KML using different feature descriptors on the KinFaceW-I dataset.

Feature	F–S	F–D	M–S	M–D	Mean
LBP	81.7	73.8	74.1	76.8	76.6
HOG	82.7	75.3	77.1	80.3	78.9
SIFT	82.7	75.4	76.2	80.7	78.7
VGG	**83.8**	**81.0**	**81.2**	**85.0**	**82.8**

Table 4
Mean verification rate (%) of KML using different feature descriptors on the KinFaceW-II dataset.

Feature	F–S	F–D	M–S	M–D	Mean
LBP	82.8	79.6	81.2	78.8	80.6
HOG	85.2	77.8	79.2	79.0	80.3
SIFT	82.4	80.2	82.0	79.2	81.0
VGG	**87.4**	**83.6**	**86.2**	**85.6**	**85.7**

Table 5
Mean verification rate (%) of KML using different feature descriptors on the Cornell KinFace dataset.

Feature	F–S	F–D	M–S	M–D	Mean
LBP	74.7	73.9	78.0	70.0	74.1
HOG	75.5	72.7	71.3	77.3	74.2
SIFT	74.7	72.5	74.6	77.0	74.7
VGG	**78.9**	**82.6**	**78.3**	**85.7**	**81.4**

Table 6
Mean verification rate (%) of KML using different feature descriptors on the UB KinFace dataset.

Feature	UB-1	UB-2	Mean
LBP	70.0	72.0	71.0
HOG	74.7	71.7	73.2
SIFT	71.5	70.2	70.8
VGG	**75.8**	**75.2**	**75.5**

Table 7
Mean verification rate (%) of different metric learning algorithms on four kinship datasets (using VGG-Face feature).

Method	KinFaceW-I	KinFaceW-II	Cornell KinFace	UB KinFace
NRML [7]	77.5	77.0	73.2	69.0
LMNN [25]	75.4	76.0	71.1	66.3
ITML [26]	75.9	76.3	71.7	65.8
CDML [54]	79.1	79.3	75.8	70.4
CMML [55]	77.9	76.7	74.0	68.8
DDML [37]	78.8	80.4	76.7	71.6
DCML [58]	80.3	81.8	79.0	72.9
S-KML baseline	79.4	79.6	76.2	71.0
C-KML baseline	80.7	82.6	79.4	73.5
KML-C baseline	80.5	82.1	79.2	73.2
KML	**82.8**	**85.7**	**81.4**	**75.5**

representation. All of them are implemented by using the VGG Face [15] as the feature extractor for kin faces.

- Compact distance metric learning (CDML) [54]. It uses the VGG as the feature extractor for kin faces input, and learns a linear distance metric with diversity regularization.
- Cross modal metric learning (CMML) [55]. It uses the VGG as the feature extractor for kin faces input.
- Discriminative deep metric learning (DDML) [20] and deep coupled metric learning (DCML) [58]. Both DDML and DCML are implemented using VGG as the initial feature extractor f_0 for kin faces. For fair comparison, DDML, DCML and our KML adopt the same parameters m (number of the network layers) and d_ℓ (number of neural units at each layer ℓ) in training.

Experimental results of the different algorithms using different metric learning strategies on four datasets are shown in Tables 7. We can see from the table that:

(*a*) With the same deep face representation (*i.e.*, VGG [15]), our KML achieves significantly better mean verification rate than NRML, LMNN, ITML, S-KML, CMML and CDML on all the four kinship datasets. The results indicated that, even having employed the deep representation for kin faces, learning a deep transformation (metric) instead of the linear distance metric can further boost the performance of kinship verification in practice.

(*b*) KML outperforms C-KML with the gain in mean verification rate of 2.1%, 3.1%, 2.0% and 2.0% on the KinFaceW-I, KinFaceW-II, Cornell KinFace, and UB KinFace datasets, respectively. This further indicated the benefit of jointly learning the coupled transformation in KML, which explicitly exploited the domain knowledge that the parent-child pair should not share the same image transformation due to the inherent large age span and sex difference between them.

(*c*) Our KML achieves better verification performance than DCML and KML-C, with the lowest gain in mean verification rate of 2.3%, 3.6%, 2.2% and 2.3% on the KinFaceW-I, KinFaceW-II, Cornell KinFace, and UB KinFace datasets, respectively. The result validated the effectiveness of hierarchical compactness imposed on the coupled DNN by latent factor modeling in KML.

d) With the same deep face representation (VGG [15]), our KML significantly outperforms DDML with the gain in mean verification accuracy of 4.0%, 5.3%, 4.7% and 3.9% on the KinFaceW-I, KinFaceW-II, Cornell KinFace, and UB KinFace datasets, respectively. This implied the benefit of the coupled structure and compactness property of the learned deep kinship transformation.

4.2.4. Parameters analysis

We first investigate the impact of the regularization parameter λ on the verification performance of KML. We conduct experiments on the four kinship datasets, and the mean verification rate of KML versus different value of λ is illustrated in Fig. 6. From the figure we can see that, the mean verification rate is improved as λ increases in the initial stage; however, continuing to increase λ leads to decline of the verification performance. The reason is two-fold. On one hand, a larger λ often enforces the latent factors of the network weights to be more diverse and uncorrelated, and hence difference aspects of the genetic traits on human face can be captured in a more comprehensive manner. On the other hand, if the trade-off parameter λ is too large, the total loss in the objective (2) is dominated by the compactness (diversity) regularization, and thus the supervised information for kinship verification would not be effectively exploited in deep transformation learning.

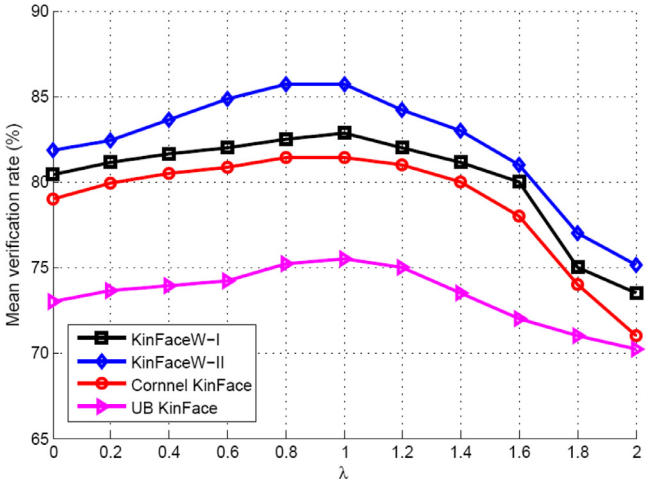

Fig. 6. Sensitivity of KML to the compactness regularization parameter λ on the four benchmark datasets.

As KML falls to the class of deep metric learning algorithm [20], we also investigate how the parameters $\{d_\ell, \ell = 1, 2\}$ affect the verification performance of KML, where d_ℓ denotes the number of the neural units at the ℓth-layer of the adaptation sub-network, and it also represents the number of the latent factors of the ℓth-layer weights. In principle, a small number of neural units (hence latent factors) may not well capture the inherent genetic similarity on human faces, while a large d_ℓ would also degrade the kinship similarity metric due to high risk of the model overfitting to limited amount of training data. Taking KinFaceW-II as a benchmark, we evaluate the mean verification rate of KML for different (d_1, d_2) in [50,500] with step-size 50, and we observe that best mean verification performance can be achieved when (d_1, d_2) is set to (300,100) in the experiment.

Since the KinNet architecture of KML receives quadruplets as input that are generated from the mini-batch of positive pairs by KNN search, we investigate the impact of the parameter K on the verification performance of KML. Fig. 7 presents the mean verification rate of KML versus different number of K on the KinFaceW-II dataset. We can

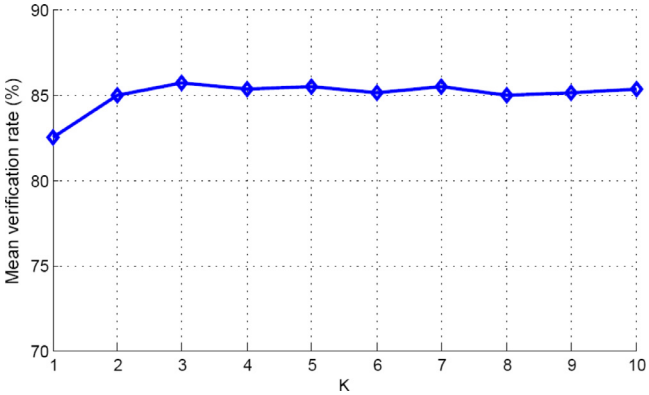

Fig. 7. Mean verification rate of KML versus different neighbor size K on the KinFaceW-II dataset.

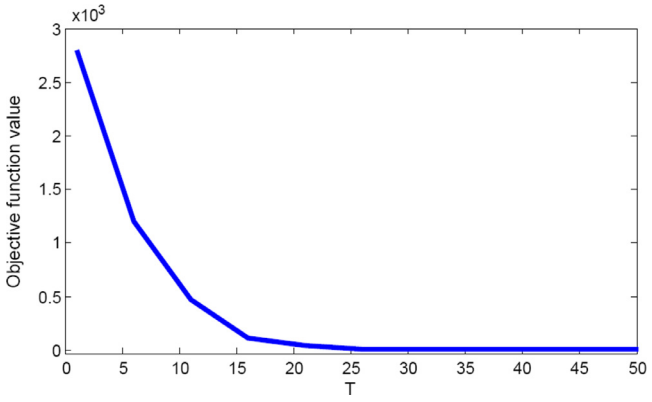

Fig. 8. Convergence curve of KML with the iteration number T on the KinFaceW-II dataset.

observe that the stable verification performance can be achieved when K is set to 3 in the experiment. This also implied that generating too many possible pairs (or triplets) from the mini-batch of positive pairs may not improve the verification performance in practice, as most of them are too easy to distinguish and would not make any contribution to the loss convergence in training.

Finally, we investigate the convergence of KML by evaluating the objective function of KML with respect to the number of iterations T. Fig. 8 presents the objective function value of KML versus different iteration number T on the KinFaceW-II dataset. It can be seen from the figure that the total loss of KML quickly decreases in the early several iterations and converges in 30 ~ 40 iterations on the dataset.

4.2.5. Discussion

We have evaluated the effectiveness of our proposed KML method in the above experiments, and we can make the following key observations:

(*a*) Despite the differences in image statistics and recognition tasks between face recognition and kinship verification, rich mid-level representation transferred from large-scale face dataset (*e.g.*, VGG Face) in our KML can achieve better performance than hand-craft descriptors for kinship verification.
(*b*) Compared to linear distance metric, nonlinear similarity metric learned by the coupled DNN can be more robust in genetic similarity measure, as it models the nonlinear nature of the real-world kin-faces distribution.
(*c*) For the coupled DNN, hierarchical compactness (*i.e.*, intra-connection diversity and inter-connection consistency) imposed on the network connections can be helpful to prevent overfitting in learning the deep similarity metric with limited amount of kinship data.
(*d*) Most recently, representation learning with deep network for kinship verification has demonstrated impressive performance [34,40]. It should be noted that the KML method proposed in this work focuses on the distance metric learning rather than feature learning, and we believe that our KML is complementary to these sophisticated representation learning method in kinship verification.

5. Conclusion

We have presented in this paper a kinship metric learning method to address kinship verification using facial images. We have shown that, despite the differences in image statistics and tasks between the datasets for face recognition and kinship verification, the transferred deep face representation leads to significantly improved accuracy in kinship verification. Also, by learning a coupled and deep compact similarity metric with the KinNet architecture tailored for kinship verification problem, our proposed KML possesses some desirable properties that help address the limitations of most existing solutions to kinship verification. Experimental results have shown that our proposed method significantly boosts the current state-of-the-art level of DML-based kinship verification.

Investigation of deep kinship metric learning with the sophisticated deep kin-face representation [34,40] and the large-scale kinship datasets (*e.g.*, FIW [66]) to further improve the kinship verification performance appears to be an interesting direction of future work.

Acknowledgment

This work is partially supported by the National Natural Science Foundation of China under grants 61373090 and 61601310.

References

[1] A. Alvergne, R. Oda, C. Faurie, A. Matsumoto-Oda, V. Durand, M. Raymond, Cross-cultural perceptions of facial resemblance between kin, J. Vis. 9 (6) (2009). 23–23
[2] G. Kaminski, S. Dridi, C. Graff, E. Gentaz, Human ability to detect kinship in strangers' faces: effects of the degree of relatedness, Proc. R. Soc. Lond. B 276 (1670) (2009) 3193–3200.
[3] M.F. Dal Martello, L.T. Maloney, Lateralization of kin recognition signals in the human face, J. Vis. 10 (8) (2010). 9–9
[4] R. Fang, K.D. Tang, N. Snavely, T. Chen, Towards computational models of kinship verification, 2010 IEEE International Conference on Image Processing, IEEE, 2010, pp. 1577–1580.
[5] S. Xia, M. Shao, Y. Fu, Kinship verification through transfer learning, IJCAI Proceedings-International Joint Conference on Artificial Intelligence, vol. 22, (2011), p. 2539.
[6] X. Zhou, J. Hu, J. Lu, Y. Shang, Y. Guan, Kinship verification from facial images under uncontrolled conditions, Proceedings of the 19th ACM International Conference on Multimedia, ACM, 2011, pp. 953–956.
[7] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, J. Zhou, Neighborhood repulsed metric learning for kinship verification, Pattern Anal. Mach. Intell. IEEE Trans. 36 (2) (2014) 331–345.
[8] R.G. Cinbis, J. Verbeek, C. Schmid, Unsupervised metric learning for face identification in TV video, 2011 International Conference on Computer Vision, IEEE, 2011, pp. 1559–1566.
[9] M. Guillaumin, J. Verbeek, C. Schmid, Is that you? Metric learning approaches for face identification, 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 498–505.
[10] M. Köstinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 2288–2295.
[11] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: closing the gap to human-level performance in face verification, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2014), pp. 1701–1708.
[12] A. Mignon, F. Jurie, Pcca: a new approach for distance learning from sparse pairwise constraints, Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 2666–2672.
[13] H.V. Nguyen, L. Bai, Cosine similarity metric learning for face verification, Asian Conference on Computer Vision, Springer, 2010, pp. 709–720.
[14] Z. Cui, W. Li, D. Xu, S. Shan, X. Chen, Fusing robust face region descriptors via multiple metric learning for face recognition in the wild, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2013), pp. 3554–3561.
[15] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, British Machine Vision Conference, (2015).
[16] W. Deng, J. Hu, J. Lu, J. Guo, Transform-invariant PCA: a unified approach to fully automatic face alignment, representation, and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 36 (6) (2014) 1275–1284.
[17] X. Cai, C. Wang, B. Xiao, X. Chen, J. Zhou, Deep nonlinear metric learning with independent subspace analysis for face verification, Proceedings of the 20th ACM international conference on Multimedia, ACM, 2012, pp. 749–752.
[18] J. Lu, Y.-P. Tan, G. Wang, Discriminative multimanifold analysis for face recognition from a single training sample per person, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 39–51.
[19] J. Lu, V.E. Liong, X. Zhou, J. Zhou, Learning compact binary face descriptor for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (10) (2015) 2041–2056.
[20] J. Hu, J. Lu, Y.-P. Tan, Discriminative deep metric learning for face verification in the wild, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2014), pp. 1875–1882.
[21] J. Hu, J. Lu, Y.-P. Tan, Deep transfer metric learning, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2015), pp. 325–333.
[22] X. Zhou, K. Jin, Q. Chen, M. Xu, Y. Shang, Multiple face tracking and recognition with identity-specific localized metric learning, Pattern Recognit. 75 (2018) 41–50.
[23] E.P. Xing, A.Y. Ng, M.I. Jordan, S. Russell, Distance metric learning with application to clustering with side-information, Advances in Neural Information Processing Systems, vol. 15, MIT; 1998, 2003, pp. 505–512.

[24] J. Goldberger, G.E. Hinton, S.T. Roweis, R. Salakhutdinov, Neighbourhood components analysis, Advances in Neural Information Processing Systems, (2004), pp. 513–520.
[25] K.Q. Weinberger, J. Blitzer, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, Advances in Neural Information Processing Systems, (2005), pp. 1473–1480.
[26] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, Proceedings of the 24th International Conference On Machine Learning, ACM, 2007, pp. 209–216.
[27] J. Hu, J. Lu, Y.P. Tan, J. Yuan, J. Zhou, Local large-margin multi-metric learning for face and kinship verification, IEEE Trans. Circuits Syst. Video Technol. (2017).
[28] H. Yan, J. Lu, W. Deng, X. Zhou, Discriminative multimetric learning for kinship verification, IEEE Trans. Inf. Forensics Secur. 9 (7) (2014) 1169–1178.
[29] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Comput. 1 (4) (1989) 541–551.
[30] G. Guo, X. Wang, Kinship measurement on salient facial features, IEEE Trans. Instrum. Meas. 61 (8) (2012) 2322–2325.
[31] X. Zhou, J. Lu, J. Hu, Y. Shang, Gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments, Proceedings of the 20th ACM International Conference on Multimedia, ACM, 2012, pp. 725–728.
[32] A. Dehghan, E.G. Ortiz, R. Villegas, M. Shah, Who do i look like? determining parent-offspring resemblance via gated autoencoders, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2014), pp. 1757–1764.
[33] H. Yan, J. Lu, X. Zhou, Prototype-based discriminative feature learning for kinship verification, IEEE Trans. Cybern. 45 (11) (2015) 2535–2545.
[34] N. Kohli, M. Vatsa, R. Singh, A. Noore, A. Majumdar, Hierarchical representation learning for kinship verification, IEEE Trans. Image Process. 26 (1) (2017) 289–302.
[35] S. Xia, M. Shao, J. Luo, Y. Fu, Understanding kin relationships in a photo, IEEE Trans. Multimedia 14 (4) (2012) 1046–1056.
[36] S. Mahpod, Y. Keller, Kinship verification using multiview hybrid distance learning, Comput. Vision Image Understanding 167 (2018) 28–36.
[37] J. Lu, J. Hu, Y.-P. Tan, Discriminative deep metric learning for face and kinship verification, IEEE Trans. Image Process. 26 (9) (2017) 4269–4282.
[38] S. Wang, J.P. Robinson, Y. Fu, Kinship verification on families in the wild with marginalized denoising metric learning, Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on, IEEE, 2017, pp. 216–221.
[39] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical Report, Technical Report 07–49, University of Massachusetts, Amherst, 2007.
[40] K. Zhang, Y. Huang, C. Song, H. Wu, L. Wang, Kinship verification with deep convolutional neural networks, Proceedings of the British Machine Vision Conference (BMVC), (2015), pp. 148.1–148.12.
[41] H. Dibeklioglu, A. Ali Salah, T. Gevers, Like father, like son: facial expression dynamics for kinship verification, Proceedings of the IEEE International Conference on Computer Vision, (2013), pp. 1497–1504.
[42] N. Kohli, R. Singh, M. Vatsa, Self-similarity representation of weber faces for kinship classification, Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on, IEEE, 2012, pp. 245–250.
[43] X. Qin, X. Tan, S. Chen, Tri-subject kinship verification: understanding the core of a family, IEEE Trans. Multimedia 17 (10) (2015) 1855–1867.
[44] X. Zhou, Y. Shang, H. Yan, G. Guo, Ensemble similarity learning for kinship verification from facial images in the wild, Inf. Fusion 32 (2016) 40–48.
[45] J. Lu, J. Hu, V.E. Liong, X. Zhou, A. Bottino, I.U. Islam, T.F. Vieira, X. Qin, X. Tan, S. Chen, S. Mahpod, Y. Keller, L. Zheng, K. Idrissi, C. Garcia, S. Duffner, A. Baskurt, M. Castrillln-Santana, J. Lorenzo-Navarro, The fg 2015 kinship verification in the wild evaluation, Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, 1 (2015). 1–7
[46] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems, (2012), pp. 1097–1105.
[47] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, (2016), pp. 770–778.
[48] Y. Sun, Y. Chen, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, Advances in Neural Information Processing Systems, (2014), pp. 1988–1996.
[49] S. Ji, W. Xu, M. Yang, K. Yu, 3D convolutional neural networks for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 221–231.
[50] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorihm for deep belief nets, Neural Comput. 18 (7) (2006) 1527–1554.
[51] Q.V. Le, W.Y. Zou, S.Y. Yeung, A.Y. Ng, Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis, Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 3361–3368.
[52] J. Wang, F. Zhou, S. Wen, X. Liu, Y. Lin, Deep metric learning with angular loss, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2017), pp. 2593–2601.
[53] C. Zhu, L. Cao, Q. Liu, J. Yin, V. Kumar, Heterogeneous metric learning of categorical data with hierarchical couplings, IEEE Trans. Knowl. Data Eng. (2018).
[54] P. Xie, Learning compact and effective distance metrics with diversity regularization, Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2015, pp. 610–624.
[55] A. Mignon, F. Jurie, CMML: a new metric learning approach for cross modal matching, Asian Conference on Computer Vision, South Korea, (2012), p. 14pages.
[56] B. Geng, D. Tao, C. Xu, Daml: domain adaptation metric learning, IEEE Trans. Image Process. 20 (10) (2011) 2980–2989.
[57] Y.C. Chen, W.S. Zheng, J.H. Lai, P. Yuen, An asymmetric distance model for cross-view feature mapping in person re-identification, IEEE Trans. Circuits Syst. Video Technol. PP (99) (2016). 1–1.
[58] V.E. Liong, J. Lu, Y.P. Tan, J. Zhou, Deep coupled metric learning for cross-modal matching, IEEE Trans. Multimedia PP (99) (2016) 1–1. doi:10.1109/TMM.2016.2646180 ..
[59] H. Liu, J. Cheng, F. Wang, Kinship verification based on status-aware projection learning, Image Processing (ICIP), 2017 IEEE International Conference on, IEEE, 2017, pp. 1072–1076.
[60] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, Y. Wu, Learning fine-grained image similarity with deep ranking, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2014), pp. 1386–1393.
[61] S. Si, D. Tao, B. Geng, Bregman divergence-based regularization for transfer subspace learning, IEEE Trans. Knowl. Data Eng. 22 (7) (2010) 929–942.
[62] J.T. Kwok, R.P. Adams, Priors for diversity in generative latent variable models, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25, 2012, pp. 2996–3004.
[63] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[64] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, IEEE, 2005, pp. 886–893.
[65] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[66] J.P. Robinson, M. Shao, H. Zhao, Y. Wu, T. Gillis, Y. Fu, Rfiw: large-scale kinship recognition challenge, Proceedings of the 2017 ACM on Multimedia Conference, ACM, 2017, pp. 1971–1973.