Kinship Verification on Families in the Wild with Marginalized Denoising Metric Learning

Shuyang Wang[1*], Joseph P. Robinson[1*], and Yun Fu[1,2]
[1]Department of Electrical & Computer Engineering,
[2]College of Computer & Information Science,
Northeastern University, Boston, MA, USA

Abstract— With our Families In the Wild (FIW) dataset, which consists of labels $1,000$ families in over $12,000$ family photos, we benchmarked the largest kinship verification experiment to date. FIW, with its quality data and labels for full family trees found worldwide, more accurately is the true, global distribution of blood relatives with a total $378,300$ face pairs of 9 different relationship types. This gives support to tackle the problem with modern-day data-driven methods, which are imperative due to the complex nature of tasks involving visual kinship recognition– many hidden factors and less discrimination when considering face pairs of blood relatives. For this, we propose a denoising auto-encoder based robust metric learning (DML) framework and its marginalized version (mDML) to explicitly preserve the intrinsic structure of data and simultaneously endow the discriminative information into the learned features. Large-scale experiments show that our method outperforms other features and metric based approaches on each of the 9 relationship types.

I. INTRODUCTION

Automatic kinship recognition capability is relevant in a wide range of practical case, with many whom reap benefit – such as consumers (*e.g.* automatic photo library management [33]), scholars (*e.g.* historic lineage & genealogical studies [3], [4]), analyzers (*e.g.* social-media-based analysis [14], [24], [9], [39]), and investigators (*e.g.* missing persons and human traffickers [25])– with advanced kinship recognition technologies. Additionally, kinship is a powerful cue for automatic understanding faces, and even the bigger picture of human-computer interaction as a whole, as the knowledge could be used as high-level evidence in bigger problems, such as an attribute in conventional facial recognition. From this, we are motivated to push the frontier of visual kinship recognition with the largest verification benchmarks for several types of relationships, running each with the data needed to properly represent that in the real-world (*i.e.*, from sources such as personal photo albums, social media outlets, Hollywood, athletic associations, places of political power, and many other unconstrained sources worldwide and *in the wild*). Amongst our top aspirations with this work is to attract more interest of researchers from the vision and related communities to help improve state-of-art and to design new problems using the new large-scale database for kinship recognition (*i.e.*, *Families in the Wild* (FIW) [25]).

Analogous to facial recognition, kinship recognition can be viewed as a broader set of subtasks to tackle different practical purposes. Of these, kinship verification was the first to be introduced back in 2010 [13]. But, even after the several years since then, automatic kinship recognition capability is still far from being mature enough to be employed in real-world uses. We believe the reasons for the delay are summarized in two-fold:

1) Current kinship recognition datasets are insufficient in terms of size, diversity and, thus under represent actual family distributions in the real-world.
2) Recognizing kinship in the visual domain is challenging, as there are hidden factors affecting the facial appearances amongst family members making for less discriminative power when compared to more conventional problems of its kind (*e.g.* facial recognition or object classification).

In response, we propose the largest benchmark for kinship verification which involves several hundred thousand additional pairs, opposed to just the couple of thousand that was available prior. We also achieve a significant boost in performance using deep learning on this generous collection of face pairs.

Several methods were proposed for kinship verification over the past decade. Of these attempts, there were several involving hand-crafted (*i.e.*, low level) features (*e.g.* SIFT [18], LBP [1], PEM [17]), face encodings (*e.g.* fisher vectors faces [26]), and metric learning methods. The common objective of metric learning is to learn a good distance metric to minimize distances between positive pairs, while pushing pairs of different types further apart. Several representative metric learning algorithms have been applied to kinship verification, including neighborhood repulsed metric learning (NRML) [20], large margin multi-metric learning (LM^{3}L) [16], discriminative multi-metric learning (DMML) [36], and online similarity learning (OSL) [35].

The goal of kinship verification is to determine whether or not a pair of faces are related; opposed to conventional facial recognition where the task is to determine if a pair of faces are the same person. Existing metric learning approaches only learn a linear transformation directly from the original feature space to project the input samples into a new representation, which may or may not be powerful enough to explicitly capture the intrinsic structure hidden in kinship knowledge, all awhile setting aside identity discrepancies and

Siblings (SIB)

B
S

P: 2,461
F: 673
S: 105,000

Father-Daughter (F-D)

F
D

P: 1,883
F: 638
S: 72,000

Father-Son (F-S)

F
S

P: 1,969
F: 655
S: 72,000

Mother-Daughter (M-D)

M
D

P: 1,836
F: 639
S: 64,000

Mother-Son (M-S)

M
S

P: 1,900
F: 646
S: 60,000

Grandfather-Granddaughter (GF-GD)

GF
GD

P: 115
F: 53
S: 1,000

Grandfather-Grandson (GF-GS)

GF
GS

P: 141
F: 62
S: 1,350

Grandmother-Granddaughter (GM-GD)

GM
GD

P: 111
F: 51
S: 950

Grandmother-Grandson (GM-GS)

GM
GS

P: 178
F: 77
S: 2,000

Total

No. Pairs (P) 10,594
No. Families (F) 996
No. Face Samples (S) 378,300

Fig. 1. Samples of the 9 relationship pair types. airs from the British Royal family are shown at different ages depicting the age variation of FIW (*i.e.*, youngest-to-oldest from left-to-right, respectfully). Total counts are provided (bottom row), along with individual counts for different types (right column). Total pairs (P) and face samples (S) was found as the sum of each, while this is not true for families (F) due to overlap between types– For instance, the pairs shown above are sampled from 1 family, not 9.

variations in age. Therefore, we set out to find a unified hidden space where we are able to learn a better, more discriminative metric.

Recently, auto-encoder and its variants [5], [28], [38], [29] have attracted many research interests with the remarkable success in many applications. The encoder is trained to produce new representations which can be reverted to the original inputs by decoder. In this study, we propose a denoising metric learning (DML) framework, which seemlessly connects denoising auto-encoding (DAE) techniques with metric learning. Specifically, the original family data are encoded into the hidden layer of the auto-encoder. Then, a robust distance metric is learned to project facial images with kinship relations as closely as possible, while moving those without further apart. Unlike previous metric learning methods, the proposed DML learns a nonlinear transformation to project face pairs to one feature space, while the transformation of DAE works as a metric to seek the discriminative information simultaneously. Thus, learned representations not only preserve the locality properties, but are also easy to revert to the original input form. This is achieved by jointly minimizing for the metric constraint and auto-encoder. Moreover, we also give an effectual marginalized version of our algorithm (mDML) based on marginalized DAE.

Our contributions in this work are as follows.

1) We benchmark the largest visual kinship verification dataset to date– increasing from a couple of thousand to several hundred thousand face pairs.
2) We propose a DAE based metric learning framework, along with its marginalized version (DML and mDML). This novel approach simultaneously maximizes the empirical likelihood to find a high-level hidden feature space and exploits the discriminative information for the verification task with DAE and metric learning, respectfully.

II. RELATED WORKS

Since 2010, great efforts have been put towards advancing automatic kinship recognition capabilities [2], [9], [12], [14], [16], [19], [20], [24], [32], [33], [35], [36], [37]. The recent works resulted from the release of KinWild I-II, which was first introduced as part of a FG 2015 challenge [19]. KinWild, with $2,000$ face pairs from 4 categories (*i.e.*, *parent-child* pairs), remained the largest image-set for kinship verification until recently (*i.e.*, until FIW). FIW is by far the largest dataset for kinship verification to date, with $378,300$ face pairs of 9 categories (*i.e.*, from 4 to 9 types) [25]. We review 3 related topics through the remainder of this section: Kinship verification, auto-encoder, and metric learning.

Kinship Verification aims to determine whether 2 people are blood relatives using facial images. Essentially, this task is a typical binary classification problem, as the pair is either relatives or not (*i.e.*, 2 choses, *kin* and *non-kin*). Prior efforts have focused on certain relationship types, which depended on the availability of labeled data. Of these, the main 4 are types of *parent-child* pairs (*i.e.*, father-son (F-S), father-daughter (F-D), mother-son (M-S), and mother-daughter (M-D)). As research in psychology and computer vision revealed, the various types kin relationships render different familial features and, hence, different relationship types are usually handled independently, and often differently. However, and

as mentioned, existing kinship datasets provides only $1,000$ positive face pairs, which are shared between 4 types (*i.e.*, just 250 positive face samples each). Considering protocols for kinship verification mimic that of conventional face recognition by splitting data in 5-folds, then, prior to FIW, each type then included just 200 positive face pairs for training and 50 for testing. Such a minimal amount of data often results in models overfitting the training data. Hence, these models do not generalize well to unseen test data and lead to unstable predictions.

Auto-encoder (AE) [5] learns an identity preserved representation in the hidden layer by setting the input and target as the same. Along the line of AE, denoising auto-encoder (DAE) [27] is trained to denoise the input, which was ultimately followed by its marginalized version to overcome high computational costs. The proposed marginalized DAE (mDAE) [7] provides a closed-form solution for the parameters optimization– opposed to using back-propagation. The idea of mDAE is to speed up the proposed DML to form its marginalized version (*i.e.*, mDML). Unlike mDAE, the encoder and decoder structures are preserved in our model. This way, the hidden layer naturally endows the discriminative property.

Metric Learning (ML) has been paid lots of attention recently from interest in finding a Mahalanobis-like distance matrix from training data. Information Geometry Metric Learning (IGML) [30] was proposed to minimize the K-L divergence between two Gaussian distributions. Information-Theoretic Metric Learning (ITML) [8] formulated the relative entropy as a Bregman optimization problem subject to linear constraints. Ding et al. developed a cross-domain metric to transfer knowledge from a well-labeled source to guide the learning of the unlabeled target [10]. Different from existing metric learning methods. Our method simultaneously learns a Mahalanobis distance metric while encoding the input data in a hidden space.

III. (MARGINALIZED) DENOISING METRIC LEARNING

We propose a new denoising auto-encoder based metric learning (DML) method for kinship verification *in the wild*. Unlike other metric learning methods, our DML first looks for a nonlinear transformation to project face pairs into a unified hidden subspace, under which we learn a discriminative metric to enforce distances of positive pairs to be reduced and that of the negatives to be enlarged. It should be noted that both tasks are handled by a single unified learning procedure, *i.e.*, the projection matrix W in our model works as both the encoder in DAE and the linear projection in metric learning simultaneously. Moreover, the resulting data cannot only be projected to a maximum margin feature space, but also be reverted to its original form.

The remainder of this section is laid out as follows. We first provide some preliminary information and motivation behind the proposed algorithm. Following this, we share details on our model jointly learning a hidden feature space and a discriminative Mahalanobis distance metric. Then, we discuss the optimization solution for our algorithm.

A. Discriminative Metric Learning

Assume $X \in \mathbb{R}^{d \times n}$ is the training data, with d as the dimensionality of the visual descriptor and n as the number of data samples. The Mahalanobis distance $d_M(x_i, x_j)$ between any two samples x_i and x_j can be computed with a learned matrix $M \in \mathbb{R}^{d \times d}$ as $\sqrt{(x_i - x_j)^{\mathrm{T}} M (x_i - x_j)}$.

Based on distance d_M's triangular inequality, non-negativity and symmetry properties, we decompose M into $W^{\mathrm{T}} W$, where $W \in \mathbb{R}^{z \times d}$, and $z \leq d$. Thus, the Mahalanobis distance converts to $d_M(x_i, x_j) = \| W x_i - W x_j \|_2$.

In the traditional supervised metric learning methods [34], [6], W is designed to have discrimination ability to pull within-class data points close together, while separating between-class data points as far as possible. With two weighted matrices S_w and S_b which respectively reflect the within-class and between-class affinity relationship [11], the formulation for discriminative metric learning is written as

$$\min_{WW^{\mathrm{T}}=I} \frac{\mathrm{tr}(WXL_wX^{\mathrm{T}}W^{\mathrm{T}})}{\mathrm{tr}(WXL_bX^{\mathrm{T}}W^{\mathrm{T}})} \qquad (1)$$

where L_w and L_b are the Laplacian matrices defined as $L_w = D_w - S_w$, where D_w is the diagonal matrix of S_w, similar for L_b and S_b. $\mathrm{tr}(A)$ represents the trace of matrix A. I is identity matrix, and the orthogonal constraint $WW^{\mathrm{T}} = I$ ensures that it is a basis transformation matrix.

Existing discriminative metric learning methods only learn a linear transformation directly from original feature space. However, we propose to integrate the role of above metric into the encoder projection of DAE or mDAE (either nonlinear or linear), in order to take both advantages from metric learning and denoising auto-encoder.

B. (marginalized) Denoising Auto-encoder Revisit

Given a input $x \in \mathbb{R}^d$, auto-encoder encourages the output of encoder and decoder to be as similar to the input as possible. That is,

$$\min_{W_1,W_2,b_1,b_2} L(x) = \min_{W_1,W_2,b_1,b_2} \frac{1}{2n} \sum_{i=1}^{n} \|x_i - \hat{x}_i\|_2^2, \qquad (2)$$

where n is the number of samples, W_1 and W_2^{T} are $z \times d$ weight matrixes, $b_1 \in \mathbb{R}^z$, $b_2 \in \mathbb{R}^d$ are offset vectors, σ is a non-linear activation function, x_i is the target and $\hat{x}_i$ is the reconstructed input. By this means, the auto-encoder can be seen to encode a good representation of the input in the hidden layer.

Recently, marginalized denoising auto-encoder (mDAE) [7] was proposed to learn a linear transformation matrix M to replace the encode and decode steps. In comparison, we still preserve encode and decode steps to make the proposed model more flexible but in a linearized way as $\frac{1}{2n} \sum_{i=1}^{n} \|x_i - W^{\mathrm{T}} W \widetilde{x}_i\|_2^2$. Where $\widetilde{x}_i$ is the corrupted version of x_i. mDAE minimized the overall squared loss of m corrupted versions to make it more reboust:

$$\frac{1}{2mn} \sum_{j=1}^{m} \sum_{i=1}^{n} \|x_i - W^{\mathrm{T}} W \widetilde{x}_{i,j}\|_2^2, \qquad (3)$$

where $\widetilde{x}_{i,j}$ is the j-th corrupted version of x_i. Define $X = [x_1, \cdots, x_n]$, its m-times repeated version $\overline{X} =$

$[X, \cdots, X]$ and its m-times different corrupted version $\widetilde{X} = [\widetilde{X}_1, \cdots, \widetilde{X}_m]$. Eq. (3) then can be reformulated as

$$\frac{1}{2mn}\|\overline{X} - W^{\mathrm{T}}W\widetilde{X}\|_{\mathrm{F}}^2, \qquad (4)$$

which has the well-known closed-form solution. When $m \to \infty$, it can be solved with expectation with the weak law of large numbers [7].

C. (marginalized) Deionising Metric Learning

In this section, we propose our joint learning framework by integrating metric learning and auto-encoder together. Specifically, we incorporate metric learning on the hidden layer units, so that endow the encoder W in auto-encoder with the same capacity of above linear transformation W in metric learning. That is, the matrix W works as the encoder in DAE and linear projection in metric learning simultaneously. Thus we have our objective function as:

$$\mathcal{L} = \min_{W_1, W_2, b_1, b_2} \frac{1}{2}\|X - \hat{X}\|_F^2 + \frac{\lambda}{2}\mathrm{tr}(\frac{HL_wH^{\mathrm{T}}}{HL_bH^{\mathrm{T}}}), \qquad (5)$$

where $H = \sigma(W_1X + B_1), \hat{X} = \sigma(W_2H + B_2)$, B_1, B_2 are the n-repeated column copy of b_1, b_2, respectively. λ is the balance parameter between auto-encoder and metric learning.

Also we can have the marginalized version as:

$$\mathcal{L}_m = \min_W \frac{1}{2mn}\|\overline{X} - W^{\mathrm{T}}W\widetilde{X}\|_{\mathrm{F}}^2 + \frac{\lambda}{2}\mathrm{tr}(\frac{WXL_wX^{\mathrm{T}}W^{\mathrm{T}}}{WXL_bX^{\mathrm{T}}W^{\mathrm{T}}}), \qquad (6)$$

where $X = [x_1, \cdots, x_n]$, $\overline{X}$ is its m-times repeated version and $\widetilde{X}$ is its corrupted version.

D. Optimization

Eq.(5) and Eq.(6) are both our proposed functions, which we name as DML and its marginalized version mDML, and we will give the optimization solutions for both respectively. Specifically, we treated the Eq.(5) as the regularized auto-encoder optimization, and Eq.(6) has a closed form solution.

1) Regularized Auto-encoder Learning: We employ the stochastic sub-gradient descent method to obtain the parameters W_1, b_1, W_2 and b_2. The gradients of the objective function in Eq.(5) with respect to the parameters are computed as follows:

$$\frac{\partial\mathcal{L}}{\partial W_2} = (X - \hat{X})\frac{\partial\hat{X}}{\partial(W_2H + B_2)}H^T, \qquad (7)$$

$$\frac{\partial\mathcal{L}}{\partial B_2} = (X - \hat{X})\frac{\partial\hat{X}}{\partial(W_2H + B_2)} = \mathcal{L}_2, \qquad (8)$$

$$\frac{\partial\mathcal{L}}{\partial W_1} = (W_2^{\top}\mathcal{L}_2 + \lambda H(L_w - \gamma L_b))\frac{\partial H}{\partial(W_1X + B_1)}X^{\top}, \qquad (9)$$

$$\frac{\partial\mathcal{L}}{\partial B_1} = (W_2^{\top}\mathcal{L}_2 + \lambda H(L_w - \gamma L_b))\frac{\partial H}{\partial(W_1X + B_1)}. \qquad (10)$$

Then, W_1, W_2 and b_1, b_2 can be updated by using the gradient descent algorithm.

2) Marginalized Version: Eq.(6) can be rewritten to:

$$\mathcal{L}_m(W) = \frac{1}{2mn}\mathrm{tr}[(\overline{X} - W^{\mathrm{T}}W\widetilde{X})^{\mathrm{T}}(\overline{X} - W^{\mathrm{T}}W\widetilde{X})]$$
$$+ \frac{\lambda}{2}\mathrm{tr}[WX(L_w - L_b)X^{\mathrm{T}}W^{\mathrm{T}}]$$
$$= \mathrm{tr}[W(Q - P - P^{\mathrm{T}} + \lambda X(L_w - L_b)X^{\mathrm{T}})W^{\mathrm{T}}], \qquad (11)$$

where $Q = \widetilde{X}\widetilde{X}^{\mathrm{T}}$ and $P = \overline{X}\widetilde{X}^{\mathrm{T}}$. We would like the repeated number m to be ∞, therefore, the denoising transformation could be effectively learned from infinitely many copies of noisy data. In this way, Q and P can be derived by their expectations $\mathbb{E}(P)$ and $\mathbb{E}(Q)$, and easy to be computed through mDAE [7]. Then, Eq.(11) can be optimized with eigen-decomposition.

IV. DATA

We next review the large-scale kinship recognition dataset used for experimentation. First, we give an overview of the database and characterize with statistics relevant to the verification task. Then, we review related datasets by comparing with FIW.

A. Families In The Wild (FIW)

FIW is the largest collection of images labeled for kinship-related tasks. It contains rich label information that describes 1,000 different families, including individuals in the family and relationships shared between them). A set of face pairs was obtained from these $1,000$ families, resulting in the largest collection of face pairs of 9 different relationship types– 4 of the types are provided through FIW for the 1st time (*i.e.*, *grandparent-grandchild* pairs).

B. Related Databases

As is true of many machine vision problems, the progress of kinship recognition has been largely influenced by data availability. To no surprise, fundalmental contributions were made possible by data made available to evaluate with. In 2012, *UB KinFace* [33], [32] proposed a different flavor of the traditional verification problem– using *parent-child* pair types, except with multiple facial images of the parents taken at different ages, ultimately allowing for a transfer learning view of the problem to be investigated. Then, *Kinship the Wild (KVW'14) Competition: the first kinship verification competition* was held as a part of the *International Joint Conference on Biometrics*, 2014. This was the first attempt to do kinship verification using unconstrained face data. The supporting dataset included hundreds of face pairs for each of the 4 *parent-child* categories. Finally, in support of the 2015 FG Kinship Verification challenge [19], *Kin-Wild I & II* were released on what was then the largest visual kinship dataset, with each containing a couple of thousand pairs for the *parent-child* types.

Thus far, attempts at kinship recognition have been predominately focused on verification. However, some attention has gone towards another task, family classification. As mentioned, the amount and quality of the work is influenced by the data available to support the given problem. Thus, *Family101* [12] led the way on tackling the kinship problem

TABLE I

VERIFICATION ACCURACY SCORES (%) FOR 5-FOLD EXPERIMENT ON FIW. NO FAMILY OVERLAP BETWEEN FOLDS.

Methods	F-D	F-S	M-D	M-S	GF-GD	GF-GS	GM-GD	GM-GS	SIB	Average
LBP	54.76	54.69	55.80	55.29	56.40	56.37	54.32	56.85	57.18	55.74
SIFT	56.13	56.34	56.30	55.36	56.90	56.07	60.32	57.95	58.80	57.13
VGG	63.92	64.02	65.99	63.70	60.80	63.11	59.89	61.85	73.21	64.05
VGG+LPP [21]	65.03	69.09	67.87	69.37	63.70	62.74	66.11	63.50	73.46	66.76
VGG+NPE [15]	64.25	63.78	64.75	64.74	59.90	61.93	64.95	61.60	73.68	64.40
VGG+LMNN [31]	65.66	67.08	68.07	67.16	63.90	60.44	63.68	60.15	73.88	65.56
VGG+GmDAE [23]	66.53	68.30	68.15	67.71	64.10	63.93	64.84	63.10	74.33	66.78
VGG+DLML [11]	65.96	69.00	68.51	69.21	62.90	62.96	64.11	64.55	74.97	66.90
VGG+ours(mDML)	67.90	70.24	**70.39**	70.40	63.20	63.78	66.11	66.45	75.11	68.18
VGG+ours(DML)	**68.08**	**71.03**	70.36	**70.76**	**64.90**	**64.81**	**67.37**	**66.50**	**75.27**	**68.79**
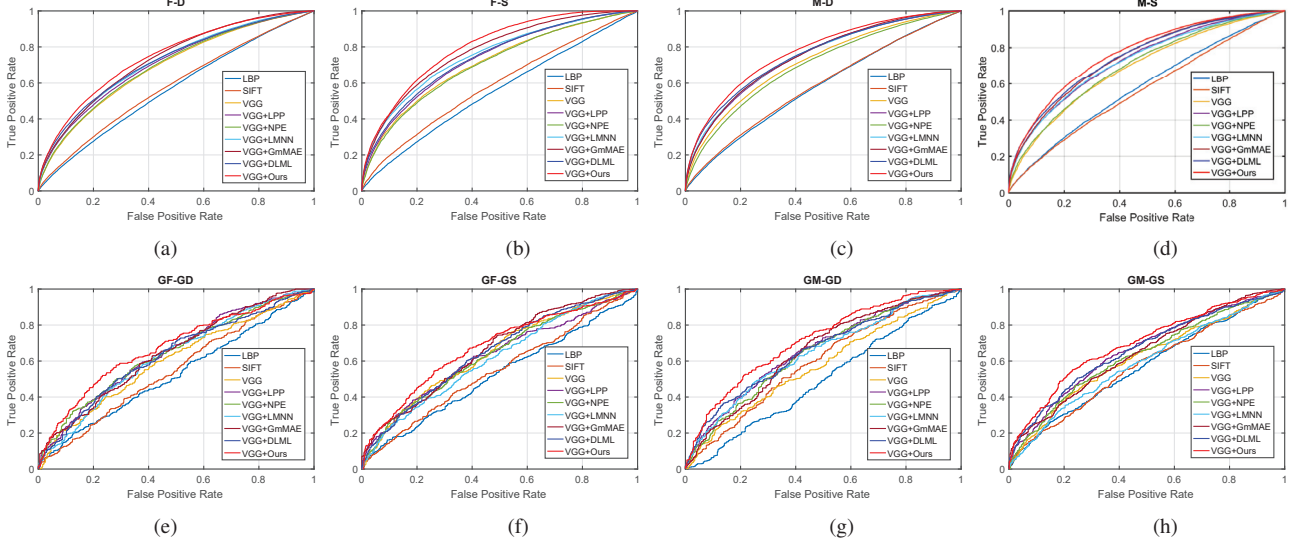
Fig. 2. Relationship specific ROC curves depicting performance of each method. Three features (LBP, SIFT and VGG-FACE) are evaluated and all methods on VGG features are reported.

at the family-level (*i.e.*, classes are families, with each made up of different family members related to one another in various ways (*e.g.* all the categories used for verification). Family classification is not included in this work, however, it certainly is doable as future work for FIW.

V. EXPERIMENTS

In this section, we evaluate our approach to the kinship verification problem on our FIW dataset. We report benchmarks using several different visual features and comparative methods. Note that our model could be based on both DAE or mDAE, which is denoted as DML and mDML, respectively.

A. Experimental Settings

As mentioned, previous reports on kinship verification mainly just used parent-child pairs, while we benchmark 9 here: 5 existing types (*i.e.*, parent-child and sibling pairs), but at scales much larger than in the past; additionally, 4 types unseen until the release of FIW (*i.e.*, grandparent-grandchild). In total, we benchmark 378,300 pairs face from the 9 categories.

Feature Three features are used as baselines, including two handcrafted features and one deep feature: Local Binary Patterns (LBP), Scale Invariant Feature Transformation (SIFT), and VGG-Face feature. VGG-Face [22] is a pretrained CNN model with a "Very Deep" architecture, tiny

convolutional kernels (*i.e.*, 3×3), and small convolutional stride (*i.e.*, 1 pixel). We see it as an off-the-shelf feature extractor. Face images were resized to 224×224 and then fed-forward to the second-to-last fully-connected layer (*i.e.*, fc7), producing a $4,096$ dimension facial encoding.

Comparative algorithms Three representative metric learning and graph embedding algorithms are used to evaluate on our dataset: Locality Preserving Projections (LPP) [21], Neighborhood Preserving Embedding (NPE) [15], and Large Margin Nearest Neighbor (LMNN) [31]. Moreover, two state-of-the-art methods, graph Regularized mDAE (GmSDA) [23] and Discriminative Low-rank Metric Learning (DLML) [11], are also used.

B. Kinship Verification

For each of the five folds, positive pairs were selected from random families to fill half, then negative pairs filled in the remainder. It is important to note that there was no family overlap between folds. Besides that, the typical *leave-one-out* test setting was followed for testing. Cosine similarity was used to compute the ROC curves and verification rate. The average verification rates for the five folds are shown in TableI. It is clear that VGG-Face features produce better verification results than hand-craft features. We also compare with the representative metric learning methods and state-of-

TABLE II

VERIFICATION ACCURACY SCORES (%) FOR 5-FOLD EXPERIMENT AND
THE TRAINING TIME (SECONDS) ON RELATIONSHIP FATHER-SON.

Methods	LBP		SIFT		VGG	
	ACC	Time	ACC	Time	ACC	Time
baseline	54.69	-	56.34	-	64.02	-
LPP [21]	55.78	20.9	59.41	91.3	69.09	11.4
NPE [15]	54.57	20.2	56.51	58.7	58.3708	8.8
LMNN [31]	55.32	1052	58.93	681	67.08	363
GmDAE [23]	55.16	2.7	59.20	2.7	68.03	2.8
DLML [11]	55.25	131	58.62	112	69.00	97
mDML[Ours]	55.54	0.84	60.10	0.98	70.24	0.87
DML[Ours]	55.52	31.2	62.33	22.4	71.03	43.4

the-art algorithms for verification. The comparative results show our approaches consistently outperform all the competitors, with the nonlinear version (DML) generally more accurate than the marginalized linear version (mDML). ROC curves are shown in Figure 2, one plot for each of the 9 relationships. To better visualize the results, we included the result for only one of our models (DML's). It can be seen from the ROC curves that our approach consistently achieves the most accurate results across all relationships.

Table II illustrates the improvement of each method on all three features, along with training times. From the table, we can see that both our DML and marginalized version outperform all others. The marginalized version achieves a slightly smaller accuracy score, but at a much faster training speed, which is due to the closed-form solution of mDAE.

VI. CONCLUSION

We benchmarked the largest number of face pairs for kinship verification to date– 378,300 face pairs from 9 relationship types. This was made possible with our FIW dataset that contains labels for 10,676 family photos of 1,000 families. A new DAE based metric learning framework DML, along with its marginalized version mDML, was proposed and evaluated on the FIW dataset. Kinship verification results indicate the effectiveness of the proposed framework when compared with related state-of-the-art methods.

REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *TPAMI*, 28(12):2037–2041, 2006.
[2] P. Alirezazadeh, A. Fathi, and F. Abdali-Mohammadi. A genetic algorithm-based feature selection for kinship verification. *IEEE Signal Processing Letters*, 22(12):2459–2463, 2015.
[3] M. Almuashi, S. Z. M. Hashim, D. Mohamad, M. H. Alkawaz, and A. Ali. Automated kinship verification and identification through human facial images: a survey. *Multimedia Tools and Applications*, pages 1–43, 2015.
[4] A. G. Bottino, M. De Simone, A. Laurentini, and T. Vieira. A new problem in face image analysis: finding kinship clues for siblings pairs. 2012.
[5] Y.-l. Boureau, Y. L. Cun, et al. Sparse feature learning for deep belief networks. In *NIPS*, pages 1185–1192, 2008.
[6] Q. Cao, Y. Ying, and P. Li. Similarity metric learning for face recognition. In *ICCV*, pages 2408–2415, 2013.
[7] M. Chen, Z. Xu, F. Sha, and K. Q. Weinberger. Marginalized denoising autoencoders for domain adaptation. In *ICML*, pages 767–774, 2012.
[8] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *ICML*, pages 209–216. ACM, 2007.
[9] A. Dehghan, E. G. Ortiz, R. Villegas, and M. Shah. Who do i look like? determining parent-offspring resemblance via gated autoencoders. In *CVPR*, pages 1757–1764, 2014.
[10] Z. Ding and Y. Fu. Robust transfer metric learning for image classification. *TIP*, 26(2):660–670, 2017.
[11] Z. Ding, S. Suh, J.-J. Han, C. Choi, and Y. Fu. Discriminative low-rank metric learning for face recognition. In *FG*, volume 1, pages 1–6. IEEE, 2015.
[12] R. Fang, A. C. Gallagher, T. Chen, and A. Loui. Kinship classification by modeling facial feature heredity. In *ICIP*, pages 2983–2987. IEEE, 2013.
[13] R. Fang, K. D. Tang, N. Snavely, and T. Chen. Towards computational models of kinship verification. In *ICIP*, pages 1577–1580. IEEE, 2010.
[14] Y. Guo, H. Dibeklioglu, and L. van der Maaten. Graph-based kinship recognition. In *ICPR*, pages 4287–4292, 2014.
[15] X. He, D. Cai, S. Yan, and H.-J. Zhang. Neighborhood preserving embedding. In *ICCV*, volume 2, pages 1208–1213. IEEE, 2005.
[16] J. Hu, J. Lu, J. Yuan, and Y.-P. Tan. Large margin multi-metric learning for face and kinship verification in the wild. In *ACCV*, pages 252–267. Springer, 2014.
[17] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic matching for pose variant face verification. In *CVPR*, pages 3499–3506, 2013.
[18] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
[19] J. Lu, J. Hu, V. E. Liong, X. Zhou, A. Bottino, I. U. Islam, T. F. Vieira, X. Qin, X. Tan, S. Chen, et al. The fg 2015 kinship verification in the wild evaluation. In *FG*, volume 1, pages 1–7. IEEE, 2015.
[20] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, and J. Zhou. Neighborhood repulsed metric learning for kinship verification. *TPAMI*, 36(2):331–345, 2014.
[21] X. Niyogi. Locality preserving projections. In *NIPS*, volume 16, page 153. MIT, 2004.
[22] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *BMVC*, volume 1, page 6, 2015.
[23] Y. Peng, S. Wang, and B.-L. Lu. Marginalized denoising autoencoder via graph regularization for domain adaptation. In *ICNIP*, pages 156–163. Springer, 2013.
[24] X. Qin, X. Tan, and S. Chen. Tri-subject kinship verification: Understanding the core of A family. *CoRR*, abs/1501.02555, 2015.
[25] J. P. Robinson, M. Shao, Y. Wu, and Y. Fu. Families in the wild (fiw): Large-scale kinship image database and benchmarks. In *ACM MM*, pages 242–246. ACM, 2016.
[26] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher vector faces in the wild. In *BMVC*, volume 2, page 4, 2013.
[27] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *JMLR*, 11:3371–3408, 2010.
[28] S. Wang, Z. Ding, and Y. Fu. Coupled marginalized auto-encoders for cross-domain multi-view learning. In *IJCAI*, pages 2125–2131, 2016.
[29] S. Wang, Z. Ding, and Y. Fu. Feature selection guided auto-encoder. In *AAAI*, 2016.
[30] S. Wang and R. Jin. An information geometry approach for distance metric learning. In *AISTATS*, volume 5, pages 591–598, 2009.
[31] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *JMLP*, 10(Feb):207–244, 2009.
[32] S. Xia, M. Shao, and Y. Fu. Kinship verification through transfer learning. In *IJCAI*, volume 22, page 2539, 2011.
[33] S. Xia, M. Shao, J. Luo, and Y. Fu. Understanding kin relationships in a photo. *IEEE Transactions on Multimedia*, 14(4):1046–1056, 2012.
[34] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning with application to clustering with side-information. *NIPS*, 15:505–512, 2003.
[35] M. Xu and Y. Shang. Kinship verification using facial images by robust similarity learning. *Mathematical Problems in Engineering*, 2016, 2016.
[36] H. Yan, J. Lu, W. Deng, and X. Zhou. Discriminative multimetric learning for kinship verification. *IEEE Transactions on Information forensics and security*, 9(7):1169–1178, 2014.
[37] H. Yan, J. Lu, and X. Zhou. Prototype-based discriminative feature learning for kinship verification. *IEEE transactions on cybernetics*, 45(11):2535–2545, 2015.
[38] X. Yin and Q. Chen. Deep metric learning autoencoder for nonlinear temporal alignment of human motion. In *ICRA*, pages 2160–2166. IEEE, 2016.
[39] H. Zhao and Y. Fu. Dual-regularized multi-view outlier detection. In *IJCAI*, pages 4077–4083, 2015.