AdvNet: Adversarial Contrastive Residual Net for 1 Million Kinship Recognition

Qingyan Duan
College of Communication Engineering, Chongqing University
Chongqing, China
qyduan@cqu.edu.cn

Lei Zhang*
College of Communication Engineering, Chongqing University
Chongqing, China
leizhang@cqu.edu.cn

ABSTRACT

Kinship verification in the wild is an interesting and challenging problem. The goal of kinship verification is to determine whether a pair of faces are blood relatives or not. Most previous methods for kinship verification can be divided as hand-crafted features based shallow learning methods and convolutional neural network (CNN) based deep learning methods. Nevertheless, these methods are still posed with the challenging task of recognizing kinship cues from facial images. Part of the reason for this may be that, the family information and the distribution difference of pairwise kin-face data based kinship verification issue are rarely considered. Inspired by maximum mean discrepancy (MMD) and generative adversarial net (GAN), family ID based **Adv**ersarial contrastive residual **Net**work (AdvNet) is proposed for large-scale (1 Million) kinship recognition in this paper. The MMD based adversarial loss (AL), pairwise contrastive loss (CL) and family ID based softmax loss (SL) are jointly formulated in the proposed AdvNet for kin-relation enhancement and discovery. Further, the deep nets ensemble is used for deep kin-feature augmentation. Finally, Euclidean distance metric is used for kinship recognition. Extensive experiments on the 1st Large-Scale Kinship Recognition Data Challenge (Families in the wild) show the effectiveness of our proposed AdvNet and ensemble based feature augmentation.

CCS CONCEPTS

• **Computing methodologies** → **Computer Vision**; **Machine Learning**;

KEYWORDS

Kinship verification; convolutional neural networks; adversarial loss; feature augmentation

*The Corresponding author.

RFIW'17, October 27, 2017, Mountain View, CA, USA

ACM ISBN 978-1-4503-5511-7/17/10...$15.00
https://doi.org/10.1145/3134421.3134422

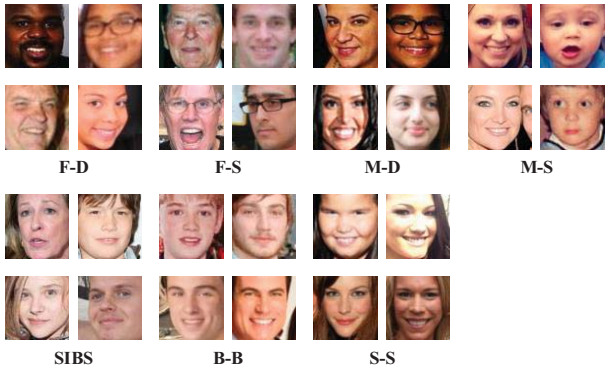


Figure 1: Some positive (with kinship relation) and negative pairs (no kinship relation) from 7 relationship types of FIW: Father-Daughter (F-D), Father-Son (F-S), Mother-Daughter (M-D), Mother-Son (M-S), Sister-Brother (SIBS), Brother-Brother (B-B), Sister-Sister (S-S). The odd rows are positive pairs and the even rows are negative pairs.

1 INTRODUCTION

The purpose of kinship verification is to recognize whether two persons are from the same family or not [2, 12]. Human face is a intuitive kin similarity measure, because the appearance of different member from the same family show more similar visual perception than others. Therefore, kinship verification in unconstrained conditions by modeling facial images has been paid more attention in recent years, and encouraging progress has been made on four typical parent-child relations. In this paper, we are facing with a new challenge which includes 7 kin-relations on a large-scale (1 million) kin-faces. Specifically, 4 parent-child relations including Father-Daughter (F-D), Father-Son (F-S), Mother-Daughter (M-D), and Mother-Son (M-S), and 3 sibling relations including Sister-Brother (SIBS), Brother-Brother (B-B), and Sister-Sister (S-S) are explored. Figure 1 shows the pairwise kin-face examples for each kin-relation. There are also many challenging applications for kinship verification, such as the human social relations exploration, social-media analysis, crime scene investigations, missing children searching, etc.

Due to the various factors in unconstrained faces, such as pose, illumination, expression, background clutters, etc. [22], kinship verification is still a challenging and unsolved topic.

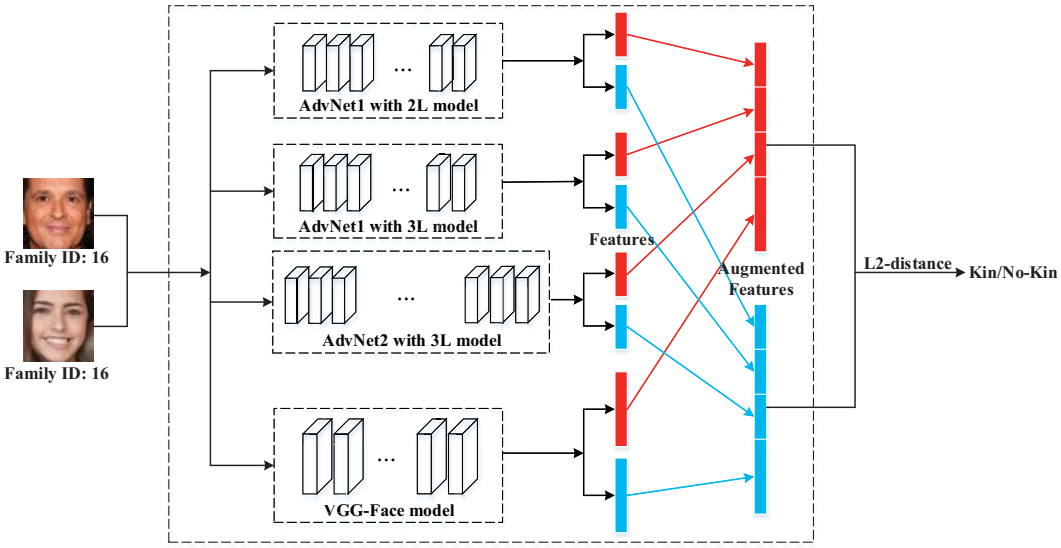


Figure 2: Overview of the proposed multiple deep nets based ensemble method for kinship verification. Three AdvNets with different loss and architecture, and one VGG-Face model are fused for feature augmentation. Euclidean distance is used for face verification.

Also, different from face recognition based discriminative feature representation, the kin-relation feature is implicit and hard to discover. Recently, many proposals based on hand-crafted low-level features (e.g. LBP, SIFT, etc.) have been tried. A representative work by Lu, et al. [12] was proposed to compress the distance of intra-class face pairs while repulse the inter-class face pairs in a neighborhood. Besides a single feature, multiple features were also jointly used for learning a discriminative metric to achieve better performance. However, these low-level features cannot well represent the underlying kin-relation implied in facial images. Thus kinship verification performance is restricted.

In recent years, convolutional neural networks (CNNs) in computer vision have obtained a huge success [1, 5, 8]. Due to its large-scale discriminative learning ability, deep CNNs have greatly boosted the face recognition performance to an unprecedent level [14, 16]. Recently, CNN has also been used in kinship verification [9, 24]. Although these CNN based algorithms greatly promote the performance of kinship verification on existing small-scale tasks such as KinFaces [12], CNN on large-scale kinship tasks remains under-studied. The existing methods usually adopt single loss function, such as softmax loss [24] or triplet loss [13], to train a CNN network from scratch. However, different from those typical face recognition, the intra-class distribution difference between pair-wise samples is not too significant, and therefore the general deep model cannot well interpret the kin-faces. In this paper, inspired by maximum mean discrepancy (MMD) [11] and generative adversarial nets (GAN) [4], a novel adversarial loss is proposed to interpret the distribution difference between pair-wise faces in the first fully-connected layer, which tends to minimize the inter-class discrepancy and maximize the intra-class variation. In contrast, a contrastive loss is formulated to maximize the inter-class distance and minimize the intra-class distance in the second fully-connected layer. As the result of the adversarial process between adversarial loss and contrastive loss, the discrimination and robustness of feature layer can be promoted. For integrating the family class, a softmax loss in the last layer can be further formulated to improve the recognition performance. Finally, the feature augmentation is achieved by the combined feature based on the different deep models, which is shown in Figure 2.

Our major contributions can be summarized as follows:

- Inspired by GAN, we propose a new MMD based adversarial loss function (AL in short), which is used to increase the learning difficulty of convolutional network model by minimizing the inter-class distribution discrepancy and maximizing the intra-class discrepancy. Thus the robustness of deep kin-features is improved.

- In order to decrease the intra-class variation while increasing the inter-class discrepancy, the proposed adversarial loss is combined with the modified contrastive loss (CL in short) and the family ID based softmax loss (SL in short), and formulates an AdvNet model which has a residual structure. The discrimination of deep kin-features is improved.

- Ensemble based feature augmentation is proposed in our method. We concatenate the deep features of multiple AdvNets with VGG-Face net, to further improve the verification performance of our method. The generality of deep kin-features is improved.

Table 1: Architectures for AdvNet1 and deeper AdvNet2. Building blocks are shown in brackets, with the numbers of blocks stacked. Down-sampling is performed by Conv1_x, Conv2_x, Conv3_x, Conv4_x, Conv5_x with a stride of 2.

CNN	Conv1_x	Conv2_x	Conv3_x	Conv4_x	Conv5_x	Conv6_x	FC1	FC2
AdvNet1	$3\times3, 32$ $3\times3, 64$	$\begin{bmatrix}3\times3,64\\3\times3,64\end{bmatrix}\times1$ $3\times3, 128$	$\begin{bmatrix}3\times3,128\\3\times3,128\end{bmatrix}\times2$ $3\times3, 256$	$\begin{bmatrix}3\times3,256\\3\times3,256\end{bmatrix}\times5$ $3\times3, 512$	$\begin{bmatrix}3\times3,512\\3\times3,512\end{bmatrix}\times3$	-	1024	512
AdvNet2	$3\times3, 32$ $3\times3, 64$	$\begin{bmatrix}3\times3,64\\3\times3,64\end{bmatrix}\times1$ $3\times3, 128$	$\begin{bmatrix}3\times3,128\\3\times3,128\end{bmatrix}\times2$ $3\times3, 256$	$\begin{bmatrix}3\times3,256\\3\times3,256\end{bmatrix}\times5$ $3\times3, 512$	$\begin{bmatrix}3\times3,512\\3\times3,512\end{bmatrix}\times3$	$\begin{bmatrix}3\times3,512\\3\times3,512\end{bmatrix}\times3$	1024	512

2 RELATED WORK

The existing work in kinship verification and deep convolutional networks are briefly introduced in the following.

2.1 Kinship Verification

Kinship verification based on facial image content is a challenging problem in computer vision. In recent years, there are many typical algorithms have been proposed [15]. Some of them are based on hand-crafted features, such as histogram of gradient (HOG) [3], scale-invariant feature transform (SIFT) [22], and local binary pattern (LBP) [12, 23]. These methods aim to extract discriminative features to discover the kin-characteristic in facial images. Some algorithms aim to learn an effective metric or model for determining human kinship relations via different learning strategies, such as neighborhood repulsed metric learning (NRML) [12], transfer subspace learning [21], large margin multi-metric learning [7], ensemble similarity learning (ESL) [25], and scalable similarity learning (SSL) [26]. Those previous work have achieved significant progress over the challenging kinship verification tasks. However, the essential flaw is that the extracted image features are general low-level representation of faces, which can not mining the implied kin-relation in facial images well.

2.2 Deep Convolutional Networks

Deep learning, proposed by Hinton and Salakhutdinov [6], has become the most popular machine learning algorithm for discovering the discriminative high-level representation in a hierarchical manner. CNN is an end-to-end supervised learning method from pixel to high-level semantic. The features from bottom to top in the network architecture can be recognized from low-level to high-level image representation. Several popular CNN models are summarized as follows. FaceNet [14] constructed a triplet-loss model and trained on nearly 200 million face images, which achieved state-of-the-art face recognition performance. A 3D-aligned method was also proposed by Deepface [18]. Recently, in order to decrease the intra-class variation and enhance the feature discrimination, the center-loss model was proposed in [20]. Further, the Sphereface is proposed to obtain within-class separable features [10] based on angle modeling. Different from the conventional CNN architecture, GAN [4] referred to generative and discriminative model, which can generate abundant labeled data via an adversarial process. Several researchers have applied the CNN methods on kinship verification task. For example, SMCNN [9] achieved the kin-relation verification through a similarity metric based cost function. The facial key-points were exploited to improve the accuracy of verification. DKV [19] integrated excellent deep learning architecture into metric learning to improve the performance of kinship verification. The triplet loss was used to fine-tune the VGG-Face model [13].

Compared with conventional metric learning, CNN based methods have achieved surprisingly good performance for kin-relation verification. However, the dataset (e.g. KinFaceW-I) used to train the CNN model is still small and insufficient for deep learning. In this paper, we propose a novel AdvNet and ensemble based feature augmentation method for large-scale kinship verification.

3 THE PROPOSED FAMILY ID BASED ADVNET

Families in the Wild (FIW) [13] is by far the largest and most comprehensive kinship dataset available in computer vision and multimedia communities. Different from previous kinship datasets, which only has kinship pair-wise mode (e.g. KinFaceI), FIW also provides the family tree to reflect the true data distribution of a family and their members. In order to improve the performance of our method, the family ID is also used in our model to obtain more discriminative deep features, that can better interpret the kin-relation than before. ResNet [5] has shown great performance in many computer vision tasks. Therefore, the proposed AdvNet follows a siamese residual network architecture with different depths. The inputs of the two residual CNN models are a pair of 224×224 RGB kinship facial images, which are slightly aligned by the developer of FIW.

3.1 The Family ID based Contrastive Loss

There are three fully-connected layers in AdvNet, and the second one is used to extract the deep kinship features (i.e. feature layer). The details of the two AdvNets with different depths are described in Table 1.

Let x_n^1 and x_n^2 represent the deep features of the left and right kinship image in the n^{th} pair, respectively. The

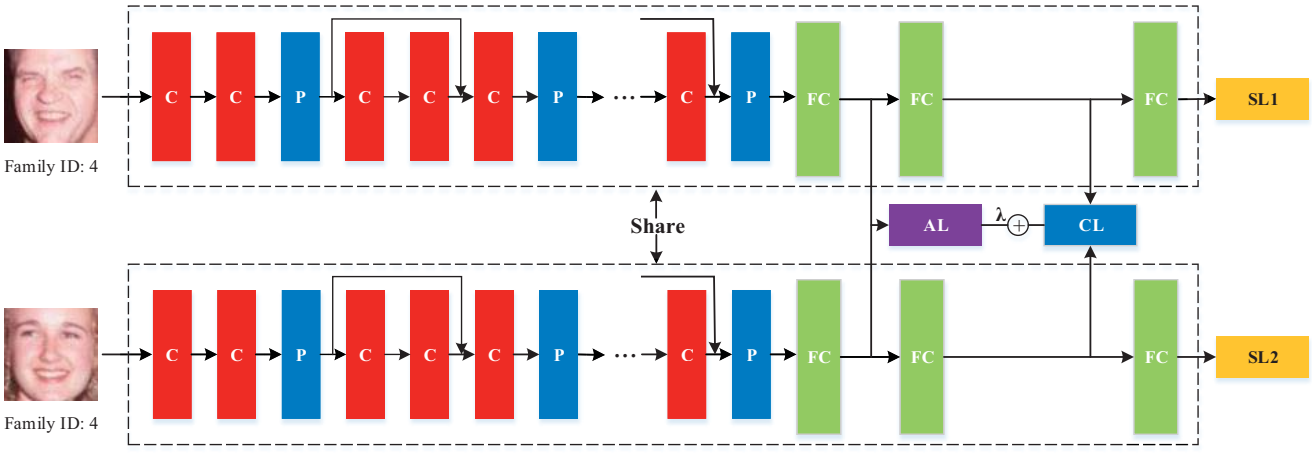


Figure 3: The Family ID based AdvNet with multiple losses residual architecture

contrastive loss function is presented as follows.

$$L_C = \frac{1}{2N}\sum_{n=1}^{N}(y_n d^2 + (1-y_n)\max(margin - d, 0)^2) \quad (1)$$

where N denotes the batch size, $d = ||x_n^1 - x_n^2||_2$ is the Euclidean distance between x_n^1 and x_n^2, and y_n denotes the label of the n^{th} pair of kinship faces. The label is 1 if there is kinship relation between two persons, otherwise 0.

The family IDs have been provided in FIW, which means that the pair of kin-relation samples must have the same family ID, and vice versa. Thus, in order to adopt the family ID to verify the kin-relation, the contrastive loss can be modified as

$$L_C = \frac{1}{2N}\sum_{n=1}^{N}(\delta(y_n^1 = y_n^2)d^2 - \delta(y_n^1 \neq y_n^2)\max(margin - d, 0)^2) \quad (2)$$

where $\delta(condition) = 1$ if the condition is satisfied. y_n^1 and y_n^2 are the family IDs of x_n^1 and x_n^2, respectively.

We observe that contrastive loss aims to train a model by pulling the positive pair as close as possible, and repulsing the negative pair as far as possible, simultaneously, but the distribution difference of pair-wise faces is neglected. However, this distribution difference of pais-wise kin faces in the wild is especially significant, which may influence the performance of kinship verification.

3.2 Family ID based Adversarial Loss

MMD is used to calculate the domain distribution discrepancy. It is usually employed to minimize the distribution difference between different domains in transfer learning [11].

Let $\hbar$ be the reproducing kernel Hilbert space (RKHS). Given two distributions s and t, which are mapped to a reproducing kernel Hilbert space by using an implicit function $\phi(\cdot)$. The MMD between s and t is defined as

$$\text{MMD}^2(s,t) = \sup_{||\phi||_\hbar \leq 1} ||E_{\mathbf{x}^s \sim s}[\phi(\mathbf{x}^s)] - E_{\mathbf{x}^t \sim t}[\phi(\mathbf{x}^t)]||_\hbar^2 \quad (3)$$

where $E_{\mathbf{x}^s \sim s}[\phi(\cdot)]$ denoted the expectation with regard to the distribution s, and $||\phi||_\hbar \leq 1$ defines a set of functions in the unit ball of RKHS $\hbar$. The most important property is that, we have $\text{MMD}(s,t) = 0$ if $s = t$.

Inspired by MMD, the distribution difference can be reduced by minimizing the distribution difference between pairwise faces (e.g. father-son). Therefore, a MMD based contrastive loss is proposed. In RKHS, the proposed loss function minimizes the intra-class variations (kin) while keeping the inter-class features separable (non-kin), which is formulated as

$$L_{\text{MMD}} = \frac{1}{2N}\sum_{n=1}^{N}(\delta(y_n^1 = y_n^2)||\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)||_\hbar^2 - \delta(y_n^1 \neq y_n^2)||\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)||_\hbar^2) \quad (4)$$

We can see that the MMD based contrastive loss is a straightforward method for decreasing the distribution difference across kin domains. Besides, some indirect approaches are used to augment the network. For example, CNN robustness can be improved by introducing additive noise. In Generative Adversarial Nets (GAN) [4], the generative model aims to generate the labeled samples as similar as possible with the source data, while the discriminative model aims to distinguish the generated data and source data as much as possible. It can be considered that, the objectives of generative model and discriminative model are exactly reverse, and the performance of GAN is promoted by this adversarial process. Inspired by the adversarial characteristic of GAN, in order to further improve the discrimination and robustness of deep features, a MMD based adversarial loss is proposed

Table 2: Accuracy of 2L based AdvNet with different trade-off weight λ

Loss	λ	M-D	M-S	S-S	B-B	SIBS	F-S	F-D	Mean
2L	0	**61.06**	61.95	62.45	**65.32**	**62.05**	61.33	59.18	61.91
2L	0.2	60.50	**64.07**	**64.17**	63.76	61.99	**62.23**	**60.53**	**62.46**
2L	1.0	50.69	49.58	63.60	62.20	61.89	60.27	59.23	58.21

Table 3: Accuracy of AdvNet with different loss

Loss	M-D	M-S	S-S	B-B	SIBS	F-S	F-D	Mean
CL	61.06	61.95	62.45	65.35	62.05	61.33	59.18	61.91
2L	60.50	64.07	64.17	63.76	61.99	62.23	60.53	62.46
3L	**64.11**	**65.65**	**64.53**	**65.80**	**64.82**	**63.42**	**63.18**	**64.50**

Table 4: Accuracy of AdvNet with different depth

Loss	CNN	M-D	M-S	S-S	B-B	SIBS	F-S	F-D	Mean
3L	AdvNet1	**64.11**	65.65	64.53	**65.80**	64.82	63.42	63.18	64.50
3L	AdvNet2	63.65	**66.80**	**65.48**	65.77	**65.35**	**64.14**	**63.59**	**64.97**

as

$$L_A = -\frac{1}{2N}\sum_{n=1}^{N}(\delta(y_n^1 = y_n^2)||\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)||_{\hbar}^2 - \delta(y_n^1 \neq y_n^2)||\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)||_{\hbar}^2) \quad (5)$$

By comparing Eq.(5) with Eq.(4), the only difference is the minus sign. It means that the adversarial loss plays an opposite role that the MMD based contrastive loss does. The adversarial loss is added on the fully-connected layers (1st layer) before the deep features layer (2nd layer), so that the adversarial process can be constructed between the first two fully-connected layers. Further, the AdvNet can be trained by combining the adversarial loss and the contrastive loss as follows,

$$\begin{aligned} L &= L_C + \lambda L_A \\ &= \frac{1}{2N}\sum_{n=1}^{N}(\delta(y_n^1 = y_n^2)d^2 - \delta(y_n^1 \neq y_n^2)\max(margin - d, 0)^2) \\ &- \lambda(\frac{1}{2N}\sum_{n=1}^{N}((\delta(y_n^1 = y_n^2) - \delta(y_n^1 \neq y_n^2))||\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)||_{\hbar}^2) \end{aligned} \quad (6)$$

where λ is a scalar used for balancing the two functions. The contrastive loss can be considered as a special case of this joint supervision, if λ is set to 0. Induced by the game between adversarial loss and contrastive loss like Eq.(6), the robustness of deep feature layer can be further improved.

In Eq.(6), $\phi(\cdot)$ denotes the implicit feature map, which can be solved by using kernel function $k(\mathbf{x}_n^1, \mathbf{x}_n^2) = \langle\phi(\mathbf{x}_n^1), \phi(\mathbf{x}_n^2)\rangle$. Thus, the Eq.(4) can be rewritten as

$$L_A = \frac{1}{2N}\sum_{n=1}^{N}(\delta(y_n^1 \neq y_n^2) - \delta(y_n^1 = y_n^2))(k(\mathbf{x}_n^1, \mathbf{x}_n^1) + k(\mathbf{x}_n^2, \mathbf{x}_n^2) - 2k(\mathbf{x}_n^1, \mathbf{x}_n^2)) \quad (7)$$

In this paper, we adopt the Gaussian kernel function as

$$k(\mathbf{x}, \mathbf{y}) = \exp(-\frac{||\mathbf{x} - \mathbf{y}||_2^2}{2\sigma^2}) \quad (8)$$

where σ^2 denotes the bandwidth (kernel parameter). In this way, Eq.(7) can be substituted into Eq.(9) as

$$L_A = \frac{1}{N}\sum_{n=1}^{N}(\delta(y_n^1 \neq y_n^2) - \delta(y_n^1 = y_n^2))(1 - \exp(-\frac{||\mathbf{x}_n^1 - \mathbf{x}_n^2||_2^2}{2\sigma^2})) \quad (9)$$

The gradients of L_A with respect to $\mathbf{x}_n^1$ and $\mathbf{x}_n^2$ are computed respectively as:

$$\frac{\partial L_A}{\partial \mathbf{x}_n^2} = \frac{1}{N\sigma^2}(\delta(y_n^1 \neq y_n^2) - \delta(y_n^1 = y_n^2)) \exp(-\frac{||\mathbf{x}_n^1 - \mathbf{x}_n^2||_2^2}{2\sigma^2})(\mathbf{x}_n^1 - \mathbf{x}_n^2) \quad (10)$$

$$\frac{\partial L_A}{\partial \mathbf{x}_n^2} = \frac{1}{N\sigma^2}(\delta(y_n^1 \neq y_n^2) - \delta(y_n^1 = y_n^2)) \exp(-\frac{||\mathbf{x}_n^1 - \mathbf{x}_n^2||_2^2}{2\sigma^2})(\mathbf{x}_n^2 - \mathbf{x}_n^1) \quad (11)$$

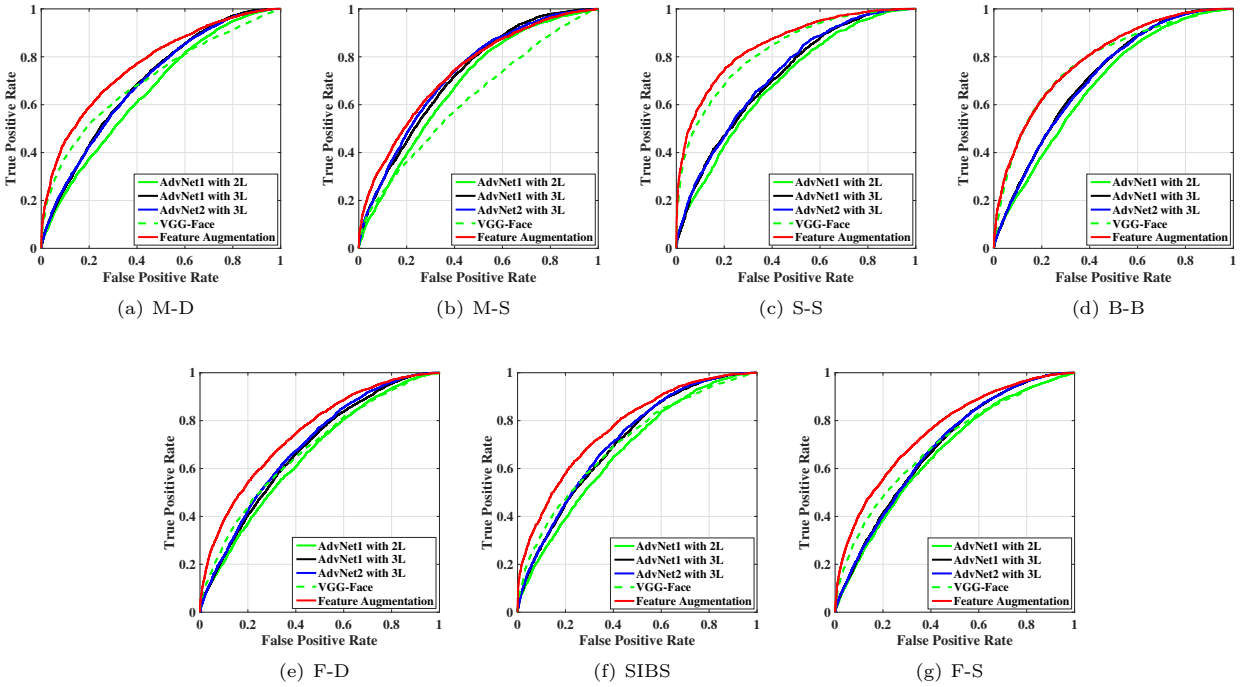
Therefore, the mini-batch SGD algorithm can be used to optimize the AdvNet.

3.3 Family based Joint Loss

In terms of the training protocol, the family ID for each kin-face is provided in large-scale FIW task which contains 300 families. Therefore, it is reasonable to integrate the most common supervisory signals by using softmax loss in

Table 5: Accuracy of different model, loss and feature augmentation

Index	Loss	Model	M-D	M-S	S-S	B-B	SIBS	F-S	F-D	Mean
1	2L	AdvNet1	60.50	64.07	64.17	63.76	61.99	62.23	60.53	62.46
2	3L	AdvNet1	64.11	65.65	64.53	65.80	64.82	63.42	63.18	64.50
3	3L	AdvNet2	63.56	66.80	65.48	65.77	65.35	64.14	63.59	64.97
4	SL	VGG-Face	65.99	58.88	74.59	**71.99**	64.69	64.71	62.87	66.25
1+2+3	Joint	Feature Augmentation	64.20	**67.55**	65.71	66.82	66.45	64.78	64.04	65.65
2+3+4	Joint	Feature Augmentation	**70.07**	65.60	**77.52**	71.88	69.72	**68.79**	67.56	70.16
1+2+3+4	Joint	Feature Augmentation	69.93	67.33	77.44	71.76	**69.80**	68.77	**67.82**	**70.41**



(a) M-D (b) M-S (c) S-S (d) B-B

(e) F-D (f) SIBS (g) F-S

Figure 4: ROC curves of different models on 7 types of kin-relation

AdvNet training. Different from the contrastive loss and the adversarial loss, the softmax loss aims to improve the family class separability of deep features. With this motivation, we also combine the softmax loss in AdvNet to further discover the implicit kin-relation of deep features. Considering the pairwise structure of AdvNet, two softmax loss functions will be formulated. Specifically, the joint loss is formulated as

$$L = L_C + \lambda L_A + L_{S1} + L_{S2}$$

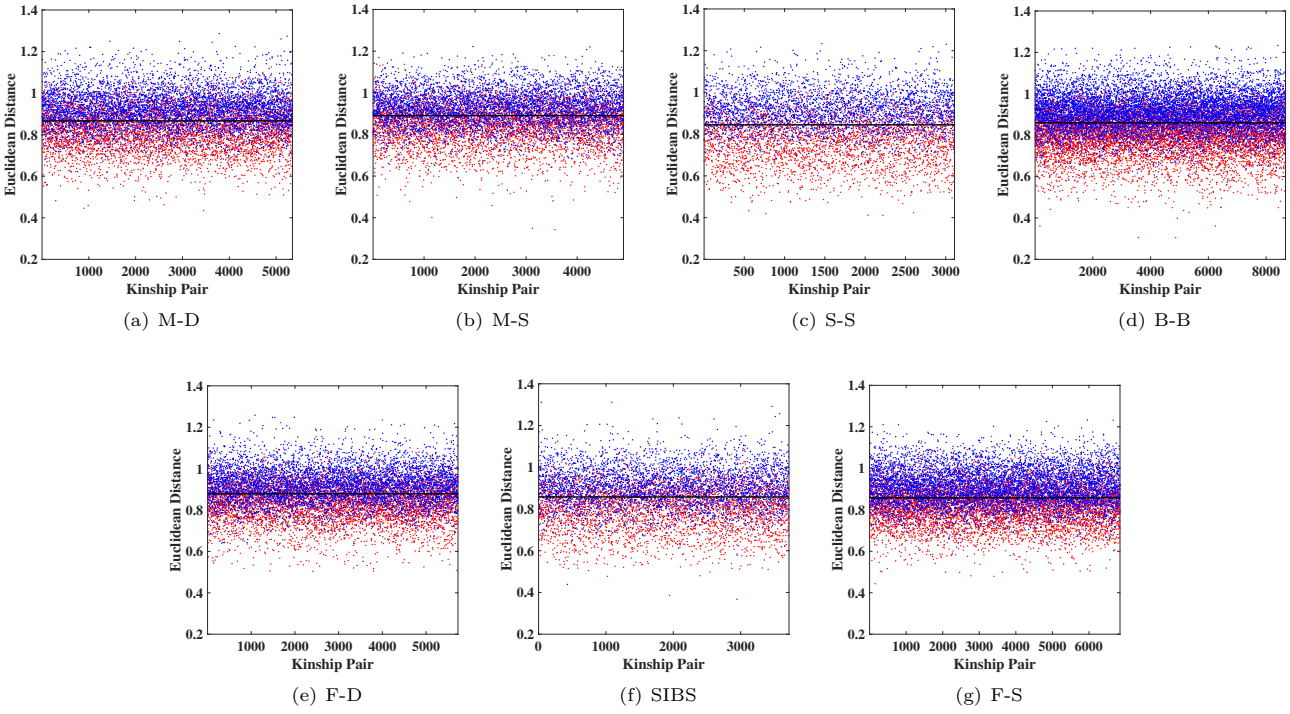
where L_{S1} and L_{S2} denote the softmax loss (cross entropy) for x_n^1 and x_n^2, respectively.

In order to adopt the softmax loss, a new output layer (softmax layer) with 300 neurons (i.e. 300 families) is added after the feature layer (contrastive loss layer). The architecture of AdvNet with joint loss is shown in Figure 3.

4 EXPERIMENTS

4.1 Experimental Data and Setup

FIW is the largest and most comprehensive image database for automatic kinship recognition, with over 12,000 family photos of 1,001 families. FIW closely reflects the true data distribution of families. The dataset used in this paper comes from the 1st Large-Scale Kinship Recognition Data Challenge. The challenge supports 2 laboratory style evaluation protocol: Kinship Verification (Track1) and Family Classification (Track2), we only focus the Track1. In Track1, FIW includes a total of 644,000 pairs , from which 538,518 pairs (i.e. over 1 Million kin-facial images) of 7 different kin-relations will be used for this data challenge. These datasets are partitioned into 3 disjoint sets referred to as Train, Validation, and Test sets. The ground truths for train and validation sets are provided, and the test set is "blind". In order to verify the



(a) M-D (b) M-S (c) S-S (d) B-B

(e) F-D (f) SIBS (g) F-S

Figure 5: L2-distances of kinship pairs on 7 types of kin-relation. The points in red and blue denote the kinship pairs and non-kinship pairs, respectively. The black line denote the searched threshold for verification.

proposed AdvNet in this paper, the validation set is used for testing the performance. Finally, the result on test set in this competition is reported.

In the competition, 7 different types of kinship: Father-Daughter (F-D), Father-Son (F-S), Mother-Daughter (M-D), Mother-Son (M-S), Sister-Brother (SIBS), Brother-Brother (B-B), Sister-Sister (S-S) are explored. The distribution of each kin in Train, Validation, and Test is shown as follows.

- In the Train set, 282,186 kinship pairs are included, which consists of 42,458, 53,974, 34,828, 38,312, 40,846, 52,482 and 19,286 pairs for 7 different types, respectively.

- In the Validation set, 76,664 kinship pairs are included, which consists of 11,460 13,696, 10,698, 9,816, 7,434, 17,342, 6,218 pairs for 7 different types, respectively.

- In the Test set, 179,668 kinship pairs are included, which consists of 23,506, 45,988, 20,674, 47,954, 15,076, 19,946 and 6,524 pairs for 7 different types, respectively.

In experiments, the proposed AdvNet with different loss is trained from scratch on the Train set, and finally Euclidean distance is used for kinship verification on the Validation set.

4.2 Parameter Analysis

In the joint loss function shown in Eq.(6), λ is a scalar used for balancing the contrastive loss (i.e. CL) and adversarial loss (i.e. AL). The joint CL and AL loss is called 2L in short. For showing the performance variation w.r.t. different loss weight λ, Table 2 reports the accuracy of our method versus λ. We can see that the 2L based AdvNet obtains the best classification performance when λ is set as 0.2.

4.3 Comparison with Different Loss Functions of AdvNet

In order to learn more separable features, the softmax loss (i.e. SL) is combined with 2L loss, which is called 3L in short. The loss weight is set as 1. As can be seen from Table 3, the results of 2L outperform the CL, which means that the adversarial loss can improve the discrimination and robustness of the kin-relation features. Besides, the results of 3L outperform the 2L, which means that the softmax loss can improve the separability of feature and kinship verification performance, by feeding the families information into the network.

4.4 Comparison with Different Depth of AdvNet

Depth is a very important factor of CNN model in classification performance [17]. In order to demonstration the influence of depth of AdvNet, the comparative results of different depth (AdvNet1 vs. AdvNet2) are listed in Table 4. It can be seen that the deeper AdvNet2 outperforms the AdvNet1.

4.5 Performance of Feature Augmentation

Many researches show the fact that, the performance of algorithm can be improved by feature augmentation and fusion [12, 16]. Therefore, feature augmentation based on AdvNets and VGG-Face also have been adopted in this paper. The performances of single model and multiple models is shown in Table 5. The features from index 1, 2, 3 and 4 represent the single feature (without augmentation). The last three rows denote the performance after feature augmentation by concatenating the features from each model together. The dimension of the augmented feature (e.g. 1+2+3) is 1536 (512×3), and the accuracy is improved. In addition, considering the excellent performance of the VGG-Face model, which is used as the feature extractor for FIW dataset in this paper, and the dimension of features extracted from VGG-Face model is 4096. After ensemble of the 4 networks (e.g 1+2+3+4), we can observe significant performance improvement. Notably, the L2-normalization is used twice before and after feature augmentation. It is noteworthy that, although the performance of VGG-FACE model is a little better than our AdvNets, the number of training data of AdvNets (i.e 0.01 Million faces) is far less than the VGG-Face model (i.e. 2.6 Million faces). Therefore, direct comparison is unfair. Note that the claimed 1 million data in this work denote the formulated kinship pairs based on the 0.01 million faces.

To better visualize the performance of different methods, the receiving operating characteristic (ROC) curves of different methods are shown in Figure 4, where Figure 4(a)-Figure 4(g) describe the ROC curves of the results on 7 types of kin-relation, respectively. We can see from this figure that the augmented features can yield the best performance in terms of the ROC curves.

In addition, the Euclidean distances of augmented features of kinship pairs are also visualized in Figure 5. It can be seen that the kin pairs and non-kin pairs are easy to be distinguished via a threshold (black line), however, there are still many pairs incorrectly recognized with L2-distance. In the future, some metric learning models may be utilized for better discrimination.

4.6 Competition on Blind Test Set

For competition on the blind test set, the feature augmentation method of AdvNets and VGG-Face model is finally used. With the help of the developers of this challenge, our final result on test set is 70.6588%, 65.2229%, 72.1030%, 63.5867%, 66.5097%, 63.3839% and 64.5963% for M-D, M-S, S-S, B-B, SIBS, F-S and F-D, respectively. The mean accuracy of the 7 kinship verification tasks is 66.5802%.

5 CONCLUSIONS

In this paper, we propose an adversarial contrastive residual network (AdvNet) and feature augmentation for large-scale Kinship verification over 1 Million faces. In AdvNet, the family ID based adversarial loss, motivated by the MMD and GAN, is proposed for feature robustness. Also, the family ID based contrastive loss is formulated for feature similarity measure. Further, two softmax losses with family class of the pairwise inputs are integrated for feature discrimination. With the joint loss, AdvNet is trained from scratch on the FIW faces based on mini-batch SGD optimization. In order to further promote the performance of our method, the feature augmentation is also adopted in this paper. Extensive experiments on the FIW challenge show the effectiveness of our proposed deep convolutional network.

ACKNOWLEDGMENTS

This work was supported by the National Science Fund of China under Grants No.: (61771079, 61401048) and the Fundamental Research Funds for the Central Universities No.: 106112017CDJQJ168819.

REFERENCES

[1] Y. Bengio. 2012. Deep learning of representations for unsupervised and transfer learning. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*. 17–36.

[2] A. Dehghan, E. G. Ortiz, R. Villegas, and S. Mubarak. 2014. Who do I look like? determining parent-offspring resemblance via gated autoencoders. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1757–1764.

[3] R. Fang, K. D. Tang, S. Noah, and T. Chen. 2010. Towards computational models of kinship verification. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. 1577–1580.

[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

[5] K. He, X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[6] G. E. Hinton and R. R. Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *science*. 313(5786) (2006), 504–507.

[7] J. Hu, J. Lu, J. Yuan, and Y. Tan. 2014. Large margin multi-metric learning for face and kinship verification in the wild. In *Asian Conference on Computer Vision*. 252–267.

[8] S. Karen and A. Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[9] L. Li, X. Feng, X. Wu, Z. Xia, and A. Hadid. 2016. Kinship Verification from Faces via Similarity Metric Based Convolutional Neural Network. In *International Conference Image Analysis and Recognition*. 539–548.

[10] W. Liu, Y. Wen, Z. Yu, Li M, B. Raj, and L. Song. 2017. SphereFace: Deep Hypersphere Embedding for Face Recognition. *arXiv preprint arXiv:1704.08063*.

[11] M. Long, Y. Cao, J. Wang, and M. Jordan. 2015. Learning transferable features with deep adaptation networks. In *International Conference on Machine Learning*. 97–105.

[12] J. Lu, X. Zhou, Y. Tan, Y Shang, and J. Zhou. 2014. Neighborhood repulsed metric learning for kinship verification. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 36(2) (2014), 331–345.

[13] J. Robinson, M. Shao, Y. Wu, and Y. Fu. 2016. Families in the Wild (FIW): Large-Scale Kinship Image Database and Benchmarks. In *Proceedings of the 2016 ACM on Multimedia Conference*. 242–246.

[14] F. Schroff, D. Kalenichenko, and J. Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 815–823.

[15] M. Shao, S. Xia, and Y. Fu. 2011. Genealogical face recognition based on ub kinface database. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 60–65.

[16] Y. Sun, X. Wang, and X. Tang. 2014. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1891–1898.

[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.

[18] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf. 2014. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1701–1708.

[19] M. Wang, Z. Li, X. Shu, and J. Wang. 2015. Deep kinship verification. In *Proceedings of the IEEE International Workshop on Multimedia Signal Processing*. 1–6.

[20] Y. Wen, K. Zhang, Z. Li, and Y. Qiao. 2016. A Discriminative Feature Learning Approach for Deep Face Recognition. *Computers & Operations Research*. 47(9) (2016), 11–26.

[21] S. Xia, M. Shao, and Y. Fu. 2011. Kinship verification through transfer learning. In *IJCAI*. 2539–2544.

[22] H. Yan, J. Lu, W. Deng, and X. Zhou. 2014. Discriminative multimetric learning for kinship verification. *IEEE Transactions on Information forensics and security*. 9(7) (2014), 1169–1178.

[23] H. Yan, J. Lu, and X. Zhou. 2015. Prototype-based discriminative feature learning for kinship verification. *IEEE Transactions on cybernetics*. 45(11) (2015), 2535–2545.

[24] K. Zhang, Y Huang, C. Song, H. Wu, and L. Wang. 2015. Kinship Verification with Deep Convolutional Neural Networks. In *BMVC*.

[25] X. Zhou, Y. Shang, H. Yan, and G. Guo. 2016. Ensemble similarity learning for kinship verification from facial images in the wild. *Information Fusion*. 32 (2016), 40–48.

[26] X. Zhou, H. Yan, and Y. Shang. 2016. Kinship verification from facial images by scalable similarity fusion. *Neurocomputing*. 197 (2016), 136–142.