Adv-Kin: An Adversarial Convolutional Network for Kinship Verification

Qingyan Duan[1], Lei Zhang[1]*, Wei Jia[2]

[1] College of Communication Engineering, Chongqing University, No.174 Shazheng street, Shapingba district, Chongqing, 400044, China
[2] School of Computer and Information, Hefei University of Technology, No.193 Tunxi road, Baohe district, Hefei, 230009, China
{qyduan,leizhang}@cqu.edu.cn
china.jiawei@139.com

Abstract. Kinship verification in the wild is an interesting and challenging problem, which aims to determine whether two unconstrained facial images are from the same family. Most previous methods for kinship verification can be divided as low-level hand-crafted features based shallow methods and kin data trained generic convolutional neural network (CNN) based deep methods. Nevertheless, these general methods cannot well mining the potential information implied in kin-relation data. Inspired by MMD and GAN, Adv-Kin method is proposed in this paper. The discrimination of deep features can be improved by introducing MMD loss (ML) to minimize the distribution difference between parents domain and children domain. In addition, we propose the adversarial loss (AL) can further improve the robustness of CNN model. Extensive experiments on the benchmark KinFaceW-I, KinFaceW-II, Cornell KinFace and UB KinFace show promising results over many state-of-the-art methods.

Keywords: Kinship verification · Convolutional neural networks · Maximum mean discrepancy · Adversarial Loss

1 Introduction

Human face carries with lots of individual information, and most human characteristics such as identity, age, gender, emotion etc. can be distinguished by facial images. Facial analysis has been widely studied in computer vision. Face verification aims to verify whether the two persons belong to the same family [1]. Biologists find that human facial appearance is an important cue for genetic similarity measurement. The purpose of kinship verification is to recognize whether the two persons are from the same family. It has many potential applications, such as missing children searching and social media mining, etc. [2]. In this work, the parent-child based kinship is studied, such as father-daughter, father-son, mother-daughter and mother-son. Some facial image pairs with kinship and no kinship have been shown in Fig.1, from which the difficulty of kin-relation discovery is shown.

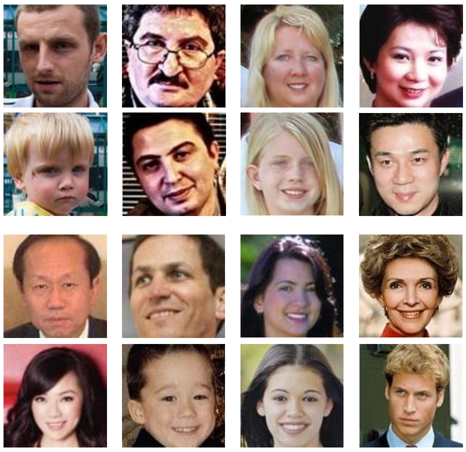

Fig. 1. Some positive (with kinship relation) and negative pairs (no kinship relation) from KinFaceW-I, KinFaceW-II, Cornell KinFace and UB KinFace, respectively. The first two rows are positive pairs and the last two rows are negative pairs. The kinship relation types from left to right are: father-daughter, father-son, mother-daughter and mother-son, respectively.

There are many algorithms proposed for kinship verification. Most of these work follow the technical routine from hand-crafted low-level feature extraction to large-margin metric learning. A representative work can be referred to as [2], in which a neighborhood repulsed metric learning (NRML) was proposed by learning a projection based metric with large margin and achieved excellent performance on kinship verification. Also, the hand-crafted features (e.g. LBP, HOG) are often used for general face analysis. However, this kinship verification algorithm strongly depends on the choice of metric learning, not the kin-relation specific features. As a result, the implicit and abstract kinship information cannot be adequately represented [3].

Deep learning, proposed by Hinton and Salakhutdinov [4], has become the most popular machine learning algorithm for discovering discriminative intermediate and high-level representations in a hierarchical manner [5]. In particular, convolutional neural networks have recently been shown to achieve great success in various computer vision tasks, such as face recognition [1, 6], object recognition, etc. Recently, CNNs have also been used for kinship verification [7, 3]. Although these work greatly promote kinship verification, they adopted a conventional CNN architecture. The loss functions are normally connected on the last fully-connected layer, but the distribution difference between the different input domains is not considered. The accuracy of kinship verification will be affected by this distribution difference. Maximum mean discrepancy (MMD) can be used to solve this problem, motivated by this fact, a MMD based loss is proposed in this paper. In addition, inspired by GAN, an adversarial loss is pro-

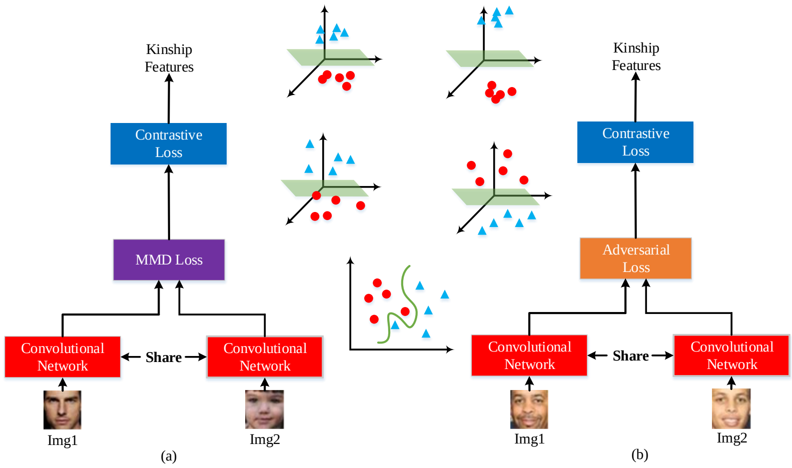

Fig. 2. Pipeline of our proposed approach. Circle denotes kinship pair, triangle denotes no-kinship pair.

posed to further improve the robustness and avoid overfitting. The pipeline of the proposed Adv-Kin methods is shown in Fig.2.

The key contributions of this work are threefold.

- We propose a new loss function (called MMD loss) to solve the problem of distribution difference in high-level features. With the joint supervision of the MMD based loss and the contrastive loss, the highly discriminative features can be obtained.

- In order to further improve the robustness of CNN model, inspired by GAN, an adversarial loss is proposed in this paper. The discrimination and robustness of deep features can be further enhanced by the game between the contrastive loss and the adversarial loss.

- Experimental comparisons with shallow and deep learning methods demonstrate that our methods outperform many state-of-the-art methods, and the gap of human-machine performance is further narrowed.

2 Related Work

In this section, we review two closely related topics with this paper: kinship verification and deep convolutional networks.

2.1 Kinship Verification

Kinship verification via facial image analysis is an challenging problem in computer vision. Existing feature representation approaches for kin-relation data include histogram of gradient (HOG) [8], scale-invariant feature transform (SIFT)

[2], and local binary pattern (LBP) [2]. Some algorithms aim to learn an effective metric or model for distinguishing whether two face images are with kinship relation, such as neighborhood repulsed metric learning (NRML) [2], prototype-based discriminative feature learning (PDFL) [9], transfer subspace learning [10, 11], support vector machine (SVM) [9], large margin multi-metric learning [12], ensemble similarity learning (ESL) [13], and scalable similarity learning (SS-L) [14]. Those previous works have achieved great progress over the challenging kinship verification. However, the common shortcoming is that the extracted image features are general representation of faces and lack of structural kin-relation meaning.

2.2 Deep Convolutional Networks

Deep learning has shown its effectiveness in face recognition. CNN is an end-to-end supervised learning methods from pixel based images to the high-level semantic. The features from the bottom to top in the network architecture can be identified from low-level and high-level image representation. Several popular CNN models are summarized as follows. MTCNN [15]used the candidate CNNs to detect facial landmarks. A Deepface [16] was proposed to solve 3D-align issue. FaceNet [1] constructed a triplet-loss model to improve the face verification accuracy. Recently, the center-loss model proposed in [6] aims to obtain within-class separable features. GAN [17] is a hot framework with generative and discriminative model via an adversarial process. SMCNN [3] achieved the kin-relation verification through a similarity metric based cost function. Although these algorithms achieved surprisingly good performance for computer vision, the progress of kinship verification is still insufficient.

3 Adv-Kin method

3.1 The Contrastive Loss

Table 1. Baseline Configuration.

Conv1	Pool1	Conv2	Pool2	Conv3	Pool3	Conv4	FC
conv11-6 conv12-6	max-2	conv21-16 conv22-16	max-2	conv31-30 conv32-30	max-2	conv4-60	FC1-128 FC2-80

In order to obtain kinship specific features, a siamese CNN that contains 4 convolutional layers is adopted. Each convolutional layers are followed by a max pooling layer. The input is a pair of 64×64 RGB kinship images. There are two fully-connected layers, and the discriminative deep features are drawn from the last fully-connected layer. For clarity, the CNN model with contrastive loss is termed as the baseline. The details of the baseline model are described in Table 1.

In baseline model, contrastive loss is acted as a supervisory signal. Let x_n^1, x_n^2 are the nth features of left and right kinship image, respectively. The contrastive loss function is presented as follows.

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

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 nth pair of kinship samples. The label is 1 if there is a kinship relation between two persons, otherwise 0. $margin$ is a adjustable parameter, which can control the maximal distance of negative pair.

Hence, it is concluded that the aim of contrastive loss is to train a model by pulling the positive pair as close as possible, while repulsing the negative pair as far as possible, simultaneously. In generic CNN model, contrastive loss normally acts on the last fully-connected layer only, but the distribution difference between the two fully-connected layers is not considered. The discrimination of deep features cannot be further improved under the influence of this distribution difference.

3.2 MMD based Adversarial Loss

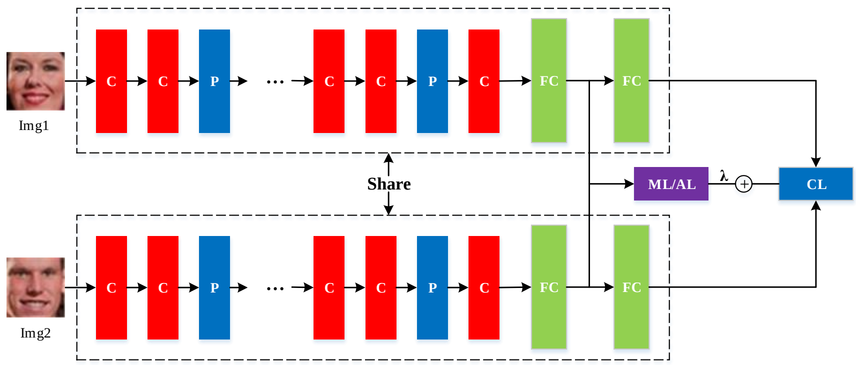

Fig. 3. The proposed Adv-Kin architecture.

MMD is a straight-forward test statistic to calculate the difference between distribution embeddings. It can be used to minimize the distribution difference between different domains on the domain adaptive issue [18]. 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 function $\phi(\cdot)$. The MMD between s and t is defined as

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

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 a RKHS $\hbar$. The most important property is that, we have $\mathrm{MMD}(s,t) = 0$ if and only if $s = t$.

Inspired by MMD and GAN, we propose the Adv-Kin method, as shown in Fig.3. The input of our CNN model is a pair of face images, one comes from parents, and the other from children. Thus, the distribution difference exists between parents domain and children domain. In order to minimize this difference, inspired by MMD, a MMD based loss is proposed as

$$L_M = \frac{1}{2N}\sum_{n=1}^{N}(y_n||\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)||_\hbar^2 - (1-y_n)||\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)||_\hbar^2) \tag{3}$$

However, some indirect approaches are also used to optimize the property of a system or a network. For example, robustness can be improved by introducing the additive interference. This thought is also applied to the CNN model, the performance of Generative Adversarial Nets (GAN) has been improved just by the adversarial process between generative model and discriminative model [17]. Inspired by GAN, in order to further improve the discrimination and robustness of deep features, the adversarial loss is proposed as

$$L_A = -\frac{1}{2N}\sum_{n=1}^{N}(y_n||\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)||_\hbar^2 - (1-y_n)||\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)||_\hbar^2) \tag{4}$$

We adopt the joint supervision of contrastive loss and adversarial loss to train the CNN model for kin-relation features learning, as formulated in Eq.(5).

$$\begin{aligned} L &= L_C + \lambda L_A \\ &= \frac{1}{2N}\sum_{n=1}^{N}(y_n d^2 + (1-y_n)\max(margin - d, 0)^2 \\ &+ \lambda(1-2y_n)(||\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)||_\hbar^2) \end{aligned} \tag{5}$$

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. Effected by game between adversarial loss and contrastive loss like Eq.(5), the robustness of deep features layer can be further improved.

By comparing MMD loss with adversarial loss, the only difference is the minus sign. So only the optimizion of adversarial loss is explained as follows.

In Eq.(4), $\phi(\cdot)$ denotes the feature map associated with the kernel map $k(\mathbf{x}_n^1, \mathbf{x}_n^2) = \langle \phi(\mathbf{x}_n^1), \phi(\mathbf{x}_n^2) \rangle$. Thus, the Eq.(4) can be rewritten as

$$L_A = \frac{1}{2N}\sum_{n=1}^{N}(1-2y_n)(k(\mathbf{x}_n^1, \mathbf{x}_n^1) + k(\mathbf{x}_n^2, \mathbf{x}_n^2) - 2k(\mathbf{x}_n^1, \mathbf{x}_n^2)) \tag{6}$$

Here, we adopt the Gaussian kernel function to optimize the proposed loss. The gradients of L_A with respect to $\mathbf{x}_n^1$ and $\mathbf{x}_n^2$ are computed respectively as:

$$\frac{\partial L_A}{\partial \mathbf{x}_n^1} = \frac{1}{N\sigma^2}(1-2y_n)\exp(-\frac{||\mathbf{x}_n^1-\mathbf{x}_n^2||_2^2}{2\sigma^2})(\mathbf{x}_n^1-\mathbf{x}_n^2) \tag{7}$$

$$\frac{\partial L_A}{\partial \mathbf{x}_n^2} = \frac{1}{N\sigma^2}(1-2y_n)\exp(-\frac{||\mathbf{x}_n^1-\mathbf{x}_n^2||_2^2}{2\sigma^2})(\mathbf{x}_n^2-\mathbf{x}_n^1) \tag{8}$$

4 Experiments

In this section, in order to demonstrate the effectiveness of our proposed approach, four benchmark kinship datasets are used.

4.1 Datasets

In experiments, KinFace data (4K) is considered, which includes four publicly available datasets, such as KinFaceW-I, KinFaceW-II [2], Cornell KinFace [8] and UB KinFace [19].

- Both KinFaceW-I and KinFaceW-II include four different types of kin relationships: father-son (F-S), father-daughter (F-D), mother-son (M-S) and mother-daughter (M-D). KinFaceW-I consists of 156, 134, 116, and 127 pairs, respectively. KinFaceW-II consists of 250 pairs for each relationship.

- Cornell KinFace contains totally 150 parent-child pairs.

- UB KinFace contains 200 triplets and each triplet is structured by child, young parent and old parent.

Table 2. Accuracy of different methods. ML and AL denote the Adv-Kin method with MMD loss and adversarial loss, respectively.

Methods	KinFaceW-I					KinFaceW-II					UB			Cor
	F-S	F-D	M-S	M-D	Mean	F-S	F-D	M-S	M-D	Mean	0-1	0-2	Mean	-
Human A	62.0	60.0	68.0	72.0	65.6	63.0	63.0	71.0	75.0	70.9	-	-	-	-
Human B	68.0	66.5	74.0	75.0	70.9	72.0	72.5	77.0	80.0	75.4	-	-	-	-
MNRML	72.5	66.5	66.2	72.0	69.6	76.9	74.3	77.4	77.6	76.5	<u>67.3</u>	<u>66.8</u>	<u>67.1</u>	71.6
MPDFL	73.5	67.5	66.1	73.1	70.1	77.3	74.7	77.8	78.0	77.0	**67.5**	**67.0**	**67.3**	71.9
SMCNN	75.0	75.0	68.7	72.2	72.7	75.0	79.0	78.0	85.0	79.3	-	-	-	-
DKV	71.8	62.7	66.4	66.6	66.9	73.4	68.2	71.0	72.8	71.3	-	-	-	-
Baseline	74.7	**77.6**	72.4	81.1	76.5	<u>85.8</u>	<u>85.8</u>	84	83.8	84.5	58.3	60.0	59.2	76.2
ML	**77.3**	74.6	**78.0**	<u>83.6</u>	<u>78.4</u>	<u>85.8</u>	84.6	<u>86.6</u>	**88.0**	<u>86.3</u>	59.8	61.0	60.4	<u>78.3</u>
AL	<u>76.9</u>	<u>77.3</u>	<u>75.8</u>	**85.9**	**80.0**	**86.2**	**86.2**	**87.4**	<u>87.0</u>	**86.9**	60.3	63.8	62.1	**79.6**

Table 3. Accuracy of ML and AL with different bandwidth σ^2.

Methods	σ^2	KinFaceW-I					KinFaceW-II				
		F-S	F-D	M-S	M-D	Mean	F-S	F-D	M-S	M-D	Mean
ML	0.5	**77.3**	74.6	**78.0**	**83.6**	**78.4**	**85.8**	**84.6**	**86.6**	**88.0**	**86.3**
ML	1.0	75.0	**76.9**	73.7	83.5	77.3	84.4	84.0	85.4	87.0	85.2
ML	2.0	75.4	76.8	75.9	82.3	77.6	84.8	83.2	85.2	**88.0**	85.2
AL	0.5	**75.3**	74.3	75.4	79.5	76.1	85.2	82.6	84.0	86.8	84.7
AL	1.0	74.4	**76.9**	75.9	**85.0**	**78.1**	**87.4**	85.0	**86.8**	**87.0**	**86.6**
AL	2.0	74.7	74.3	**77.1**	81.4	76.9	85.8	**85.4**	83.4	**87.0**	85.4

4.2 Experimental Setup

In experiments, the proposed models are trained on KinFace via 5-fold cross validation, and finally NRML metric [2] is used for kinship verification. The mini-batch Stochastic Gradient Descent (SGD) based error back propagation algorithm is used for training, with an initial learning rate of 10^{-2}. The Batch size is 151 images, and the *margin* of contrastive loss is set as 1.

We have compared Adv-Kin method with four state-of-the-art methods in kinship verification, including two shallow learning methods such as MNRML [2] and MPDFL [9], and two deep learning methods such as SMCNN [3] and DKV [7]. Additionally, the performance comparison with human score [9] is also analyzed. Notably, for all algorithms, 5-fold cross-validation is used by following the standard setting.

4.3 Comparison with Previous Methods

The verification results of the proposed MMD Loss (i.e. ML) and Adversarial Loss (i.e. AL) on four benchmark kinship datasets have been shown in Table 2. Specifically, from the results listed in Table 2, we can observe that:

- The proposed Adv-Kin methods consistently outperform state-of-the art face verification methods, i.e. MNRML and MPDFL based on feature ensemble and metric learning. The effectiveness of high-level kin-relation semantic discovery is demonstrated.

- The proposed Adv-Kin methods also outperform the deep learning based face verification methods, i.e. SMCNN and DKV which are modeled under the generic loss.

- By comparing our method with human knowledge on the KinFaceW-I and KinFaceW-II, the results show that our methods achieve even better performance than human.

- By comparing ML with AL, we get that AL based Adv-Kin shows superiority to ML based Adv-Kin. Thus it can be seen that the robustness of deep features can be further improved by introducing adversarial characteristic.

- For UB dataset, the accuracy of our methods is lower than MNRML and MPDFL. The reason may be that UB datasets consists of triplet samples, the contrastive loss may not distinguish the positive samples and negative samples in this dataset. However, the result of Adv-Kin methods are better than contrastive loss based baseline, it means that our methods still work in UB data.

4.4 Parameter Analysis

The parameter σ^2 is first to be investigated. Table 3 shows the accuracy of ML and AL versus different bandwidth σ^2. It can be seen that ML and AL based Adv-Kin can obtain the best classification performance when σ^2 is 0.5 and 1.0, respectively. We can also observe that the proposed methods demonstrate a stable recognition performance with different bandwidth σ^2.

After fixing σ^2, we also evaluate the performance with different loss weight λ. Table 4 shows the accuracy of ML and AL versus different loss weight λ. We can see that ML and AL based Adv-Kin can obtain the best classification performance when λ is set as 2.0 and 0.2, respectively.

Table 4. Accuracy of ML and AL with different loss weight λ.

Methods	λ	KinFaceW-I					KinFaceW-II				
		F-S	F-D	M-S	M-D	Mean	F-S	F-D	M-S	M-D	Mean
ML	0.2	76.0	73.9	73.6	80.7	76.1	83.8	83.8	82.6	83.4	83.4
ML	1.0	75.0	**76.5**	75.0	81.9	77.1	84.0	83.8	**87.0**	86.2	85.3
ML	2.0	**77.3**	74.6	**78.0**	**83.6**	**78.4**	**85.8**	**84.6**	86.6	**88.0**	**86.3**
AL	0.2	**76.9**	**77.3**	**75.8**	**85.9**	**80.0**	85.2	85.4	**88.6**	**88.0**	86.8
AL	1.0	76.3	**77.3**	74.6	84.6	78.2	**86.2**	**86.2**	87.4	87.0	**86.9**
AL	2.0	74.4	76.9	75.9	85.0	78.1	87.4	85.0	86.8	87.0	86.6

5 Conclusion

In this paper, we propose two loss functions as supervisory signals for kinship verification, which is motivated by the MMD and GAN. The performance of CNN model for kinship verification can be improved by use of the MMD loss to minimize the distribution difference between parents domain and children domain. In order to improve the discrimination and robustness of deep features, inspired by GAN, the adversarial loss is proposed. Extensive experiments on the benchmark KinFaceW-I, KinFaceW-II, Cornell KinFace and UB KinFace show the promising results compared to many state-of-the-art methods. In future, the combined deep learning and metric learning will be studied.

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

References

1. F. Schroff, D. Kalenichenko, and J. Philbin.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp:815-823 (2015)
2. J. Lu, X. Zhou, Y. P. Tan, Y. Shang, and J. Zhou.: Neighborhood repulsed metric learning for kinship verification. In: IEEE Transactions on Pattern Analysis & Machine Intelligence, pp:331-345 (2014)
3. L. Li, X. Feng, X. Wu, Z. Xia, and A. Hadid.: Kinship Verification from Faces via Similarity Metric Based Convolutional Neural Network. In: Image Analysis and Recognition, pp: 539-548 (2016)
4. G. E. Hinton and R. R. Salakhutdinov.: Reducing the dimensionality of data with neural networks. In: Science, pp: 504-507 (2006)
5. X. Glorot, A. Bordes, and Y. Bengio.: Domain adaptation for large-scale sentiment classification: A deep learning approach. In: Proceedings of the 28th International Conference on Machine Learning, pp: 513-520 (2011)
6. Y. Wen, K. Zhang, Z. Li, and Y. Qiao.: A discriminative feature learning approach for deep face recognition. In: Computers & Operations Research, pp: 11-26 (2016)
7. M. Wang, Z. Li, X. Shu, and J. Wang.: Deep kinship verification. In: IEEE International Workshop on Multimedia Signal Processing, pp: 1-6 (2015)
8. R. Fang, K. D. Tang, N. Snavely, and T. Chen. Towards computational models of kinship verification. In: 17th IEEE International Conference on Image Processing, pp: 1577-1580 (2010)
9. H. Yan, J. Lu, and X. Zhou.: Prototype-based discriminative feature learning for kinship verification. In: IEEE Transactions on Cybernetics, pp: 2535-2545 (2015)
10. M. Shao, S. Xia, and Y. Fu.: Genealogical face recognition based on ub kinface database. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp: 60-65 (2011)
11. S. Xia, M. Shao, and Y. Fu.: Kinship verification through transfer learning. In: International Joint Conference on Artificial Intelligence, pp: 2534-2544 (2011)
12. J. Hu, J. Lu, J. Yuan, and Y. P. Tan.: Large margin multimetric learning for face and kinship verification in the wild. In: Asian Conference on Computer Vision. Springer International Publishing, 252-267 (2014)
13. X. Zhou, Y. Shang, H. Yan, and G. Guo.: Ensemble similarity learning for kinship verification from facial images in the wild. In: Information Fusion, pp:40-48 (2016)
14. X. Zhou, H. Yan, and Y. Shang.: Kinship verification from facial images by scalable similarity fusion. In: Neurocomputing, pp:136-142 (2016)
15. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao.: Joint face detection and alignment using multi-task cascaded convolutional networks. In: IEEE Journal of Solid-State Circuits, pp:1161-1173 (2016)
16. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf.: Deepface: Closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp: 1701-1708 (2014)
17. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D.: Warde-Farley, S. Ozair, , C. Aaron, Y. Bengio. Generative adversarial nets. In: Advances in neural information processing systems, pp: 2672-2680 (2014)
18. M. Long, Y. Cao, J. Wang, M. Jordan.: Learning transferable features with deep adaptation networks. In: International Conference on Machine Learning, pp:97-105 (2015)
19. S. Xia, M. Shao, and Y. Fu.: Kinship verification through transfer learning. In: International Joint Conference on Artificial Intelligence, pp: 2539-2544 (2011)