Kinship Verification with Deep Convolutional Neural Networks

Kaihao Zhang[12]
kaihao.zhang@nlpr.ia.ac.cn
Yongzhen Huang[2]
yzhuang@nlpr.ia.ac.cn
Chunfeng Song[2]
chunfeng.song@nlpr.ia.ac.cn
Hong Wu[1]
hwu@uestc.edu.cn
Liang Wang[2]
wangliang@nlpr.ia.ac.cn

[1] Statistical Machine Intelligence & LEarning (SMILE)
Big Data Research Center
University of Electronic Science and Technology of China
Chengdu, China

[2] National Laboratory of Pattern Recognition (NLPR)
Institute of Automation,
Chinese Academy of Sciences
Beijing, China

Abstract

Kinship verification from facial images is an interesting and challenging problem. The current algorithms on this topic typically represent faces with multiple low-level features, followed by a shallow learning model. However, these general manual features cannot well discover information implied in facial images for kinship verification, and thus even current best algorithms are not satisfying. In this paper, we propose to extract high-level features for kinship verification based on deep convolutional neural networks. Our method is end-to-end, without complex pre-processing often used in traditional methods. The high-level features are produced from the neuron activations of the last hidden layer, and then fed into a soft-max classifier to verify the kinship of two persons. Considering the importance of facial key-points, we also extract key-points-based features for kinship verification. Experimental results demonstrate that our proposed approach is very effective even with limited training samples, largely outperforming the state-of-the-art methods. On two most widely used kinship databases, our method achieves 5.2% and 10.1% improvements compared with the previous best one, respectively.

1 Introduction

Face verification in unconstrained conditions has obtained increasing attention and encouraging progress in recent years [2, 3, 11, 18, 25]. Biologists find that human facial appearance is an important cue for genetic similarity measurement [5, 6, 12]. Motivated by this finding and related applications such as social media analysis, missing children searching, children adoptions and finding imitation [10, 23, 26], kinship verification through facial image analysis has attracted more and more attention over the past few years [7, 8, 10, 14, 15, 21, 22, 24, 26, 27]. Some kinds of kinship from the KinFaceW database [14] are shown in Figure 1.

Pages 148.1-148.12
DOI: https://dx.doi.org/10.5244/C.29.148

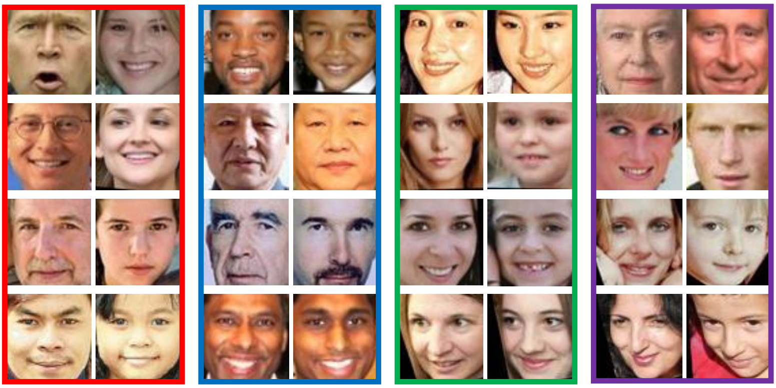


Figure 1: An illustration of some kinds of kinship from two databases. Face images from the first to second row are from the KFW-I database. Face images from the third to fourth row are from the KFW-II database.

The current best-performing kinship verification algorithms typically represent faces with over-complete low-level features and strongly depend on the choice of metric learning algorithms. However, such low-level features cannot well represent the underlying information of kinship implied in facial images, and thus the performance of current kinship verification is still unsatisfying. Meanwhile, we find that deep models play a key role in effectively extracting high-level features for face identification [2, 11, 18]. Based on the existing studies on kinship verification and the great progress of general face identification, we attempt to use deep models for addressing the problem of face-based kinship verification. In particular, we propose to use the powerful deep convolutional neural networks (CNN) for kinship verification, in which a group of CNNs are simultaneously learned to verify different kinds of kinship. In our method, a pair of facial images from intra-class (with kinship relations) or inter-class samples (without kinship relations) are sent to CNNs, with a fully-connected layer to generate a feature representation, which is finally fed into a two-way soft-max classifier to predict the kinship. Experimental results demonstrate that our methods for kinship verification largely outperform both the previous best machine algorithms and human prediction. Besides, we find that key-points of faces, e.g., eyes, mouth, and nose, are especially useful for kinship verification. This is possibly due to the phenomenon that if two persons are with kinship relations, they are often similar in some facial key-points instead of the whole face. This is different from general face recognition. Due to the importance of key-points, we also design key-points-based feature representation based on the deep CNNs. The results show that facial key-points can further improve the accuracy of kinship verification.

The main contributions of this paper are three folds. First, to the best of our knowledge, we are the first to study kinship verification with deep learning, and demonstrate that deep CNNs is a very effective solution even with very limited samples. Second, key-points-based face representation is introduced for kinship verification, which further improves the accuracy of kinship verification. Third, we extensively evaluate various methods, and find that the proposed method largely boosts the state-of-the-art level of kinship verification, with 5.2% and 10.1% enhancements over the previous best methods on two most widely used databases, respectively.

2 Related Work

Our work in this paper is closely related with kinship verification and deep convolutional neural networks, which are briefly introduced as follows, respectively.

2.1 Kinship Verification

In the past few years, many vision researchers have investigated the problem of kinship verification via facial image analysis, and many papers have been published in top journals/ conferences [7, 14, 15, 22, 24, 26, 27]. The earliest attempt to solve kinship verification is based on local facial feature extraction and selection [8]. Firstly, the key parts of a face are localized, and then multiple features (e.g., color, gray value, histogram of gradients) are extracted with classic classifiers such as k-nearest-neighbor (KNN) or support vector machine (SVM) for kinship classification. In order to obtain better features, some improvements are proposed, like Gabor gradient orientation pyramid [26], salient part and self-similarity [7] and dynamic expressions [10]. Also, some researchers try to solve kinship verification with more effective learning models. For example, Yan et al. learn multiple discriminative metrics for kinship verification, and achieve the best results [24]. Though the performance keeps increasing, these studies are mainly based on low-level features and shallow models, whose results are not sufficiently satisfying. And to the best of our knowledge, the powerful deep learning has not yet been explored in kinship verification based on facial image analysis.

2.2 Deep Convolutional Neural Networks

Recently, deep learning has shown its effectiveness in various vision tasks, such as image classification, object recognition and face verification. Krizhevsky et al. [13] propose an 8-layer network which significantly outperforms other methods. Szegedy et al. [19] replace the network with a deeper convolution neural network called GoogLeNet. Girshick et al. [9] propose a method called R-CNN to solve the problem of object recognition. They use an image segmentation technique to find candidate image regions, then use an AlexNet to classify those candidates. Szegedy et al. [20] improve R-CNN by increasing the selective search proposals with candidate image regions and replace AlexNet with GoogLeNet. For the task of face verification, Huang et al. propose a deep model without supervision [11]. Cai et al. apply deep nonlinear metrics to enhance verification [2]. Some researchers use the structure with more than one network for face verification. The Siamese network [1] is a discriminative network using two deep convolution neural networks to extract features from two input facial images respectively. The distance between the outputs of the two sub-networks is defined as the dissimilitude. Different from [1], Sun et al. adopt two deep neural networks to solve face verification [18], in which one neural network is used to learn features through face identification, and another neural network is for face verification with the generated features as input. These studies achieve surprisingly good performance for face identification and verification, which motivates us to exploit a deep convolutional neural network structure for kinship verification.

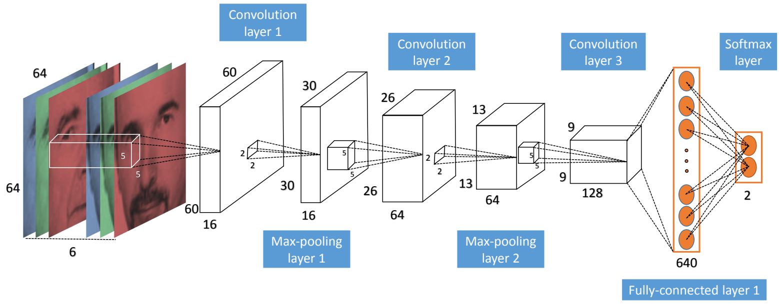


Figure 2: The proposed architecture of basic CNN for kinship verification. For all layers, the length of each cuboid is the map number, and the width and height of each cuboid are the dimension of each map. The inside small cuboids and squares denote the 3D convolution kernel sizes and the 2D pooling region sizes of convolutional and pooling layers. The input is a pair of RGB images and the output is a two-value label.

3 Our method

In our method, we first propose a basic structure of CNN. Considering the importance of facial key-points on kinship verification, we modify the basic structure to a new structure of CNN based on key-points.

3.1 Basic Structure of CNN

Overview of the structure. The basic structure of CNN (CNN-Basic) used in this work contains three convolutional layers, followed by a fully-connected layer and a soft-max layer. As shown in Figure 2, the input is a pair of 64×64 images with three channels (RGB). Following the input, the first convolutional layer is generated after convolving the input via 16 filters with a stride of 1. Each filter is with the size $5 \times 5 \times 6$. The second convolutional layer filters the input of the previous layer with 64 kernels of size $5 \times 5 \times 16$. The third convolutional layer contains 128 kernels of the size $5 \times 5 \times 64$. After the convolutional layers, a fully-connected layer projects the extracted features into a subspace with 640 neurons. Max-pooling layers follow the first and second convolutional layers. We adopt the ReLU function as the activation function of the convolution layers. Finally, this network is trained via a two-way soft-max classifier at the top layer. Some operations in CNN are explained as follows.

Activation function. The widely used activation function to model a neuron's output is the sigmoid function. However, considering the training time with the gradient descent algorithm, the non-saturating nonlinearity is much faster than this kind of saturating nonlinearity. And thus we adopt the ReLU function as the activation function of neurons, which has been shown to achieve better performance than the sigmoid function. With ReLU, the convolution operation is formulated as

$$y^{j(r)} = \max\left(0, b^{j(r)} + \sum_i w^{ij(r)} * x^{i(r)}\right), \tag{1}$$

where x^i and y^j are the i-th input map and the j-th output map, respectively. w^{ij} denotes the

Figure 3: The ten face regions used in our network. The images on the top row are five key-point regions. The images on the bottom row are the original image and its four local regions, i.e., the top-left corner, the top-right corner, bottom-left corner, bottom-right corner.

weight between the i-th input map and the j-th output map. b^j is the bias of the j-th output map, and $\times$ denotes the convolutional operation.

Max-pooling. In general, the pooling layer summarizes the outputs of neighboring groups of a feature map via down-sampling. The two frequently used pooling methods are max-pooling and average-pooling. In this paper, we choose max-pooling with a neighboring region size of 2×2. Max pooling is helpful to increase the translation invariance and avoid over-fitting, which is defined as

$$y^i_{j,k} = \max_{0\leq m,n\leq s} \left\{ x^i_{j\cdot s+m,k\cdot s+n} \right\}, \tag{2}$$

where $y^i_{j,k}$ denotes the outputs of the i-th feature map in the location of (j,k). Similarly, $x^i_{j,k}$ denotes the value of location (j,k) in the i-th feature map.

Implementation details. The CNN is trained by back-propagation with logistic loss over the predicted scores using the soft-max function. To initialize weights, we use a Gaussian distribution with zero mean and a standard deviation of 0.01. The biases are initialized as zeros. In each iteration, we update all the weights after learning the mini-batch with the size of 128. In all layers, the momentum is set as 0.9 and the weight decay is set as 0.005. To expand the training set, we also randomly flip images and add grayscales during training.

3.2 Key-points Structure of CNN

When a subject is demanded to verify the kinship from two face images, it is highly possible that the key-points are focused, such as their eyes, mouth and nose. We consider that the facial key-points have a significant impact on kinship analysis, and thus design a key-points-based feature representation for kinship verification. In particular, we detect the centers of two eyes, the corners of the mouth and the nose with a facial point detection algorithm [17]. Then each face image is cropped and aligned according to the five key-points. To extract more complementary information, we also crop other five face regions without key-points detection. The five images are the original image and its four local regions, i.e., the top-left corner, the top-right corner, bottom-left corner, bottom-right corner. Figure 3 shows the ten face regions.

In order to improve kinship verification with these face regions, we propose a new structure (CNN-Points) which is shown in Figure 4. The new structure contains 10 basic CNNs (see Figure 2), each of which receives a pair of face regions. Ten sets of 640-dimensional features are produced from the last hidden layer of the basic CNNs. The last hidden layer of

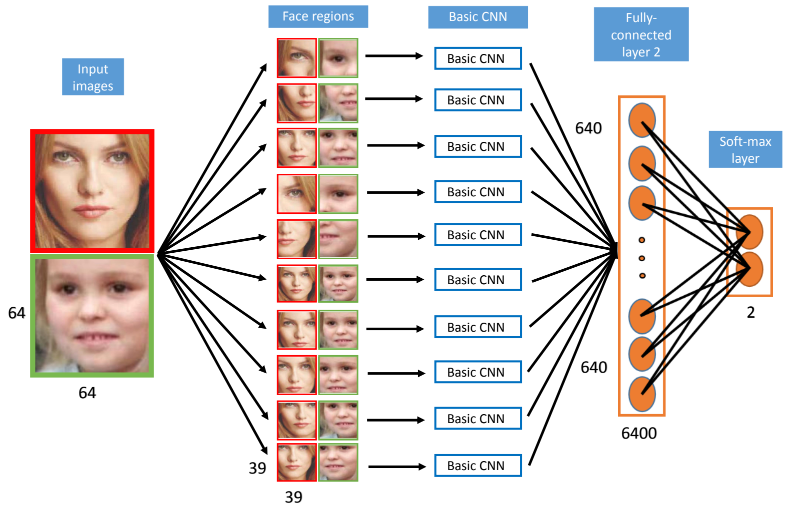


Figure 4: Overview of the proposed CNN-Points structure for kinship verification. The input is a pair of RGB images, which are cropped into ten face regions and fed into different basic CNNs. Then ten 640-dimensional feature representations are generated. After a fully-connected layer, the final representation of the relationship becomes to be a 6400-dimensional representation and followed by a soft-max layer for predicting their kinship.

the CNN-Points is fully-connected to the ten basic CNNs, which is defined as

$$y^i = f\left(\sum_{k=1}^{10}\sum_{j=1}^{640} w^i_{j,k} * x_{j,k} + b^i\right), \tag{3}$$

where y^j is the output of the i-th neuron activation, $w^i_{j,k}$ denotes the weight between the input features and the i-th neuron, and $f(\cdot)$ is chosen to be the sigmoid function. The final representation is 6400-dimensional features, and fed into a soft-max classifier to predict the kinship of two persons.

4 Experiments

To show the effectiveness of the proposed deep CNNs, we conduct a series of experiments on two publicly available kinship datasets.

4.1 Dataset

Two publicly available kinship datasets (KFW-I, KFW-II) [14] are used to evaluate our method. These two datasets are the most widely used databases for kinship verification. The difference between them is that each pair of facial images in KFW-I is collected from different pictures whereas that in KFW-II is collected from the same picture. In addition, KFW-II is larger than KFW-I in size. There are four types of kinship in the two datasets: Father-Son (FS), Father-Daughter (FD), Mother-Son (MS), and Mother-Daughter (MD). In

Fold	KFW-I				KFW-II
	FD	FS	MD	MS	all subset
1	[1,27]	[1,31]	[1,25]	[1,23]	[1,50]
2	[28,54]	[32,64]	[26,50]	[24,46]	[51,100]
3	[55,81]	[65,96]	[51,75]	[47,69]	[101,150]
4	[82,108]	[97,124]	[76,101]	[70,92]	[151,200]
5	[109,134]	[125,156]	[102,127]	[93,116]	[201,250]

Table 1: The face index of the five folds cross-validation on the KFW-I and the KFW-II databases.

KFW-I dataset, there are 156, 134, 116, and 127 pairs of facial images for these four relations, respectively. In KFW-II dataset, each kinship relationship contains 250 pairs of facial images. Some examples from KFW-I and KFW-II are shown in Figure 1.

4.2 Experimental Setups

According to the provided eyes' positions in the two datasets, face images are aligned and cropped into 64×64 pixels. Although there are several validation methods on the two datasets [15, 16], we follow the standard protocol provided in [15]. In this standard protocol, the datasets are equally divided into five folds. We perform five-fold cross validation experiments on the datasets.

Table 1 shows the index of face pairs on the two datasets in our experiments. For face images in each fold of these datasets, all pairs of face images are picked out to generate positive and negative samples. Namely, the positive samples are the true pairs of face images (one from the parent and one from others' children), and the negative samples are the false pairs of face images (one from the parent and the other from a child whose true parent is not him/her). Obviously, the number of negative samples is much larger than that of the positive samples. So we adopt a sample balancing operation, by which the numbers of positive and negative samples are almost the same in each batch sent to CNNs.

We have compared our method with four widely used metric learning methods and multi-metric learning methods which are proposed to address kinship verification, namely Concatenated Metric Learning (CML), Local Discriminative Distance Metrics (LDDM), Multi-feature Neighborhood Repulsed Metric Learning(MNRML), and Discriminative Multi-Metric Learning (DMML) [4, 15, 24]. To show an important baseline, we also compare our method with the human for kinship verification from facial images [14].

4.3 Results and Analysis

Table 2 and Table 3 show the verification results of different kinship verification algorithms on KFW-I and KFW-II, respectively. Our method significantly outperforms the state-of-the-art method. The previous best method used for kinship verification is DMML [24]. Results listed in Table 2 indicate that our method increases the previous best method by 2.3% for the FD subset, 1.6% for the FS subset, 8.6% for the MD subset, 8.5% for the MS subset and 5.2% for the mean accuracy on the KFW-I dataset. The results listed in Table 3 indicate that our method increases the previous best method by 5.4% for the FD subset, 10.9% for the FS subset, 12.9% for the MD subset, 11.4% for the MS subset and 10.1% for the mean accuracy on the KFW-II dataset.

Table 2 and Table 3 also show comparison between different machine algorithms and humans for kinship verification. The human performance is reported by Lu et al. [14], in which they firstly select 50 positive samples and 50 negative samples from each of the four subsets of KFW-I and KFW-II, and then choose 10 humans (5 males and 5 females) for kinship verification. All of them are 20-30 years old and have never received training for kinship verification before the experiment. In Table 2 and Table 3, HumanA means that only the cropped face regions are presented to humans which are the same as that sent to our CNNs. HumanB denotes that the whole original facial images are presented to observers, so that observers can use additional information such as hair and backgrounds.

According to the results shown in Table 2 and Table 3, several conclusions could be drawn as follows.

1. Our proposed method largely outperforms the other four machine algorithms including the state-of-the-art method, which implies that the deep convolution neural networks is a feasible and effective method to solve the problem of kinship verification via facial images.

2. The method of CNN-Points further improves kinship verification of CNN-Basic. The reason is that they can make good use of facial key-points to learn richer facial features for kinship verification. The results show that facial key-points contain effective information for kinship verification.

3. It is clear that the accuracy on KFW-II is much higher than that on KFW-I for all machine algorithms. However, the improvement for human performance is not so obvious. The reasons are two-folds. On the one hand, the data size of KFW-II is larger than that of KFW-I, which is beneficial to improve the performance of machine algorithms. On the other hand, humans have a lot of prior knowledge about face recognition before kinship verification, and thus their performance does not strongly rely on the data size as that of machine algorithms.

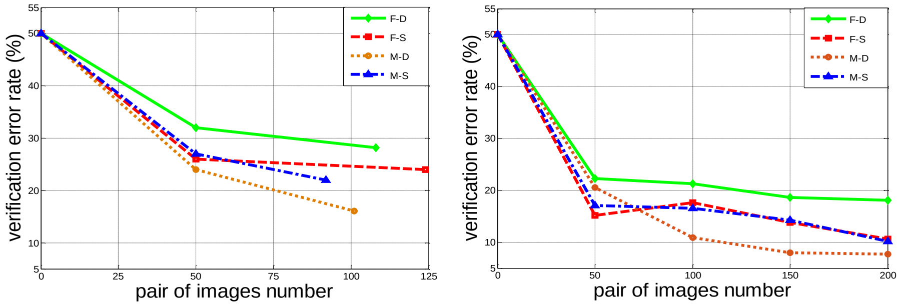
We also investigate the influence of the data size on our proposed methods. Figure 5 shows the error rate of CNN-Points with respective to different kinds of kinship verification on KFW-I and KFW-II, respectively. We can observe from the figure that the performance

Method	FD	FS	MD	MS	Mean	V.S. DMML
CML [24]	65.5%	69.5%	72.0%	64.5%	67.9%	-4.4%
IML [24]	67.5%	70.5%	72.0%	65.5%	68.9%	-3.4%
MNRML [15]	66.5%	72.5%	72.0%	66.2%	69.3%	-3.0%
DMML [24]	69.5%	74.5%	75.5%	69.5%	72.3%	0.0%
HumanA [14]	58.0%	61.0%	70.0%	66.0%	63.8%	-8.5%
HumanB [14]	65.0%	67.0%	77.0%	75.0%	71.0%	-1.3%
CNN-Basic	**70.8%**	**75.7%**	**79.4%**	**73.4%**	**74.8%**	**+2.5%**
CNN-Points	**71.8%**	**76.1%**	**84.1%**	**78.0%**	**77.5%**	**+5.2%**

Table 2: Accuracy of different methods on KFW-I. The previous best method is DMML [24].

Method	FD	FS	MD	MS	Mean	V.S. DMML
CML [24]	73.0%	73.5%	76.5%	76.0%	74.8%	-3.5%
IML [24]	74.0%	74.5%	78.5%	76.5%	75.9%	-2.4%
MNRML [15]	74.3%	76.9%	77.6%	77.4%	76.6%	-1.7%
DMML [24]	76.5%	78.5%	79.5%	78.5%	78.3%	0.0%
HumanA [14]	61.0%	61.0%	73.0%	69.0%	66.8%	-11.5%
HumanB [14]	68.0%	70.0%	80.0%	78.0%	74.0%	-4.3%
CNN-Basic	**79.6%**	**84.9%**	**88.5%**	**88.3%**	**85.3%**	**+7.0%**
CNN-Points	**81.9%**	**89.4%**	**92.4%**	**89.9%**	**88.4%**	**+10.1%**

Table 3: Accuracy of different methods on KFW-II. The previous best method is DMML [24].



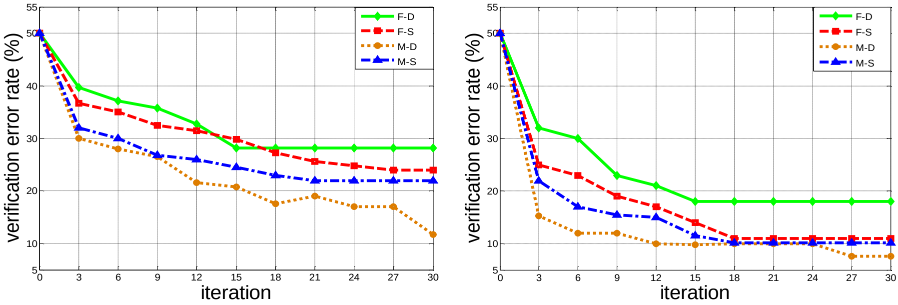
(a) (b)

Figure 5: I. Four kinds of kinship verification accuracy of CNN-Points versus different numbers of images on the KFW-I (a) and KFW-II (b), respectively.

becomes better as the data size increases. The experimental results imply that the data size has a significant influence on the final accuracy. Since the optimization of CNN is based on iterative optimazation, we also examine the performance with different numbers of iterations. Figure 6 shows the error rate of four kinds of kinship verification as the iteration increases on KFW-I and KFW-II, respectively. It is clear that the error rates of all cases quickly decline in the early several times of iterations and become stable after about 20 times of iterations on two datasets. The experimental results demonstrate that the CNNs are well trained for all cases of kinship verification.

5 Conclusion and Future Work

In this paper, we have proposed deep CNNs to address kinship verification via facial image analysis. The proposed model generates effective high-level features related with key-points-based representations. Experimental results demonstrate that the proposed method largely enhances the state-of-the-art performance, and outperforms human performance. In future, we intend to explore more deep neural networks structures, and more data for pre-training.



(a) (b)

Figure 6: The relationship between training sizes and the error rate of four kinds of kinship verification with CNN-Points on the KFW-I (a) and KFW-II (b), respectively.

6 Acknowledgement

This work is jointly supported by National Basic Research Program of China (2012CB316300), National Natural Science Foundation of China (61135002), SAMSUNG GRO Program, CCF- Tencent Open Fund, 360 OpenLab Program and Open Project Program of the State Key Laboratory of Mathematical Engineering and Advanced Computing.

References

[1] J. Bromley, J. W. Bentz, L. Bottou, I. Guyon, Y. LeCun, C. Moore, E. Säckinger, and R. Shah. Signature verification using a "siamese" time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(04):669–688, 1993.

[2] X. Y. Cai, C. H. Wang, B. H. Xiao, X. Chen, and J. Zhou. Deep nonlinear metric learning with independent subspace analysis for face verification. In *Proceedings of the 20th ACM International Conference on Multimedia*, pages 749–752. ACM, 2012.

[3] X. D. Cao, D. Wipf, F. Wen, G. Q. Duan, and J. Sun. A practical transfer learning algorithm for face verification. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3208–3215. IEEE, 2013.

[4] R. G. Cinbis, J. Verbeek, and C. Schmid. Unsupervised metric learning for face identification in tv video. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1559–1566. IEEE, 2011.

[5] M. F. Dal Martello and L. T. Maloney. Lateralization of kin recognition signals in the human face. *Journal of Vision*, 10(8):9, 2010.

[6] L. M. DeBruine, F. G. Smith, B. C. Jones, S. Craig Roberts, M. Petrie, and T. D. Spector. Kin recognition signals in adult faces. *Vision Research*, 49(1):38–43, 2009.

[7] H. Dibeklioglu, A. A. Salah, and T. Gevers. Like father, like son: Facial expression dynamics for kinship verification. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1497–1504. IEEE, 2013.

[8] R. G. Fang, K. D. Tang, N. Snavely, and T. Chen. Towards computational models of kinship verification. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1577–1580. IEEE, 2010.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 580–587. IEEE, 2014.

[10] G. D. Guo and X. L. Wang. Kinship measurement on salient facial features. *Instrumentation and Measurement, IEEE Transactions on*, 61(8):2322–2325, 2012.

[11] G. B. Huang, H. Lee, and E. Learned-Miller. Learning hierarchical representations for face verification with convolutional deep belief networks. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2518–2525. IEEE, 2012.

[12] G. Kaminski, S. Dridi, C. Graff, and E. Gentaz. Human ability to detect kinship in strangers' faces: effects of the degree of relatedness. *Proceedings of the Royal Society B: Biological Sciences*, 2009.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[14] J. W. Lu, J. L. Hu, X. Z. Zhou, Y. Y. Shang, Y. P. Tan, and G. Wang. Neighborhood repulsed metric learning for kinship verification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2594–2601. IEEE, 2012.

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

[16] J. W. Lu, J. L. Hu, V. E. Liong, X. Z. Zhou, A. Bottino, I. Islam, T. F. Vieira, X. Q. Qin, X. Y. Tan, S. C. Chen, et al. The fg 2015 kinship verification in the wild evaluation. *Proc. FG*, 2015.

[17] Y. Sun, X. L. Wang, and X. O. Tang. Deep convolutional network cascade for facial point detection. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3476–3483. IEEE, 2013.

[18] Y. Sun, X. G. Wang, and X. O. Tang. Deep learning face representation from predicting 10,000 classes. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1891–1898. IEEE, 2014.

[19] C. Szegedy, W. Liu, Y. Q. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*, 2014.

[20] Christian Szegedy, Scott Reed, Dumitru Erhan, and Dragomir Anguelov. Scalable, high-quality object detection. *arXiv preprint arXiv:1412.1441*, 2014.

[21] X. L. Wang and C. Kambhamettu. Leveraging appearance and geometry for kinship verification. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 5017–5021. IEEE, 2014.

[22] S. Y. Xia, M. Shao, and Y. Fu. Kinship verification through transfer learning. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, page 2539, 2011.

[23] S. Y. Xia, M. Shao, J. B. Luo, and Y. Fu. Understanding kin relationships in a photo. *Multimedia, IEEE Transactions on*, 14(4):1046–1056, 2012.

[24] H. B. Yan, J. B. Lu, W. H. Deng, and X. Z. Zhou. Discriminative multimetric learning for kinship verification. *Information Forensics and Security, IEEE Transactions on*, 9 (7):1169–1178, 2014.

[25] Q. Yin, X. O. Tang, and J. Sun. An associate-predict model for face recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 497–504. IEEE, 2011.

[26] X. Zhou, J. Lu, J. Hu, and Y. Shang. Gabor-based gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments. In *ACM International Conference on Multimedia*, pages 2322–2325, 2012.

[27] X. Z. Zhou, J. L. Hu, J. W. Lu, Y. Y. Shang, and Y. Guan. Kinship verification from facial images under uncontrolled conditions. In *Proceedings of the 19th ACM International Conference on Multimedia*, pages 953–956. ACM, 2011.