Kinship verification based on hand-craft features regression and multi-level feature fusion

Shizhe Cai 36920220156883*, Hanyu Guo 23020221154086*, Wenyan He23020221154087*, Xiying Zheng 23020221154152*, Wanchuan Yu 23020221154140*

Department of Computer Science, Xiamen University, China
wenyanhe928@163.com

Abstract

Kinship verification based on face images is an important subfield of face image analysis. It is valuable in many practical applications, such as smuggling children investigation and criminal tracking. Traditional kinship verification methods usually only use hand-craft face features. However, these methods often fail to make good use of the deep information hidden in the face image. In recent years, with the rapid development of deep learning, various kinship verification models, which are all based on convolutional neural networks, have been proposed. These models can extract the deep-level and more abstract features of face images, but most of them directly ignore the shallow features of face such as edge and texture features.Aiming at the shortcomings of traditional hand-craft features and the latest deep learning based kinship verification methods, we propose a multi-level feature knowledge mining model ($\mathrm{HF^2KM^2}$) for kinship verification. By introducing the hand-craft feature regression module and the self-attention module, and adopting the feature fusion strategy in the feature extraction part, the model considers both the shallow geometric features and the deep semantic features of the face, and excavates the more discriminative multi-level features in the face image. Experiments show that the model has improved the recognition accuracy.

Introduction

In the past few decades, with the rapid development of deep learning, face image analysis has become a research hotspot and an active research field in the field of computer vision and biometric recognition. Many facial analysis tasks have been well studied and achieved remarkable success, especially in facial recognition(Taigman et al. 2014), facial verification(Chen, Patel, and Chellappa 2016), facial expression analysis(Yang et al. 2018), etc.

In recent years, experimental studies by some biologists and psychologists have also shown that genes have heritability and similarity. Individual genes are inherited from parents, and faces can be used as the embodiment of the genetic similarity of different individuals in the same family tree(DeBruine et al. 2009). These findings provide strong theoretical support for kinship verification based on face images, so it becomes an important sub-field of face image analysis. Affinity identification is valuable in many practical applications, such as searching for lost children(Deb et al. 2019), social network analysis(Qin et al. 2015), and criminal tracking(Jain et al. 2012), which can help the police, the Ministry of Justice, social scientists, and the masses. Compared with face recognition and verification, it is more challenging to identify kinship through face images. The existing research results show that the relevant face recognition algorithms and models achieve 99.15% higher accuracy than people in the recognition performance of a single face(Sun et al. 2014)but still face many problems in the processing of kinship recognition tasks, because the face images of people with parent-child relationships may have high genetic appearance variation due to age, gender, environment, and other complex reasons.

In this paper, we will summarize the methods and processes of kinship recognition tasks, and propose a kinship recognition model based on manually designed feature regression and multi-level feature fusion ($\mathrm{HF^2KM^2}$). Understand and learn the characteristics of typical manual design features of images, as well as the principles of convolutional neural network and self-attention mechanism, and analyze the advantages and disadvantages of these technologies. Collect the data set related to kinship recognition and preprocess it. Carry out a large number of comparative experiments on the same data set between the model proposed in this paper and other excellent models related to kinship recognition, and analyze the results. Ablation experiments were carried out on the proposed model to verify the rationality of the model.

Related work

In recent decades, many algorithms have been proposed for facial image-based kinship verification. Most existing works can be divided into two categories: feature-based approaches and metric-based approaches.

Feature-based Kinship Verification

The feature-based approaches focus on developing the discriminative facial feature. The early works often utilize the hand-crafted features, such as Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP), Scale-Invariant Feature Transform (SIFT), Gabor Wavelet and their variants,

*These authors contributed equally.

to better capture the low-level geometric, color, textual facial characteristics for kinship verification.

Recently, there has been a growing trend for researchers to use the learning-based feature, particularly the deep learning features inspired by the remarkable success in deep learning. For example, Dehshibi et al.(Dehshibi et al. 2019)present a kernelized bidirectional PCA conjunct with the cubic norm for learning discriminative feature space for kinship verification. Regarding the deep learning feature, Zhang et al.(Zhang et al. 2015)utilize Convolutional Neural Networks (CNN) for feature learning and achieved impressive verification performance.

Integrating multi-features or multi-level features is also a common strategy for kinship verification. In deep learning, exploiting multi-level features is also a common way for performance improvement. Such strategies have been extensively proven its effectiveness in many computer vision tasks, since it is well-known that feature maps from shallower layers encode low-level geometric details while the ones from deeper layers encode the spatial information, which can be further exploited for better outlining the structure. A very recent kinship verification work uses two convolutional neural networks that share parameters to extract different scales of deep features and are expected to provide global contextual information of face images(Yan et al. 2021). Inspired by these advances, our work also adopts multi-level deep features for presenting a better facial image representation.

Metric-based Kinship Verification

After feature extraction, the follow-up step needs to learn a discriminative metric for distinguishing whether the given two facial images have a kin relationship. In the recent decade, a large number of metric-based approaches have been developed. Lu et al.(Lu et al. 2013)proposed a Neighborhood Repulsed Metric Learning (NRML) method to learn a distance metric that can pull the pairs with kin relation close while pushing those without kin relation away simultaneously. Zhou et al.(Zhou et al. 2016)present an ensemble metric learning method, which utilizes a sparse bilinear similarity function to delineate the relative characteristics of kin samples.

The majority of these previous methods learn the linear mappings or its kernelized version from the feature space to the metric space. however, such mappings are essentially nonlinear. Deep learning is a well-known technique for learning complex nonlinear relationships. In kin verification, many researchers have worked in this direction. For example, Deep Discriminative Metric Learning (DDML) aims at learning a set of hierarchical nonlinear transformations to project face pairs into the same latent feature space, under which the distance of each positive pair is reduced and that of each negative pair is enlarged(Lu et al. 2017). The main merit of the deep metric learning methods is that they can be trained in an end-to-end manner, in which feature learning and metric learning are jointly optimized.

In verification tasks, the cross-pair information is very useful to enlarge the margin between the positive samples and negative samples. The most influential approach should be the triplet loss, which has been widely successfully applied to dozens of supervised learning tasks(Hermans et al. 2017). However, only a few kinship verification approaches have been explored in this direction, and all of them just directly apply the triplet loss for learning the discriminative metric space. In(Yu et al. 2020), two deep Siamese networks are integrated into a deep triplet network for tri-subject kinship verification. Kinnet adopts a soft triplet loss to further learn a nonlinear metric space where related pairs distribute closely and unrelated pairs distribute remotely in the fine-tuning phase(Li et al. 2017).

Although the triplet loss is a powerful deep metric learning technique, some studies still show that it suffers from the weak generalization ability to the test set and the slow convergence due to the reason that only one negative pair is considered in each update(Laiadi et al. 2020). We mitigate these issues via generalizing the triplet loss, which allows the positive example to compare with multiple negative examples. Moreover, our work also adaptively highlights the hard-negative examples via considering their learned relation scores as the metric weights for better optimizing the model.

Method

Overall architecture

Here, we present a novel end-to-end deep learning model named HF^{2}KM^2 for kinship verifification. The model is mainly composed of two parts, namely feature extraction part and metric network part, which correspond to feature learning and metric learning steps respectively. In this paper, $\Gamma_\zeta(\cdot)$ and $\theta_\phi(\cdot)$ are used to represent the feature extractor and metric network respectively, where ζ and ϕ are the parameters in the corresponding network. A pair of facial images are fed into the feature extractor, learn the corresponding face features, and then input the obtained features into the metric network to verify whether there is a kinship between the face images.

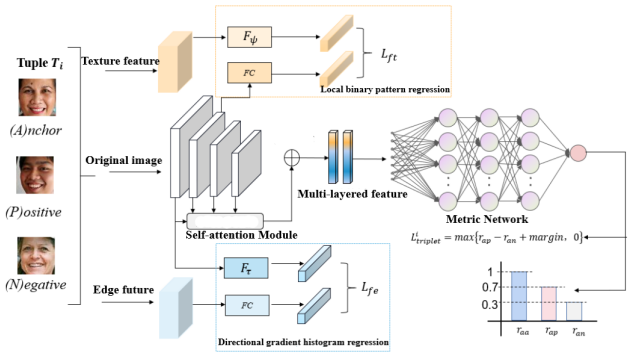

Figure 1: The overall architecture of HF^{2}KM^2.

Feature extraction

In the feature extraction part, we uses a lightweight four layer convolutional neural network (CNN) as the backbone network of feature learning. It is worth noting that this paper

can also use other more classic CNN models as the backbone network of feature learning. However, the following experimental results show that the performance of this lightweight four layer CNN network is better than VGGNet and ResNet. This is because the parameters of the above two networks are too many, and the data used to train the model's kinship verification dataset is too small, which makes the model training very easy to fall into over fitting.

Manually design feature regression module. $\mathrm{HF^2KM^2}$ adds a manually designed feature regression module to the backbone network of the feature extraction part. It uses the HOG feature and LBP feature of the face image to perform linear regression on the features output from the first and second layers of CNN respectively, and performs the feature fusion operation of splicing the features output from each layer of the four layers of CNN.

This model extracts the HOG feature of face image, and then L_2 regularizes the extracted feature. In order to ensure that these features contain as much detail as possible and align with the HOG features in scale, the HOG conversion function $\Gamma_\psi(\cdot)$ is introduced, where ψ is the corresponding parameter of the function, which will project the features of CNN corresponding to the shallow network output to the HOG space. In this space, we expect to make the model pay more attention to the edge information of the face image by minimizing the similarity between the output feature and the sample's HOG feature, so this paper introduces the edge consistency loss L_{fe} in the model

$$L_{fe} = ||\Gamma_\psi(\Gamma_\zeta^t(x)) - H(x)||_2^2 \tag{1}$$

Where, $\Gamma_\zeta^t(\cdot)$ represents the output of the convoluted layer of layer t and $H(\cdot)$ is the HOG feature extractor proposed in literature.

We extract the LBP features of the face image. After LBP features are extracted, we also perform L_2 regularization. Earlier, we mentioned that LBP features are more advanced than HOG features and are complementary. In order to combine edge shape information and texture information at the same time to better capture the details of face images, this paper introduces a local binary pattern regression module on the network layer different from the directional gradient histogram regression. We introduce an LBP transformation function F_τ, where τ is the corresponding parameter of the function. This function projects the output characteristics of the shallow network into the LBP space and aligns them with the LBP characteristics on a scale. Finally, this paper introduces texture consistency loss L_{ft} to achieve the goal of enhancing texture information at the network layer:

$$L_{ft} = ||F_\tau(\Gamma_\zeta^\eta(x)) - B(x)||_2^2 \tag{2}$$

Where $\Gamma_\zeta^\eta(\cdot)$ represents the output of the convoluted layer of layer η, where η is not equal to t in Formula 1; $B(\cdot)$ is the LBP feature extractor.

Self-attention module. We added a self-attention module to the model backbone network to further screen the features of different channels in the feature map output by each convolution layer of CNN, so the network can focus on the key areas that help to identify the kinship, and improve the expression ability of the features extracted by the network. The structure of the self-attention module is shown in Figure 2.

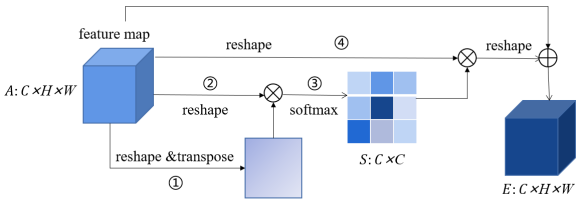

Figure 2: The architecture of self-attention module.

Metrics Network

After completing feature extraction, each face image pair $s_t = (x_i, x_j)$ can be expressed as the extracted feature pair $p_t = (f_i, f_j)$, and we project the feature pairs into a metric space, and we can verify the affinity by finding the similarity of the feature pairs by metric learning. Traditional methods usually learn a linear metric space to verify the affinity of a given image pair of affinities. However, the mapping of the face image feature pairs to the metric space should actually be nonlinear. Therefore, this paper chooses to use the recently proposed relational network (RN)(Sung et al. 2017)to accomplish this nonlinear mapping task.

Loss Function Design

The loss function of the whole model is divided into two parts, the loss generated by the hand-designed feature regression module L_{fe} with L_{ft} and the loss generated by the metric learning network. Losses from manually designed feature regression modules have been discussed above and will not be repeated here. This section focuses on the losses generated by the metric network.

The key to metric learning is to find a way to expand the distance between positive examples and negative examples. Triplet Loss is the most representative loss function to achieve this goal. Triplet loss is the most representative loss function to achieve this goal. Triplet loss was first proposed in FaceNet(Schroff et al. 2015), and its goal is to make the features with the same labels as close as possible in the metric space, while different features are as close as possible to each other. The goal is to make the features with the same label as close as possible in the metric space, while the features with different labels as far as possible in the metric space, and the learning process is shown in Figure 3.

Unlike the ternary loss function in FaceNet, we do not calculate the Euclidean distance between the feature pairs $p_t = (f_i, f_j)$ of the face images, but directly use the relationship fraction output by the metric network as $r_{ij} \in (0, 1)$ as the distance between the face image pairs, which can avoid unnecessary calculations and reduce the complexity of the model. According to the definition of the triadic loss function, the loss $Ltriplet$ of the part of the metric network is

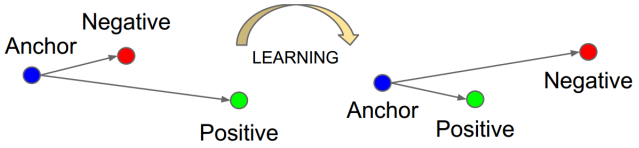

Figure 3: Optimization process of the triplet loss function.

assumed to be fed to the network with the triad T_i as follows:

$$L^i_{triplet} = max\{r_{ap} - r_{an} + margin, 0\} \quad (3)$$

where r_{ap} is the relationship score between the anchor point and the positive example, and r_{ap} is the relationship score between the anchor point and the negative example. The total loss of the metric network component is given as:

$$L_{triplet} = \sum_{i=1}^{N} L^i_{triplet} \quad (4)$$

The loss L_{fe}, L_{ft} and $triplet$ of the integrated manual part design module part, the overall loss function of the whole model can be be expressed as:

$$L_{total} = L_{triplet} + L_{fe} + L_{ft} \quad (5)$$

By jointly minimizing L_total to optimize the parameters of the model, on one hand the feature extraction part of our model can focus on more regions with manual prior knowledge (edges, textures), thus outputting a more multi-level feature representation. On the other hand On the other hand, the metric network projects the input feature pairs into the metric space, which can achieve the goal of distancing the feature pairs that do not have affinity and bringing them closer to those that have affinity. The metric network projects the input feature pairs into the metric space, so that feature pairs that are not related can be distanced and feature pairs that are related can be brought closer.

Experiments

This chapter will introduce the experimental results of the HF^2KM^2 model proposed in this paper on the three mainstream datasets of kinship recognition tasks and the results of comparison with the mainstream kinship recognition methods in recent years. In addition, this chapter will introduce and analyze the results of the ablation experiment conducted on the model.

Datasets and Experimental Settings

The datasets used in this experiment are KinFaceW-I, KinFaceW-II(Zhang et al. 2015)and TSKinFace(Qin et al. 2015). Search for learning rates in {0.0001,0.0005,0.001,0.005} and three in {0.7,0.8,0.9}, The size of the required hyperparameter margin in the tuple loss.The ratio of the training set to the test set is 4:1.In the experiment, in order to make fair comparison with other kinship recognition model methods, the parameters were adjusted to achieve the best performing model was run five times and recorded with average recognition accuracy and overall average recognition accuracy on the four relatives.

Experiment Results and Analysis

In order to better evaluate our model, we compare the model presented in this paper with the most advanced baselines in recent years. A comparison was made. For the two datasets of KinFaceW-I and KinFaceW-II, we select WGEML(Jiangqing et al. 2018), DCBFD(Yan and Haibin 2019), KML(Zhou et al. 2019), KinMix(Song et al. 2020), and DRN(Yan et al.2021) models as baselines. The TSKinFace dataset was chosen separately DDMML(Yan et al.2021), DKMR(Wang et al. 2020), TXQDA(Laiadi et al. 2020) as our baseline. Table 1, Table 2, and Table 3 are shown Experimental results of the model on the KinFaceW-I, KinFaceW-II, TSKinFace datasets.

Experimental results show that our model compares the baseline on the KinFaceW-II and TSKinFace datasets to optimal performance. More specifically, on the KinFaceW-II and TSKinFace datasets, the second-best performing methods are KinMix and WGEML, respectively, and the average recognition accuracy of our proposed model has improved by 1.2% and 2.6%. Our model achieves 80.2% accuracy on the KinFaceW-I dataset. Although there is still some gap compared to the new DRN model, our model is still one of the most advanced models, superior to the newly proposed KinMix method in 2020.

Although there is still some gap compared to the new DRN model, our model is still one of the most advanced models, superior to the newly proposed KinMix method in 2020. The KML model uses pre-production on large-scale face datasets the trained VGGNet performs feature extraction, and obviously the model is basically not affected by the small size of the dataset. DRN significantly enhances the representation ability by introducing dense samples in local areas and scales in the CNN feature space, which has great advantages when the training data is not rich, but the time cost of the work carried out in the feature space is very high, and our end-to-end one-step training model is actually more advantageous than it overall. The experimental results on the KinFaceWII dataset actually validate our inference. The average recognition accuracy of our model on the KinFaceW-II dataset is 5.2% and 2.1% higher than that of the KML model and the DRN model, respectively, which highlights the superiority of our model.

Ablation study

In order to further verify the improvement of recognition performance of each module of the model, a series of ablation experiments are set up in this paper.

Figure 4 shows the difference in recognition performance of the model in this paper when using only the fourth layer output of the convolution layer, that is, using a single feature (conv4) for recognition and integrating the features output from the fourth layer convolution layer, that is, using

method	F-S	F-D	M-S	M-D	MEAN
WGEML	78.5	73.9	80.6	81.9	78.8
D-CBFD	79.6	73.6	76.1	81.5	77.6
KML	83.8	81.0	81.2	85.0	82.8
KinMix	76.5	75.6	83.5	78.5	78.5
DRN	85.8	87.5	88.1	80.9	85.6
Ours(HF^2KM^2)	76.1	80.1	80.1	86.3	80.2

Table 1: The evaluation of the results of HF^2KM^2 and other models on KinFaceW-I dataset.

method	F-S	F-D	M-S	M-D	MEAN
WGEML	88.6	77.4	83.4	81.6	82.8
D-CBFD	79.6	73.6	76.1	81.5	77.6
KML	87.4	83.6	86.2	85.6	85.7
KinMix	87.2	89.6	90.6	91.2	89.7
DRN	90.4	86.6	91.0	87.2	88.8
Ours(HF^2KM^2)	85.2	92.6	95.4	90.2	90.9

Table 2: The evaluation of the results of HF^2KM^2 and other models on KinFaceW-II dataset.

multi-level features (conv1234) for fusion. The experimental results show that the performance of multi-level feature recognition is higher than that of single feature recognition because it takes into account the geometric, edge, texture features of the shallow layer of the face image as well as the high-level semantic abstract features of the deep layer.

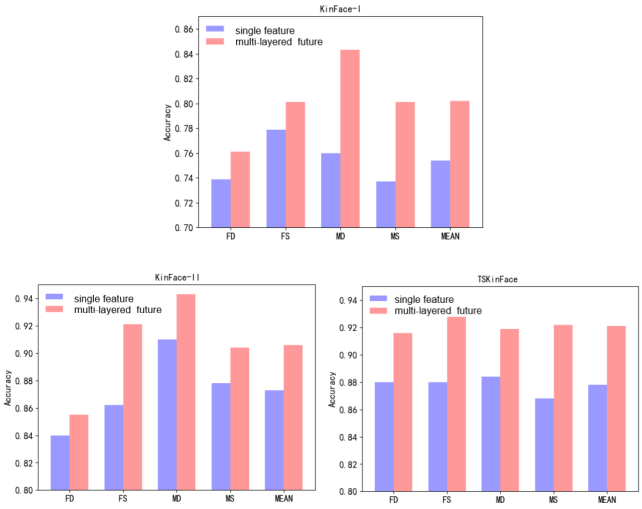

Figure 4: Using a single feature versus using a multi-layered feature.

At the same time, in order to further verify the role of the manual design feature regression module and the self attention module of the model in facial feature extraction, we respectively add the hang-craft feature regression module (HR) and the self-attention module (SA) to the backbone network for kinship experiments. The experimental results shown in the figure 5.

method	F-S	F-D	M-S	M-D	MEAN
DDMML	86.6	82.5	83.2	84.3	84.2
WGEML	90.3	89.8	91.4	90.4	90.5
DKMR	81.3	77.8	79.2	77.7	79.0
TXQDA	89.3	90.7	90.3	91.0	90.3
Ours(HF^2KM^2)	93.0	92.5	92.7	94.1	93.1

Table 3: The evaluation of the results of HF^2KM^2 and other models on TSKinFace dataset.

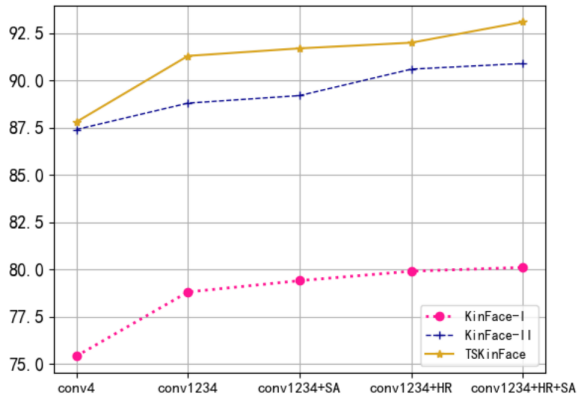

Figure 5: The effect of the hand-craft feature regression module and the self-attention module on model performance.

By analyzing the above experimental chart, we can easily draw the following conclusions: (1) hang-craft feature regression module and self-attention module will improve the recognition performance of the model, which verifies the effectiveness of the two modules. (2) Compared with the HR module and SA module, the feature fusion strategy using four-layer CNN splicing can greatly improve the performance of model recognition. Compared with the self-attention module, the hang-craft feature regression module can greatly improve the performance of model recognition.

Conclusion

Kinship recognition based on face image pairs is nowadays widely used in various fields such as finding lost children, social network analysis and criminal tracking. In this paper, we propose a multi-level feature knowledge mining model for kinship recognition, HF^2KM^2, which mines more discriminative multi-level features in face images by introducing a hand-designed feature regression module and a self-attentive module in the feature extraction part, and employing a feature fusion strategy. In order to verify the effectiveness of the proposed model, it is compared with existing good algorithms. The experimental results show that the performance of the model proposed in this paper obtains significant results compared to the baseline.

References

Chen, J.-C.; Patel, V. M.; and Chellappa, R. 2016. Unconstrained face verification using deep cnn features. In *2016 IEEE winter conference on applications of computer vision (WACV)*, 1–9. IEEE.

Deb, D.; Aggarwal, D.; and Jain, A. K. 2019. Finding missing children: Aging deep face features. *arXiv preprint arXiv:1911.07538*.

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

Dehshibi, M. M.; and Shanbehzadeh, J. 2019. Cubic norm and kernel-based bi-directional PCA: toward age-aware facial kinship verification. *The Visual Computer*, 35(1): 23–40.

Hermans, A.; Beyer, L.; and Leibe, B. 2017. In defense of the triplet loss for person re-identification. arXiv 2017. *arXiv preprint arXiv:1703.07737*, 4.

Jain, A. K.; Klare, B.; and Park, U. 2012. Face matching and retrieval in forensics applications. *IEEE multimedia*, 19(1): 20.

Jianqing; Liang; Qinghua; Hu; Chuangyin; Dang; Wangmeng; and Zuo. 2018. Weighted Graph Embedding based Metric Learning for Kinship Verification. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*.

Laiadi, O.; Ouamane, A.; Benakcha, A.; Taleb-Ahmed, A.; and Hadid, A. 2020. Tensor cross-view quadratic discriminant analysis for kinship verification in the wild. *Neurocomputing*, 377: 286–300.

Li, Y.; Zeng, J.; Zhang, J.; Dai, A.; Kan, M.; Shan, S.; and Chen, X. 2017. Kinnet: Fine-to-coarse deep metric learning for kinship verification. In *Proceedings of the 2017 workshop on recognizing families in the wild*, 13–20.

Lu, J.; Hu, J.; and Tan, Y.-P. 2017. Discriminative deep metric learning for face and kinship verification. *IEEE Transactions on Image Processing*, 26(9): 4269–4282.

Lu, J.; Zhou, X.; Tan, Y.-P.; Shang, Y.; and Zhou, J. 2013. Neighborhood repulsed metric learning for kinship verification. *IEEE transactions on pattern analysis and machine intelligence*, 36(2): 331–345.

Qin, X.; Tan, X.; and Chen, S. 2015. Tri-subject kinship verification: Understanding the core of a family. *IEEE Transactions on Multimedia*, 17(10): 1855–1867.

Schroff, F.; Kalenichenko, D.; and Philbin, J. 2015. FaceNet: A Unified Embedding for Face Recognition and Clustering.

Song, C.; and Yan, H. 2020. KINMIX: A data augmentation approach for kinship verification. In *2020 IEEE international conference on multimedia and expo (ICME)*, 1–6. IEEE.

Sun, Y.; Chen, Y.; Wang, X.; and Tang, X. 2014. Deep learning face representation by joint identification-verification. *Advances in neural information processing systems*, 27.

Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P.; and Hospedales, T. M. 2017. Learning to Compare: Relation Network for Few-Shot Learning.

Taigman, Y.; Yang, M.; Ranzato, M.; and Wolf, L. 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.

Wang, M.; Shu, X.; Feng, J.; Wang, X.; and Tang, J. 2020. Deep multi-person kinship matching and recognition for family photos. *Pattern Recognition*, 105: 107342.

Yan, H. 2019. Learning discriminative compact binary face descriptor for kinship verification. *Pattern Recognition Letters*, 117: 146–152.

Yan, H.; and Song, C. 2021. Multi-scale deep relational reasoning for facial kinship verification. *Pattern Recognition*, 110: 107541.

Yang, H.; Ciftci, U.; and Yin, L. 2018. Facial expression recognition by de-expression residue learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2168–2177.

Yu, J.; Li, M.; Hao, X.; and Xie, G. 2020. Deep fusion siamese network for automatic kinship verification. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, 892–899. IEEE.

Zhang12, K.; Huang, Y.; Song, C.; Wu, H.; Wang, L.; and Intelligence, S. M. 2015. Kinship verification with deep convolutional neural networks. British machine vision conference. BMVA Press.

Zhou, X.; Jin, K.; Xu, M.; and Guo, G. 2019. Learning deep compact similarity metric for kinship verification from face images. *Information Fusion*, 48: 84–94.

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