KinfaceNet: A New Deep Transfer Learning based Kinship Feature Extraction Framework

Pallabi Ghosh
Univ. of Florida
pallabighosh@ufl.edu

Sumaiya Shomaji
Univ. of Kansas
shomaji@ku.edu

Damon L. Woodard
Univ. of Florida
dwoodard@ece.ufl.edu

Domenic Forte
Univ. of Florida
dforte@ece.ufl.edu

Abstract

Advances in vision and deep learning have revolutionized feature extraction for face recognition and verification systems, yet, performing kinship verification from such features is still challenging. Ongoing research attempts to imitate a human by identifying features for kinship verification. In this paper, we propose KinfaceNet, a deep learning based kinship feature extractor, capable of extracting kinship features from a single input image independently without requiring its kin pair image. The base model of the method is adopted from face recognition domain which is then transfer learned in the domain of kinship by learning a distance mapping from face images to a compact Euclidean space where distances directly correspond to a measure of kinship similarity. Thus, unlike most of the works in deep learning based kinship domain, the extracted features can be used in many other applications such as image generation and family based clustering, etc. Training is performed by rearranging the data into classes of kin pairs and using a state-of-the-art triplet mining algorithm to address the unbalanced kinship data problem which causes overfitting. Also, one of the major advantages of our framework is that training can be performed on any face feature extractor model pre-trained on large face recognition data, thereby reducing training time by a considerable amount. Comparable verification accuracy is obtained from simple MLP network at only 20th epoch with KinfaceNet features extracted from the Family-In-the-Wild dataset, the largest in the wild kinship dataset available, as well as KinfaceW-I and II datasets.

1. Introduction

Human faces convey a pool of information starting from identity, age, and gender to emotional state and intentions. One such information is kinship. Kinship cues are properties that convey resemblance between human faces which helps in recognizing the relatedness between people. They are traits that are transferred from one generation to another, like from parents to their children. It is well known that DNA can help in determining the kin relationship and ancestry with high accuracy. However, the procedure to extract kinship information from DNA is complex. A recent study reveals that more than 130 regions in human DNA play an important role in sculpting facial features [1–3]. These visual capacities to detect relatedness become weaker with lower degrees of relatedness [4]. Hence it is evident that in future, an accurate mapping of the kinship facial features from images and the regions in human DNA can help in determining the kin relationship without undergoing the complexities of DNA testing. Facial kinship cues identification has the potential to reduce, if not eliminate, the number of experiments needed for DNA-based analysis in critical situations like forensics.

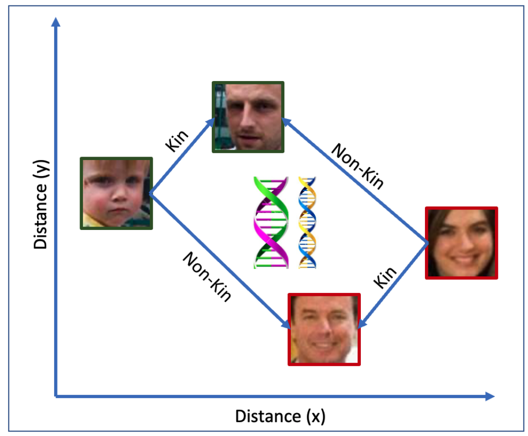


Figure 1. KinfaceNet features are such that distance between kin pairs is less than non-kin pairs.

Another study provides a detailed investigation on the relationship between human recognition of kinship and degree of relatedness [4]. It states a possible effect of age and gender on kinship clues. [5] compares the accuracy of kinship recognition from face images obtained by humans and computers. The results showed that humans identified kin pairs of KinFaceW-I and KinFaceW-II [6] with an accuracy of 78.6% and 83.5%, respectively while the state-of-the-art method on the same datasets were 82.7% and 86.0%, respectively. Another paper [7] states that for FIW dataset the mean accuracy by humans is 57.5%. Over the years, many automated computer vision methods have been developed, where machines were shown to be better in determining kin relation than humans [8–17]. For our proposed method too the accuracy is higher than kinship verification by humans. Also, the accuracy depends largely on the degree of relatedness used for evaluation. For example, the kinship cues will be larger between a parent and a child compared to grandparent and grandchild.

In this paper, we have focused on extracting a vector loaded with kinship information. To maximize the kinship cues in the features we have used only the highest degree of relatedness, i.e., Parent-Child for training and testing of our feature extraction models. Similarly, to minimize bias due to gender, training is done on all combinations of Father and Mother as Parent and Daughter and Son as Child. An introductory image of the goal of the paper is shown in Figure 1. While applications of face recognition in various fields are exploding, kinship identification and analysis, being one of the inherited domains, has significant influence on real-life applications starting from family photo album organizations to critical applications like surveillance, security, tracking, forensic etc. The main advantage of our KinfaceNet framework over existing network focused deep learning based kinship verification systems is that it uses a simple data mining and training to achieve state of the art accuracy and can support a plethora of applications and not just kinship verification without requiring complex resources and time. In Section 5.4, we have discussed more about the applications of our proposed framework.

1.1. Related Work

Research performed in the domain of kinship cues identification using computers can be broadly classified into two types: *network-focused* and *feature-focused*. Network-focused systems are those methodologies where the proposed method puts more emphasis on deriving a network architecture which, through training, finds a mapping between kin and non-kin pairs and provides higher accuracy. Most of the deep learning works done in the kinship domain are network focused [15, 18–23]. Over the years, the input features used for training such network focused deep learning methods are either Local Binary Pattern (LBP), face images, parts of face images like eyes, nose, etc. or face recognition features like arcface, facenet, etc. Some of the notable works that uses each of these as input feature is given in Table 1. They have mainly relied on the architecture to train and find the mapping. Hence the number of computations for each of these models is very high.

On the other hand, feature-focused methods mainly concentrate on computing the perfect kinship features containing maximum kinship cues. These features are such that, even simple distance measures can differentiate between kin and non-kin pairs without the need of training deep network architectures. Most of these features in kinship domain are handcrafted features or statistical features without involving any training. Some of the work done in this domain are [5, 24–26]. Although some of the papers have reported very high verification accuracy [25], these feature computations will become more computationally expensive with larger amount of data. Thus, we see the main computationally expensive part of a network focused method is training and for feature focused method it is feature extraction. Also, the network focused methods work best with more data while the feature focused ones work best with less data. Moreover, the main utility of the state-of-the-art network and feature focused methodologies in the domain of kinship are mainly concentrated only on kinship verification.

In this work, the shortcomings of the network focused and feature focused methodologies are addressed using a simple transfer learning based framework. The proposed KinfaceNet framework can be computed both on large as well as small amount of data, is computationally efficient, and outputs a feature vector which can be used in a plethora of applications including kinship verification. It has all the advantages of deep learning, can be used in the existing kinship verification architectures, does not require extensive training and outputs a feature vector instead of kin or non-kin classification results. Two of the notable surveys in this domain that provides a complete list of methodologies are [8] and [27].

Also, an interesting observation from all the existing works, is that although the kinship verification problem is inherited from the research problem domain of face verification, the accuracies obtained are not as high as the reported accuracy in the domain of face verification. The reported accuracy is also worse for larger datasets like Family-In-the-Wild (FIW) [28]. Hence most of the state-of-the-art works focus on improving the accuracy by using different deep learning architectures or statistical feature extraction. But the goal of this work is to find a feature vector containing kinship cues which can be used in many more applications in this domain, as proposed in Section 5.4, and not just achieving the highest verification accuracy. For proof of concept and to show the application of the feature in verification, it is shown that without involving any rigorous training or computations, unlike the methods in the domain of network focused and feature focused, the extracted features provide comparable high verification accuracy at a very optimum time, even for the large FIW dataset, using a simple Multi-Layer Perceptron network after only 20 epochs. The contributions of this paper are summarized follows:

- Proposes a deep learning based kinship feature extractor,

Table 1. Related work done in the domain of deep learning based kinship verification.

Paper	Input	Dataset	Accuracy	Method	Epochs	Remarks
[15]	Full Face	Kinfacew-I KinfaceW-II	74.8 85.3	CNN	30	Not performed on large data Only verification Trained from scratch More time complexity
[19]	Parts of Face	KinfaceW-I KinfaceW-II	76.5 88.7	Attention Network	60 100	Complex network compared to data size Trained from scratch More time complexity
[18]	Local Binary Pattern	KinfaceW-I KinfaceW-II	66.9 71.3	Stacked Auto-encoder	-	Much less accuracy, complex network, less data
[20]	Face Recognition Feature	Family-In-the-Wild (FIW)	79.6	Unified Multi-task learning	60	Uses face recognition system for encoding, results can be improved using our approach

where the kinship cues of an image can be mapped to a 512 dimensional feature vector. Thus our method can be used beyond kinship verification. It can be used in problems like simple document tagging to complex problems like GAN generated synthetic kin image generation.

- Proposes a framework that can blend any pre-trained face feature extractor model, which are available in abundance in the current literature. Thus it can reduce training time to a considerable amount, reduce computational resource requirement and also address the balanced data availability bottleneck in kinship.
- Performs feature analysis and computes verification accuracy of the features commonly used in the state-of-the-art deep learning based methods (LBP, face image, parts of face image, face embedding using a face recognition embedding extraction network) and proposed framework features with and without transfer learning.
- Lists the utility of the feature vector in a plethora of applications including kinship verification.

The rest of the paper is organized as follows. In Section 2, the necessary background on kinship and kinship verification systems along with algorithms and networks used in the paper is provided. In Section 3, the proposed framework and its methodology is described in detail. The kinship datasets used are mentioned in Section 4. The experimental setup and results are presented in Section 5. Finally, the paper is concluded with directions for future research in Section 6.

2. Background

An exact objective definition of kinship is the traits that are transferred from one generation to another through our genes. A human can identify these traits and verify if they are related or not. Kinship mostly has three degrees; namely, primary, secondary, and tertiary kinship. The primary degree refers to kin that are closely and directly related to one another. The secondary and tertiary degree refers to the primary and secondary kin of the first degree kin respectively. Most of the available kinship face datasets contain only four types of primary kinship paired entities: mother-son (MS), father-son (FS), mother-daughter (MD) and father-daughter (FD). These four categories of image

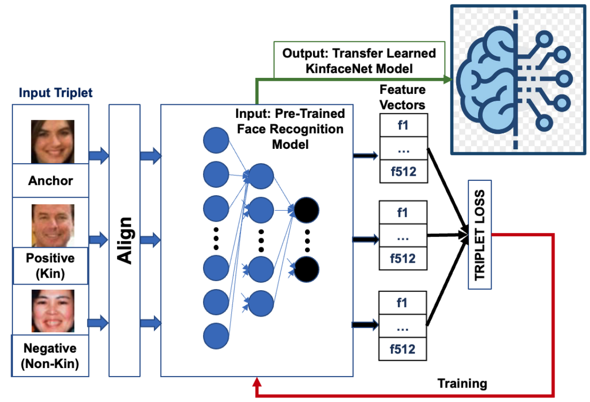


Figure 2. Proposed Framework. A triplet set of input image with Anchor, Positive, and Negative images are given as input to a pre-trained face recognition model and trained. The output of the framework is the transfer learned Kinface Model capable of extracting kin features.

pairs contain most of the kinship information. In our work we will only concentrate on these four pairs. In this work, we have focused on creating a feature vector that will help a computer to know the traits and efficiently identify kin pairs.

2.1. Face vs. Kinship Verification Feature Extractor

Over the last century, biometrics have been used to identify humans. Faces are one of the most common modalities used to do so. In any automated biometric verification system, a series of features are extracted from a face before performing face verification. Just like kinship verification, in this domain too, the first features explored were simple handcrafted features like parts of the face. Later came the more advanced statistical features like Local Binary Pattern. (LBP) [29], Histogram of Gradients (HOG), Speeder Up Robust Feature (SURF), etc. However, the major boom was observed with the advent of deep learning based feature vector extraction systems, the most prominent one being FaceNet which outputs a 128D feature vector [30]. In this work we aim to create a similar deep learning based metric feature vector extraction system which the kinship domain lacks. An immediate utility of our work will to generate synthetic kinship data, which is another bottleneck in this domain.

2.2. Triplet Loss

Triplet loss is a loss function for machine learning algorithms where a reference input (called *anchor)* is compared to a matching input (called *positive*) and a non-matching input (called *negative*). In our problem statement, we have considered two images of the same family, mainly pairs from one of the four primary kin relations mentioned in Section 2, as the anchor and the positive images (i.e., kin pair), and an entity from a different family as a negative (i.e., non-kin) input. The loss function is given as

$$Loss(A,P,N) = \max(\|f(A) - f(P)\|_2 - \|f(A) - f(N)\|_2 + \alpha, 0), \quad (1)$$

where A, P, and N are the anchor, positive (or kin with anchor), and negative (non-kin with anchor), respectively. The feature vector embedding is represented as f and α is the margin between the positive and negative pairs.

This loss is used to train a neural network. To ensure convergence, it is crucial to select triplets that violate the triplet constraint in Equation (1). For this, online hard triplet generation method for triplet selection is adopted from [30]. Here, the triplets are chosen in such a way that the point with maximum anchor and the positive distances and minimum anchor and negative distances are chosen. In this selection methodology, the hard positive and hard negative exemplars are selected from within a minibatch. More details about the loss and the triplet selection method can be found in [30].

2.3. Deep Learning based Verification Systems

To prove the utility of our framework, we have used a simple MLP network [31] as our kinship verification system, to identify kin and non-kin relationship using only the extracted KinfaceNet features. MLP is the most basic form of neural network often referred as "vanilla" neural network [32]. It is a fully connected class of feedforward artificial neural networks (ANNs). This paper focuses on the feature quality and MLP is used here to illustrate how comparable high accuracy can be achieved without training any complex network for a prolonged time compared to the other features commonly used in the domain of deep learning.

3. KinfaceNet Implementation & Methodology

In this paper, we have first proposed a simple KinfaceNet feature extraction framework that utilizes transfer learning from face recognition domain. Then the efficiency of the features extracted using the proposed framework is compared with the features of the framework without transfer learning, as well as other commonly used features in the deep learning based kinship verification works. The methodology can be broadly divided into two steps, namely feature extraction and kinship verification.

3.1. Image Alignment

The first step of our framework involves refactoring the dataset folders in such a way that every dataset contains multiple sub-folders, and each sub-folder contains a parent and child image. Faces in each of these images are then detected and aligned using MTCNN model [33, 34].

3.2. Feature Extraction

Feature extraction using proposed framework contains three main components. They are discussed in detail in this section.

3.2.1 Input Triplet

This is a set of three images, namely anchor, positive and negative. The images are chosen in such a way that in a triplet image input the anchor and positive images are taken from the same family, say the ith family, and the negative image is a random image from any other family, say the jth family, and $i \neq j$. This input triplets are chosen using a triplet selection algorithm discussed in Section 2. Batches of hard triplets are formed before every epoch.

3.2.2 Pre-Trained Model

In this part of the proposed framework, any efficient face feature extractor models trained for face recognition system that outputs a feature vector can be used. In this work, a pretrained Resnet34 model is used. It has a face recognition accuracy of 98.45% on the LFW dataset [35]. Any other state-of-the-art model can be used as well. The initial weights for our framework model are transfer learned from the pre-trained model. If any other face feature extraction model is used, then the framework model architecture should be same as the pre-trained model, so that the initial weights can be transferred.

3.2.3 Triplet Loss and Training

As discussed in Section 2, at the beginning of every epoch a batch of triplet inputs are selected. At every epoch, the framework model, is used to extract 512D feature vector from each input triplet, i.e., anchor, positive and negative. From these batch of input triplets, a unique type of loss called Triplet Loss is calculated using Equation (1). This loss is then propagated back through the network.

3.3. Kinship Verification System

To determine the efficiency of our proposed framework, a simple Multi-Layer Perceptron (MLP) network is used. The purpose of the verification system is to determine kin or non-kin relationship between two test images using only the extracted face features. For comparison, we have used features extracted using three deep learning models: 1) proposed framework without transfer learning: a Resnet34 network trained from scratch on kinship dataset without transfer learning; 2) the pre-trained face feature extractor model: a Resnet34 network trained on glint360k dataset; and 3) proposed framework with transfer learning: a Resnet34 network trained on glint360k dataset, further trained on kinship dataset. The other features used for comparison are LBP, full face image, and parts of face image (eyes and nose).

The dlib package is used to extract the boundary landmark points for eyes and nose. Mean and standard deviation were calculated from the extracted pixel values for each feature set, reducing input dimensions and chances of underfitting. For each case, features extracted from two test images are passed through the trained verification systems and the test accuracy is noted. For our verification system, the two feature vectors are concatenated into a single vector. For unequal length vectors like pixels of eyes and nose, mean and standard deviation are used as features of the feature vector. This vector is given as input to an MLP model with two hidden layers, with 512 and 256 nodes respectively, which returns a 1-dimensional binary output, where 0 represents non-kin and 1 represents kin.

4. Datasets

4.1. Family-in-the-Wild

This is the largest and most comprehensive dataset available in kinship literature [28,36]. FIW is made up of 11,932 natural family photos of 1,000 families. It has 656,954 image pairs split between the 11 relationships, starting from the primary kin relationship of parent and child to secondary kin relationship like grandparent and grandchild. Since kinship properties are most prominent amount the four primary kin relationship of parent and child, as mentioned in Section 2. In this work, we have worked with only these four parts for ease of comparison and kinship cue identification. For this dataset, the train/test split for verification accuracy is adopted from [37]. Although this dataset is of the largest size, the quality of the images is not good enough. Hence, in the current works done in the kinship verification domain, the accuracy obtained on the FIW dataset is much worse compared to the KinfaceW dataset.

4.2. KinfaceW

This is another popular in-the-wild kinship dataset used in kinship verification literature [8, 38]. There are mainly four primary kin relations in the two datasets, as mentioned in Section 2: Father-Son (F-S), Father-Daughter (F-D), Mother-Son (M-S), and Mother-Daughter (M-D). There are mainly two subgroups, namely KinfaceW-I and KinfaceW-II. The difference of KinFaceW-I and KinFaceW-II is that face images with a kin relation were acquired from different photos in KinFaceW-I and the same photo in KinFaceW-II in most cases. In the KinFaceW-I dataset, there are 156, 134, 116, and 127 pairs of kinship images for these four relations. For the KinFaceW-II dataset, each relation contains 250 pairs of kinship images. These datasets contain images collected from the internet, including some public figure face images and their parents' or children's face images. Face images are captured under uncontrolled environments in two datasets with no restriction in terms of pose, lighting, background, expression, age, ethnicity, and partial occlusion. The train, validation and test splits taken as 80%, 10%, and 10%.

Table 2. Resnet34 models used to compare the extracted kinship features. The proposed framework model is M3.

Model Name	Epochs	Pre-trained Model used
KinfaceNet without Pre-trained (M1)	150	No
Pre-trained Face Feature Extractor (M2)	0	Yes
Proposed Transfer-Learned KinfaceNet (M3)	60	Yes

5. Experiments and Results

5.1. Experiments for Kin Feature Extractor

In this paper, we have proposed an efficient kinship feature extractor framework using a pre-trained ResNet34 model used in a face recognition system. The model without transfer learning was trained for 150 epochs, but with transfer learning the best model was achieved at 60 epochs with no much change in loss after this. For each of these training a batch size of 16, iterations per epoch is 1000, number of human identities per batch is 16, Adagrad optimizer value of 0.075 is used. We have shown how higher accuracy can be achieved at a lower training epoch if the training is started from a trained facial feature extractor model from face recognition domain. This is important in the domain of kinship verification since the amount of balanced training data is still not adequate to train models from scratch. So, a solution should be devised, such that the data requirement bottleneck can be addressed to some extent, mainly by using transfer learning. We used a pre-trained Resnet34 model [39]. This model is trained on glint360 dataset [40] trained for 90 epochs, is readily available, and has high accuracy in a face recognition system. It takes $140X140$ size image as input and outputs a 512D vector. Any other pre-trained face feature extractor model used in face recognition system can be used. This model is then transferred in the domain of kinship by training on kinship data using triple loss, as explained in Section 3.

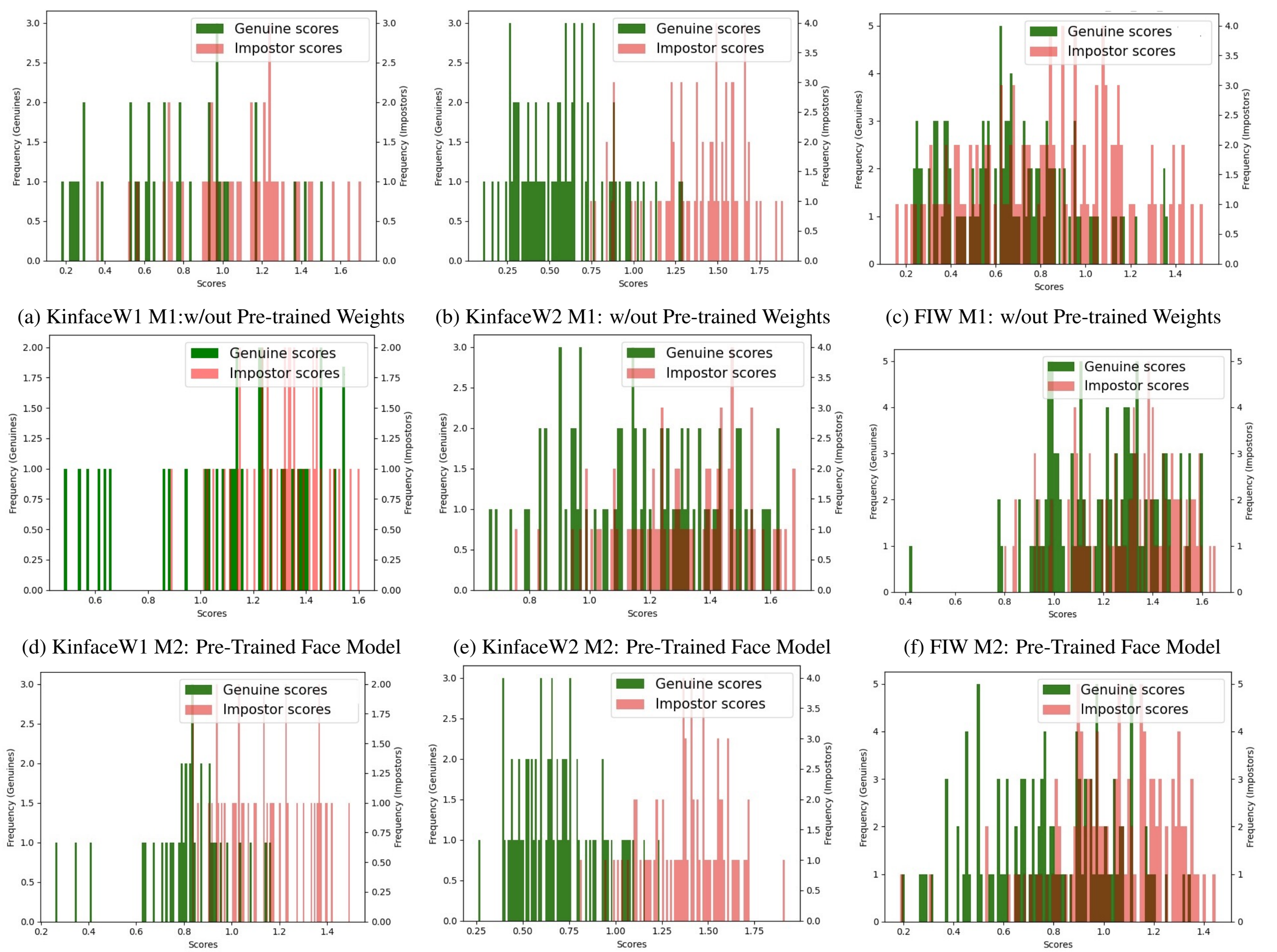
The list of models used for evaluation can be found in Table 2. We have also extracted features that are used in previous deep learning based works, like LBP, full face image, parts of face images, and face embeddings from face recognition system for comparison. As mentioned in Section 1 and 2, human eyes and nose contains the maximum information hence we have used only these two parts. For analysis and verification of these models, features are extracted from each of the datasets, i.e., FIW, KinfaceW-I, and KinfaceW-II, and their distribution and verification results are compared.

5.2. Extracted Kin Feature Distributions

This part of the experiments forms the main contribution of the paper. In the framework, the experiments performed

Table 3. Equal Error Rate (EER) and Area Under Curve (AUC) values for the different features and datasets. TL=Transfer Learning

Model Used	No. of Epochs Trained	Pre-Trained Model Used	FIW		KinfaceW-I		KinfaceW-II	
			EER	AUC	EER	AUC	EER	AUC
LBP	-	-	0.47	0.5509	0.47	0.5279	0.39	0.6400
Full Face Image	-	-	0.44	0.5647	0.35	0.7109	0.43	0.6254
Parts of Face: Eyes	-	-	0.44	0.5853	0.30	0.7143	0.32	0.7422
Parts of Face: Nose	-	-	0.43	0.5833	0.31	0.7374	0.32	0.7127
M1(KinfaceW without TL)	150	No	0.36	0.6889	0.31	0.7891	0.09	0.9765
M2 (Pre-trained Face-Recognition model)	0	Yes	0.41	0.6277	0.36	0.6598	0.4	0.6350
Proposed M3(KinfaceW with TL)	60	Yes	**0.35**	**0.7067**	**0.17**	**0.9090**	**0.07**	**0.9808**



(a) KinfaceW1 M1:w/out Pre-trained Weights (b) KinfaceW2 M1: w/out Pre-trained Weights (c) FIW M1: w/out Pre-trained Weights

(d) KinfaceW1 M2: Pre-Trained Face Model (e) KinfaceW2 M2: Pre-Trained Face Model (f) FIW M2: Pre-Trained Face Model

(g) KinfaceW1 M3: Transfer Learned Model (h) KinfaceW2 M3: Transfer Learned Model (i) FIW M3: Transfer Learned Model

Figure 3. Euclidean distance distribution plots between Genuine (kin) vs Imposter (non-kin) pairs of images from subsets of different datasets. Plots shows the distribution for our proposed method with and without kinship. (a), (b), and (c) are plots of proposed KinfaceNet framework feature extractor without transfer learning (M1); (d), (e), and (f) are associated with a pre-trained face model without KinfaceNet; (g), (h), and (i) are plots of proposed KinfaceNet framework feature extractor with transfer learning (M3). M1 trained for 150 epochs and M3 trained for 60 epochs only.

in this part belong to the **Feature Extraction** section. The features used in state-of-the-art methods like LBP, full face image, parts of face image, as mentioned in Table 1, are extracted from the images of all three datasets. For the Face Recognition feature we have used a pre-trained resnet34 network given in Table 2 as M2. The models M1 and M3 are the proposed framework models with and without transfer learning respectively. Features are extracted using these two models as well. All these feature distributions are then compared using biometric score distribution analysis. To plot the genuine vs. imposter distribution, a popular plotting method used in biometrics to visualize the feature space, the distance between a parent and a child is considered a genuine score and the distance between two images of different families is considered an imposter score. Euclidean distance is used for score evaluation. Score distribution plots are evaluated for each and every type of feature for the same subset of each of the dataset. This subset is taken from the test set and is not used in training. The genuine (kin) vs. imposter (non-kin) distribution plots for our proposed framework features, with and without transfer learning, as well as, features from the pre-trained network(M2) are shown in Figure 3. The Equal Error Rate (EER) and Area Under Curve (AUC) values for all the features, including the state-of-the-art features, mentioned in Table 1 are given in Table 3. These parameters help in determining the quality of the features.

In Figure 3, the first row provides the score distribution for the model trained from scratch (M1), i.e., no transfer learning from face recognition model is used. A resnet34 model is trained from scratch on the kinship datasets FIW, KinfaceW-I and KinfaceW-II separately. The last row shows the score distribution for the embeddings extracted using our proposed framework (M3). The middle row shows how the distribution looks for M2, i.e., just a pre-trained network, which is very common in state-of-the-art methods. Unlike face recognition systems, in kinship problem the genuine scores are from different people sharing the kinship traits. Still the score distribution clearly shows a separation between the genuine or kin pairs and imposter or non-kin pairs for all the three types of models for each dataset. But the distribution is maximally separated for our proposed technique (M3), indicating that the KinfaceNet extracted features contains the maximum kinship information when trained for the same number of epochs as M1.

Table 3 provides similar observation where the error rate for our proposed model is minimum and AUC is maximum compared to model M1 and M2 for each and every dataset. The EER and AUC values for the proposed kinfaceW features are much better than the state-of-the-art features too.

5.3. Verification Accuracy of Extracted Features

This set of experiments show the performance of the extracted features on a simple MLP based verification system after 20 iterations. The best verification model may be different for different features. Since the input features are different, for comparison purpose we have used the same MLP network with similar parameters as the verification system, except for the input size. For dissimilar sized input vectors like pixel values of eyes or nose, the mean and standard divisions are used as features of the input feature vector. Any other state-of-the-art verification systems can be used in place of MLP. But the verification accuracy achieved using a simple MLP network by the features extracted using our framework is higher than most of the deep learning based state-of the art complex networks. Thus, it is shown, if we can take advantage of transfer learning, training can be reduced to a huge extend and comparable accuracy can be achieved even if there is limited balanced data like kinship. Deep complex network training from scratch is not always the best solution, if we can take advantage of the existing trained networks. Thus, even with limited access to GPUs and large datasets, one can achieve comparable accuracy in the domain of kinship.

In Table 4, an exhaustive search for best training and testing feature is done in terms of verification accuracy using a 2-layer MLP network with 512 and 256 hidden nodes respectively. Accuracy is computed for features used in state-of-the-art deep learning systems as well as our proposed framework features for all the three datasets. A total of 63 experiments are performed. The most important observation in the table is that our proposed KinfaceNet extracted features have higher verification accuracy for both inter as well as intra dataset systems. The intra-dataset verification accuracies are the ones where an MLP model is trained and tested on the same dataset, whereas the inter-dataset verification accuracies are the ones where the MLP model is trained on one dataset and tested on a different dataset. The purpose of experimenting on the inter dataset is to check for which features the bias due the image properties is minimum. For example, in KinfaceW-I dataset, the kin pairs are extracted from the same family photo. So, a high inter dataset accuracy on KinfaceW-I dataset but very low intra dataset accuracy may show that the model is learning the non-kinship properties of the image as well. In the table the intra dataset accuracies for each type of features are marked in bold. The red color represents the highest intra-dataset accuracies observed. For all the three datasets the highest intra as well as inter dataset accuracy is observed using our proposed framework. Our framework without the transfer learning model, i.e., M1, failed to show such high accuracy in 150 epochs. The accuracies are even lower than using a pre-trained network.

Another notable observation is that the verification accuracy of our proposed framework features using a simple MLP network in the 20th epoch is much higher than the deep learning based state-of-the-art verification accuracies using features like LBP, face image, parts of face image, face recognition embedding as given in Table 1. In future, we plan to develop a verification system, which can help understand the features better and improve the accuracies.

5.4. Discussion

In this paper, we have proposed a framework, which is faster to train, better in performance and has many applications. The main advantage of this framework is that it can blend in any trained model from its parent face recognition domain and transfer the weights from the trained face

Table 4. Intra and Inter dataset Kinship Verification accuracies after 20 epochs of training a simple Multi Layer Perceptron network on the FIW, KinFaceW-I and KinFaceW-II datasets. LBP, Full face, parts of face(eyes and nose) and feature extracted using M2 are the features used in state-of-the-art deep learning based kinship verification. M3 is our proposed framework and M1 is proposed framework method without transfer learning.

Features	Datasets (Test)	FIW (Train)	KinFaceW-I (Train)	KinFaceW-II (Train)
LBP	**FIW**	**50.22**	48.88	48.88
	KinFaceW-I	54.54	**58.44**	51.94
	KinFaceW-II	49.65	62.75	**61.37**
Full Face	**FIW**	**52.08**	51.66	50.00
	KinFaceW-I	54.48	**60.45**	58.18
	KinFaceW-II	49.65	54.90	**52.23**
Parts of Face: Eyes	**FIW**	**50.22**	56.44	58.22
	KinFaceW-I	54.54	**62.33**	63.63
	KinFaceW-II	49.65	71.03	**68.96**
Parts of Face: Nose	**FIW**	**50.22**	51.11	57.77
	KinFaceW-I	54.54	**68.83**	70.12
	KinFaceW-II	49.65	66.89	**71.03**
KinfaceNet (M1)	**FIW**	**52.92**	50.00	50.00
	KinFaceW-I	41.50	**70.78**	53.93
	KinFaceW-II	56.00	50.00	**85.00**
Pre-trained face recognition model(M2)	**FIW**	**59.58**	59.58	63.69
	KinFaceW-I	68.53	**68.53**	50.56
	KinFaceW-II	59.50	58.50	**63.00**
KinfaceNet (M3)	**FIW**	**69.16**	65.00	64.16
	KinFaceW-I	62.92	**75.28**	66.29
	KinFaceW-II	73.00	80.50	**88.00**

recognition model, thus reducing the overall training time. Since, face recognition is a popular field, a plethora of such face feature extracting models are already available. Not only time complexity, another major necessity of the transfer learning technique in kinship recognition domain is to address balanced data shortage bottleneck. There are very few large kinship datasets, and none of the datasets have balanced kin and non-kin data. Our framework has proven to be better in performance than most of the state-of-the-art methods, even using a simple MLP verification system. In future, we plan to explore other complex verification systems to achieve higher accuracy. With the advent of face feature extractor in the face recognition domain, similar trend was observed.

While most of the work in the kinship domain is focused only on face verification, our work is focused on solving a bigger set of problems along with kinship verification. A set of applications are as follows:

1. **Kinship Verification, Finding missing /trafficked /smuggled family members and children.** Kinship verification can be useful in cases of missing elderly people with reduced cognitive capabilities, as well as in kidnapping cases. Also, in cases of human tracking, if a person is discovered many years after their disappearance, then it may help in reuniting with parents.
2. **Implementing new consumer product features.** Today, technology has come to a point where a camera can focus on only the humans faces and blur the rest. In the future, a camera that can focus on family faces rather than other faces in the background in group photos may be developed. Also, software and social robots may be able to distinguish family members from impostors.
3. **Document tagging.** Identifying people and their relationship from images has significant social and business values. It can provide a useful tool in social media clustering and analysis, similar trait categorization, etc.
4. **Forensics, historical, and genealogical research.** Although modern biological measures are available, most of them are inapplicable on a large scale. For example, DNA testing is widely used in paternity and crime scene investigations, but the tedious process takes days to generate results and is expensive. As discussed above, there is significant correlation between regions of DNA and facial features [1–3]. Our proposed framework proposes a simple way to amplify such cues, thus improving verification accuracy.
5. **Generating kin images.** Nowadays, unsupervised GAN-based models have helped in generating high quality synthetic facial images. The age, gender and demographics can be varied in the generated synthetic images by playing with the traits in the facial feature vector responsible for a particular soft-biometric. Kinship is one such trait which is present in all these facial feature vectors used in face recognition, but still needs to be explored to a large extent for better accuracy.

Our proposed feature extractor will be particularly helpful in generating the kin synthetic images and forensics, as similar trend is observed in its parent domain. In future, we plan to explore each of these domains using our extracted feature space.

6. Conclusion

This paper proposes a fast, better performing kinship feature extraction framework. The framework can blend in any trained face recognition model from its parent domain face recognition, and hence does not need prolong training or huge face dataset. This method also helps in addressing imbalanced dataset problem in kinship domain. Further, the independent kinship feature vector has a plethora of utility in different research domains. The trained models also provided high intra- and inter-dataset verification accuracies for the different datasets, even using a simple verification network. In future, many different applications of the representation of kinship cues as feature vector shall be explored.

Acknowledgement

This work has been supported in part by the US Army Research Office (ARO) under award number W911NF-19-1-0102.

References

[1] Karlijne Indencleef, Jasmien Roosenboom, Hanne Hoskens, Julie D White, Mark D Shriver, Stephen Richmond, Hilde

Peeters, Eleanor Feingold, Mary L Marazita, John R Shaffer, et al. Six nscl/p loci show associations with normal-range craniofacial variation. *Frontiers in genetics*, 9:502, 2018.

[2] John R. Shaffer Seth M. Weinberg. We scanned the dna of 8,000 people to see how facial features are controlled by genes. *The Conversation*.

[3] Stephen Richmond, Laurence J Howe, Sarah Lewis, Evie Stergiakouli, and Alexei Zhurov. Facial genetics: a brief overview. *Frontiers in genetics*, 9:462, 2018.

[4] Gwenaël Kaminski, Slimane Dridi, Christian Graff, and Edouard Gentaz. Human ability to detect kinship in strangers' faces: effects of the degree of relatedness. *Proceedings of the Royal Society B: Biological Sciences*, 276(1670):3193–3200, 2009.

[5] Miguel Bordallo Lopez, Abdenour Hadid, Elhocine Boutellaa, Jorge Goncalves, Vassilis Kostakos, and Simo Hosio. Kinship verification from facial images and videos: human versus machine. *Machine Vision and Applications*, 29:873–890, 2018.

[6] Jiwen Lu, Xiuzhuang Zhou, Yap-Pen Tan, Yuanyuan Shang, and Jie Zhou. Neighborhood repulsed metric learning for kinship verification. *IEEE transactions on pattern analysis and machine intelligence*, 36(2):331–345, 2013.

[7] Joseph P Robinson, Ming Shao, Yue Wu, Hongfu Liu, Timothy Gillis, and Yun Fu. Visual kinship recognition of families in the wild. *IEEE Transactions on pattern analysis and machine intelligence*, 40(11):2624–2637, 2018.

[8] Wei Wang, Shaodi You, Sezer Karaoglu, and Theo Gevers. A survey on kinship verification. *Neurocomputing*, 525:1–28, 2023.

[9] I Serraoui, Oualid Laiadi, Abdelmalik Ouamane, Fadi Dornaika, and Abdelmalik Taleb-Ahmed. Knowledge-based tensor subspace analysis system for kinship verification. *Neural Networks*, 151:222–237, 2022.

[10] Haibin Yan and Chaohui Song. Multi-scale deep relational reasoning for facial kinship verification. *Pattern Recognition*, 110:107541, 2021.

[11] Naman Kohli, Mayank Vatsa, Richa Singh, Afzel Noore, and Angshul Majumdar. Hierarchical representation learning for kinship verification. *IEEE Transactions on Image Processing*, 26(1):289–302, 2016.

[12] Wanhua Li, Yingqiang Zhang, Kangchen Lv, Jiwen Lu, Jianjiang Feng, and Jie Zhou. Graph-based kinship reasoning network. In *2020 IEEE international conference on multimedia and expo (ICME)*, pages 1–6. IEEE, 2020.

[13] Jiwen Lu, Junlin Hu, and Yap-Peng Tan. Discriminative deep metric learning for face and kinship verification. *IEEE Transactions on Image Processing*, 26(9):4269–4282, 2017.

[14] Xiaojing Chen, Le An, Songfan Yang, and Weimin Wu. Kinship verification in multi-linear coherent spaces. *Multimedia Tools and Applications*, 76:4105–4122, 2017.

[15] Kaihao Zhang12, Yongzhen Huang, Chunfeng Song, Hong Wu, Liang Wang, and Statistical Machine Intelligence. Kinship verification with deep convolutional neural networks. In *In British Machine Vision Conference (BMVC)*, 2015.

[16] Wanhua Li, Shiwei Wang, Jiwen Lu, Jianjiang Feng, and Jie Zhou. Meta-mining discriminative samples for kinship verification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16135–16144, 2021.

[17] Xiaopan Chen, Changlong Li, Xiaoke Zhu, Liang Zheng, Ya Chen, Shanshan Zheng, and Caihong Yuan. Deep discriminant generation-shared feature learning for image-based kinship verification. *Signal Processing: Image Communication*, 101:116543, 2022.

[18] Mengyin Wang, Zechao Li, Xiangbo Shu, Jingdong, and Jinhui Tang. Deep kinship verification. In *2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6, 2015.

[19] Haibin Yan and Shiwei Wang. Learning part-aware attention networks for kinship verification. *Pattern Recognition Letters*, 128:169–175, 2019.

[20] Eran Dahan and Yosi Keller. A unified approach to kinship verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8):2851–2857, 2020.

[21] Andrei Shadrikov. Achieving better kinship recognition through better baseline. In *2020 15th IEEE international conference on automatic face and gesture recognition (FG 2020)*, pages 872–876. IEEE, 2020.

[22] Stefan Hörmann, Martin Knoche, and Gerhard Rigoll. A multi-task comparator framework for kinship verification. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 863–867. IEEE, 2020.

[23] Lei Zhang, Qingyan Duan, David Zhang, Wei Jia, and Xizhao Wang. Advkin: Adversarial convolutional network for kinship verification. *IEEE transactions on cybernetics*, 51(12):5883–5896, 2020.

[24] Siyu Xia, Ming Shao, and Yun Fu. Toward kinship verification using visual attributes. In *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*, pages 549–552. IEEE, 2012.

[25] Aarti Goyal and Toshanlal Meenpal. Patch-based dual-tree complex wavelet transform for kinship recognition. *IEEE Transactions on Image Processing*, 30:191–206, 2020.

[26] Zhipeng Luo, Zhiguang Zhang, Zhenyu Xu, and Lixuan Che. Challenge report recognizing families in the wild data challenge. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 868–871. IEEE, 2020.

[27] Mohammed Almuashi, Siti Zaiton Mohd Hashim, Dzulkifli Mohamad, Mohammed Hazim Alkawaz, and Aida Ali. Automated kinship verification and identification through human facial images: a survey. *Multimedia Tools and Applications*, 76:265–307, 2017.

[28] Joseph P. Robinson, Ming Shao, Yue Wu, and Yun Fu. Families in the wild (fiw): Large-scale kinship image database and benchmarks. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 242–246. ACM, 2016.

[29] Masoud Faraki, Mehrtash T Harandi, and Fatih Porikli. More about vlad: A leap from euclidean to riemannian manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4951–4960, 2015.

[30] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[31] T. Andersen and T. Martinez. Cross validation and mlp architecture selection. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339)*, volume 3, pages 1614–1619 vol.3, 1999.

[32] David Ruppert. The elements of statistical learning: data mining, inference, and prediction, 2004.

[33] Ning Zhang, Junmin Luo, and Wuqi Gao. Research on face detection technology based on mtcnn. In *2020 international conference on computer network, electronic and automation (ICCNEA)*, pages 154–158. IEEE, 2020.

[34] David Sandberg. Face recognition using tensorflow. *github*, 2018.

[35] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[36] Joseph P. Robinson, Zaid Khan, Yu Yin, Ming Shao, and Yun Fu. Families in wild multimedia (fiw-mm): A multi-modal database for recognizing kinship. *Transa*, 2020.

[37] RFIW. Rfiw2020 tri-subject verification. 2020.

[38] J. Lu, J. Hu, X. Zhou, Y. Shang, Y.-P. Tan, and G. Wang. Neighborhood repulsed metric learning for kinship verification. In *IEEE International Conference on Computer Vision and Pattern Recognition, 2012*, pages 2594–2601. IEEE, 2012.

[39] Tamer Tahamoqa. facenet-pytorch-glint360k. *github*, 2021.

[40] Xiang An, Xuhan Zhu, Yuan Gao, Yang Xiao, Yongle Zhao, Ziyong Feng, Lan Wu, Bin Qin, Ming Zhang, Debing Zhang, et al. Partial fc: Training 10 million identities on a single machine. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1445–1449, 2021.