Kinship verification via ear images: A comparative study

Phuong Quang Luu[1], Bay Van Nguyen[1], Huy Quoc Nguyen[1*]

[1]Ho Chi Minh City Open University, Ho Chi Minh City, Vietnam
[*]Corresponding author: huy.nq@ou.edu.vn

ARTICLE INFO

DOI:10.46223/HCMCOUJS.tech.en.15.1.3683.2025

Received: August 23rd, 2024

Revised: September 23rd, 2024

Accepted: September 27th, 2024

Keywords:

deep learning; ear biometric; kinship verification

ABSTRACT

Kinship verification is crucial in daily life, especially in the legal field. Nowadays, most kinship verification methods utilize the advantages of human DNA and facial features. However, these methods require a lot of complex procedures, so they are unsuitable for real-time application. Therefore, researchers started to propose other promising biometrics, and the human ear is one of the most potential. The human ear has long been recognized as a robust biometric trait, comparable to others, such as face, iris, and fingerprint. This paper proposes using ear images to identify human kinship based on several well-known deep-learning networks. Moreover, an ear image set is presented to tackle the lack of a kinship-annotated dataset.

1. Introduction

Kinship verification has various applications in daily life, including inheritance of assets, finding long-lost relatives, and resident analysis. With the advantages of modern medicine, kinship verification has become amazingly accurate, and most methods are based on human DNA structure. However, these methods involve numerous steps and complex procedures that are time-consuming and unsuitable for real-time applications. In certain critical cases, prolonged delays in determining kinship can have significant negative consequences. In forensic investigations, untimely identification of familial relationships could impede the progress of criminal cases, delaying justice. Moreover, in medical scenarios, the inability to establish biological relationships could prevent timely medical interventions, such as organ transplants or treating hereditary conditions. To speed up kinship verification progress, researchers leverage the advantages of computer vision. They apply images of biometric traits, such as face, iris, and fingerprint, to identify kinship between two people. For instance, Yan and Song (2020) presented a relational network based on CNN for kinship identification using facial images. Guo and Wang (2012) deployed the DAISY method to extract essential information from facial images to support kinship prediction.

However, in practice, the mentioned biometric traits are inappropriate for kinship verification. Faces can age through time, which causes the loss of several crucial kinship characteristics. Fingerprints and iris do not carry much kinship information. Consequently, researchers have started to search for other potential biometric traits. In recent years, the human ear, particularly the shape of the helix, has been proven to have uniqueness comparable to other biometric traits (Emeršič et al., 2017; Jain et al., 2006). As a result, ear biometrics has gained the attention of various researchers and has been applied to tackle several computer vision tasks, such as recognition, classification, detection, and verification. Alshazly et al. (2020) deployed several standard CNN models for ear recognition, and the experimental results show that ResNeXt outperformed other models with an accuracy of 93.45% (Alshazly et al., 2020). Nguyen and Hoang (2021a) introduced a modern ear detection pipeline based on the YOLO detector and RetinaFace to support ear-related tasks.

Moreover, Dvorsak et al. (2022) introduced a kinship verification system based on human ear images, demonstrating that deep learning models can effectively extract valuable kinship-related information from ear images. Inspired by previous research, we propose a robust kinship verification system using human ear images. Our system is built on the foundation of the Siamese network architecture and advanced deep learning models, enabling precise verification of the relationship between two input images. Furthermore, our system is trained to address the binary classification problem (kinship-related or non-related) and tackle a more extended multi-class classification task. This allows for accurately identifying specific relationships between the two input ear images, such as father-son, mother-son, and brother-sister.

The contribution of our work can be summarized as follows:

- Analyzed and evaluated the performance of several CNN-based and Transformer-based models to find the most optimized method for kinship verification tasks.

- The experimental results proved that human ear features have promising potential for computer vision tasks.

- Presented a new dataset of unconstrained ear images labeled for kinship verification tasks.

2. Theoretical basis

2.1. Kinship verification

Kinship verification is a common problem that has existed for a long time. A kinship verification system takes the inputs of two individuals. These inputs are usually the characteristic traits of two people, such as DNA, face, and fingerprint. After analyzing and extracting essential features of these inputs, the system predicts whether these two people are blood-related or non-related.

Before the development of artificial intelligence and machine learning, most kinship verification methods are based on genetic techniques, including DNA analysis and DNA sequencing. Until 2010, when machine learning began to flourish and be widely applied across various fields, the approach of kinship verification through machine learning gained the attention of many researchers. Remarkably, researchers have started to explore kinship information in the shapes of human biometric traits, such as face, fingerprint, iris, and ear, rather than just analyzing human DNA patterns. This shift expanded the kinship verification task from the genetic field to the realm of computer vision. At first, the face was the most popular biometric trait (Mzoughi et al., 2024), and the proposed methods then were mainly based on handcrafted algorithms. For example, Chergui et al. (2020) deployed several handcrafted features, such as LBP, SIFT, LTP, and LPQ, combined with Pyramid Multilevel to extract the kinship information from face images and predict the blood relationship between two representation vectors (Chergui et al., 2018; Chergui et al., 2020). Bessaoudi et al. (2020) combine handcrafted features with several deep-learning models to improve the performance of their kinship verification system via face images. Lots of comparative studies of kinship verification via face images were proposed, providing statistics and analysis of several preceding kinship verification methods, including handcrafted algorithms and deep learning methods (Mzoughi et al., 2024; Wang et al., 2022). The datasets applied in these works include Cornell KinFace, UB KinFace, KinFaceW-I, and KinFaceW-II.

However, handcrafted methods became obsolete and inappropriate for real-time applications. Therefore, researchers proposed using deep learning methods instead, specifically CNN-based models. Dahan and Keller (2020) created a kinship verification model based on Sphereface and features fusion, and the authors proved that the proposed model not only can verify the kinship between two people but also can identify which relationship between them with an accuracy of 79.6%, higher than other methods in the past. Nandy and Mondal (2019) proposed using Siamese network architecture for kinship verification via face images; SqueezeNet is applied as the backbone model to extract features from input images; the authors also utilized the trained weight of VGGFace2 model to improve the training performance of the proposed model, experimental result show that the proposed method scored 67.66% of Cosine Similarity evaluation.

Transformer is also a potential approach for kinship verification. For example, Li and Jiang (2023) present GLANet based on Vision Transformer (ViT) with an accuracy of 79.6% on the FIW dataset. Zhu et al. (2022) modified the original ViT to be suitable for face kinship verification. However, the variety of Transformer-based methods for kinship verification is poor.

In our observation, most proposed kinship verification methods are based on the Siamese network. This architecture has been proven to be a robust template for all verification tasks, including kinship verification. Therefore, we also leverage the advantages of the Siamese network in our proposed method.

2.2. Ear biometric

The human ear has been long recognized as one of the unique characteristic features of the human body. In 1963, the first ear identification system was created by Manuel (1963). This system has proved the uniqueness of the shape of the human ear, comparable to other biometric traits, such as face fingerprints. After that, researchers successfully analyzed and showed that the helix, anti-helix, and other parts of the ear have formed several curves that create the discriminate shape of the human ear (Jain et al., 2006). Moreover, the shape of the left ear and right ear of the same person are also different from each other.

On the other hand, biometric traits, such as face, are noticeably aged through time, which causes heavy changes to facial features and kinship information. The ear's shape also suffers from aging but is less and slower. Ear images are easier to collect for training models.

Due to the potential of ear biometrics, many researchers have applied ear to tackle several computer vision tasks. Hassaballah et al. (2018) used an LBP descriptor to extract essential features from ear images for recognition. Emeršič et al. (2018) presented an end-to-end ear recognition system based on CNN. For gender classification, Nguyen and Hoang (2020) analyzed and compared various methods, including handcrafted algorithms and deep learning models for gender classification via ear images on the EarVN1.0 dataset. For detection, Wahab et al. (2012) proposed HEARD, an automatic ear detection technique. Nguyen and Hoang (2021b) introduce an ear detection pipeline based on YOLOv5 (the latest version of YOLO in 2021); the proposed method can also classify the left ear and right ear (Nguyen & Hoang, 2021b).

Additionally, ear biometrics have the potential to be integrated with other biometric traits to enhance performance in solving more complex problems. This multimodal approach could leverage the strengths of various biometric features, such as facial recognition, fingerprints, or iris scans, to increase accuracy and robustness, particularly in scenarios where a single biometric modality may not be sufficient. Li et al. (2020) combined images of hands and ears to detect driver distraction. This approach highlights the potential of using multiple biometric inputs to enhance the accuracy and reliability of detecting driver distraction, improving road safety, and preventing accidents caused by inattentiveness. Ma et al. (2020) combined facial and ear images to enhance the accuracy of recognition models. This multimodal strategy demonstrates the effectiveness of combining different biometric traits and the potential of ear images in computer vision tasks.

Inspired by these significant works, we proposed using ear images for kinship verification to show that ear biometrics also carry each individual's useful kinship information.

2.3. Kinship verification via ear images

To our knowledge, ear biometrics had not been commonly applied in kinship verification. Dvorsak et al. (2022) propose the most recent ear-based work. In their work, the authors used Siamese architecture to verify the kinship of a pair of ear images. The backbones used in their works consisted of VGG16, ResNet152, USTC-NELSLIP (ResNet50 variant), Attentional Feature Fusion (AFF) and Contextual Transformer Network (CoTNet) (Dvorsak et al., 2022). After evaluating the proposed method on the KinEar dataset, the VGG16 model gave the best performance with an accuracy of 64.01%. The author believes this result is auspicious, and the proposed method can be improved further due to the potential of ear biometrics. This work is the basis for our proposed method. However, our proposed method can achieve higher performance with several crucial modifications.

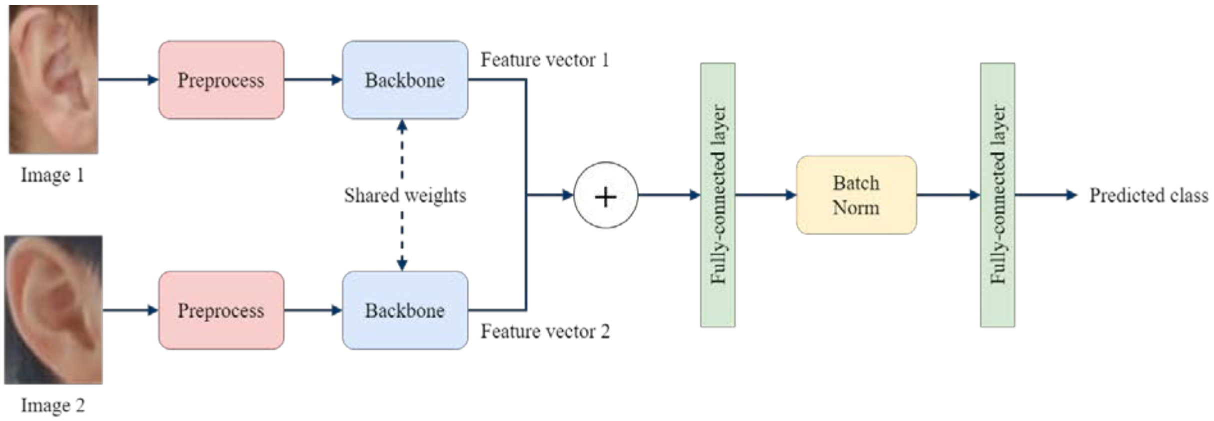
3. Methodology

3.1. Proposed architecture

In this paper, we proposed using Siamese network architecture for kinship verification. A basic Siamese network consists of two neural networks that work simultaneously. Each network takes the input data to extract the feature vector, and the resulting vectors are then combined and fed to a fully connected network for classification. It is important to note that these networks share the same weights and biases. Figure 1 illustrates the proposed kinship verification system.

Figure 1

An Illustration of Proposed Kinship Verification System



Source. The researcher's data analysis

At first, the input images are preprocessed, such as resizing images to $224 \times 224 \times 3$ for training deep learning models. Image augmentation is also applied to increase the variety and complexity of the dataset and avoid overfitting. We utilize the functionality from the Torchvision library and use random zoom, random translation by a factor of 0.1, random rotation, random horizontal flip, and random adjust sharpness by a factor of 2. Moreover, these augmentation techniques also help us solve Transformer models' data starvation, such as ViT.

After that, the input images are fed to our proposed architecture. The experimental backbones consisted of VGG, ResNet, DenseNet, and ViT to extract feature vectors. In our Siamese architecture, output vectors of both CNN models are combined and provided to two fully connected layers to predict the kinship relationship of the input pair of ear images. We add a batch normalization layer between these fully connected layers to avoid overfitting. The process of the proposed kinship verification system is formulated as follows:

$$x'_1 = f(x_1) \tag{1}$$

$$x'_2 = f(x_2) \tag{2}$$

$$y = FC_2(BN\left(FC_1(x'_1 + x'_2)\right)) \tag{3}$$

Where x_1 and x_2 are the input images, x'_1 và x'_2 are the extracted feature vectors, f is a deep learning model, FC_1 and FC_2 are fully connected layers, BN is the batch normalization layer, y is the prediction of the system.

3.2. Deep learning backbones

VGG is first introduced by Simonyan and Zisserman (2014). The improved version of the preceding AlexNet with more convolutional layers gives the model more information about the input image. VGG won ImageNet in 2014 with an accuracy of 92.7%. Its architecture utilized the advantages of the ReLU activation function and dropout techniques to avoid overfitting and vanishing gradients. VGG16 and VGG19 are the most common versions of VGG, where the numbers stand for the number of layers of each model.

ResNet, proposed by He et al. (2016), stands for Residual Network. The authors' first idea is to create a model as deep as possible since the more profound the model is, the more information the model can theoretically extract from the input data. However, deep models may cause vanishing gradients in practice, leading to lousy training performance. To tackle this limitation, the authors proposed a residual block with a skip connection. The input data is reutilized and combined with the extracted data across the model to keep the crucial information of the input image and avoid vanishing gradient throughout the training process. Moreover, ResNet introduced several optimizing techniques, such as batch normalization and bottleneck block.

DenseNet is a CNN model that ResNet inspires. DenseNet, a Densely Connected Convolutional Network, is first presented by Huang et al. (2017). DenseNet consists of multiple dense blocks. Each block reuses the input data and combines all features of the previous block in the model. However, this work requires lots of computations. Therefore, the authors add several 1×1 convolutional layers to reduce the size of the processing data. Hence, DenseNet can be optimally deeper than ResNet. The most profound version of ResNet has 152 layers, but DenseNet has 169-layer and 201-layer versions.

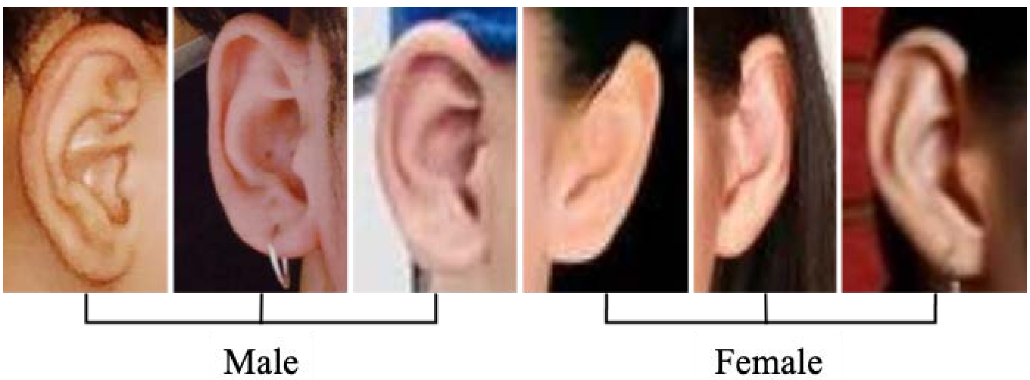
ViT is a deep-learning model for image classification based on Transformer (Dosovitskiy et al., 2020). ViT is proposed by Dosovitskiy et al. (2020). The Transformer is designed for natural language processing and contains an encoder and a decoder. At first, Transformer showed an outstanding performance that outperformed the NLP algorithms in text classification, sentiment analysis, and machine translation. Therefore, Dosovitskiy et al. (2020) modified the Transformer to suit computer vision tasks while maintaining the original architecture's robustness. The most crucial part of ViT is patch embedding, where the model divides the input image into patches (smaller images from the original one); each patch has the exact size of 16×16. These patches are then flattened into vectors (input sequence) to meet the requirements of the Transformer, as the original Transformer works with sequence data. After that, the input sequence is assigned several important information, such as the position of each patch (positional embedding) and a [CLS] token to extract the model classification information. ViT architecture is mainly based on BERT, specifically the encoder of the original Transformer. After several Transformer blocks, including Multihead Self-Attention, Layer Normalization, and Multi-layer Perceptron (a simple, fully connected network), ViT extracts the local feature of each patch and produces an output sequence. The authors only extracted [CLS] tokens for prediction in the output sequence. ViT is a new step in the field of computer vision. Nowadays, most computer vision tasks are applied or inspired by ViT. However, because the nature of ViT is a Transformer, for the model to operate as optimally as possible, researchers need to train the model with a lot of data.

3.3. Experimental dataset

In this paper, we utilized three ear image sets, including KinEar, EarVN1.0, and our self-collected dataset, which we temporarily named KinEarVN. EarVN1.0 is an ear dataset of Vietnamese celebrities with 28,412 images annotated for identification and gender classification (Hoang, 2019). We applied EarVN1.0 to pre-train the backbone models with gender classification tasks. We aim to get these models used to the shape of the human ear, and then we take the trained weight to initialize the training kinship verification task. Figure 2 illustrates several samples of the EarVN1.0 dataset.

Figure 2

An Illustration of several Samples of EarVN1.0 Dataset



Source. The researcher's data analysis

KinEar and KinEarVN are used to train and evaluate the performance of the proposed kinship verification system. KinEar is a human ear image set that is annotated for kinship information. KinEar is first introduced in the work of Dvorsak et al. (2022). This dataset contains 1,477 images of 19 families. The number of available pairs of images is 44,627 pairs, such as father-son, father-daughter, mother-son, mother-daughter, brother-brother, brother-sister, sister-sister, and father-mother (non-related). Therefore, the number of classes is 08, according to the above relationships. We divided these pairs into three subsets: the train set contains 60% of the dataset (26,783 pairs), the validation set includes 10% of the dataset (4,460 pairs), and the test set contains 30% of the dataset (13,384 pairs). Figure 3 illustrates several samples of the KinEar dataset.

Figure 3

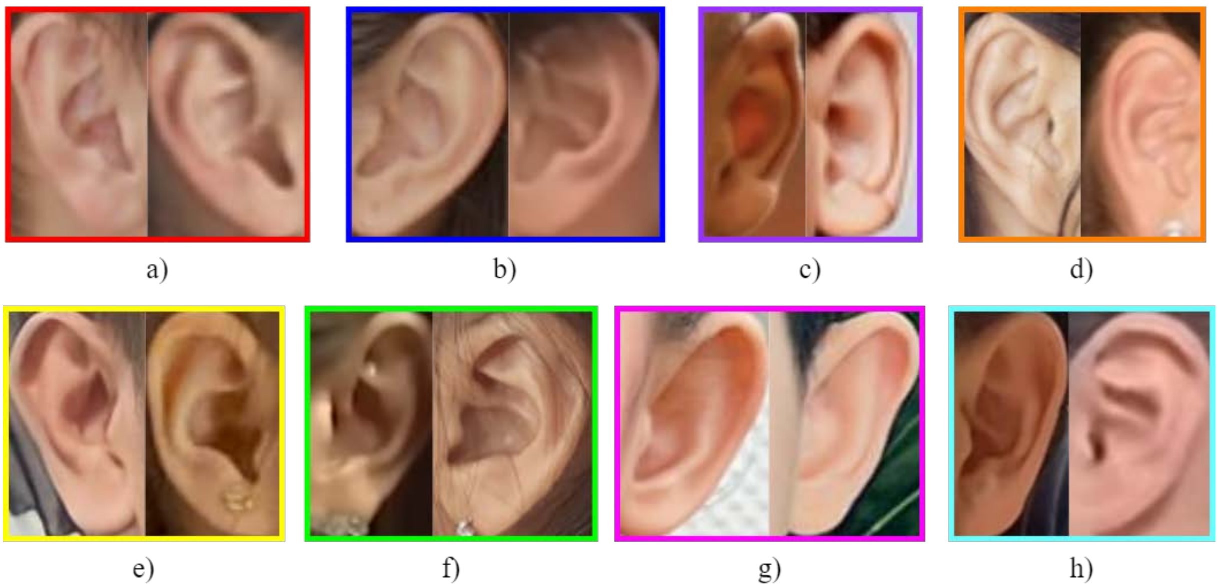
An Illustration of several Samples of KinEar Dataset

Source. The researcher's data analysis

We noticed that the number of pairs in the KinEar dataset is tiny, which is not the best circumstance for deep learning models, especially Transformer. Therefore, we collect and annotate a larger ear image set from open sources and social media. Our dataset contains 4,876 images of 156 families. The number of available pairs of ears is 85,423 pairs, with 08 classes similar to the KinEar dataset. We divided these pairs into three subsets: the train set contains 60% of the dataset (51,263 pairs), the validation set includes 10% of the dataset (8,538 pairs), and the test set contains 30% of the dataset (25,622 pairs). Figure 4 illustrates several samples of the KinEarVN dataset.

Figure 4

An Illustration of several Samples of the KinEarVN Dataset



Note. a) Father-son b) Mother-son c) Brother-brother d) Brother-sister
e) Father-daughter f) Mother-daughter g) Sister-sister h) Father-mother

Source. The researcher's data analysis

4. Results and discussion

4.1. Results

In the experiment, we train each proposed model by 40 epochs. We deploy the Adam optimizer with a learning rate 0.0001, and the loss function is Cross-Entropy (CE). These hyperparameters are considered after several experiments. For the number of epochs, we first initialize by 100 epochs, but we soon recognize that the performance of the deep learning models does not increase much after 40 epochs. Therefore, we reduce the number of training epochs to save more time. All experiments are deployed on a computer with 256GB RAM and Nvidia RTX A5000 24GB GPU. Table 1 shows experimental results on the KinEar dataset and KinEarVN dataset.

Table 1

Experimental Results on KinEar Dataset and KinEarVN Dataset (%)

Models	**KinEar**		**KinEarVN**	
	Train	Test	Train	Test
VGG16	84.79	84.99	75.05	74.39
ResNet50	98.86	98.08	90.03	89.38
ResNet152	98.57	98.17	91.66	91.36
DenseNet121	99.28	98.89	89.70	88.40
DenseNet169	99.43	98.76	95.03	94.19
ViT-b-16	93.50	93.48	81.17	80.85
ViT-b-32	87.93	87.82	75.60	75.79

Source. Data analysis result of the research

4.2. Discussion

Due to the pre-trained weights on EarVN1.0, the experimental results are enhanced by 10% to 20% of accuracy. These results prove the importance and benefit of a large amount of training data and transfer learning techniques. Our observation shows that DenseNet models perform best (Table 1). Specifically, on the KinEar dataset, DenseNet121 achieves an accuracy of 98.89% on the test set. On the KinEarVN dataset, we obtain the highest accuracy from the DenseNet169 model on the test set. ViT models are expected to outperform the others but can only achieve 93.48% accuracy on KinEar and 80.85% on KinEarVN, which is lower than the performance of DenseNet models. To our knowledge, this ineffective result is partly because of the lack of training data. This problem is shown in the performance of the ViT-b-32 model, which has a more profound architecture but lower accuracy than the ViT-b-16 model. For the VGG16, this model performed the best in the previous work of Dvorsak et al. (2022), but with our proposed architecture, this model has the lowest accuracy. The reason is mainly because the complexity of the kinship verification system has extended from binary to multi-class classification tasks.

In conclusion, we believe that DenseNet is the most optimized backbone for our kinship verification system. On the other hand, the accuracy of the ResNet model is not as high as the DenseNet model. Still, it has fewer parameters, so the computational cost is lower and more suitable for real-time applications.

5. Conclusions & recommendations

Kinship verification is an essential problem in many aspects of life. Especially in the digital transformation process, fast kinship verification via images is highly urgent and understandable. Therefore, we proposed a kinship verification system based on Siamese architecture and deep learning models. We evaluate several well-known backbones to achieve the most optimized model that can combine with our system and give the best performance. After several experiments, we choose DenseNet with the highest accuracy of 98.89% on the benchmark KinEar and 94.19% on our larger KinEarVN dataset. Our work has proved that human ears not only contain information about gender and identity, comparable to other biometric traits, but also carry valuable information about blood relations between people. This is one of the most critical contributions affecting not only the field of computer science but also the fields of biology and genetics. We plan to integrate additional functions, such as ear detectors and ear identifiers, into the system pipeline to create a complete application and deploy it in practice.

References

Alshazly, H., Linse, C., Barth, E., & Martinetz, T. (2020). Deep convolutional neural networks for unconstrained ear recognition. *IEEE Access*, *8*, 170295-170310. https://doi.org/10.1109/access.2020.3024116

Bessaoudi, M., Chouchane, A., Ouamane, A., & Boutellaa, E. (2020). Multilinear subspace learning using handcrafted and deep features for face kinship verification in the wild. *Applied Intelligence*, *51*(6), 3534-3547. https://doi.org/10.1007/s10489-020-02044-0

Chergui, A., Ouchtati, S., Mavromatis, S., Bekhouche, S. E., Sequeira, J., & Dornaika, F. (2020). Kinship verification through facial images using multiscale and multilevel handcrafted features. *Journal of Electronic Imaging*, *29*(2), 023017-023017. https://doi.org/10.1117/1.jei.29.2.023017

Chergui, A., Ouchtati, S., Sequeira, J., Bekhouche, S. E., & Bougourzi, F. (2018). Kinship verification using BSIF and LBP. In *2018 International Conference on Signal, Image, Vision and their Applications (SIVA)* (pp. 1-5). IEEE. https://doi.org/10.1109/siva.2018.8661085

Dahan, E., & Keller, Y. (2020). A unified approach to kinship verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(8), 2851-2857. https://doi.org/10.1109/tpami.2020.3036993

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). *An image is worth 16x16 words: Transformers for image recognition at scale*. https://doi.org/10.48550/arxiv.2010.11929

Dvorsak, G., Dwivedi, A., Struc, V., Peer, P., & Emersic, Z. (2022). Kinship verification from ear images: An explorative study with deep learning models. In *2022 International Workshop on Biometrics and Forensics (IWBF)* (pp. 1-6). IEEE. https://doi.org/10.1109/iwbf55382.2022.9794555

Emeršič, Ž., Križaj, J., Štruc, V., & Peer, P. (2018). Deep ear recognition pipeline. In *Studies in computational intelligence* (Vol. 804, pp. 333-362). Springer. https://doi.org/10.1007/978-3-030-03000-1_14

Emeršič, Ž., Štruc, V., & Peer, P. (2017). Ear recognition: More than a survey. *Neurocomputing*, *255*, 26-39. https://doi.org/10.1016/j.neucom.2016.08.139

Guo, G., & Wang, X. (2012). Kinship measurement on salient facial features. *IEEE Transactions on Instrumentation and Measurement*, *61*(8), 2322-2325. https://doi.org/10.1109/tim.2012.2187468

Hassaballah, M., Alshazly, H. A., & Ali, A. A. (2018). Ear recognition using local binary patterns: A comparative experimental study. *Expert Systems with Applications*, *118*, 182-200. https://doi.org/10.1016/j.eswa.2018.10.007

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). IEEE. https://doi.org/10.1109/cvpr.2016.90

Hoang, V. T. (2019). EarVN1.0: A new large-scale ear images dataset in the wild. *Data in Brief*, *27*, Article 104630. https://doi.org/10.1016/j.dib.2019.104630

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4700-4708). IEEE. https://doi.org/10.1109/cvpr.2017.243

Jain, A., Bolle, R., & Pankanti, S. (2006). Introduction to biometrics. In *Springer eBooks* (pp. 1-41). https://doi.org/10.1007/0-306-47044-6_1

Li, D., & Jiang, X. (2023). Kinship verification method of face image deep feature fusion. *Academic Journal of Science and Technology*, *5*(1), 57-62. https://doi.org/10.54097/ajst.v5i1.5348

Li, L., Zhong, B., Hutmacher, C., Liang, Y., Horrey, W. J., & Xu, X. (2020). Detection of driver manual distraction via image-based hand and ear recognition. *Accident Analysis & Prevention*, *137*, Article 105432. https://doi.org/10.1016/j.aap.2020.105432

Ma, Y., Huang, Z., Wang, X., & Huang, K. (2020). An overview of multimodal biometrics using the face and ear. *Mathematical Problems in Engineering*, *2020*, 1-17. https://doi.org/10.1155/2020/6802905

Manuel, Z. (1963). U.*S. Patent No. 3,102,459*. U.S. Patent and Trademark Office.

Mzoughi, M. C., Aoun, N. B., & Naouali, S. (2024). A review on kinship verification from facial information. *The Visual Computer*, 1-21. https://doi.org/10.1007/s00371-024-03493-1

Nandy, A., & Mondal, S. S. (2019). Kinship verification using deep siamese convolutional neural network. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)* (pp. 1-5). IEEE. https://doi.org/10.1109/fg.2019.8756528

Nguyen, H. Q., & Hoang, V. T. (2020). Gender recognition based on ear images: A comparative experimental study. *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, *38*, 451-456. https://doi.org/10.1109/isriti51436.2020.9315366

Nguyen, H. Q., & Hoang, V. T. (2021a). Real-time human ear detection based on the joint of Yolo and RetinaFace. *Complexity*, *2021*(1). https://doi.org/10.1155/2021/7918165

Nguyen, H. Q., & Hoang, V. T. (2021b). Human ear-side detection based on YOLOv5 detector and deep neural networks. In *2021 International Conference on Decision Aid Sciences and Application (DASA)* (pp. 819-823). IEEE. https://doi.org/10.1109/dasa53625.2021.9682306

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. https://doi.org/10.48550/arxiv.1409.1556

Wahab, N. K. A., Hemayed, E. E., & Fayek, M. B. (2012). HEARD: An automatic human EAR detection technique. In *2012 International Conference on Engineering and Technology (ICET)* (pp. 1-7). IEEE. https://doi.org/10.1109/icengtechnol.2012.6396118

Wang, W., You, S., Karaoglu, S., & Gevers, T. (2022). A survey on kinship verification. *Neurocomputing*, *525*, 1-28. https://doi.org/10.1016/j.neucom.2022.12.031

Yan, H., & Song, C. (2020). Multi-scale deep relational reasoning for facial kinship verification. *Pattern Recognition*, *110*, Article 107541. https://doi.org/10.1016/j.patcog.2020.107541

Zhu, X., Li, Y., Li, D., Dong, L., & Chen, X. (2022). Lightweight transformer network and self-supervised task for Kinship verification. In *2022 IEEE 8th International Conference on Computer and Communications (ICCC)* (pp. 1506-1511). IEEE. https://doi.org/10.1109/iccc56324.2022.10066034