BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2025
http://excel.fit.vutbr.cz

Person Identification via Ear Biometrics

Bc. Gregor Karetka*

Abstract
Unconventional methods of personal biometrics are gaining popularity not only in academic circles but also in the commercial sphere. This paper focuses on the human ear as an alternative biometric modality and builds on top of the current trends in ear recognition. In this paper, we present a method for generating a dataset for ear recognition, and we trained multiple deep-learning models on an existing ear recognition dataset. Furthermore, we thoroughly evaluate and compare the models and training methods used and current state-of-the-art research in ear recognition.

*xkaret00@stud.fit.vut.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Ear recognition is a key area of research in the field of biometrics. The benefits of ear recognition are that the ear structure is unique to every human, taking a sample of an ear is non-invasive and contactless, and the ear changes only slightly with age, making it a long-lasting biometric trait. [1]

Ear recognition can be applied to person verification and identification tasks. In the identification scenario, the ideal system would be able to correctly assign an identity to the input image of an ear. In contrast, in the verification scenario, the system, given two images of an ear, is able to determine whether the ears belong to the same person or a different person. Furthermore, as noted in the [2], the system should not present a demographic bias. In this paper, we present a method for generating a synthetic dataset for ear recognition, and we present the results and setup of trained models on the UERC 2023 [2] dataset and on the newly generated synthetic dataset and provide a comparison with the current state-of-the-art models. The main improvement over previous state-of-the-art solutions is the usage of more aggressive augmentations, multiple training steps, each with a different loss function, and utilization of a newer vision model, which resulted in better performance in EER and GINI. Unfortunately, the synthetic dataset did not improve the performance. Therefore, the training was conducted with just the UERC 2023 dataset, but it presents a possibility for future research in this domain.

2. Commentary

2.1 Training setup architecture

All the experiments are based on a single training setup architecture, as seen in *Figure 1*, where the main difference between each setup is in the input data, augmentations, vision transformer used (which also impacts the embedding size), and loss function.

2.1.1 ArcFace

An ArcFace [3] loss function is used as a baseline training method. This type of angular margin loss minimizes the angular distance between embedding vectors in the same class (person) while keeping a margin from other classes. After training with this loss function, the model is able to generate discriminative embeddings for each class, which can be seen in *Table 1* (denoted as ArcFace only).

2.1.2 Triplet loss

A triplet loss [4] is used as the second step after training the model with ArcFace [3]. In order to improve intra-gender and intra-ethnical performance, the triplets are generated such that a negative example is from the same gender and ethnical group as the anchor and its corresponding positive example. This training procedure slightly improved AUC and F1F metrics. The model results can be seen in *Table 1* (denoted as ArcFace + triplet).

2.1.3 Self-knowledge distillation

A self-knowledge distillation [5] is employed as a third step to generate more robust embeddings. In this

step, the teacher and student are the same model. The teacher is fixed and receives the original image as input, while the student receives an augmented image as input. The objective is to minimize the distance between the augmented and original image. We experimented with several loss functions (distance metrics); their respective results can be seen in *Table 1* (denoted as self-learning).

2.1.4 Open Set Loss

In order to improve identification metrics (such as F1F), an open set loss [6] is integrated into training with ArcFace. This training method did not significantly improve F1F, EER, or AUC when used with a small batch size. On the other hand, when used with a smaller model, which allowed for a larger batch size, the improvements were significant. The training experiments with this loss function are denoted in *Table 1* as osl.

2.2 Data preparation

As data preprocessing (eg. face alignment) is a common technique in face recognition and ear recognition, we implemented a similar approach. First, we trained a model to predict key points of the ear. Then, using the obtained key points, we rotated the ear and stretched the bounding box to cover the same image area. The process can be seen in *Figure 2*. This approach did not yield better results than using original images, yet it presents a promising research direction. The results are in *Table 1*, denoted as norm + cov.

2.3 Results

The evaluation of the models closely follows UERC 2023 [2] competition to assess the performance with the current SOTA methods. The UERC 2023 competition compares not only the raw verification and identification performance but also the bias (and performance) in different gender and ethnical groups.

2.3.1 Metrics

The main metrics evaluated in the UERC 2023 [2] competition are EER (equivalent error rate), AUC (area under ROC curve), F1F (False Non-Match Rate at 1% False Match Rate – FNMR @ 1% FMR), R1% (Rank-1 accuracy), GINI index computed over EER, representing demographic bias and UERC Ranking, which is a combination of all the metrics.

2.3.2 Results description

All results can be seen in *Table 1*. The best results (in baseline models and in newly trained models) per each metric (eg. AUC) are highlighted in **bold**. The first eight results are taken directly from UERC 2023 [2] paper and are considered baseline. Newly trained models that perform the best in any metric are marked with a light red color and included compared to baseline models.

2.3.3 Different demographic groups

The performance of the models in different demographic groups is presented in *Figure 3*, where it can be observed that some models outperform other models in some gender-ethnical groups. For example, MEM-EAR outperforms all the other models in Female - Black category, EVA02-Base (EER) outperforms all other models in Male - White category by a large margin. On the other hand, ViTEar and EVA02-Tiny perform poorly in Male - Black category.

2.4 Variance in performance per model

The variance of performance of individual models can be seen in *Figure 4*. While models such as UERC Baseline, RecogEAR, and EVA02-Tiny present poor performance overall, the performance of these models across different ethnical and gender groups does not vary as much as in better models, which results in a good bias performance.

2.5 Size of the model vs EER and GINI

The size of a machine learning model greatly affects its usability, mainly whether it can be used on embedded devices. A GINI vs EER tradeoff plot is shown in *Figure 5*. The most optimal solution lies in the coordinates system's origin. It can be seen that newly trained models such as EVA02-Tiny, EVA02-Base, and EVA02-BASE (EER) are pareto-optimal models, presenting very good EER performance while maintaining good GINI performance. The size of the individual dots represents a number of parameters of the model.

Acknowledgements

I want to express my gratitude to my advisor, Ing. Jakub Špaňhel Ph.D., for his invaluable guidance, patience, and support throughout the course of this thesis. I am also thankful to Dr. Žiga Emeršič and Prof. Vitomir Štruc for providing us with the UERC 2023 dataset and evaluation scripts.

References

[1] Amir Benzaoui, Yacine Khaldi, Rafik Bouaouina, Nadia Amrouni, Hammam Alshazly, and Abdeldjalil Ouahabi. A comprehensive survey on ear recognition: Databases, approaches, comparative analysis, and open challenges. *Neurocomputing*, 537:236–270, 2023.

[2] Z. Emersic, T. Ohki, M. Akasaka, T. Arakawa, S. Maeda, M. Okano, Y. Sato, A. George, S. Marcel, I. I. Ganapathi, S. S. Ali, S. Javed, N. Werghi, S. G. Isik, E. Saritas, H. K. Ekenel, V. Hudovernik, J. N. Kolf, F. Boutros, N. Damer, G. Sharma, A. Kamboj, A. Nigam, D. K. Jain, G. Camara-Chavez, P. Peer, and V. Struc. The unconstrained ear recognition challenge 2023: Maximizing performance and minimizing bias. In *IEEE International Conference on Biometrics, Theory, Applications and Systems*, pages 1–10. IEEE, 2023.

[3] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[4] Daniel Ponsa Vassileios Balntas, Edgar Riba and Krystian Mikolajczyk. Learning local feature descriptors with triplets and shallow convolutional neural networks. In Edwin R. Hancock Richard C. Wilson and William A. P. Smith, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 119.1–119.11. BMVA Press, September 2016.

[5] Lucas Beyer, Xiaohua Zhai, Amélie Royer, Larisa Markeeva, Rohan Anil, and Alexander Kolesnikov. Knowledge distillation: A good teacher is patient and consistent. *CoRR*, abs/2106.05237, 2021.

[6] Yiyang Su, Minchul Kim, Feng Liu, Anil Jain, and Xiaoming Liu. Open-set biometrics: Beyond good closed-set models, 2024.