Ear-based Person Recognition using Pix2Pix GAN augmentation

Eman Abdullah M Alomari[12], Su Yang[1], Sanaul Hoque[3], and Farzin Deravi[3]

Abstract: This study presents a robust framework that leverages advanced deep-learning techniques for ear-based human recognition. Faced with the challenge of dataset sizes, our approach is developed based on a generative adversarial network (GAN) method namely Pix2Pix to augment the dataset. It is demonstrated that this approach offers the ability to produce complementary images for ear recognition. To be more specific, Pix2Pix GAN is employed to generate missing sides in ear image pairs (i.e., creating corresponding left ear images for right ear images and vice versa). As such, this augmentation could substantially increase the dataset size, making it more diverse and of significantly greater use for training purposes. The employed dataset consisted of several images of the right ear and only one left ear for each individual. A series of corresponding synthetic left-ear images is generated using Pix2Pix GAN as a tool for augmenting the available data and mitigate the dataset's lack of left ear images. The experiment framework used the EarNet model and conducted comparative evaluations before and after Pix2Pix GAN augmentation using the AMI Ear dataset. By employing the Pix2Pix GAN, the proposed approach can effectively double the size of a dataset and in the process, provide significantly greater utility regarding how that data can be utilised in real-world applications scenarios. The resulting accuracy reaches 98% on the AMI dataset, demonstrating that this technique can improve model performance for ear-based human recognition.

Keywords: Deep Learning, Generative Adversarial Networks (GAN), Ear biometrics, data augmentation.

1 Introduction

With the escalating security needs in the era of information proliferation, researchers have started to recognise the value offered by methods that use automated identity recognition based on alternative biometric characteristics to confirm an individual's identity. Despite extensive research having been dedicated to traditional modalities such as facial traits, these processes continue to experience challenges due to factors such as ageing, people's changing facial expressions, occlusions, people adopting different poses, and distorted images [JRP04]. As an additional trait often captured together with face, the human ear could be an effective and reliable option. The human ear maintains a stable structure and shape unique to each individual since birth [YBR17]. As a biometric

[1] Faculty of Science, Department of Computer Science, Swansea University Bay Campus, Swansea, UK, eaalomari@iau.edu.sa / su.yang@swansea.ac.uk
[2] Department of Information Science, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia.
[3] School of Engineering, University of Kent, Canterbury, UK. S.hoque@kent.ac.uk

trait, the camera-based nonintrusive and contactless data acquisition eliminates the need for the subject's cooperation during recognition. This characteristic position ears a versatile supplement to other modalities, offering identity cues when alternative information is unreliable or unavailable. Previous studies indicate that the shape of the human ear can change before the age of eight years and after the age of 70 years [YBR17], however, the population of interest is typically between these age ranges. The prominent positioning of the ears on the human face means that they can be readily viewed without the subject being conscious of what is happening, and images captured on CCTV provide almost uniform colour distribution [JRP04].

The key aim of this project is to demonstrate that the effectiveness of using Pix2Pix GAN for data augmentation, specifically by generating synthetic ear images to address dataset limitations, and to showcase the resulting enhancement in model performance for ear-based human recognition using the EarNet model. The remainder of this paper is structured as follows: Section 2 presents a review of the existing literature on ear-based person identification. Section 3 provides details of the proposed approach for ear-based human recognition incorporating the Pix2Pix Generative Adversarial Network (GAN) for dataset augmentation. Section 4 presents the evaluation results of the proposed system, along with a comprehensive discussion and comparison with relevant studies in the field. The final section concludes the paper and points out directions for future work.

2 Literature review

The use of ear images for human identification has gained significant attention due to its distinct characteristics compared to alternative biometric modalities. This section reviews the existing literature that used deep learning for ear-based identification, particularly the advances that have been made in terms of ear generation and reconstruction.

Previous research has addressed issues regarding the loss of colour information [KB20, KB21] when training using greyscale and colour images. It was suggested to use conditional deep convolutional generative adversarial networks (GAN) to colourise the test images. The authors used Pix2Pix GAN to train the segmented regions of interest, the model was trained using the Annotated Web Ears (AWE) dataset [Es17]. Another novel method for recognising individuals based on their ears was devised by Omara et al. [Om21] using Mahalanobis distance metric learning in conjunction with deep Convolutional Neural Network (CNN). More specifically, this approach entails using pre-trained ResNet and Visual Geometry Group (VGG) models to extract deep features from pictures of ears. The proposed approach achieved remarkable rank-1 recognition rates, particularly outperforming other methods on AWE [Es17]. and USTB II [Mu04] databases. Aiadi et al. [AKS23] devised a new unsupervised lightweight network for recognising ear prints called Magnitude and Direction Fusion Network (MDFNet). MDFNet makes use of data-driven filters as well as gradient magnitude and direction, offering an effective method for feature extraction, alignment, and pre-processing. This method outperformed many of the alternatives and proved to be highly resilient in terms

of the effects of occlusion. Alshazly et al. [Al19] applied transfer learning methods and deep CNNs to undertake a range of thorough investigations using the WPUT [FT10], AMI [EG] and AMI cropped (AMIC) [Al19] databases, achieving high recognition accuracy under diverse conditions across different datasets. Their study emphasises the effectiveness of transfer learning and ensemble techniques in enhancing ear recognition accuracy, contributing valuable insights to recognition systems.

Whilst it is apparent that substantial advances have been achieved, there remain numerous shortcomings that need to be addressed. For instance, it is still quite challenging to recognise features from both ears, synthetic images are still unrealistic, and many datasets are small. Our proposed novel approach intends to tackle these shortcomings head-on by leveraging Pix2Pix GAN for realistic data augmentation, specifically targeting the generation of synthetic ear images. This augmentation not only will enhance dataset richness but also improve the model's ability to recognize features from both ear sides accurately.

3 Methodology & proposed approach

This section provides details of the proposed deep learning approach for ear-based human recognition. At present, the available data lacks an evenly distributed left and right ear image pairs, which presents a challenge when attempting to train a deep learning system. By synthetically creating realistic images to compensate the missing ears to maintain the balanced left-right pairs, it is possible to significantly boost the effectiveness of the model training. Therefore, in this study we propose to employ Pix2Pix GAN to augment the data by producing images of left ears from corresponding camera angles. By producing synthetic images of people's left ears, Pix2Pix GAN helps to ensure that the model is able to better identify landmarks on either ear. Figure 1 illustrates several crucial steps involved in the experiment design. During the data pre-processing stage, image re-sizing and conventional augmentation were conducted first, the Pix2Pix GAN was then employed to balance the dataset for improved model training.

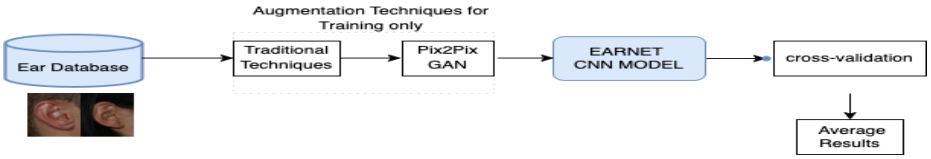

Fig. 1: The proposed approach for ear-based human recognition.

Subsequently, the deep EarNet model is used to extract and train the extracted features, the cross-validation performances on with and without Pix2Pix GAN augmentation are compared to see the effectiveness of generated ears samples for person recognition. It worth to emphasise that the all the test samples are the original ear images in the AMI dataset.

3.1 AMI Dataset Overview

We used the AMI dataset for this study due to its noticeable left-right imbalance. The AMI Ear Database [EG] features images of students, lecturers, and other members of staff at the Computer Science Department of Universidad de Las Palmas de Gran Canaria (ULPGC), Las Palmas, Spain. This indoor dataset includes images from 100 subjects within the age range between 19 and 65. There are six images of each person's right ear and just a single image of their left ear. All images were taken using a Nikon D100 camera in inconsistent lighting conditions, with subjects seated approximately 2 meters from the camera and looking at predefined marks. The profile image for the left-hand side was referred to as BACK, whereas the images for the right-hand side were included in different angles. A selection of images taken from the AMI dataset is presented in Figure 2.

Fig.2: Few examples of images from the AMI Database.

The dataset involves variations in gender, head orientation, and ethnicity, offering a comprehensive foundation for our research. The dataset shows a significant bias, with only one left ear image compared to six right ear images per subject as can be seen in Figure 2. Consequently, this makes it significantly more challenging for a model to learn representative features, could resulting in biased predictions. To address this challenge, Pix2Pix GAN was applied to balance the dataset and mitigate the impact of the imbalance on the subsequent training of using the EarNet model. In addition, augmentation effectively increases the size of the dataset, thereby helping to improve the accuracy of the deep learning model. The imbalance in the AMI dataset offers the potential to help validate the efficacy of Pix2Pix GAN. By producing images of the missing side of the ear, it demonstrates the proposed method in overcoming the limitations of the dataset, effectively increasing the size of the dataset.

3.2 Data Preprocessing

3.2.1 Traditional Data Augmentation

To prepare the dataset for training, several preprocessing steps were employed. The original images with a size of 702×492 pixels were resized to 206×144 pixels to align with the input requirements of EarNet. The smaller size contains sufficient detail to reliably recognize the features of the ears and identify individuals. Traditional augmentation is considered an essential step before training the model, as it makes the model more robust and allows it to see more images in every iteration. Data augmentation techniques were applied to enrich the training dataset and improve model generalization. These techniques included random resized cropping, color jittering,

affine transformations, horizontal flipping, random rotation, and ten-crop extraction. The ten-crop technique involved extracting crops of varying positions from each image, further diversifying the dataset. In each training iteration, the training images undergo random changes based on the data augmentation pipeline, ensuring that the images vary within a specific range in every loop. Augmentation provides additional versions of an image, thereby helping the model identify patterns and enhance its ability to generalize to unseen data.

3.2.2 Pix2Pix GAN for Data Augmentation

Augmentation using Pix2Pix GAN was deemed necessary due to the significant disparity in the number of images of the left and right ears. The training enabled the GAN to generate corresponding left-ear images for each right-ear image. This resulted in six additional images being generated of each individual's left ear to ensure that there were six images of each ear for the sample of 100 people. Consequently, each subject comprised a total of 13 images (six left-right pairs and one back angle left-ear), consisting of seven genuine and six synthetic images.

The Pix2Pix GAN [Is17] has previously demonstrated its versatility when undertaking image-to-image translation. Furthermore, it illustrates the ability of Pix2Pix GAN to colourise images, reconstruct objects based on edge maps, synthesise photographs using label maps, and other tasks. In addition to learning the mapping from input images to output images, this approach is able to learn a loss function that serves to guide the mapping. In Figure 3, the process of training a conditional GAN to map right ear images to left ear images is illustrated. The U-Net architecture is incorporated into the generator [RFB15] and there are encoder and decoder pathways in the generator. With this approach, high-level features are captured by the encoder, whereas the output image is reconstructed by the decoder using the features identified. The U-Net framework was selected for the current study because of its track record of reliably undertaking image-to-image translation which fits well with the stated objective of using right ear images to generate the left ear. This approach was deemed to be appropriate because of the need for a high-resolution input grid to be mapped to an output grid of equally high-resolution.

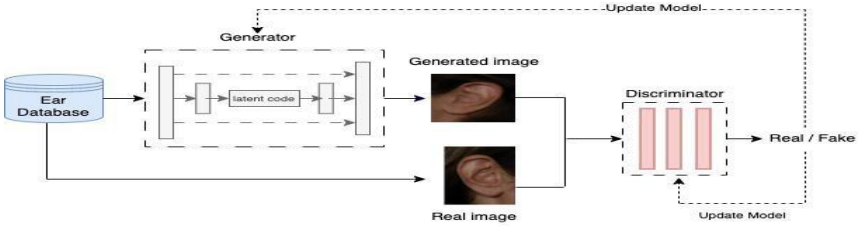

Fig. 3: Training of a conditional GAN to map right ear → left ear. The discriminator learns to classify between fake (synthesised by the generator) and real right ear, left ear tuples. The generator learns to spoof the discriminator. Unlike an unconditional GAN, both the generator and

discriminator observe the input edge map.

3.3 EarNet Model Architecture

The EarNet is a CNN-based model which was developed to discern unique features within ear images, to facilitate the identity of the individuals concerned [APAA21]. Highly detailed patterns are extracted from the RGB images with the help of the convolutional layers (Conv1 to Conv6). Each of these layers has a kernel size of three and they utilise various output channels: 8, 16, 32, 64, 128, 256. In order to speed up the convergence process and improve stability whilst training, every convolutional layer is followed by batch normalisation layers. By applying this approach, when the model is faced with diverse patterns, it is better able to adapt. The strategic positioning of a max-pooling layer helps to retain important features whilst minimising computational complexity. Following the sixth convolutional layer, a dropout layer with a probability of 0.5 is introduced, thereby helping to address the issue of overfitting. By doing so, this helps to ensure that any unseen data can be generalised by the model. Non-linearity is introduced by utilising the Tanh activation function across the network and this process also supports the ability of the model to identify complicated relationships in the dataset. Meanwhile, the forward pass incorporates the fully connected layers, dropout, max pooling, batch normalization, and convolutional layers.

3.4 Model Training and Evaluation

The training phase incorporated a cross-validation process in order to enhance the EarNet model's generalisation and robustness. Once the AMI dataset had been augmented with the images produced using Pix2Pix, it was split to produce seven folds to facilitate training and validation. Seven original images were used: six of them are right ears, and one is a left ear. Doing so enabled the model to learn from a wide range of data subsets, thereby making it better able to adapt to changes in ear images. Training of the EarNet model was performed prior to augmentation and again following augmentation. The purpose of conducting comparisons was to explore how the performance of the model was impacted by the augmentation process. Rank-N accuracy was chosen to provide a straightforward measurement of the model's overall correctness in subject identification.

Cumulative Matching Characteristic (CMC) curves and training accuracy curves are selected for comprehensive comparison and evaluation of the EarNet model's performance. CMC curves illustrate the model's ability to rank subjects correctly across different thresholds, offering insights into its identification accuracy at varying levels.

4 Experiment

4.1 Traditional Augmentation Techniques

The available data can be manipulated using traditional augmentation, resulting in variations which effectively enhance the ability of the model to predict accurately and

generalise based on unseen data. In this work, it is used as a baseline to compare with the more advanced technique, i.e., the Pix2Pix GAN.

The EarNet model is trained with a batch size of 32, with the size of the input images being resized to 176 × 123 pixels. To avoid overfitting, the last CNN layer is assigned a dropout probability of 0.4. The Adam optimizer is used with a learning rate of 0.001.A 7-fold cross-validation approach is applied, enabling the model to be trained using different subsets in each iteration across 150 epochs. Despite traditional augmentation achieving a promising result with an accuracy of 93%, it was apparent that the dataset was imbalanced due to the availability of left-ear images.

To address this issue, it was designed to purposely exclude the left ear images from the testing set and reserve them along with the five images of each subject's right ears for training. Notably, this resulted in accuracy increasing to 97.67%. The model's performance was found to be improved because it was able to capitalise on the greater amount of data relating to the right ear. This finding pointed out the dominating ear during training and the possible approach further improve the accuracy, once this imbalance issue of the dataset is resolved. Using advanced augmentation method, it is possible to further boost the performance.

4.2 Pix2Pix GAN

The Pix2Pix GAN is used to generate pairs of ears for each individual. There are six right ear images with different angles per subject; each of them is used along with the one left ear to generate the corresponding left ear with different angles(see Fig. 3). The model trained for 20000 epochs, with a batch size of 8 and images resized to 256 × 256 pixels. Meanwhile, the U-Net generator was optimised using the Adam optimiser, with the learning rate being 0.0002. Whilst the Pix2Pix GAN was being trained, we experienced a significant challenge related to the potential for mode collapse, wherein the generator output and the discriminator output became identical once there had been a certain number of iterations. This problem occurs when the generator fails to capture the diversity and complexity of the target distribution, thereby causing information contained within the generated samples to be lost. To address this problem, one of the approaches applied was combining original images with cropped versions obtained by removing the background in different levels to augment the training dataset. In addition, various image transformations were applied to further augment the training data, such as random affine transformations, zooming in and out, and random changes in saturation (colour jittering), contrast and brightness to provide the Pix2Pix GAN with a more comprehensive and varied training dataset. This also helped to reduce the possibility of mode collapse because the model was able to focus on the ear's structural landmarks instead of being forced to memorise patterns from the original images. Pix2Pix GAN was used as a transformative augmentation approach, to generate synthetic images of the left ear for each corresponding right ear image, thereby augmenting the dataset. It further generated 600 ear images and addressed the imbalanced dataset. When the right ears were mapped to generate their respective left ears, the different angles played a significant role in ensuring the robustness of the results for every individual.

Encouraging results were achieved for subjects with sufficient training examples, indicating that the images generated were similar to the appearance of the real left ears. Figure 4 illustrates examples of successful ear mapping.

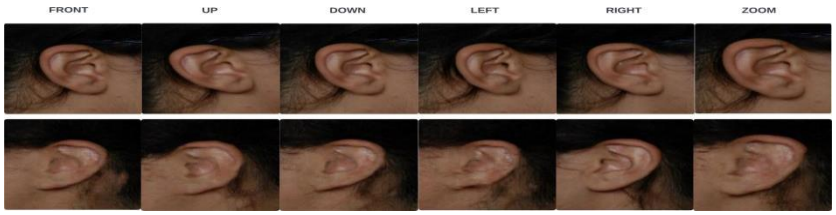

Fig. 4: Pix2Pix GAN transforms the initial right ear images (top row) from a selected subject into their corresponding left ear counterparts (bottom row), supervised by the initial left ear image.

The synthetic images of left ears are quite realistic, captured unique shapes and features in the real photos of the left ears. This helps to visually demonstrate the value of using Pix2Pix GAN as an effective augmentation technique to boost the performance of deep learning models. The EarNet model benefited significantly from employing the Pix2Pix GAN technique. Indeed, test accuracy was recorded at 98%, confirming Pix2Pix GAN's ability to mitigate the problems associated with unbalanced datasets, and consequently enhancing the model's recognition accuracy.

4.3 Comparative Analysis

The effectiveness of augmentation strategies plays a pivotal role in overcoming dataset imbalances and enhancing model performance. The study further compared the efficacy of two augmentation techniques: Traditional augmentation and Pix2Pix GAN Technique using the EarNet model. Traditional Augmentation serves as the baseline, the Pix2Pix GAN exhibits promising improvement. The goal is to unveil the collaborative impact of these augmentation strategies, offering insights into their relative efficacy in the realm of ear-based recognition. When traditional augmentation was employed to train the EarNet model, it achieved a testing accuracy of 93% when both ears were used. It worth to note that, in order to test the resilience and reliability of the generate images, tests were conducted based on EarNet model and an average accuracy of 88% was achieved through cross-validation, when the model was trained using only 600 synthetic images generated by Pix2Pix GAN. This initial test highlighted a notable level of reliability in the generated synthetic images.

Subsequently, the model underwent evaluation using real AMI right ear dataset images. The comparison revealed that while the synthetic images attained 88% accuracy, using the original right ear images achieved a significantly higher accuracy of 97.67%. This encourages a comparison of the model's performance using both synthetic and real images. Therefore, whilst the performance when using exclusively synthetic images was commendable, it fell well short of what could be achieved under ideal conditions, and this demonstrates the need for datasets to be augmented with a diverse array of examples. Consequently, Pix2Pix GAN was used as an augmentation technique to

increase the number of images, which significantly improved the performance of the EarNet model. With the addition of synthetic images generated by Pix2Pix GAN, testing accuracy reached 98%. Testing accuracy is represented in Table 1.

Table 1: Comparison for Testing accuracy for each experiment.

Training set with which side of ears images	**Test set**	**Accuracy**
600 (Right + left) with traditional augmentation	100 Right	97.67%
600 (Right + left) with traditional augmentation	100 (Right + left)	93%
Right (exclude Left ears)	Right only	96.66%
Right (exclude Left ears)	Left only	87.8%
All synthetic left ear (600 generated only)	100 (Right + left)	88%
Include other variations 1200 images (Right + left+ synthetic)	100 (Right + left)	98%

As represented in Figure 5, the comparison of validation accuracies before and after using Pix2Pix GAN as data augmentation clearly demonstrates the model's improved accuracy. Figure 6 presents the CMC (Cumulative Match Characteristic) curve, illustrating the rank- 1 through rank-10 accuracy on the test set before and after using Pix2Pix GAN. These figures provide visual evidence of the effectiveness of Pix2Pix GAN in improving model performance through advanced data augmentation techniques.

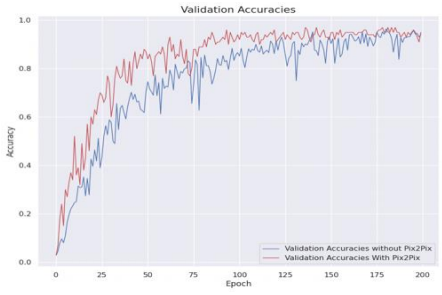

Fig. 5: The EarNet model training and validation accuracies during training before and after using Pix2Pix GAN as data augmentation.

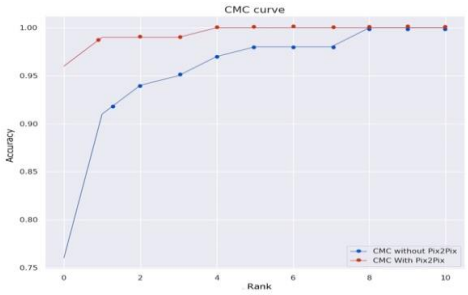

Fig. 6: CMC curve shows rank-1 through rank-10 on the test set before and after using pix2pix GAN as data augmentation.

Based on the findings of this comparative experiment, it is possible to conclude that the synthetic images of ears generated by Pix2Pix GAN are highly reliable. Comparisons of the accuracy results when using synthetic and real images of ears emphasise the importance of close accuracy alignment to evaluate how reliable the images are. In addition, comparing the performance of the EarNet model when tested using real left images of ears after being trained using synthetic images confirms that the synthetic images are reliable when used in real-world scenarios. To set our work in the broader landscape, Table 2 compares the EarNet model with state-of-the-art methods applied in the previous literature based on the rank-1 recognition rate (Accuracy). The comparative analysis with existing literature demonstrates the ability of EarNet to outcompete the

alternative methods. Particularly with the EarNet-Pix2Pix scheme, which outperforms existing methods with a remarkable accuracy of 98%. Using Pix2Pix GAN method to enhance the performance of the EarNet model offers an accurate and robust approach which effectively sets the standard for other ear-based human recognition systems.

Table. 2 Comparison with state-of-the-art methods based on rank-1 recognition rate (Accuracy)

Summary of related work	**Method**	**Accuracy (%)**
R. Raghavendra ea al.2016 [RRB16]	Hybrid Fusion Scheme using HoG and LPQ	86.36
Omara et al.2018 [Om18]	Efficient Metric Learning Method	95.50
Alshazly et al.2019 [Al19]	Ensembles of VGG-13-16-19	97.50
Khaldi and Benzaoui 2020 [KB20]	Colorizing with conditional GANs for Addressing Color Information Loss	96.00
Priyadharshini et al.2021 [APAA21]	Six-layer Deep CNN Architecture	96.99
Aiadi et al.2023 [AKS23]	MDFNet(Unsupervised Lightweight Network)	97.67
EarNet + Pix2Pix GAN (ours)	**EarNet-Pix2Pix**	**98.00**

The proposed work surpasses the accuracy achieved by various state-of-the-art methods, including Aiadi et al.'s[AKS23] MDFNet (97.67%). This indicates that our approach not only outperforms existing methods but also establishes a new state- of-the-art benchmark on the AMI Ear Database.

5 Conclusion

In this study, we present a novel approach for recognising people based on their ears utilising deep learning and Pix2Pix GAN to augment a dataset. Our methodology centres around the EarNet model, a Convolutional Neural Network, complemented by Pix2Pix GAN technique for data augmentation. We employed Pix2Pix GAN to generate synthetic ear images, effectively enlarging the dataset and improving its balance. Our experiment results indicate the significant improvement in testing accuracy achieved by Pix2Pix GAN, thereby helping to overcome the identified limitations with the dataset and generating realistic ear images that could enhance the model's ability to capture more features. The performance of the EarNet model was significantly enhanced by both Pix2Pix GAN. From a baseline accuracy of 93.0% when applying standard augmentation methods, utilising Pix2Pix GAN increased this figure to 98%. Despite this promising performance, it is important to recognise that there remains scope for further advances to be made in the future. For example, Future research should seek to create larger and more extensive datasets in order to make the model more robust and better able to generalize. Furthermore, Improving Pix2Pix GAN's training process would enable better quality images to be produced. Lastly, Future research should consider deploying the model in real-world settings with careful consideration of the ethical implications, the need for real-time processing, and computational efficiency.

6 Acknowledgements

This work was supported by the Saudi Arabian Cultural Bureau in London, Imam Abdulrahman Bin Faisal University in Saudi Arabia, and Swansea University.

References

[AKS23] Aiadi, Oussama; Khaldi, Belal; Saadeddine, Cheraa: MDFNet: An unsupervised lightweight network for ear print recognition. Journal of Ambient Intelligence and Humanized Computing, 14(10):13773–13786, 2023.

[Al19] Alshazly, Hammam; Linse, Christoph; Barth, Erhardt; Martinetz, Thomas: Ensembles of deep learning models and transfer learning for ear recognition. Sensors, 19(19):4139, 2019.

[APAA21] Ahila Priyadharshini, Ramar; Arivazhagan, Selvaraj; Arun, Madakannu: A deep learning approach for person identification using ear biometrics. Applied intelligence, 51:2161–2172, 2021.

[EG] Esther Gonzalez, Luis Alvarez, Luis Mazorra: , AMI Ear Database for Ear based human identification licensed under a Creative Commons Reconocimiento-NoComercial-SinObraDerivada 3.0 Unported License.

[Es17] Emeršić, V., Struc, P., Peer, P.: Ear recognition: More than a survey. Neurocomputing, 255(2017),26–39.doi:10.1016/j.neucom.2016.08.139.URL: https://doi.org/10.1016/j.neucom.2016.08.139.

[FT10] Frejlichowski, D., Tyszkiewicz, N.: The West Pomeranian University of Technology ear database – a tool for testing biometric algorithms. In: 2010 International Conference Image Analysis and Recognition (ICIAR), Povoa de Varzim, Portugal, 2010, p. 227–234. doi:10.1007/978-3-642-13775-4_23. URL: https://doi.org/10.1007/978-3-642-13775-4_23.

[Is17] Isola, Phillip; Zhu, Jun-Yan; Zhou, Tinghui; Efros, Alexei A: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. S. 1125–1134, 2017.

[JRP04] Jain, Anil K; Ross, Arun; Prabhakar, Salil: An introduction to biometric recognition.IEEE Transactions on circuits and systems for video technology, 14(1):4–20, 2004.

[KB20] Khaldi, Yacine; Benzaoui, Amir: Region of interest synthesis using image-to-image translation for ear recognition. In: 2020 International Conference on Advanced Aspects of Software Engineering (ICAASE). IEEE, S. 1–6, 2020.

[KB21] Khaldi, Yacine; Benzaoui, Amir: A new framework for grayscale ear images recognition using generative adversarial networks under unconstrained conditions. Evolving Systems, 12(4):923–934, 2021.

[Mu04] Mu, Z., Yuan, L., Xu, Z., Xi, D., Qi, S.: Shape and structural feature based ear recognition, pp. 663–670. Springer, Berlin, Heidelberg (2004).

[Om18] Omara, Ibrahim; Zhang, Hongzhi; Wang, Faqiang; Hagag, Ahmed; Li, Xiaoming; Zuo, Wangmeng: Metric learning with dynamically generated pairwise constraints for ear

recognition. Information, 9(9):215, 2018.

[Om21] Omara, Ibrahim; Hagag, Ahmed; Ma, Guangzhi; Abd El-Samie, Fathi E; Song, Enmin: A novel approach for ear recognition: learning Mahalanobis distance features from deep CNNs. Machine Vision and Applications, 32:1–14, 2021.

[RFB15] Ronneberger, Olaf; Fischer, Philipp; Brox, Thomas: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, S. 234–241, 2015.

[RRB16] Raghavendra, Ramachandra; Raja, Kiran B; Busch, Christoph: Ear recognition after ear lobe surgery: A preliminary study. In: 2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA). IEEE, S. 1–6, 2016.

[YBR17] Yoga, S.; Balaih, J.; Rangdhol, V.; Vandana, S.; Paulose, S.; Kavya, L.: Assessment of age changes and gender differences based on anthropometric measurements of ear: A cross-sectional study. In: Journal of Advanced Clinical & Research Insights 4 (4), S. 92–95, 2017. DOI: 10.15713/ins.jcri.167.