UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

School of Electronics and Computer Science

On Distinctiveness in Ear Biometrics

by

Di Meng

A thesis submitted for the degree of

Doctor of Philosophy

April 2021

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

School of Electronics and Computer Science

Doctor of Philosophy

On Distinctiveness in Ear Biometrics

by Di Meng

Ear biometrics have developed rapidly in last decade. Ears have distinct advantages over face and fingerprint, such as invariant structure over time, and ear images can be captured without subject's participation. There are also some considerations when ears are used as a biometric, such as rotation variance, varying illumination and occlusion by hair. Wider application has been hindered by these problems. Previous works show that human ears can be used for identification, gender classification, and age classification. In this thesis, we propose a new model-based approach to ear biometrics, which contains geometric features based on ear anatomy. The keypoints of our model are determined by scale-invariant feature transform (SIFT), and we consider the rotation of ear images under an affine transformation, by modelling the ear as a flat plane attached to the head. Then, we extend our model with image pre-processing step that using the force field transform to remove the noise.

We apply the model and fine-tuned convolutional neural networks on ear recognition, gender classification and ear symmetry. In ear symmetry, we address the question as to whether it is possible that given an image of one ear, a person can then be recognized from his/her other ear. Such a symmetry-based strategy could reduce constraints on applications of ear biometrics. To investigate symmetry, we compare one ear with a mirrored version of the other ear.

In addition, we consider the important parts of ear recognition, gender classification and ear bilateral symmetry on ear images, in these three cases we aim to determine the ear parts from which recognition is derived. For analysing the model-based, we use accuracies of different ear regions to evaluate the significant parts for ear recognition, gender classification and ear symmetry. Moreover, we are the first to apply the heatmaps on ear images to determine the contributions of different parts of ear, and this is the first study to analyse the differences between male and female. Also, we have *compared the model-based method with deep learning, and the contributions of different parts based on different approaches.*

Furthermore, we are the first to exploit ear for kinship verification, and we collect SOTEAR dataset for the kinship verification experiments. We compare the influence of father with that of mother by the accuracies of kinship verification.

Contents

List of Figures

List of Tables

Declaration of Authorship

Print name: Di Meng

Title of thesis: On Distinctiveness in Ear Biometrics

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

   1.4 List of Publications

Signature: .... Date:.............................

Acknowledgements

I would like to express my gratitude to my supervisors Dr. Sasan Mahmoodi and Professor Mark Nixon for their support and guidance, and thanks them for introducing me to an interesting research and exciting approach to research. I am very interested about my research.

I would also like to thank my parents and Zhen, who have supported me, and whose love have accompanied me all the time. Thanks all my friends for helping me to my research.

Chapter1

Context and Contributions

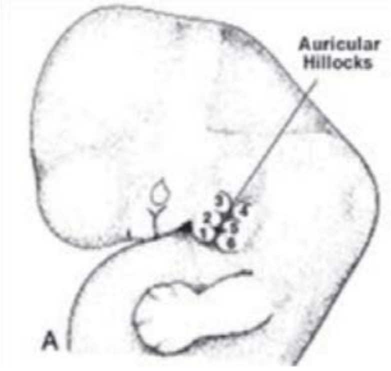
1.1. Context

The ear might appear to be an unusual biometric, but it actually has a long history. During the investigation of a crime, Sherlock Holmes examined some ears that had been found in a box and focused on their anatomical peculiarities. He then said to Watson:

> 'As a medical man, you are aware, Watson, that
> there is no part of the body which varies so much as
> the human ear. Each ear is as a rule quite distinctive
> and differs from all other ones.'
>
> Conan-Doyle, 'The Adventure of the Cardboard Box', 1893

Ears have long been considered unique to their owners and therefore constitute a reliable biometric. Ear recognition has been based on holistic and model-based approaches, with more recent work using deep learning. Many approaches have been targeted at, and evaluated on, standardised datasets where the ear position is controlled (including early datasets like XM2VTS [18] and later and larger ones like SC face [17]). This is a necessary step in any biometric since the prime consideration in biometrics is uniqueness to a subject.

The initial structure of the external ear in a human embryo can be divided into six individual hillocks [13]. The first three hillocks are derived from the lower part of the first pharyngeal arch and form the tragus, crus of the helix, and helix, respectively. The final three hillocks are derived from the upper part of the second pharyngeal arch and form the antihelix, antitragus, and ear lobe [12][14]. Figure 1. 1 illustrates the six auricula in a six-week-old human embryo. Figure 1. 2 presents the mandibular and the hyoid hillocks. The hillocks develop into the final shape of the auricle. The structure of the ear is not random; it has a regular structure.



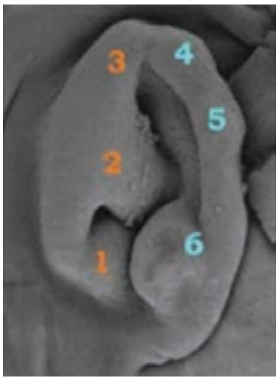


Figure 1. 1 The six auricular hillocks in a six-week-old human embryo [11]

Figure 1. 2 Mandibular hillocks and hyoid hillocks [12]

Ears have many features that can be used to identify people. Their advantages include:

i) a rich structure.

ii) their appearances do not alter a great deal between the ages of 8–70.

iii) they remain the same when facial expressions change, unlike the face.

Also, ear images can be obtained using simple tools without the coordination of subjects and in an unconstrained environment [15]. Although the recognition performance of ear biometrics continues to improve, there are still some problems that need to be overcome, such as occlusion, noise, pose changes, scale variations, illumination diversifications and poor resolution [16].

Figure 1. 3 illustrates the anatomy of the external ear. The labels are represented as follows:

1. (a-d) helix
2. lobe
3. antihelix
4. cavum conchae
5. tragus
6. antitragus
7. helix
8. triangular fossa
9. incisura

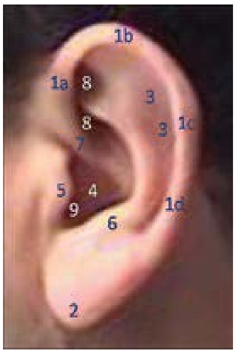


Figure 1. 3 The terminology of the ear [1]

The recent COVID-19 outbreak and the increased wearing of masks might hinder the development of using the face as a biometric method for recognition. Two examples for ear identification are demonstrated in Figure 1. 4. As shown here, a rioter might conceal his/her face but not his/her ear, and a mask obscures much of the face but not the ear. In these images, it might be preferable to use the ear for identification since it is the only biometric that can be seen clearly.

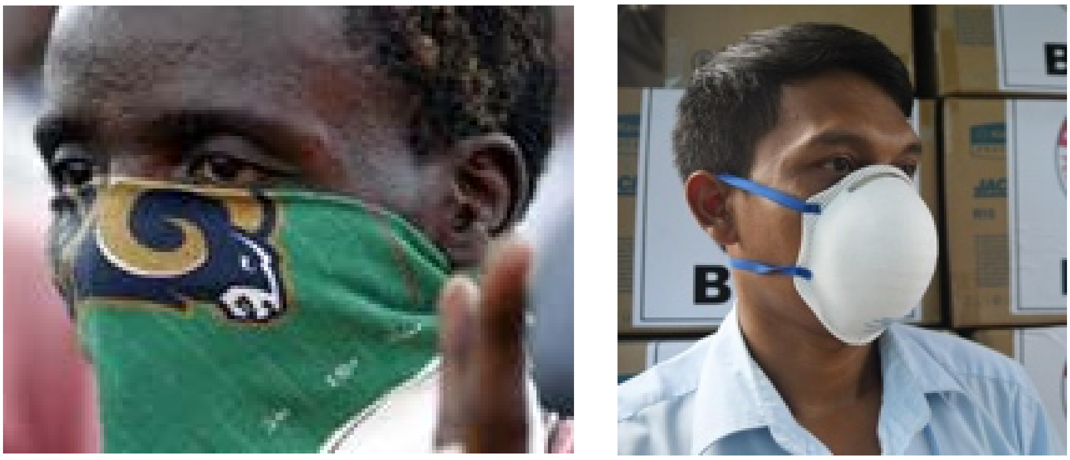


Figure 1. 4 Face concealed

Ear biometrics are becoming increasingly popular. This has developed over the last two decades, and a history of ear biometrics is shown in Figure 1. 5. The first person to suggest that people can be recognised by their ears was a French criminologist, Alphonse Bertillon, in 1890 [1]. Iannarelli later demonstrated that ears are unique to their owner by examining over 10,000 ear samples by developing a manual system for ear identification [2]. Burge et al. were the first to develop an automatic ear biometric approach [3], though recognition was only demonstrated later [4]. In addition, [5] proposed to identify humans from 3D ear images, and [6] used a geometric approach for ear recognition, and deep learning has been applied for ear recognition. [44] used local descriptors for ear recognition, and [125] proposed ear recognition based on deep convolutional networks. In 2017, [67] presented the challenge of unconstrained ear recognition. [65] proposed a hybrid method in 2021, which used deep learning together with Mahalanobis distances. Meanwhile, ear images have been used for extracting soft traits, such as gender [7][8][9][10][86], age[9][10], and kinship[7].

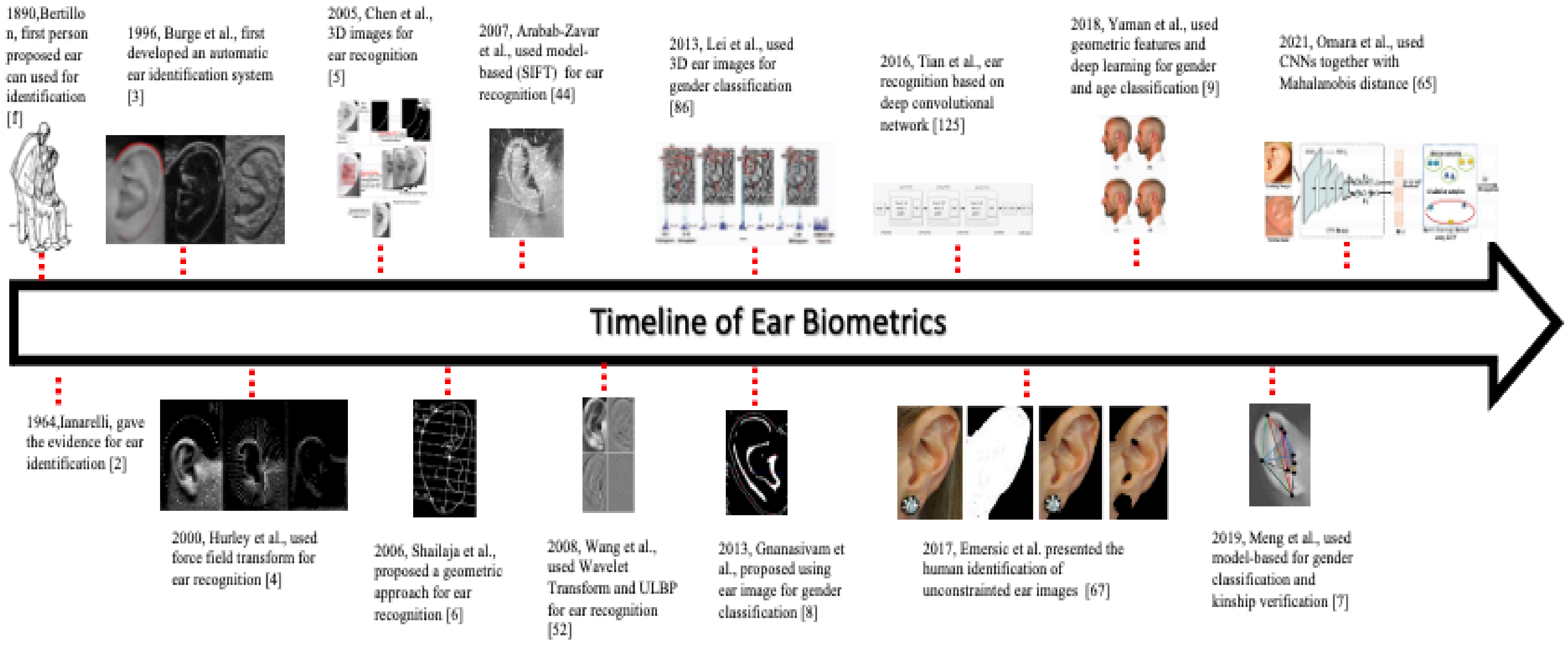

1890,Bertillon, first person proposed ear can used for identification [1]
1996, Burge et al., first developed an automatic ear identification system [3]
2005, Chen et al., 3D images for ear recognition [5]
2007, Arabab-Zavar et al., used model-based (SIFT) for ear recognition [44]
2013, Lei et al., used 3D ear images for gender classification [86]
2016, Tian et al., ear recognition based on deep convolutional network [125]
2018, Yaman et al., used geometric features and deep learning for gender and age classification [9]
2021, Omara et al., used CNNs together with Mahalanobis distance [65]
Timeline of Ear Biometrics
1964,Ianarelli, gave the evidence for ear identification [2]
2000, Hurley et al., used force field transform for ear recognition [4]
2006, Shailaja et al., proposed a geometric approach for ear recognition [6]
2008, Wang et al., used Wavelet Transform and ULBP for ear recognition [52]
2013, Gnanasivam et al., proposed using ear image for gender classification [8]
2017, Emersic et al. presented the human identification of unconstrainted ear images [67]
2019, Meng et al., used model-based for gender classification and kinship verification [7]

Figure 1. 5 The history of ear biometrics

With the advancement of biometrics, more and more people are concerned about privacy. The International Organization for Standardization (ISO) is a standard-setting body composed of representatives from various national standards organisations around the world. [131] 'ISO/IEC JTC 1/SC 27' [132] was created in 1989 and provided standards for the protection of information, communication and technology (ICT). It introduced generic methods, techniques and guidelines for security and privacy, such as methodology for the capture security requirements; management of information and ICT security; cryptographic and other security mechanisms; security management support documentation; security aspects of identity management, biometrics and privacy; conformance assessment, accreditation and auditing requirements in the area of information security management systems and security evaluation criteria and methodology. 'ISO/IEC JTC 1/SC 17' [133] established the standardisation of generic biometric technologies for data interchange among applications and systems in 2002. Generic human biometric standards include common file frameworks, biometric application programming interfaces, biometric data interchange formats, related biometric profiles and the application of evaluation criteria to biometric technologies.

1.2. Contributions

The main contributions in this this, related to publications arising from this research [A]-[F] include:

1) Model-based ear biometrics

There are clear advantages of a model-based approach to ear biometrics, including potential robustness to occlusion and scale, as well as the intimate relationship with ear structure. Our new approach focuses on the ear's structure and appearance rather than on its embryonic nature. We develop a new model-based approach derived from the ear structure that has practical advantages [A][D]. The advantages of our model include robustness to rotation since the subject's head may rotate in yaw and pitch. We assume that the ear can be approximated as a planar structure and use the affine transformation to register the image dataset.

2) Gender classification by ears

There is yet limited work on gender classification from ears, though it is common in many other biometrics. This is perhaps because, like a model, it does not appeal to intuition that gender can be ascertained from the loose and apparently random structure of the inner ear. Figure 1. 6 shows some examples of ear images. Is it easy to determine the subject's gender from these images?

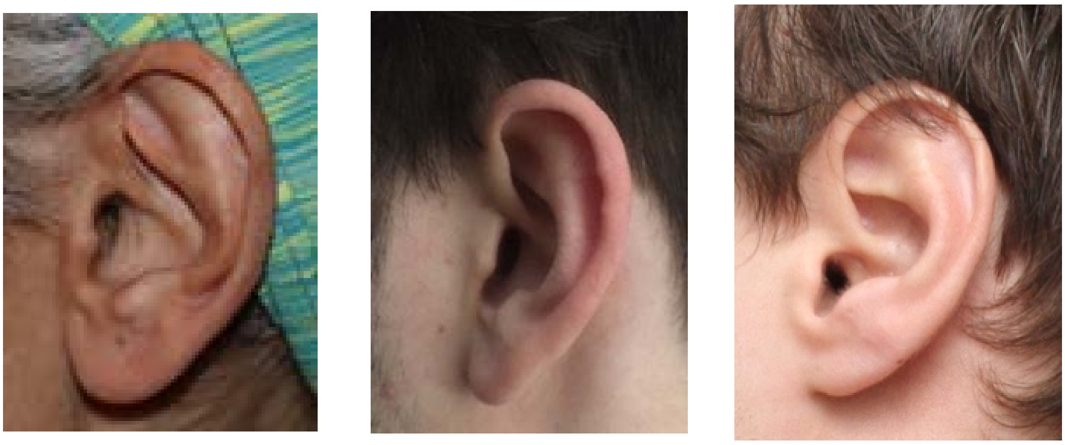


Figure 1. 6 Samples of ear images

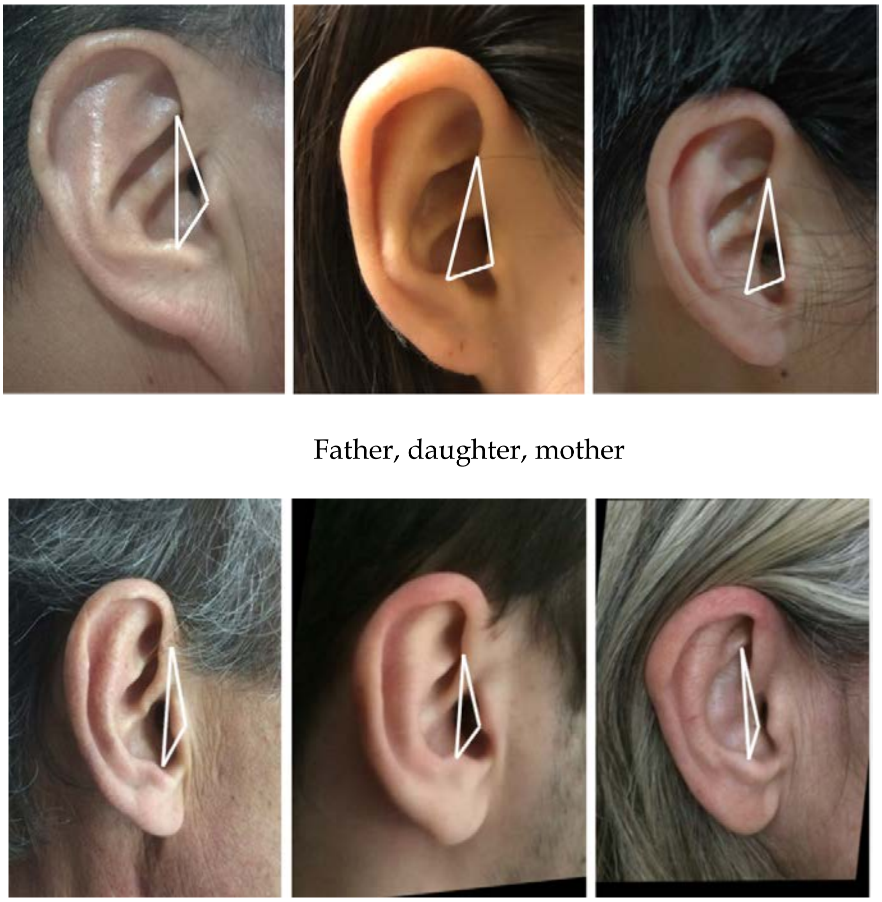
Most of this research capitalises on general properties and the overall appearance of ear images. One study used geometric features (distance between landmark points), but the landmarks were marked manually. However, the points that we use for our model are detected automatically [B][D]. Moreover, we add an image pre-processing step to improve the performance of our model. In addition, deep learning approaches are also used for gender classification from ear images. There are three convolutional neural networks used for gender classification, and the accuracies and Equal Error Rates are used to evaluate the networks.

Despite various approaches to gender classification from ear images, the difference between different genders has not been explicitly understood. What are the differences in ears between females and males? We have analysed the differences between different genders based on different approaches. The performances of different regions represent the contributions of different parts. Although some papers present the gender classification from ear images through deep learning, why can these networks make the decision for the classification? We apply Gradient-weighted Class Activation Mapping (Grad-CAM) on these networks, and the heatmaps of networks show which parts of the ears are significant for gender classification. We can explain the contributions of different parts of ears for gender classification and the disparity of the performances of these networks through the heatmaps of different networks.

3) Kinship analysis by ears

Our research is the first study on kinship verification from ears [A][E]. There is some kinship verification from facial images, but there are the traditional difficulties in face recognition: expression, pose, occlusion and age. By way of contrast, ears do not suffer from change in expression and appear largely invariant with age (except to enlarge), though change in pose is innate.

Figure 1. 7 shows the ears of two parents either side of their daughter's ear, and in the other a son. By overall structure, the daughter's ear appears to resemble that of her mother, and the son's ear resembles his father's. As we will later describe, the triangles are prescribed by the locations of interest points and reflect these similarities. Naturally, these are just one of the geometric features, and there are many more. For the mother and daughter, the triangle formed by the fossa, incisura and antitragus appears to be more similar than that of the father; in contrast the triangle for the son appears closer to that of his father. The difference in gender is less intuitive, though on these images the helix (the outer rim) appears wider for the female subjects.



Father, daughter, mother

Father, son, mother

Figure 1. 7 On Kinship and Gender from Ear Images

We use our model on the kinship verification to find the relationship between two ears within a family. We therefore expect to get significantly higher classification accuracies for kinship once we fuse other features with our existing geometrical features. Moreover, the ear dataset of families (SOTEAR [21]) is collected by us, which contains 21 families now.

4) Ear Symmetry

Because human anatomy dictates that only one ear can be seen in a single image, is it possible to use one side ear to recognize the other side? There is an example shown in Figure 1. 8.

Figure 1. 8 shows different, low-quality surveillance images recorded at a crime scene suggesting a need for the capability to handle noisy images. In all of these images, it might be preferable to use the ear for identification since it is the only biometric that can be seen clearly. To achieve that, we need to reduce constraints on the deployment of surveillance/security systems to be able to handle images where the ear is not presented in line with a camera's view and might also be rotated. Given that often one ear only can be seen, it is also prudent to investigate symmetry. Although it will always be best to store two images per subject, it is conceivable that a match might be required when only one ear can be seen in different situations.

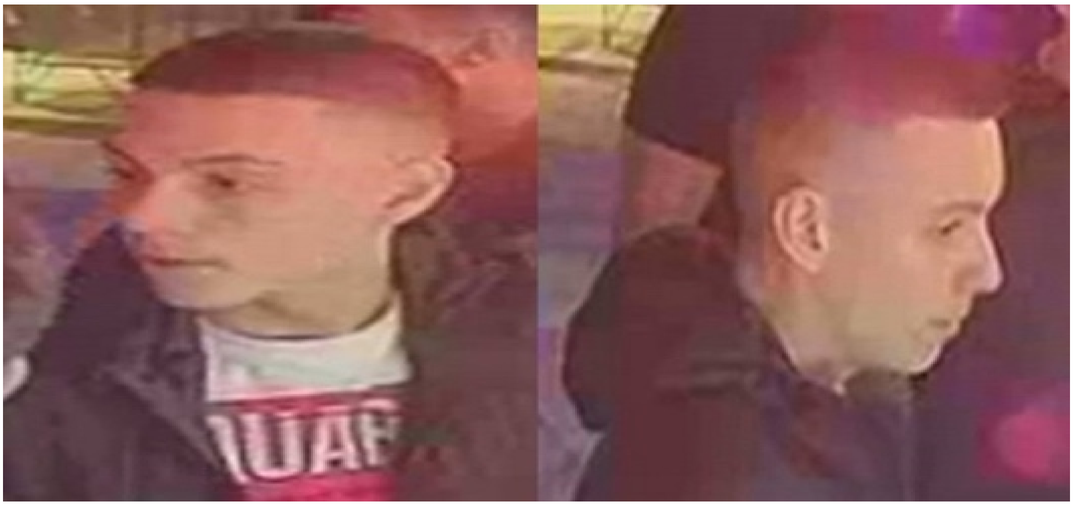


Figure 1. 8 Different views of a subject

There is limited study on ear symmetry, and previous studies do not consider rotation. Ears are also recorded in laboratory conditions under which a subject looks straight ahead. Figure 1. 8 shows how a subject's head might be inclined or rotated, especially when committing a crime. Only three angles are discriminated in [20], where ear symmetry is considered based on the appearance of ears rather than their structure. However, some subjects' heads are rotated along the yaw or pitch axes, which can change the appearance of the ears.

We demonstrate [D] that there is an implicit similarity between left and right ears. Therefore, ear recognition can be achieved regardless of which ear is used for analysis as one can be compared with the other. Also, there is limited research on ear symmetry, but previous results have been less than satisfactory. We are the first to utilise deep learning to study ear symmetry. We also use a model-based technique for comparison with deep learning, and we consider unconstrained ear images. Meanwhile, we measure performance by training noisy ear images and using the affine transformation to register the two ears for comparison. Furthermore, we use heatmaps to show which ear regions are significant for evaluating ear symmetry, and we make comparisons for ear recognition, gender classification and ear symmetry through the heatmaps of different networks.

5) Associated Contributions

We collect the SOTEAR dataset [F], which could be used on a variety of tasks, such as kinship recognition, gender classification and ear symmetry. We are the first to use heatmaps to study ears, and also we are the first to analyse the difference between two genders through ear biometrics. Moreover, we propose the notion of kinship verification from ear images and two-sided ears for symmetry analysis.

1.3. Thesis Overview

- ***Chapter 2: Ears as a Biometric and Identity Science***

In this chapter, we will review the main approaches for ear biometrics, including ear detection, ear recognition and the other information from ear images. The techniques of ear identification include 2D and 3D ear recognition. 2D ear recognition will be introduced with different approaches, which contain holistic approaches, local approaches, geometric approaches, hybrid approaches and deep learning. We will also present the other information from ear images, such as gender classification, ear symmetry and age classification. Furthermore, the use of ear print will also be presented in this chapter.

- ***Chapter 3: Model-based Approach for Ear Biometrics***

This chapter will present our model based on geometric features. The algorithms used for our model-based approach will also be presented in this chapter, which include ear detection, finding the key points and aligning the ear images. We will use Hough

Transform for ear detection, Histogram of Gradients and Scale Invariant Feature Transform for feature extraction, while the affine transformation will be used for aligning the ear images. Also, force field transform will be used for ear image analysis.

- ***Chapter 4: Deep Learning for Ear Biometrics***

There are some networks in recent research. We will choose VGG network, GoogleNet network and Deep Residual Learning for ear biometrics. This chapter will describe the architectures of these three networks. Because of limited ear images, we apply transfer learning for these three networks. Moreover, Gradient-weighted Class Activation Mapping will be used for evaluating visualisations, which represents the contributions of different ear parts.

- ***Chapter 5: Ear Databases***

This chapter introduces some available ear databases. We present the datasets which we have used for ear recognition, gender classification and ear symmetry. We have collected the ear dataset for kinship verification, and it is now available on Github.

- ***Chapter 6: Results and Analysis of the Model-Based Methodology***

The results of the model-based methodology for ear biometrics will be presented in this chapter, to contain accuracies of ear recognition, gender classification, kinship verification and ear recognition by using symmetry. Also, we will present the analysis of our model and the accuracies of different regions for ear biometrics used to analyse the contribution of different parts.

- ***Chapter 7: Results and Analysis of the Deep Learning Approaches for Ear Biometrics***

In this chapter, we will demonstrate the results and analysis of deep learning for ear biometrics. The VGG-16, GoogleNet and ResNet-50 networks will be utilised for ear biometrics, and the noisy ear images and affine transformed ear images will also be used for training to test the networks. The accuracies of different networks will be used for evaluation. Also, we will use the Equal Error Rate to evaluate the performance of each network. The heatmaps of the networks show the contribution of different ear regions.

- ***Chapter 8: Conclusions and Future Works***

Finally, in this chapter, the conclusions are drawn, and potential future work avenues are discussed.

1.4. List of Publications

[A] Meng, D., Nixon, M.S. and Mahmoodi, S., 2019, September. Gender and Kinship by Model-Based Ear Biometrics. In *2019 International Conference of the Biometrics Special Interest Group (BIOSIG)* (pp. 1-5). IEEE.

[B] Meng, D., Mahmoodi, S. and Nixon, M.S., 2020, April. Which Ear Regions Contribute to Identification and to Gender Classification?. In *2020 8th International Workshop on Biometrics and Forensics (IWBF)* (pp. 1-6). IEEE.

[C] Meng, D., Nixon, M.S. and Mahmoodi, S., 2020. Ears as a biometric and identity science. In, Jajodia, Sushil, Samarati, Pierangela and Yung, Moti (eds.) Encyclopedia of Cryptography, Security and Privacy. Third ed. Springer.

[D] Meng, D., Nixon, M.S. and Mahmoodi, S., 2021. On Distinctiveness and Symmetry in Ear Biometrics. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *3*(2), pp.155-165.

[E] Meng, D., Nixon, M.S. and Mahmoodi, S., 2021. Kinship verification from ear images, In preparation for submission.

[F] SOTEAR dataset, Github reference (https://github.com/dm-vlc/dataset.git)

Chapter2

Ears as a Biometric and Identity Science

Developing algorithms for ear biometrics requires databases, which should contain enough quantity (include many subjects and several samples from each subject) and variables, such as different angles, illumination variants, hair occlusion etc. Also, detecting ears automatically is a necessary step for ear biometrics, and many published recognition approaches have achieved this manually. This chapter will give a summary of the state-of-the-art in ear detection techniques. Furthermore, there has been a great deal of research into ear recognition. Techniques for identifying humans by ear images can be classified as 2D ear recognition and 3D ear recognition. In addition, there is some soft information which can be extracted by using ear images, such as gender, age and kinship. We will also present a summary of ear biometrics in this chapter.

2.1. Ear Databases

There are many ear datasets that can be used in experiments involving ear biometrics. The first such analysis was conducted by Iannarelli [2] and contained over 10,000 ear samples obtained from randomly selected samples in California. Burge et al. [107] captured the subjects' head profiles (300×500 pixels, greyscale image) using a CCD camera. The Carreira-Perpinan (CP) [108] ear dataset was one of the earliest publicly available datasets of this type. Figure 2. 1 shows some samples from the CP ear dataset, which contains a total of 102 images from 17 subjects. This chapter introduces the datasets that were used in this study's experiments and summarises the applicability of existing ear data for ear biometrics.

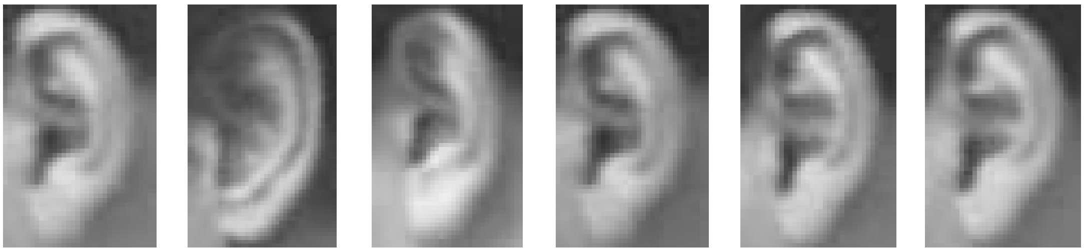


Figure 2. 1 Samples from the CP ear dataset

2.1.1. The University of Science and Technology Beijing Ear Datasets (USTB Ear Datasets)

Three ear datasets were collected by The University of Science and Technology Beijing (USTB) [109].

1. USTB Dataset I

This dataset was compiled using information from 60 subjects; there are three images for each subject, making a total of 180 images. The images are normal frontal images with trivial angle rotation and illumination variants. Each image has 256 grey scales, and all are sheared. This dataset was used for ear recognition purposes. Figure 2. 2 presents some samples from USTB dataset I, in which three consecutive images belong to a single person.

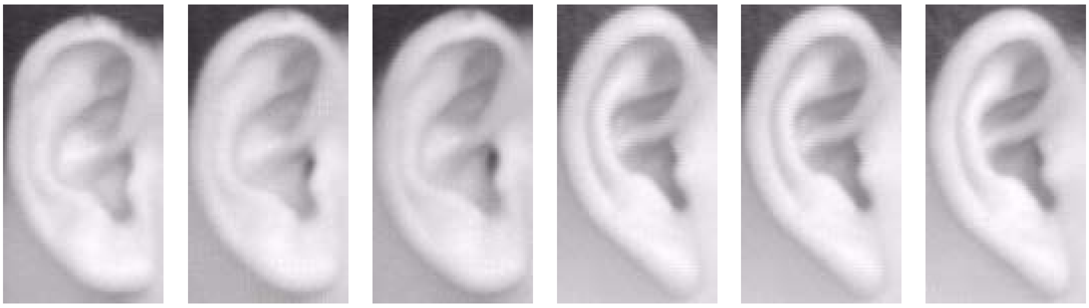


Figure 2. 2 Samples from USTB dataset I

2. USTB Dataset II

USTB dataset II contains 308 images from a total of 77 volunteers (four images for each subject). The images in this database were taken in profile at a resolution of 300×400 pixels and included variations in illumination and rotation. The first image is the frontal ear image under standard illumination conditions, and the second and third images are the respective ear images after a rotation of ± 30°. The fourth image is presented under weak illumination conditions. This dataset supported the research into ear recognition under variations in illumination and rotation. Figure 2. 3 presents some of the dataset's images; here, the images in each row belong to a single person.

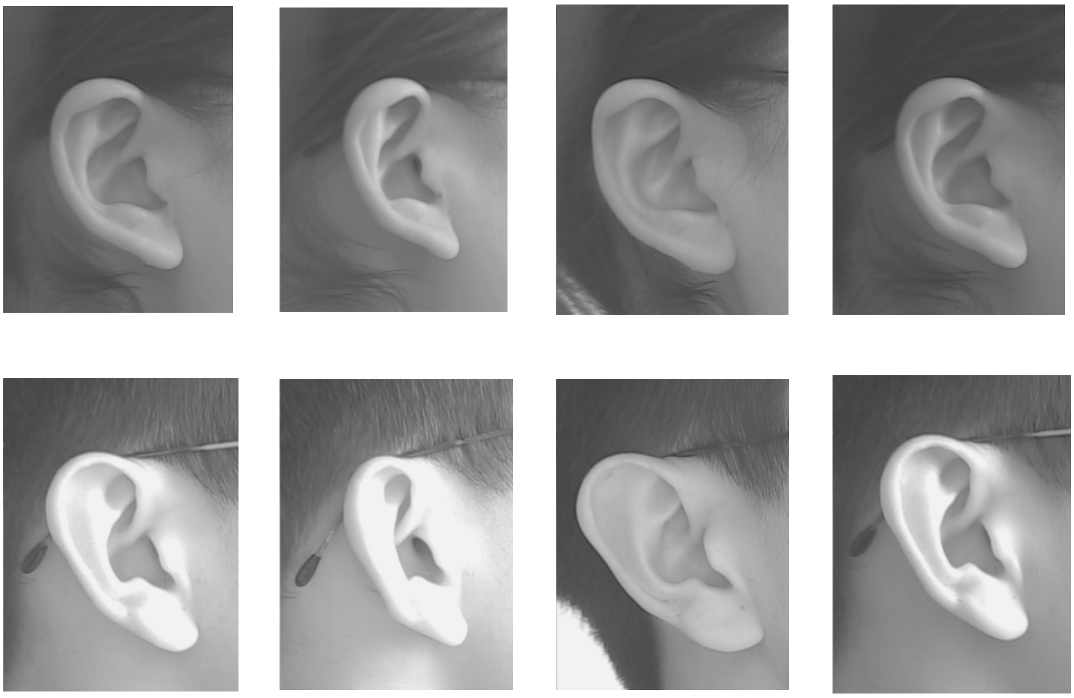


Figure 2. 3 Samples from USTB dataset II

3. USTB Dataset III

USTB dataset III contains ten images from each of its 79 different subjects. The images are right-facing profiles, and no illumination variations are considered. The profiles were captured from a distance of 1.5 metres and show a rotation to the left by 5°, 10°, 15° and 20°. There are two images for every rotation, and the dataset includes some partly occluded images. Figure 2. 4 presents some samples from a single subject in USTB dataset III, which was used for gender classification.

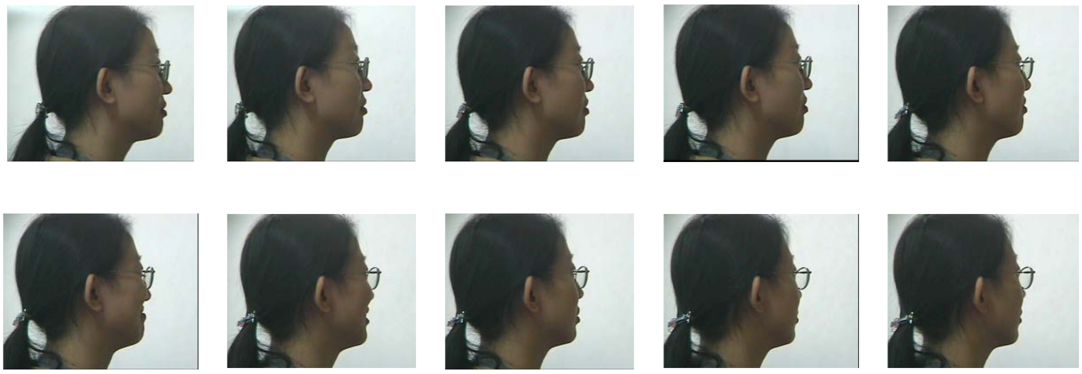


Figure 2. 4 Samples from USTB dataset III

2.1.2. SC Face Dataset

The SC face dataset [17] as collected by The Technical University of Zagreb and contains 4,160 images from 130 subjects. A subset of this database contains images of each subject. Each of the 130 subjects in the SC face database has nine images in different poses. One is a mugshot taken at 0°, while the other images are rotated from − 90° to + 90° (four left-side images and four right-side images). However, as the ears of some of the subjects were entirely occluded by hair, 100 subjects without total occlusion were selected for the experiments in this study. Figure 2. 5 presents some samples from the SC face dataset, which was used in this study's ear symmetry experiments. The images show the subjects and the derived ear images.

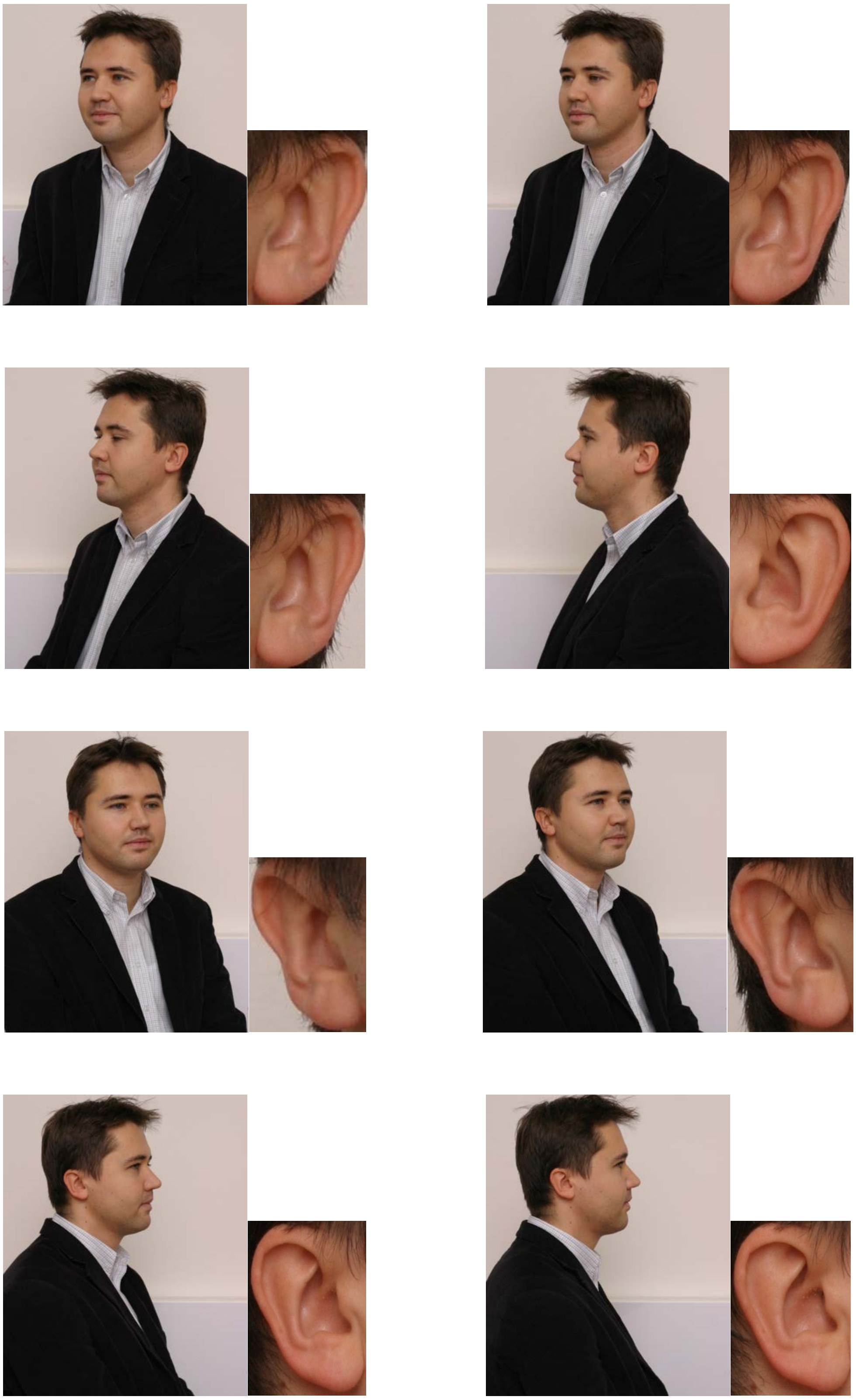


Figure 2. 5 Samples from the SC face dataset

2.1.3. Annotated Web Ears Dataset

The Annotated Web Ears (AWE) [110] database was compiled by the University of Ljubljana. This dataset contains 1,000 images of the ears of 100 different subjects and includes 520 images of left ears and 480 images of right ears. All were obtained from the web by compiling a list of people for whom it was reasonable to assume that a large number of online images could be found. Similar to other datasets collected in the wild, this list mostly featured actors, musicians, politicians, etc. Figure 2. 6 presents some samples from the AWE dataset, which was used for the gender classification and ear symmetry experiments in this study. All the images belong to the same subject.

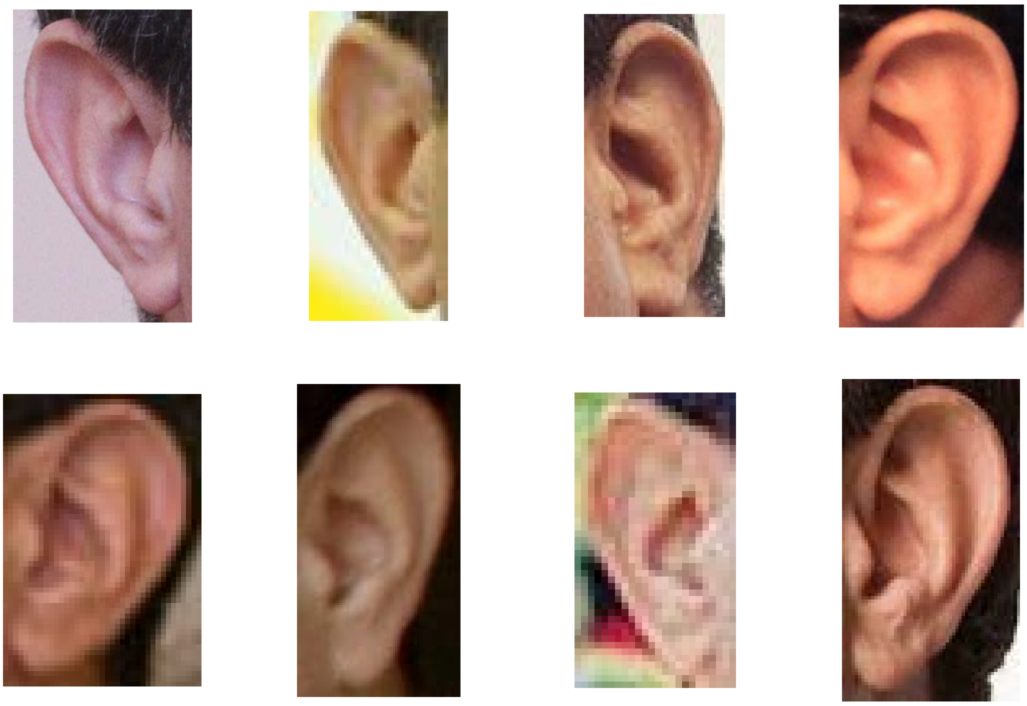


Figure 2. 6 Samples from the AWE dataset

2.1.4. SOTEAR Dataset

The SOTEAR dataset [21] was collected by us, for kinship recognition and gender classification, and this is the first dataset that contains kinship information. The family ear dataset was collected for this study and contains 134 images and was collected for the kinship verification and gender classification experiments in this study. It includes 21 families (21 parents) and 25 children (16 sons and nine daughters). Each subject has two ear images taken from the left and right sides, and all the images were captured from any angle and under any illumination conditions. The database includes two children under the age of eight and two between the ages of eight and 12. This dataset is available on GitHub [21]; however, because there are not many subjects yet, additional images of family ears will be added in the future for wider use. Figure 2. 7 shows some samples from two families (each line contains a single family).

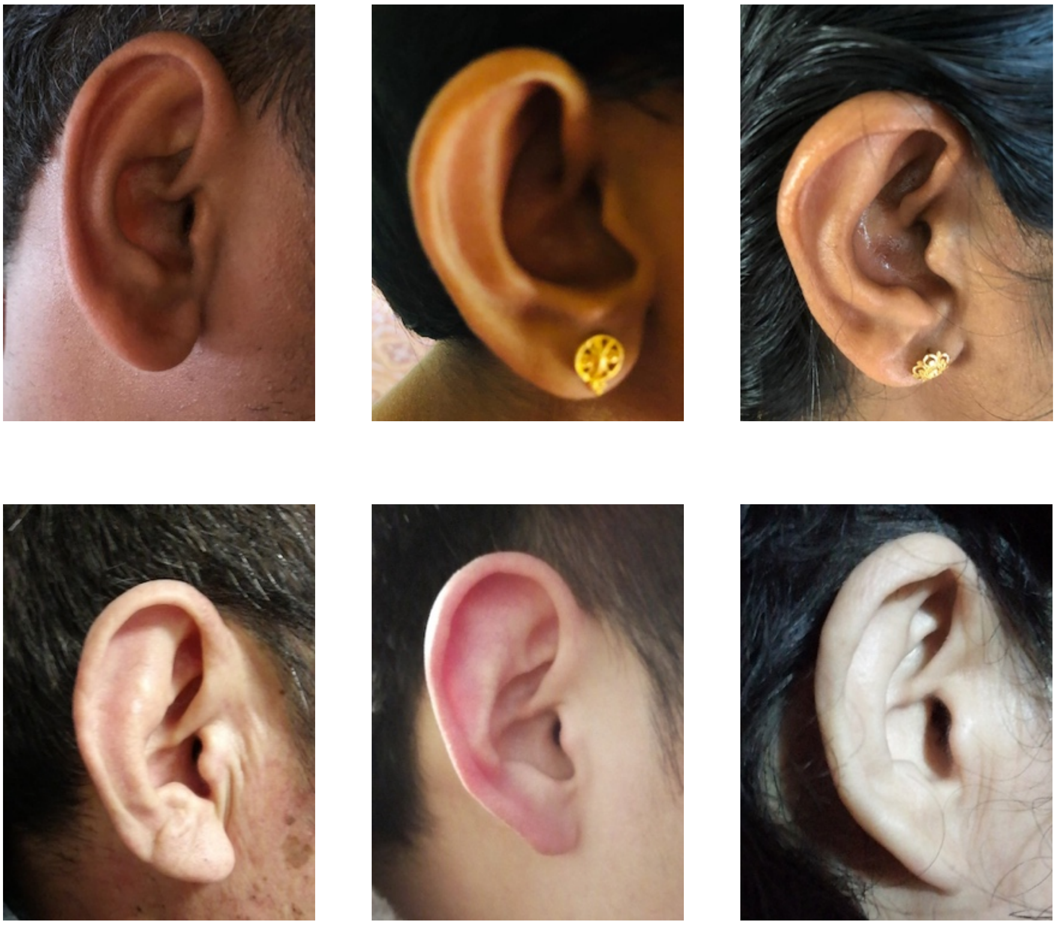


Figure 2. 7 Samples from the SOTEAR dataset

2.1.5. Other Ear Datasets

1. The Institute of Technology Delhi Dataset

The Institute of Technology (IIT) Delhi dataset [111] contains two sub-datasets. The first includes 493 images from 125 subjects, each with at least three images. All the images were taken from a distance using a simple imaging setup, and the imaging was performed in an indoor environment. All the subjects were aged between 14 and 58 years. The second dataset contains 793 images from 221 volunteers, and its images were automatically cropped and normalised.

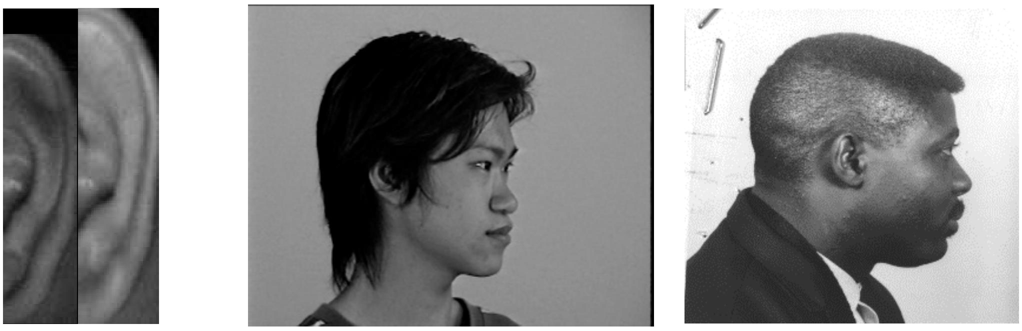
2. The National Cheng Kung University Dataset

The dataset from the National Cheng Kung University (NCKU) in Taiwan [112] contains 90 subjects, each with 74 images. Each subject was captured at different rotation angles between − 90° (left profile) and 90° (right profile); profiles were captured at incremental rotations of 5°, and two images were obtained for each angle. All images were taken under the same lighting and distance conditions.

3. National Institute of Standards and Technology Mugshot Identification Dataset

The National Institute of Standards and Technology (NIST) [113] Mugshot Identification (MID) special database contains both front and side (profile) views. It includes 131 cases with front views and profiles, and each case has two or more subjects. The information for each image includes gender and age.

Additionally, other ear or head profile datasets for ear biometrics are available. The FEARID [114] database was compiled by the FEARID project and includes 7,364 images from 1,229 subjects. It differs from other datasets in that it contains ear prints rather than ear images. The XM2VTS dataset [18] comprises frontal and profile facial images from 295 subjects; however, this dataset is not freely available. The University of Notre Dame (UND) database [115] contains both 2D and 3D datasets, which can be used to compare 2D and 3D techniques. The UBEAR dataset [116] includes 4,429 profile images captured from both the left and right sides of 126 subjects and provides a good indication of intra-subject variation. EarVN1.0 dataset [117] is new a large-scale ear images dataset which contains over 28,412 ear images of 164 subjects. The images have variation in pose, illumination, occlusion, resolution, light condition. Figure 2. 8 shows some samples from these additional datasets.



IIT Delhi NCKU MID

Figure 2. 8 Samples from other datasets

We have described all the datasets that were used in this study's ear biometrics experiments and presented a summary of the ear datasets available for use in ear biometrics in chapter 2.1. There are many datasets available for evaluating the different approaches to ear biometrics. The family ear dataset was collected for the kinship verification of ear images, and all images were unconstrained. However, because this dataset includes only a limited number of families, additional images of family ears will be collected in the future.

2.2. Ear Detection

Ear detection is a necessary step for automated ear biometrics. Many published recognition approaches have achieved this manually. This section summaries state-of-the-art techniques in automatic ear detection. Generally, all the ear detection approaches rely on mutual properties of the ear's morphology [22]. Table 2. 1 shows the results of different ear detection approaches.

Table 2. 1 Summary of ear detection approaches

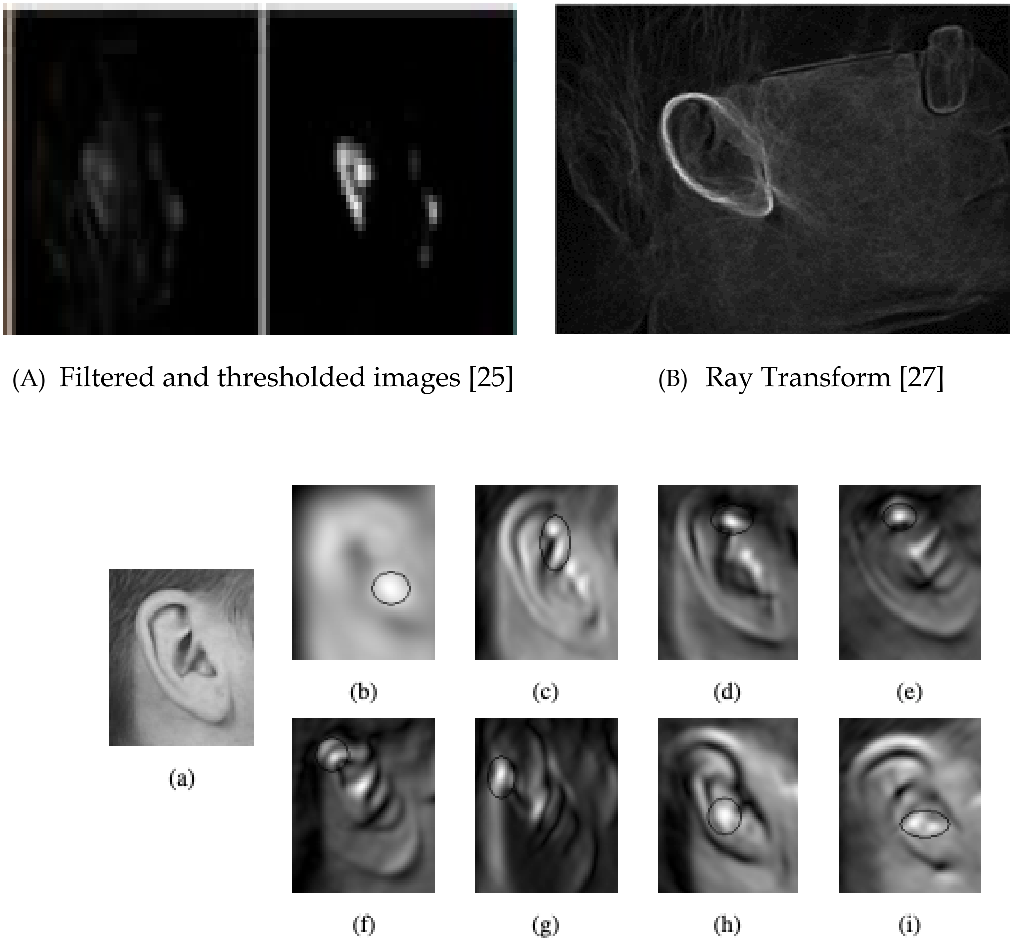
Database	Publication	Technique	Accuracy
UND	Islam et al.[23]	AdaBoost	100%
	Arbab-Zavar et al. [26]	Hough transform	91%
	Resmi et al. [36]	Banana wavelets and Circular Hough transform	85.56%
UND-E	Sarangi et al. [29]	Modified Hausdorff Distance (MHD)	94.54%
	Kamboj et al. [121]	CED-Net-2	95.47%
UND-J2	Zhang et al. [31]	Multiple Scale Faster Region-based Convolutional Neural Network	100%
	Mursalin et al. [37]	EpNet	93.09%±2.67%
	Zhu et al. [73]	PointNet++	93.0%
	Kamboj et al. [121]	CED-Net-2	98.0%
XM2VTS	Arbab-Zavar et al. [26]	Hough transform	100%
	Cummings et al.[27]	Image ray transform	99.6%
	Ibrahim et al.[28]	Banana wavelets	100%

UBEAR	Zhang et al. [31]	Multiple Scale Faster Region-based Convolutional Neural Network	99.22%
	Kamboj et al. [121]	CED-Net-2	99.84%
USTB	Yuan et al.[24]	Improved AdaBoost	99.54%
	Yuan et al. [32]	YOLO	98.67%
	Kamboj et al. [121]	CED-Net-2	99%
AWE	Emeršič et al. [33]	PED-CED	99.4%±0.6%
	Ganapathi et al. [34]	Ensemble model (3 models)	99.52%
	Bizjak et al. [35]	Mask R-CNN	99.7%
	Zhang et al. [31]	Multiple Scale Faster Region-based Convolutional Neural Network	98%
CVL	Sarangi et al. [29]	Modified Hausdorff Distance (MHD)	91%
	Cintas et al. [122]	CNN	94.1%
IITK	Prakash et al. [123]	Distance transform template	95.2%
	Kamboj et al. [121]	CED-Net-2	99%
IITD	Kumar et al.[124]	Morphological operators	NA
	Kamboj et al. [121]	CED-Net-1	72%
WVU	Said et al. [25]	Morphological operators	90%
Various Databases	Naggar et al. [30]	Fast R-CNN	98%

Islam et al. [23] used an automatic and fast technique based on the AdaBoost algorithm which is used to detect a subject's ear from his/her 2D image. Yuan et al. [24] improved AdaBoost algorithm by two stage: off-line cascaded classifier training and on-line ear detection. Otherwise, Said et al. [25] used a low computational cost appearance-based feature for segmentation and they used a Bayesian classifier to detect the output of segmentation. Kumar et al. [124] also used morphological operators for ear detection. The ear shapes were extracted by a series of grey scale morphological operations.

Some model-based approaches have used the Hough transform to detect ears. It is a basic approach for finding the geometric shapes from images. Arbab-Zavar et al. [26] presented an ear enrolment algorithm based on finding the elliptical shape of the ear using a Hough transform (HT) measuring tolerance to noise and occlusion. They used the Canny operator to detect the edges, and this was used in an accumulator for centre of the ellipse. Cummings et al. [27] used the image ray transform, based on an analogy to light rays, to detect ears in an image. This transformation can highlight tubular structures such as the helix of the ear but also spectacle frames. Figure 2. 9 shows the samples of ear detection.

Prakash et al. [123] proposed a novel technique for ear detection, making use of the connected components in graphs obtained from the edge maps of the profile images, and this technique is rotation, scale and shape invariant.



(A) Filtered and thresholded images [25]

(B) Ray Transform [27]

(C) Ear images with banana filters (a) Input image, and (b)-(i) after convolution with 8 banana filters [28]

Figure 2. 9 Examples of ear detection

Ibrahim et al. [28] employed a bank of curved and stretched Gabor wavelets, known as banana wavelets. Resmi et al. [36] used banana wavelets and circular Hough transform for ear detection. They also calculated the efficiency of the proposed method using automatic classification techniques. They used LBP and bank of Gabor filters to extract features from segmented ear image, and they used different classifiers to determine whether the segmented portion of the image is class Ear or Non Ear.

Sarangi et al. [29] used modifying the Hausdorff distance for ear detection. Their method first segments a skin region from the non-skin region. Ear template was created to detect various ear shapes. The Hausdorff distance was used for measuring the similarity likeness between the template and input image.

In recent years, more and more increasing numbers of studies have used deep learning to detect the ear. Naggar et al. [30] presented an ear detection system using Faster R-CNN. There were two stages in their training. In the first stage, classification is achieved by using an AlexNet model to classify ear and non-ear segment. In the second stage, the unified region proposal network with Alexnet was trained for ear detection. Zhang et al. [33] also described an approach involving Multiple Scale Faster Region-based Convolutional Neural Networks (Faster R-CNN) to detect ears from profile images automatically. The study in [31] made a breakthrough with the challenge, including pose variation, occlusion and scale. First, they detected three regions of different scales to infer the information about the ear's location in the image. Next, they proposed an approach based on ear region filtering to extract the correct ear region and remove the false positives automatically.

Yuan et al. [32] presented a real time detection for ears based on embedded systems. The framework used for ear detection is YOLO (You Only Look Once), which is an object detection system used for real time processing. They used YOLO-v2-tiny network which is an enhancement version of YOLO in terms of speed. Their approach achieved a good detection performance in case of complex conditions such as poor illumination and partial occlusion. Emeršič et al. [33] presented a pixel-wise ear detection technique based on Convolutional Encoder-Decoder networks (PED-CED). In this paper, they described the ear detection as a two-class segmentation and trained a convolutional encoder-decoder network based on the SegNet architecture to distinguish the image-pixels and determine if they are ears or not. The processing step remained the two largest regions or one (if only one was detected) and discards the rest. They compared the location of the ground truth that is marked manually with output of the PED-CED to evaluate the performance. Ganapathi et al. [34] proposed a method using an ensemble of convolution neural network (CNN). The first part of this approach was training the datasets on three models of CNN. The second part was using the weighted average of the outputs of trained models to detect the ear region. Bizjak et al. [35] proposed detecting the ear by using Mask R-CNN. This was the extension of Faster R-CNN, which used bounding box detection. For the Mask R-

CNN, they added a branch for predicting the objects and used the existing branch for bounding box detection. The advantage of Mask R-CNN is locating the pixels of each object exactly to instead of bounding boxes. Mursalin et al.[37] proposed a deep neural network for 3D ear detection which named EpNet, and this network can detect directly ear from 3D point clouds. An automatic pipeline was proposed to annotate ears in the profile face images of UND J2 public data set in this paper. Zhu et al. [73] used PointNet++ for ear segmentation. Cintas et al. [122] used convolutional neural networks together with geometric morphometrics to detect the ears with partial occlusions. Kamboj et al. [121] proposed two ear detection models: CED-Net-1 and CED-Net-2.

2.3. Ear Recognition

Techniques for identifying humans by ear images can be categorized as 2D ear recognition and 3D ear recognition. Table 2. 2 presents a summary of these techniques.

Table 2. 2 Summary of ear recognition approaches

Database	Publication	Technique	Accuracy
UND	Kumar et al. [46]	LRT	92.9%
	Prakash et al. [48]	UND	2.25% EER
	Chen et al. [55]	ICP	96.4%
	Zhou et al. [58]	Holistic and local feature	98.6%
		Keypoint description	
	Ganapat et al. [56]	ICP	98.6%
	Yan et al. [57]	PCA	98.8%
	Yan et al. [39]	ICP	63.6%
		PointNet++,	84.1%
	Zhu et al. [73]	LJS descriptors	98.82%
	Kamboj et al. [127]	CG-ERNet	100%
XM2VTS	Hurley et al. [4]	Force Field Transform	99.2%
	Arbab-Zavar et al. [44]	SIFT (non-occlusion)	91.5%
		SIFT (20% occlusion)	80.4%
		PCA (non-occlusion)	98.4%
		PCA (20% occlusion)	12.7%
IITD	Zarachof et al. [41]	PCA	51.3%
	Murukesh et al [47]	PCA, FLDA	92%
	Omara et al. [49]	Geometric feature	99.6%

	Anwar et al. [50]	Geometric feature	98%
	Hassaballah et al. [126]	ALBP	94.72±3.06%
	Kamboj et al. [127]	UALBP	94.73±2.77%
	Priyadharshini et al. [130]	CG-ERNet	96.8%
		CNN	97.36%
	Jangilla et al.[42]	BPPCA	94.73%
	Guo et al. [45]	LSPB, LBP	93.3%
USTB	Tian et al. [61]	Gabor-SIFT	93.1%
	Omara et al. [49]	Geometric feature	98.33%
	Wang et al. [52]	ULBP, wavelet transform	100%
	Sajadi et al. [63]	Generic Algorithm	100%
USTB	Sinha et al. [59]	CNN, HOG	97.9%
	Omara et al. [65]	VGG, ResNet	94.31±2.82%
	Kamboj et al. [127]	CG-ERNet	96.66%
Google DB	Oravec et al. [40]	PCA	100% 55%(rotated)
	Khaldi et al. [64]	DCGAN, CNN	50.53%
	Omara et al. [65]	VGG, ResNet	78.13±3.52%
AWE	Emersic et al. [54]	AlexNet, VGG-16, SqueezeNet	96.09%
	Hassaballah et al. [126]	ALBP	42.20 ±2.29%
		UALBP	38.40 ±3.09%
	Hassaballah et al. [126]	ALBP	71.43±1.86%
AWI	Omara et al. [65]	UALBP	69.86±2.45%
	Khaldi et al. [64]	VGG, ResNet	98.37±0.72%
	Alkababji et al. [129]	DCGAN, CNN	96.00%
	Priyadharshini et al. [130]	CNN, PCA	97.73%
		CNN	96.99%
	Oravec et al. [40]	PCA	40%
WPUT	Hassaballah et al. [126]	ALBP	38.71±4.79%
	Omara et al. [65]	UALBP	35.60±3.76%
		VGG, ResNet	72.67±1.19%
FERRET	Chang et al. [62]	PCA	71.6%
ERJU	Banerjee et al. [43]	Force Field Transform	97.71%
UCR	Chen et al. [55]	ICP	96.8%

EarVN1.0	Alshazly et al. [69]	CNNs	93.45%
-	Victor et al. [38]	PCA	40%
	Jangilla et al.[42]	BPPCA	90.32%
	Ying et al. [51]	Weighted wavelets transform and DCT	98.1%
	Ying et al. [66]	CNN	98.56%
	Eyiokur et al. [53]	VGG-16	99.7%
	Almisreb et al. [60]	AlexNet	100%
	Hassin et al. [128]	SIFT	92%

2.2.1 2D Ear Recognition

2D ear recognition approaches can be divided into five different subclasses depending on the technique for feature extraction:

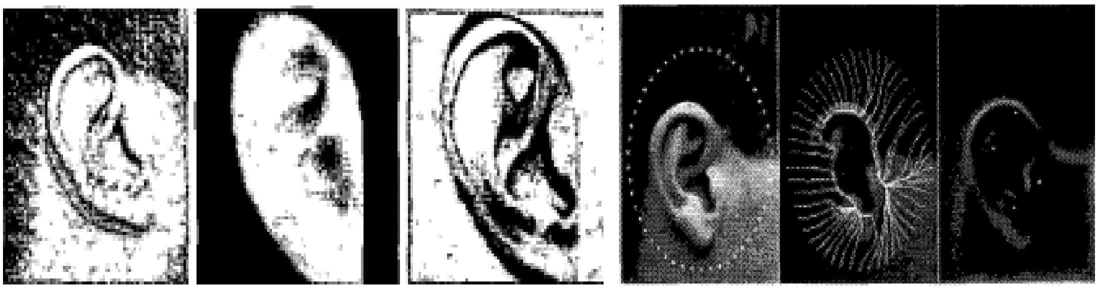
1. holistic approaches;
2. local approaches;
3. geometric approaches;
4. hybrid approaches;
5. deep learning.

Holistic approaches consider ears as a whole, which means they consider the appearance of the ear. Local approaches concentrate on the local parts of the ears. Geometric approaches utilise the geometrical features of the ear structure, such as the distance between tragus and antitragus and the shape of the ear (almost elliptical). Hybrid approaches gather the different techniques and the aim to improve the performance, even though this involves more complex computations. Deep learning is currently very popular for ear recognition, and it essentially depends on the appearance of the ear. However, contrasting with holistic approaches, which are also based on the appearance of ears, deep learning achieves good results for dealing with occlusion. Although deep learning has excellent performance for ear recognition, it needs many images for training to prevent overfitting.

1. Holistic Approaches

PCA is one of the most popular holistic techniques, and it was proved to be a very effective way to reduce storage space. Victor et al. [38] is the first study of ear recognition with PCA. Yan et al. [39] considered PCA for 2D ear recognition and ICP for 3D ear recognition. Chang et al. [62] used PCA to compare the face recognition with ear recognition. Oravec et al. [40] applied PCA to actual scenes to prove its practical feasibility. PCA was then implemented to extract 2D features to identify the person. Zarachof et al. [41] evaluated the performance of PCA based ear recognition in conjunction with super-resolution algorithms. First, the image was down sampled. Second, the resulting image was expanded to original size by various super-resolution methods. PCA was used to extract features.

Some algorithms based on PCA were developed to improve performance. Jangilla [42] used bilinear probabilistic PCA (BPPCA) for feature extraction, and Murukesh et al. [47] presented Contourlet transform and Appearance shape model (ASM) for feature extraction.



Masked mean image and normalized image for ear [38]

Force field feature extraction for an ear [4]

Figure 2. 10 Examples for feature extraction (holistic approach) for 2D ear images

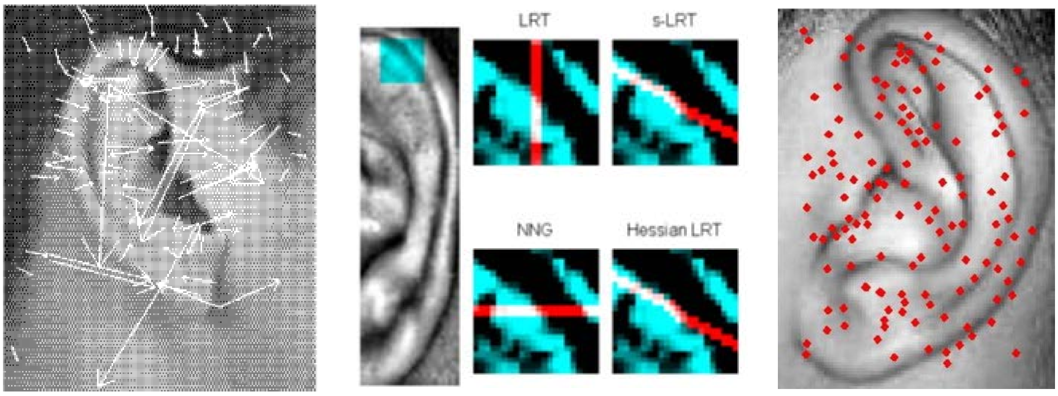
In addition, there are also some other global features, such as, force field transform. Hurley et al. [4] proposed a novel force field transform in which the image is treated as an array of Gaussian attractors that act as the source of a force field. Banerjee et al. [43] presented an approach based on the force field transform. They used force field transform to enhance ear structure from a redundant background. Figure 2. 10 shows some examples of holistic approaches for ear recognition systems.

2. Local Approaches

Local feature is a local representation of an image that reflects local image characteristics [16]. Scale-Invariant Feature Transformation [76] (SIFT) and Speeded-up Robust Feature

[77] (SURF) are very popular for local feature extraction. Arbab-Zavar et al. [44] proposed a new model-based approach, which uses the structure of the outer ear, with the advantages of being robust in noise and occlusion. The model is a constellation of generalized ear parts, and the authors used SIFT for feature extraction. The feature vectors are the sets of keypoints detected by SIFT. The model acts as a mask for keypoints selection, only using the keypoints in the model for ear recognition. Arbab-Zavar et al. [78] proposed a part-wise description of the ear derived by a stochastic clustering on a set of scale invariant features of a training set. The feature extraction used SIFT, and the model was constructed using a stochastic clustering method on the SIFT keypoints of the training set. Tian et al. [61] also used SIFT for extracting the key points of ear images, and then the features of key points are extracted with the local multi-scale analysis feature of the Gabor wavelet. Prakash et al. [48] used SURF to find some unique feature points from ear images. Hassin et al. [128] used a likelihood detector to divide the ear images into skin and non-skin pixels, and then used SIFT for the extraction of ear features. Euclidian Distance Measure used for evaluating the recognition system.

Local Binary Patterns [79] (LBP) and the extension of LBP was widely used for ear recognition. Guo et al. [45] proposed a novel approach for ear recognition that combined local similar binary pattern (LSPB) and local binary pattern (LBP). Kumar et al. [46] presented local random transform (LRT) as transform coefficient features. Hassaballah et al. [126] proposed a new variant of the traditional LBP operator for ear recognition, referred as to local binary patterns (ALBP), and they applied this method to constrained and unconstrained ear datasets.



SIFT keypoints for an ear image [44]

Normalized ear image sample [46]

SURF keypoints for an ear image [48]

Figure 2. 11 Examples of feature extraction (local approaches) from ear images

Yuan et al. [80] described a 2D ear recognition approach based on local information fusion under partial occlusion. The whole image is divided into sub-windows. For each sub-window, they used Neighbourhood Preserving Embedding to extracting features. Figure 2. 11 gives some samples of feature extraction from ear images by local approaches.

3. Geometric Approaches

The geometric feature is another approach for ear detection. Omara et al. [49] proposed a method for using geometric features for ear recognition using the Canny operator to detect the edge, and the geometric feature extraction based on maximum and minimum Ear Height Line (EHL). The Euclidian distance was used for matching.

Anwar et al. [50] presented a new algorithm for ear recognition based on geometrical feature extraction. The feature vector consisted of seven values, which are the means of some ear image measurements, such as x coordinate of centroid, y coordinate of centroid and four different distances from a matrix, which contain Euclidian distance between every pixel in the image. Figure 2. 12 shows an example for using geometric approaches to extract features from ear images.

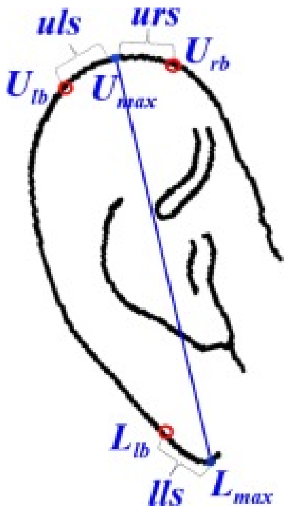


Figure 2. 12 An example of feature extraction (geometric approaches) from ear images [49]

4. Hybrid Approaches

These approaches combined the components from the other categories to increase the performance of the ear recognition.

The approach of Ying et al. [51] combined wavelet transform and discrete cosine transform (DCT). Another hybrid method based on PCA and wavelet transform, which was proposed by Nosrati et al. [81]. LBPs and Haar wavelet transform were also used for ear recognition by Wang et al. [52]. Taertulakarn et al. [82] used Gaussian curvature to determine 3D ear boundaries, and then the mapped of the 3D ear boundaries on 2D images. The hybrid approaches aim to increase the correct recognition for the ear, but they are more complex than approaches based on holistic, local or geometric approaches. Sajadi et al. [63] applied Gabor-Zernike operator to extract global features of ears, and the local phase quantization operator is used for extracting the local features of ears. Then, they used genetic algorithm to select the optimum combination of global and local features.

5. Deep Learning

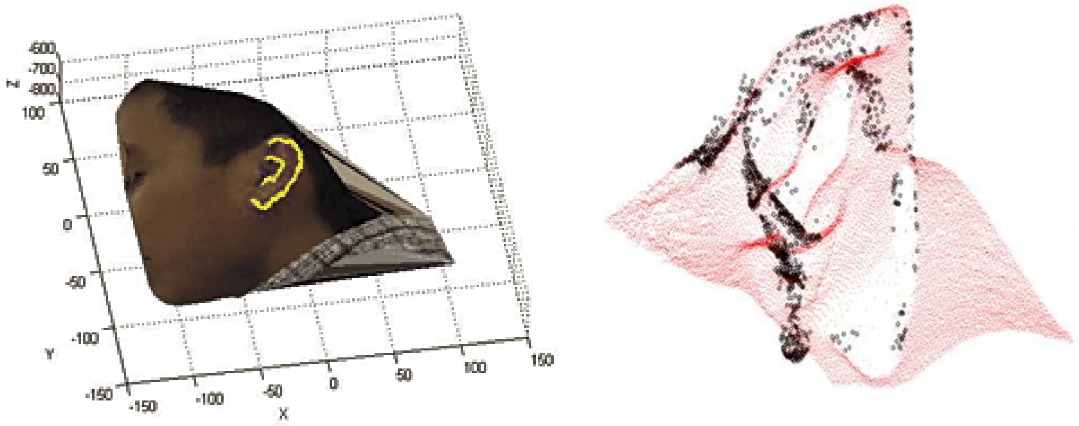
Deep learning has been used for ear recognition in recent years, motivated by its superior performance. However deep learning methods can be limited by a lack of training data. Sinha et al. [59] discussed the feasibility of deep neural networks (DNNS) in the field of ear biometrics. They used Histogram of Gradient (HOG) and support vector machines (SVMs) for ear localization, then a CNN for ear recognition. Ying et al. [66] described a human ear recognition algorithm based on a convolutional neural network. The Dropout technology was introduced in the final fully connected layer to prevent network overfitting. Khaldi et al. [64] proposed a framework that includes Deep Convolutional Generative Adversarial Networks (DCGAN), and Convolutional Neural Network (CNN) models. Omara et al. [65] extracted features by VGG and ResNet pre-trained models, and then the Mahalanobis distance is learned based on LogDet divergence metric learning. Kamboj et al. [127] presented a novel Siamese-based CNN (CG-ERNet) for ear recognition and used Curvature Gabor filters to exploit domain-specific knowledge. [129] used fast region based convolutional neural networks for ear detection. They used CNN for feature extraction, and PCA used for feature reduction. Priyadharshini et al. [130] proposed a six-layer deep convolutional neural network architecture for ear recognition. This model contains six levels of convolutional layers. A sub sampling layer was introduced to reduce the dimension of the feature maps after every convolutional layer and a batch normalisation layer was introduced to improve the stability of the network after every two convolutional layers.

There are some researches for unconstrained ear recognition problem in recent years. Emersic et al. [67][68] presented the summaries of the unconstrained ear recognition challenges in 2017 and 2019. In [67] the results are presented for five groups, and six ear recognition techniques are evaluated. The challenges of unconstrained ear images include head rotation, flipping, gallery size, and large-scale recognition and the others. They had found that the robust performance is based on a smaller part of the dataset, but there was still a significant performance drop when the entire dataset was used for testing. In [68], the results presented for 12 groups and submitted 13 recognition approaches based on descriptor-based methods and deep learning methods. Deep learning methods had better performance than the techniques relying on hand-crafted descriptors. Eyiokur et al. [53] studied the unconstrained ear recognition problem. A deep convolution neural network (CNN) model was used for ear recognition and showed the significance of domain adaptation. Emersic et al [54] built a CNN based ear recognition model. They explored different strategies towards model training with limited amounts of training data and showed that by selecting an appropriate model architecture, using aggressive data augmentation and selective learning on existing (pre-trained) models, they were able to produce an effective CNN-based model using a little more than 1300 training images. [60] applied transfer learning to the AlexNet Convolutional Neural Network (AlexNet CNN) for ear recognition from unconstrained ear images. Though deep learning requires a great deal of data, transfer learning is an approach for solving classification problems with a small amount of data. Alshazly et al. [69] used AlexNet, VGGNet, Incaption, ResNet and ResNeXt for unconstrained ear recognition., and then employed the t-SNE algorithm to visualize the learned features. Ear recognition in the wild was also discussed in [70][71][72].

2.2.2 3D Ear Recognition

In 2D ear recognition, the accuracy is affected by pose variation, and rotation variation. For a 3D model, these influences were not unconsidered because a 3D representation of the subject can be adapted to any rotation, scale and translation [74].

Chen et al. [75] proposed a complete 3D ear recognition system that included 3D ear detection, 3D ear recognition and 3D ear verification. Yan et al. [57] used an ICP (Iterative Closet Point) -based approach for 3D ear recognition. Zhou et al. [58] presented an approach combining holistic and local features for 3D ear recognition. Zeng et.al [120] presented a local feature descriptor based on 3D ear images. Ganapathi et al. [83] proposed an approach for recognition by co-registering 3D and 2D ear images. The technique is based on local feature detection and description. Ganapathi et al. [84] presented a method for 3D ear recognition based on a global 3D descriptor, and they presented a strategy to combine local and global descriptors for superior recognition performance. Feature keypoint detection and description techniques have been proposed by Ganapathi et al. [56] for 3D shape recognition. These techniques precisely discriminate different classes of shapes. Zhu et al. [73] proposed an LJS feature descriptor–pairwise surface patch cropped using a symmetrical hemisphere cut-structured histogram with an indexed shape (PSPHIS) descriptor. Figure 2. 13 depicts two samples of feature extraction from 3D images.



Reference of ear shape [75]

Detected keypoints on 3D ear images [56]

Figure 2. 13 Example of feature extraction for 3D ear images

2.4. Identity Science and Ear Biometrics

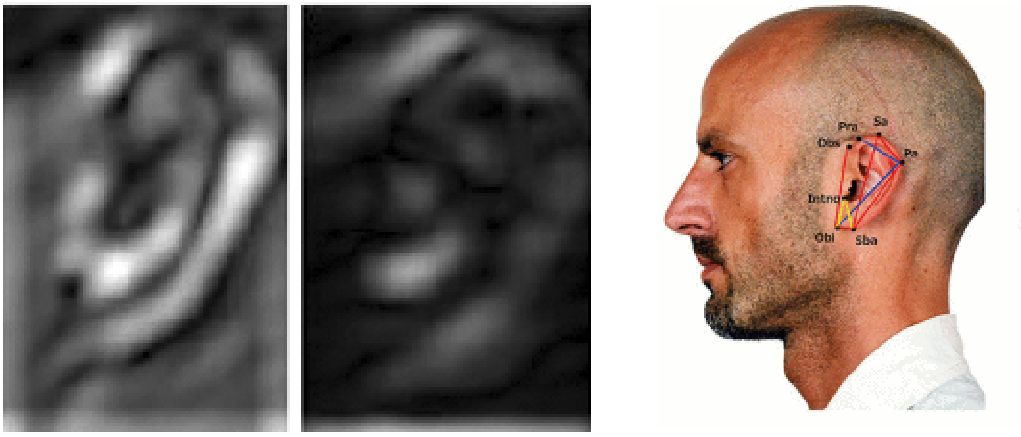
As in other biometrics, there are wider aspects to identification, and more recently it has been shown that gender can be determined automatically from ears. There are a several studies on the other soft biometrics' traits that can be determined from ear images, such as age classification. Table 2. 3 presents the summary of gender classification from ear images.

Table 2. 3 Summary of gender classification from ear images

Publication	Technique	Database	Accuracy
Khorsandi et al. [85]	Gabor Filter	UND	89.49%
Gnanasivam et al. [8]	Geometric features	-	90.42%
Lei et al. [86]	ECM, HIS, SPHIS	UND F UND J2	92.94% 91.92%
Yaman et al. [9]	Geometric Features Deep Learning	Own dataset	65% 94%
Yaman et al. [10]	Deep Learning	Own dataset	> 99%
Yaman et al. [87]	Deep learning	FERET	98.00%
Nguyen-Quoc et al. [88]	HOG, CNNs	EarVN1.0	91.12%

2.3.1 Gender Classification from Ear Images

Khorsandi et al. [85] presented the first research for the gender classification using 2D ear images based on sparse representation. In [85], the Gabor filters are used for feature extraction. Gnanasivam et al. [8] used the Euclidian distance between the ear features and ear hole. The ear features consisted of the outer lobe edge, outer and inner curve of the helix, outer and inner curves of antihelix and two edges of the concha. The Bayes classifier, K-Nearest Neighbour (KNN) classifier and the neural network classifier have been used for the classification. The first paper on the gender classification using a 3D ear database [86] used three types of features: Ear Curvedness Map (ECM), Histogram of Indexed Shapes (HIS) and Surface Patch Histogram of Indexed Shape (SPHIS). They used a Neural Network Classifier, Support Vector Machine (SVM) and Adaboost. There are two samples for gender classification from ear images in Figure 2. 14.



Gabor features of ears [85]

Geometric Features of ear images [9]

Figure 2. 14 Examples of gender classification from ear images

There is some research on gender classification from ear images that uses deep learning. The superior performance of deep learning is not only demonstrated in ear recognition, but also in gender classification from ear images. Yaman et al. [9] presented gender classification from ear images; this paper used geometric features and appearance-based features for gender and age classification, 16 geometric features from the eight landmark points. In addition, they have employed convolutional neural networks for representing and classifying the ear's appearance. They used well-known CNN architectures, namely, AlexNet, VGG-16, GoogLeNet and SqueezeNet. At the end of all these deep network architectures, SoftMax layer was used as a classifier. Yaman et al. [10] conducted another study on gender classification using ear and face profile images and proposed end-to-end multimodal deep learning frameworks. They used different multimodal strategies, such as employing data, feature, and score level fusion. They utilized domain adaptation and employed centre loss besides softmax loss to increase representation and discrimination capability of the deep neural networks. The accuracy of the gender classification in this paper exceeded 99%. This is the state-of-the-art technology for gender classification. Yaman et al. [87] applied CNN models on FERET dataset for age and gender classification from ear images. Nguyen-Quoc et al. [88] used HOG and CNN models for gender classification from ears, and EarVN1.0 dataset had been used for these experiments. Figure 2. 15 gives an overview of the multimodal, multitask age and gender classification framework.

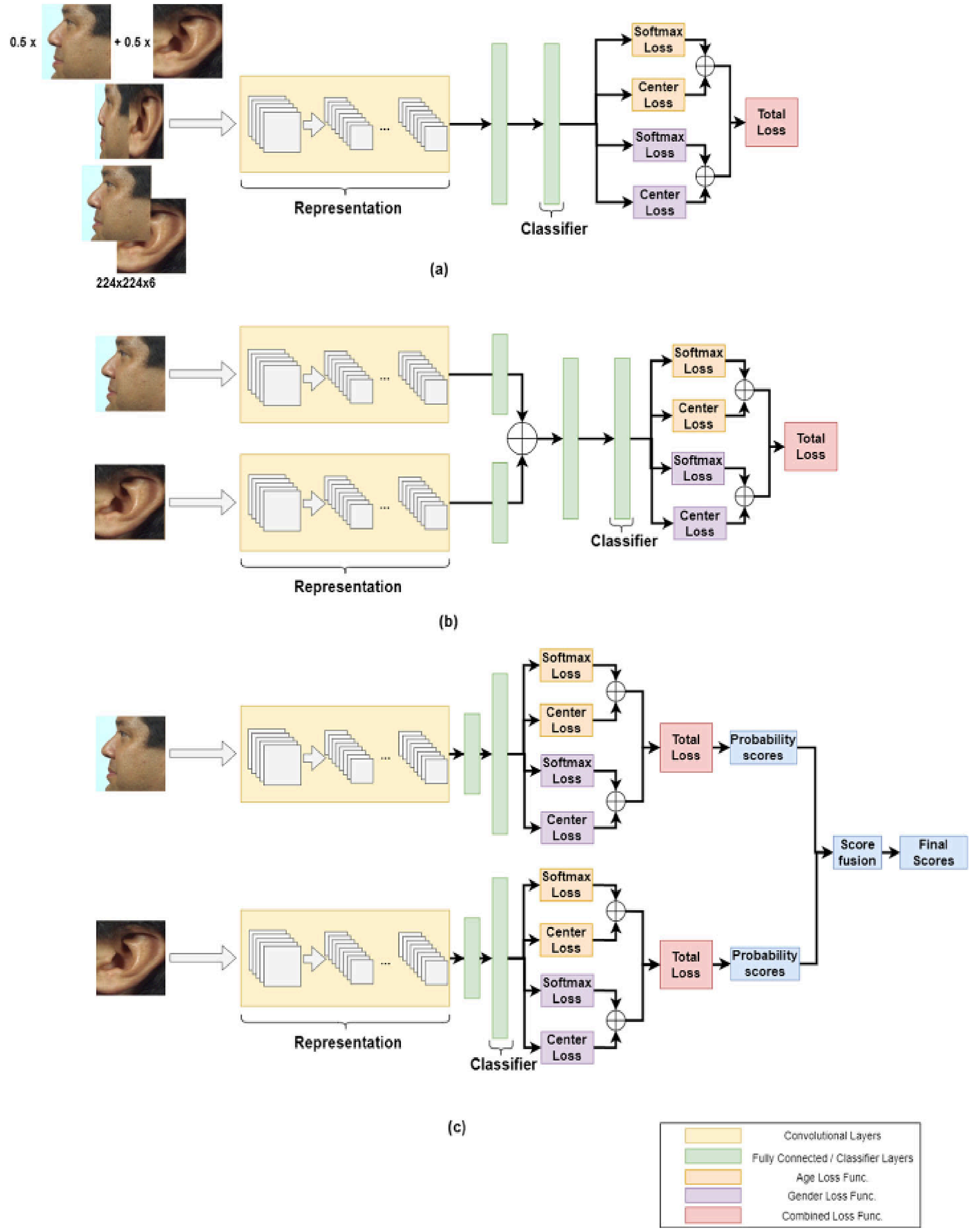


Figure 2. 15 Example of age and gender classification from ear images by deep learning [87] (a) shows data fusion method. (b) shows feature fusion method. (c) shows the score fusion method.

2.3.2 Ear Symmetry

Yan et al. [39] were the first to mention the symmetry of ears and indicated that around 90% of people's right and left ears show bilateral symmetry. They considered 3D ear symmetry and formulated results of oblique (15°, 30°, 45°) views with a 24-subject dataset. Although these are naturally welcome advances, they do not present a concerted study on a large database and compare different techniques. Abaza et al. [89] presented an analysis of symmetry of human ears, using a semi-automated mode and geometric features, with an Equal Error Rate (EER) of 16.8%. They also used Eigen-Ears (PCA) and the Shape From Shading (SFS), with the EERs of 21.1% and 17.1%, respectively. Toygar et al. [90] also indicated symmetric ear and profile face fusion. They used LBP, LPQ and BSIF algorithms, and the best performance of using ear images without profile faces was 76.1%.

Previous studies did not consider rotation, and ears were recorded in laboratory conditions with the subject looking straight ahead. Figure 1. 8 shows how a subject's head may be inclined or rotated, especially during a criminal act. Only three angles are discriminated in [39], where ear symmetry is considered based on the appearance of ears, rather than its structure. However, some subjects' heads were rotated along the yaw or pitch axes, which can change the appearance of the ears.

2.3.3 Age Classification from Ear Images

Yaman et al. [9] presented an approach for age classification from ear images. It is the first research for age classification from ear images. In [9], the authors used geometric features and appearance-based for age classification and they have used deep learning for age classification. Yaman et al. [10] conducted another study on age classification using ear and face profile images.

2.3.4 Ear Print

Hoogstrate et al. [91] investigated the possibility of ear identification for ear images derived from a surveillance camera. They used a better camera for surveillance to increase the chance of identifying an offender. Lugt [92][93][94] proposed using ear prints as evidence to identify a criminal. The claim that personal identification can be established via the ear print left by perpetrator at the scene of crime was later discredited [119]. The studies on the measures of ear and the morphology of various ear features on a sample of 500 male images. Meijerman et al. [95][96] launched the Forensic Ear Identification Project (FEARID) in Europe, that aimed at individualization of ear prints. They have been extensively researched on studying inter and intra -individual variability in ear prints, changes in ear prints on the application of different magnitude of force, and individualization of prints. Alberink et al. [97] also presented the ear prints which can be used as a criminal evidence.

2.5. Conclusions

Ear biometric systems are no longer in the initial phase of development. There has been significant progress, not only in ear recognition, but also in extracting soft traits from ear images. A survey of the ear biometrics has been presented in this chapter, we have stated two main steps of ear biometrics: the first step is the ear detection, and the second one is feature extraction for identification, gender classification, and age classification.

Ear biometric systems are improving. There has been some research addressing pose variation, rotation variation, illumination variation, occlusion and so on. In addition, ear soft biometric is largely in early stages. Gender classification was discussed by more and more researchers. Otherwise, the ear prints can be used as an evidence for accusing criminal.

For comparing with these works of 2D ear recognition, gender classification and ear symmetry, we propose a new model based on geometric features for 2D ear recognition and gender classification. For our model, we use SIFT to detect the landmarks automatically. We also apply transfer learning to VGG-16 [101], GoogleNet [103] and ResNet-50 [104] networks for human identification, gender classification and symmetry from ear images. Although there are many studies of ear recognition based on deep convolutional neural networks, there is no study about why the networks can classify the ear images. Therefore, we use heatmaps to show the contributions of different ear regions for ear recognition and ear symmetry, also, the heatmaps of genders can show the difference between male and female ears. Moreover, there are some studies on the other soft biometrics' traits that can be determined from ear images, such as gender and age classification from ear images, and we propose the kinship recognition from ear images.

Chapter 3

Model-based Approach for Ear Biometrics

Ears have rich structures for biometrics and many studies have utilised the appearance of ears. However, we consider using different parts of ears based on the ear's anatomy, using a model-based approach to denote the local parts of the ears. Considering ear occlusion by hair, local approaches can perform better than holistic approaches because the local approaches for ear recognition are based on the description of local parts. Implicitly, approaches that use structure are less susceptible to occlusion than approaches which use the whole ear. Of the model-based approaches, those that are based on local features are less prone to occlusion than those using a global model. Holistic approaches require a database showing all possible occlusions. Based on a local approach, this problem is mitigated by the fact that the occlusion occurs largely at the rear and top of the ear. By this, local approaches based on the lower parts of the ear near the front are the least susceptible to occlusion by hair.

This chapter presents the algorithms that are used for our model. First, we use Hough Transform for ear detection, and then use SIFT to extract landmarks of ear images, that are used for calculating the distances of our geometric model. Some ear images are influenced by hair, so we apply force field transform to remove the noise of ear images. The datasets used for our experiments have some rotation, we utilise affine transformation to correcting the viewpoints. Moreover, histogram of gradients is also used for feature extraction, and kNN is used for classifications that use Euclidian distance, Manhattan distance and Mahalanobis distance to measure the similarity. Figure 3. 1 shows the general steps of the this approach.

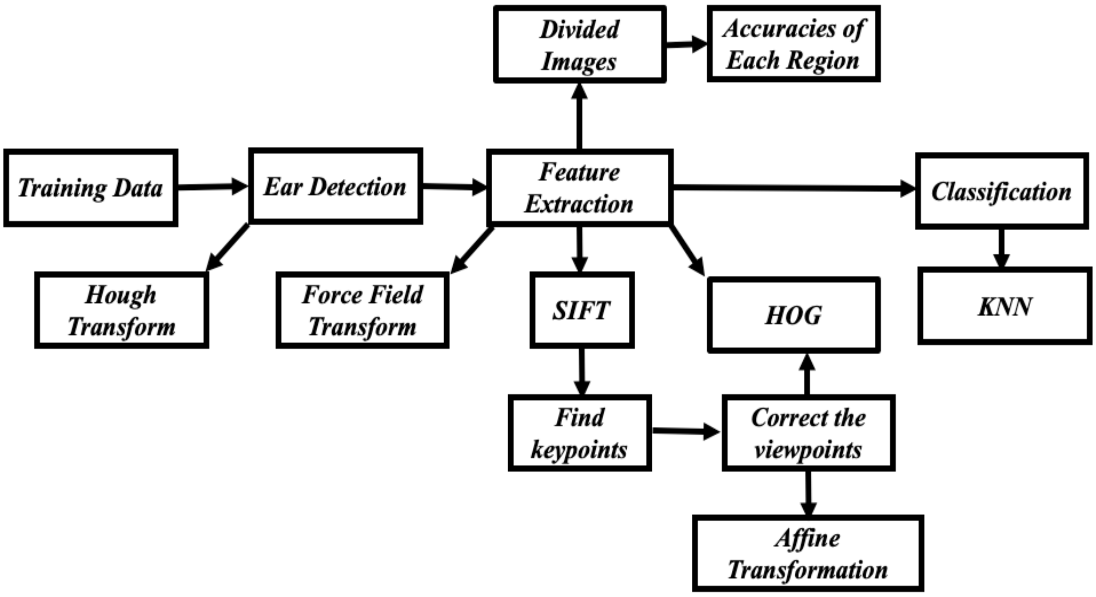


Figure 3. 1 Flow chart of model-based approach

3.1. Model for Ear Biometrics

It is possible to determine human identity, gender and kinship by appearance-based features from ear images. However, this would only suggest performance capability and would not identify salient structures for future investigation. We therefore decided to use a model-based approach. Essentially, we model the ear as a flat plane which is attached to the side of the head. This means that the affine transform can be used to correct changes in pose/ viewpoint (perspective correction is unnecessary given the scale of the ear relative to its context). We detect interest points in a guided manner, thereby identifying salient structures within the ear itself.

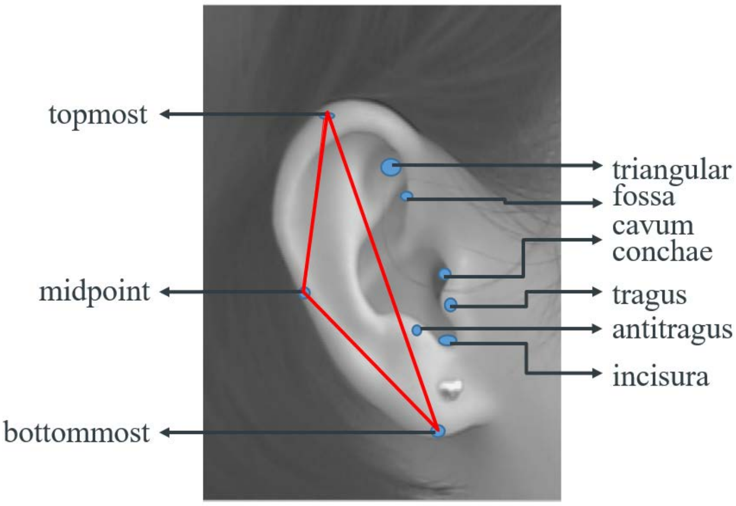


Figure 3. 2 Ear model points

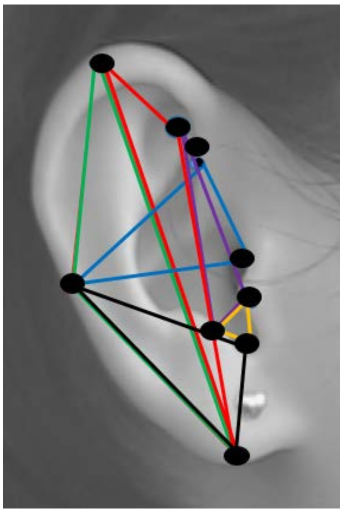


Figure 3. 3 Basis of ear model

A primary description is to use triangles, given their strong invariant properties. Our model consists of nine points, shown in Figure 3. 2, that form 84 triangles from which geometric features can be extracted. Therefore, 252 (relative) distances can be computed among these 84 triangles. The points for the geometric features are shown in Figure 3. 3 and these have reliably been detected on ear datasets. The red triangle shows one of the 84 triangles and the basis used for the affine viewpoint correction. Figure 3. 3 shows six triangles of the 84 model triangles.

We use Hough Transform to detect the ears. The ellipse (ear detection) has been detected via the edges detected by the Canny operator, labelling the apices of the ellipse (the top-most and bottom-most points). These are the first two ear points to be detected and define a major axis from top to bottom. The third ear point is defined as the rear-most edge point along a line normal to the central major axis at its centre. As shown in Figure 3. 3, this defines a point on the rear helix of the ear. These points were found to be robust except in ear images with severe occlusion by hair, when the ear can hardly be detected anyway.

The other landmarks are detected by SIFT, and each keypoint is depicted by a vector showing its scale, orientation and location. There are many keypoints in an ear image; some are detected on the hair, which are not significant for recognition. The noteworthy keypoints describe specific ear components, such as the tragus, the antitragus and the triangular fossa, as shown in Figure 3. 2. Figure 3. 4 shows a sample of ear locations detected by SIFT.

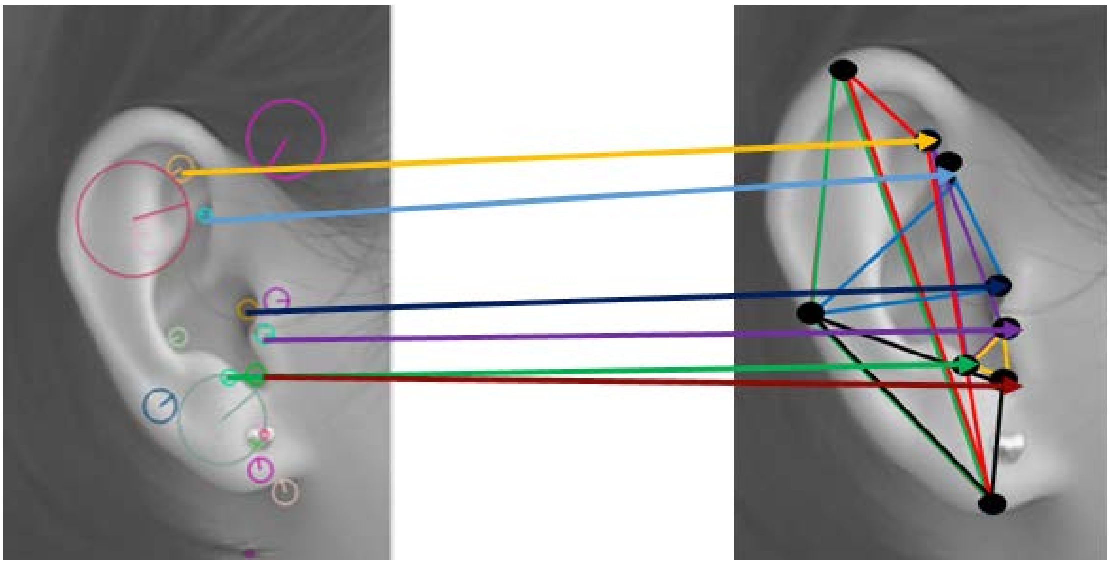


Figure 3. 4 Ear locations shown by SIFT

The images used in this study were from a database where subjects were looking in a plane normal to the camera view, so the ear appears flat. We also have another unconstrained database where subjects' ears are not necessarily normal to the camera view. These are the ears that are corrected so as to make them appear normal to the camera view, using affine transformation. When capturing the ear from the experimental subject using a camera, people's heads may have yaw axis rotation and pitch axis rotation (but not roll, assuming the head is held in a vertical position). Thus, rotations are handled using the affine transformation.

3.2. Hough Transform

In the first stage we use the Hough transform (HT) for ellipses to detect the ear, given good performance in noise and occlusion [26]. The HT implementation defines a mapping from the image points, represented in an accumulator space (Hough space) [98]. The mapping is achieved in a computationally efficient manner, based on a function that describes the target shape, but this mapping requires significant storage and high computational requirements (these problems are addressed later, since they can provide ideas for the continuing development of the HT). However, the fact that the HT is equivalent in performance to template matching has given sufficient impetus for the technique to be among the most popular of all existing shape extraction techniques. The result of applying the Hough transform for ellipses in an ear image is shown in Figure 3. 5. The mapping of the ellipse is defined as:

$$\frac{(x\cos\alpha + y\sin\alpha)^2}{a^2} + \frac{(x\sin\alpha - y\cos\alpha)^2}{b^2} = 1$$

where α is the rotation of points (x, y) with axis lengths a, b.

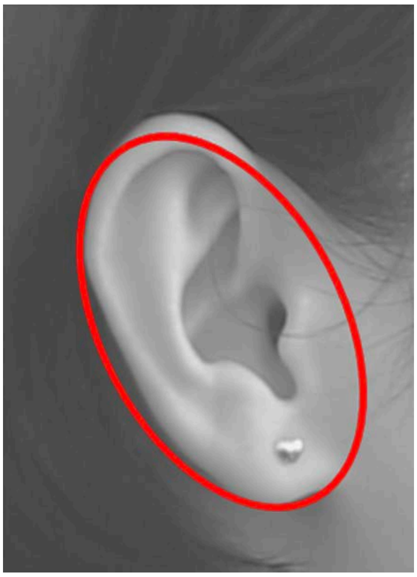


Figure 3. 5 Hough transform for ear detection

3.3. Histogram of Gradients

The Histogram of Gradients (HOG) [99] is employed to describe the ear area surrounded by points detected in the previous section. There are five steps in the computing of HOG descriptors.

1. Global image normalisation, equalization, and gamma (power law) compression, and also the calculation of the square root or log of every colour channel.

2. Calculating the first order gradient of the image. This can capture contour and some texture information, while resisting the illumination variations.

3. Computing gradient histograms. The window of the image is divided into small spatial regions, or ‘cells’. For the pixels of every cell, a 1-D local histogram of gradient or edge orientations is accumulated. This 1-D histogram forms the basic ‘orientation histogram’ representation, and the gradient angle range is divided into a fixed number of predetermined bins by every orientation histogram. The gradient magnitudes of the pixels in the cell are used to vote into the orientation histogram.

4. Normalisation across blocks. Using local groups of cells, all the responses are normalised before the next stage. The normalised block descriptors are described as HOG descriptors.

5. Flattening into a feature vector. Collecting HOG descriptors from all the blocks, and then constructing a combined feature vector for use in the window classifier.

Figure 3. 6 shows the HOG of two ear images.

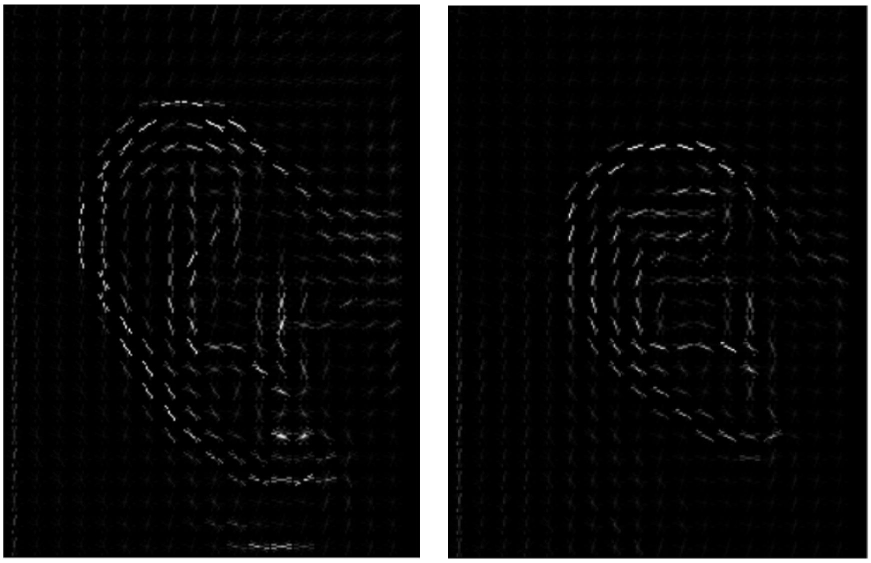


Figure 3. 6 HOG descriptions of two ear images

3.4. Scale Invariant Feature Transform

The SIFT (Scale-invariant feature transform) [76] automatically detects interest points and their descriptions. SIFT is known to be a robust way to use landmark extraction, even in images with small pose-variations and varying brightness conditions. SIFT landmarks provide a measure for local orientation; they can also be used for estimating the rotation and translation between two normalised ear images. The SIFT algorithm follows four main steps for feature extraction:

1. Scale-space extrema detection: the first step of computation processes all scales and image locations. Using the Difference of Gaussian to estimate scale-space extrema.
2. Keypoint localisation: the keypoint candidates are localised and refined by eliminating low-contrast points.
3. Orientation assignment: each keypoint location is based on the assignment of one or more locations based on local image gradient directions.
4. Keypoint descriptor: the local image descriptor for each keypoint is based on image gradient magnitude and orientation.

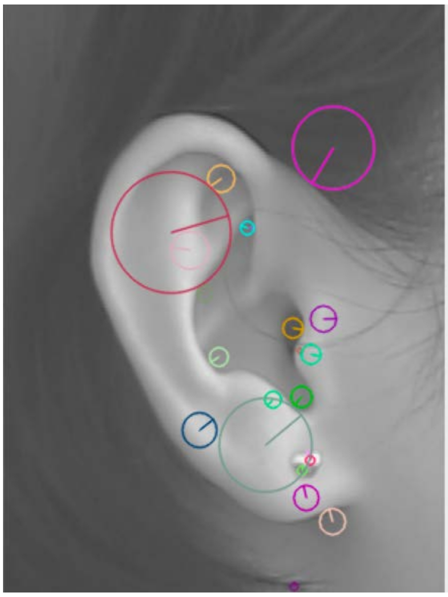


Figure 3. 7 Ear keypoints by SIFT

Figure 3. 7 shows an ear image with the keypoints detected by SIFT superimposed. Each keypoint is depicted by a vector showing its scale, orientation and location. There are many keypoints in an ear image; some are detected on the hair, which are not significant for recognition. The noteworthy keypoints describe specific ear components, such as the tragus, the antitragus and the triangular fossa, some of the structures shown earlier in Figure 3. 2 which are important for our model.

Any point found by the HT that is not within the ellipse is discarded. For the database we used, SIFT shows acceptable stability. SIFT correctly detected the desired points in 94.4% of ear images in our database. SIFT has only failed to detect one or two of the desired points in 5.6% of ear images. Our experiments therefore demonstrate the SIFT is stable enough to be used in this ear biometric study.

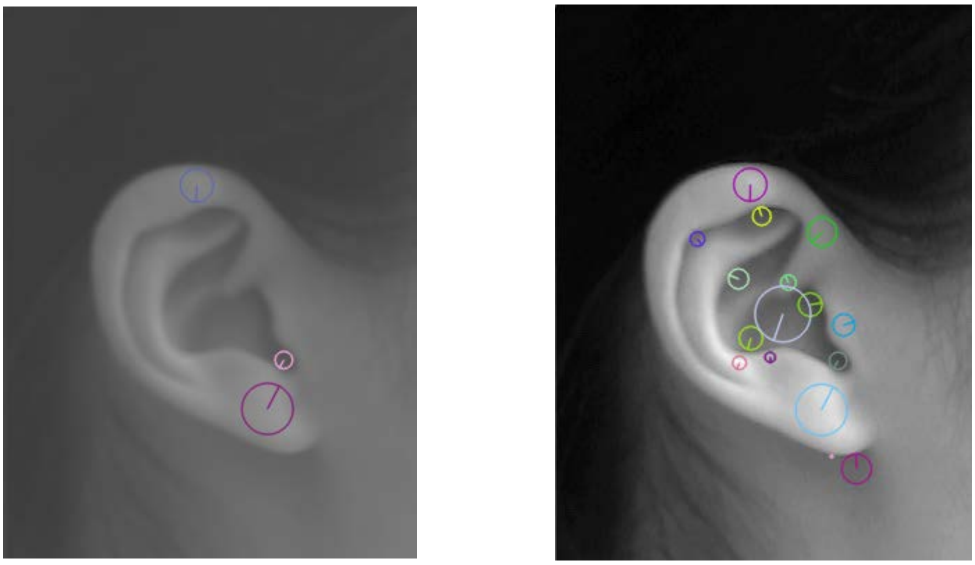
3.5. Histogram (Intensity) Normalisation

Because some images were taken under weak illumination conditions without laboratory-controlled illumination, and SIFT has partial invariance to change in illumination, it is difficult to detect the points which we want to select for the model in some of the images. Thus, to compensate for poor illumination we use histogram normalisation to obtain better contrast images.

Here, the original histogram is stretched and shifted, to cover all 256 available levels. The histogram of the input image starts at I_{min} and extends up to I_{max} brightness levels, the minimum brightness level of the output image is O_{min} and the maximum is O_{max}. The input level is defined as I and the output level is defined as O:

$$O_{x,y} = \frac{O_{max} - O_{min}}{I_{max} - I_{min}} \times \left(I_{x,y} - I_{min}\right) + O_{min} \quad (\forall x, y \in 1, N)$$

Figure 3. 9 denotes the comparison of keypoints detected by SIFT in the original image and an image using histogram normalisation (new image). Figure 3. 9 (a) is the feature extraction from the original image and Figure 3. 9 (b) is the feature extraction from the new image.



(a) original image

(b) new image

Figure 3. 8 Comparison of keypoints by SIFT

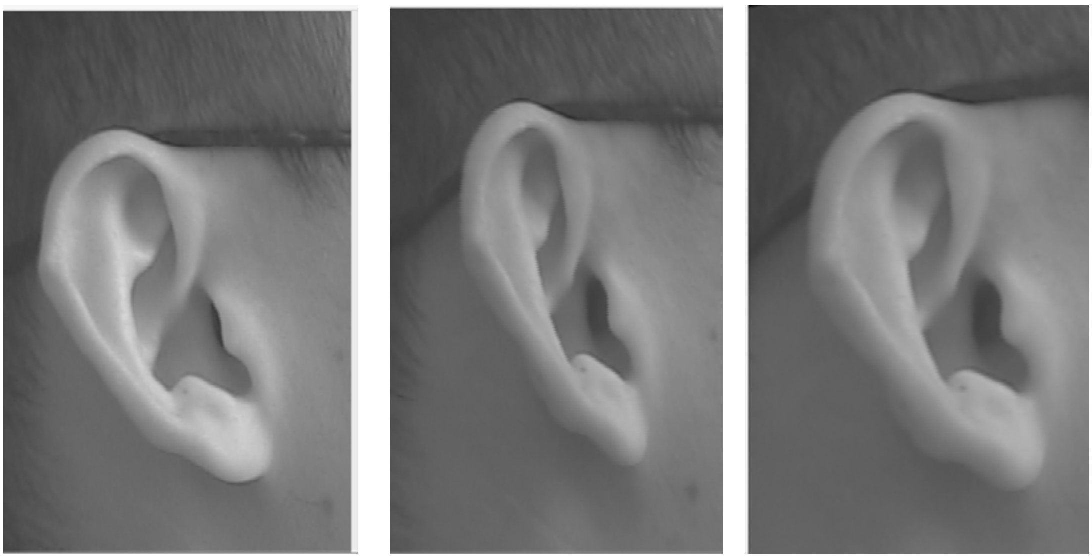
3.6. Affine Transformation

Ear rotation is an important issue in ear biometrics. Planar rotations can effectively be solved by alignment techniques. Manual ear alignment is the general method, and automatic ear alignment studies have thus far found limited use. Thus, we propose that ear rotation can be addressed by using affine transformation.

The affine transformation is defined as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_5 \\ a_6 \end{bmatrix}$$

As observed from the above equation, x, y are the original coordinates and x', y' are the coordinates applied affine transformation. The affine transformation [100] has six degrees of freedom corresponding to the six parameters $(a_1, a_2, a_3, a_4, a_5, a_6)$ of the transformation (scales, rotation and translation parameters), which can be calculated using non-collinear correspondences. Figure 3. 10 shows an example of the affine transformation. Figure 3. 10 (a) is a reference image, and Figure 3. 10 (b) is the registered image, which rotates around the yaw axis. Figure 3. 10 (b) changes to Figure 3. 10 (c) after affine transformation.



(a) gallery (b) probe (c) after affine

Figure 3. 9 Applying the affine transformation

3.7. Force Field Transform

The force field transform was inspired by electrostatics and electromagnetics and was used in one of the earliest approaches to ear biometrics [4]. The ear image is transformed into a force field image by assuming that each pixel is applying an attractive force to the other pixel. Then, each pixel in the image is replaced by the corresponding force field created by rest of the pixels at that point. The approach has an intrinsic averaging capability which removes hair as it changes in local regions, and its basis also preserves broad features/ ear structures.

Each pixel is considered to generate a spherically symmetrical force field so that the force $F_i(r_j)$ exerted on a pixel of unit intensity at the pixel location with a position vector r_j by any other pixel with a position vector r_i and pixel intensity $P(r_i)$ is given by:

$$F_i(r_j) = P(r_i)\frac{r_i - r_j}{\left|r_i - r_j\right|^3}$$

The total force $F(r_j)$ exerted on a pixel of unit intensity at the pixel location with position vector r_j is the vector sum of all the forces due to the other pixels in the image, and is calculated by:

$$F(r_j) = \sum_{i=0,\neq j}^{N-1} F_i(r_j)$$

Associated with the force field generated by each pixel, there is a spherically symmetrical scalar potential energy field denoted by $E_i(r_j)$. This energy is the potential energy imparted to a pixel of unit intensity at the pixel location with position vector r_j by the energy field of any other pixel with position vector r_i and pixel intensity $P(r_i)$, and is determined by:

$$E_i(r_j) = \frac{P(r_i)}{|r_i - r_j|}$$

The total potential energy of all the image pixels is given by:

$$E(r_j) = \sum_{i=0,\neq j}^{N-1} E_i(r_j)$$

Because each pixel affects the other pixels with different distances, the potential energy transformation is the inverse of the gradient of the potential energy surface at that point:

$$\mathbf{F}_{(r)} = -\boldsymbol{\nabla}(E(r))$$

Figure 3. 11 depicts the results of the force field feature extraction applied to ear images.

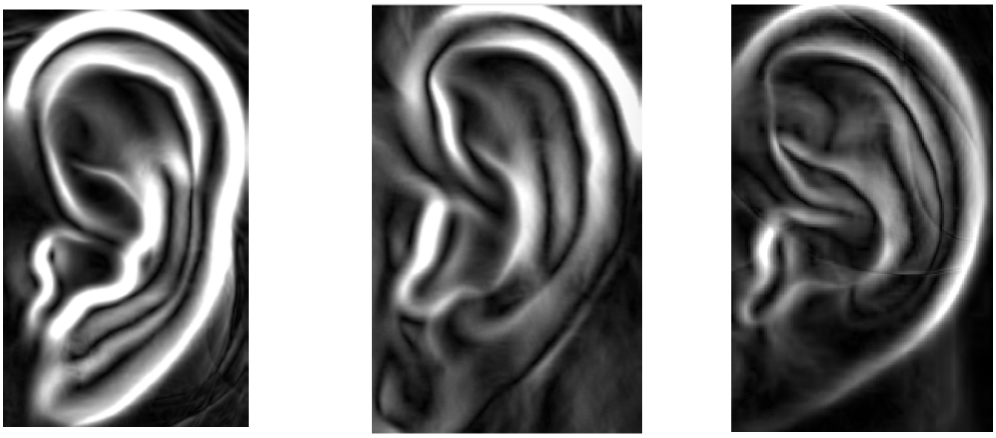


Figure 3. 10 Force field transform of different ears

3.8. Classification

In this section, the classifier used for ear biometrics is kNN (k-nearest neighbourhood algorithm), and the Euclidian, Manhattan, and Mahalanobis distance [98] measures are defined below:

$$D_E = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2}$$

$$D_M = \sum_{i=1}^{N}|x_i - y_i|$$

$$D_{MAH} = \sqrt{(p-\mu)^T \Sigma^{-1}(p-\mu)}$$

where x, y are the coordinates of image pixels and N is the number of pixels in the image, $p_i = (m1_i, m2_i, m3_i, \cdots, mN_i)^T$ is sets of vector points, which have mean values $\mu = (\mu1, \mu2, \mu3, \cdots, \mu N)^T$ and a covariance matrix Σ.

3.9. Conclusions

In this chapter, we described the general steps of our model and also presented all the approaches that used in our model. Our model is based on ear anatomy and ear structure. The advantages of our model include robustness in handling noise and rotation. Our approach uses the Hough Transform to find the apices of the ears. It also exploits the SIFT to detect the landmarks. The affine transformation is then employed to solve the rotated ear images. The histogram normalisation is also used for a better contrast in ear images. HOG and force field transform are also used to detect the local information, while kNN is exploited for classification. We will present and analyse the results in Chapter 5.

Chapter 4

Deep Learning for Ear Biometrics

Using deep learning for ear biometrics is very popular at the present. It the recognition approach essentially depends on the appearance of ears and is functionally similar to holistic approaches which are also based on the appearance of ears. Deep learning achieves a high performance for dealing with occlusion. Deep learning was motivated by its advantageous performance, as has been used for ear recognition in recent years. However, deployment of deep learning can be limited by a lack of training data. Although deep learning has an excellent performance for ear recognition, it needs to collect data from a larger number of images to avoid overfitting.

In this chapter, we will present three different networks, pre-trained models and some approaches to evaluating visualisations. Figure 4. 1 shows a general flowchart of the deep learning approaches for ear biometrics. The strategy here is motivated by the advantageous performance of deep networks and our aim is to demonstrate that this can be exploited in ear biometrics, particularly with the later agendas here of analysing kinship and gender with uncontrolled image acquisition. Currently, it is of course a very popular field of research, and it continues to advance very fast. We have selected some of the more popular architectures, given their proven performance, though there is a richer selection available. There are many studies of ear biometrics based on deep convolutional neural networks, but there are no studies on the contributions of ear regions. Therefore, we will calculate the heatmaps to show the contributions of ear regions and the difference between male and female ears.

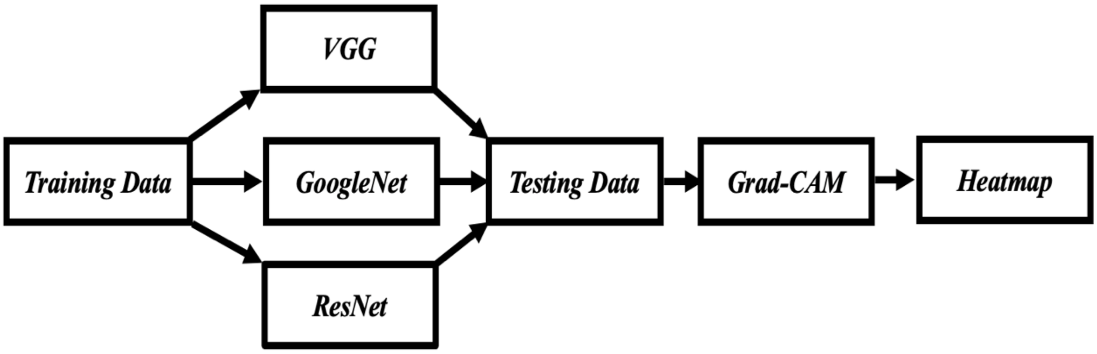


Figure 4. 1 Flowchart of deep learning for ear biometrics

4.1. VGG

Oxford's Visual Geometry Group develop the VGG network [101], which has two architectures, VGG 16 and VGG 19. The pre-processing is subtracting the mean RGB value, computed from the training set, from each pixel. The network uses filters with a small 3×3 receptive field, and 1×1 convolution filters, which is a linear transformation of the input channels. The convolution stride is fixed to one pixel. The spatial pooling includes five max-pooling layers, which are performed over 2×2 pixel windows. There are three fully connected layers of the network and the last layer is the softmax layer.

The VGG network utilises convolutional layers with small receptive field (3×3), instead of convolutional layers with larger reception field $(11 \times 11, 7 \times 7, 5 \times 5)$ used in AlexNet [102]. Thus, VGG is deeper than AlexNet and the performance is better. However, VGG has more parameters, so it needs a long time to train a model. Figure 4. 2 shows the architecture of the VGG network.

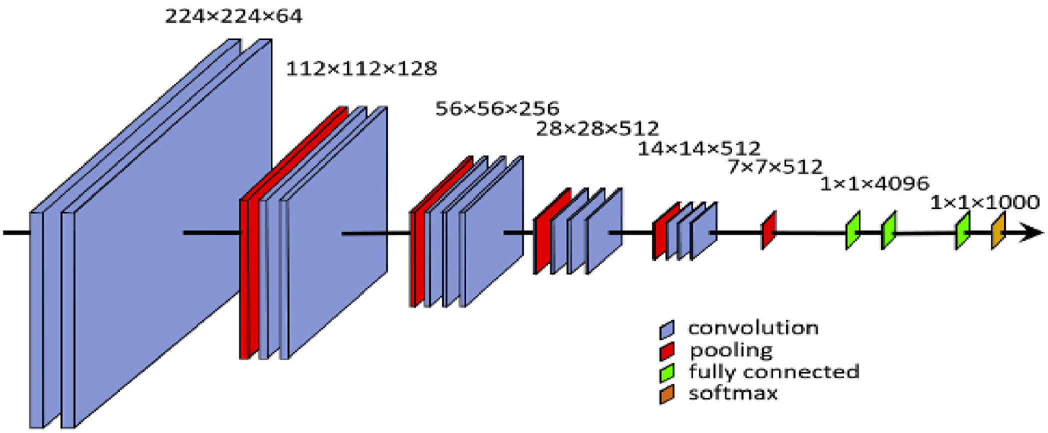


Figure 4. 2 VGG architecture [98]

4.2. GoogleNet (Inception)

Szegedy et al. [103] propose a deep convolutional neural network known as GoogleNet or Inception. AlexNet [102] and VGG [101] optimise training performances by using more layers in their architecture, whereas the main hallmark of GoogleNet is improved utilisation of computing resources within the network. The GoogleNet network uses deeper and wider architecture while keeping the computational budget constant, thus, the GoogleNet network extracts more features with the same computation.

Figure 4. 3 shows the inception module. There are two main contributions of the inception module.

1. 1×1 convolution is used for reducing dimension, which allows the decrease of computational load by lessening the layers' number of operations.
2. There are four branches in the inception module, which have different sized convolutions or pooling, and are aggregated so that the next stage can extract features from the different scales simultaneously.

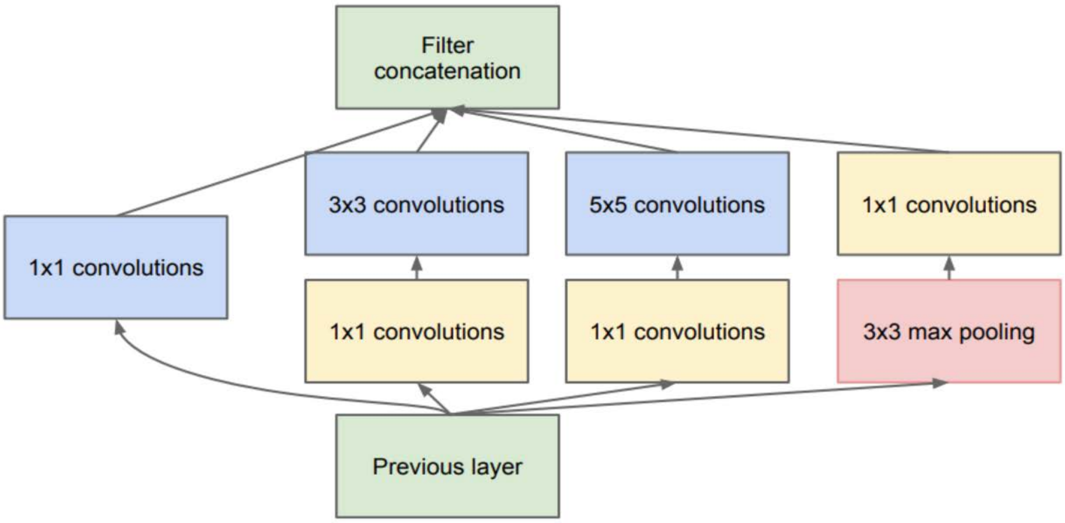
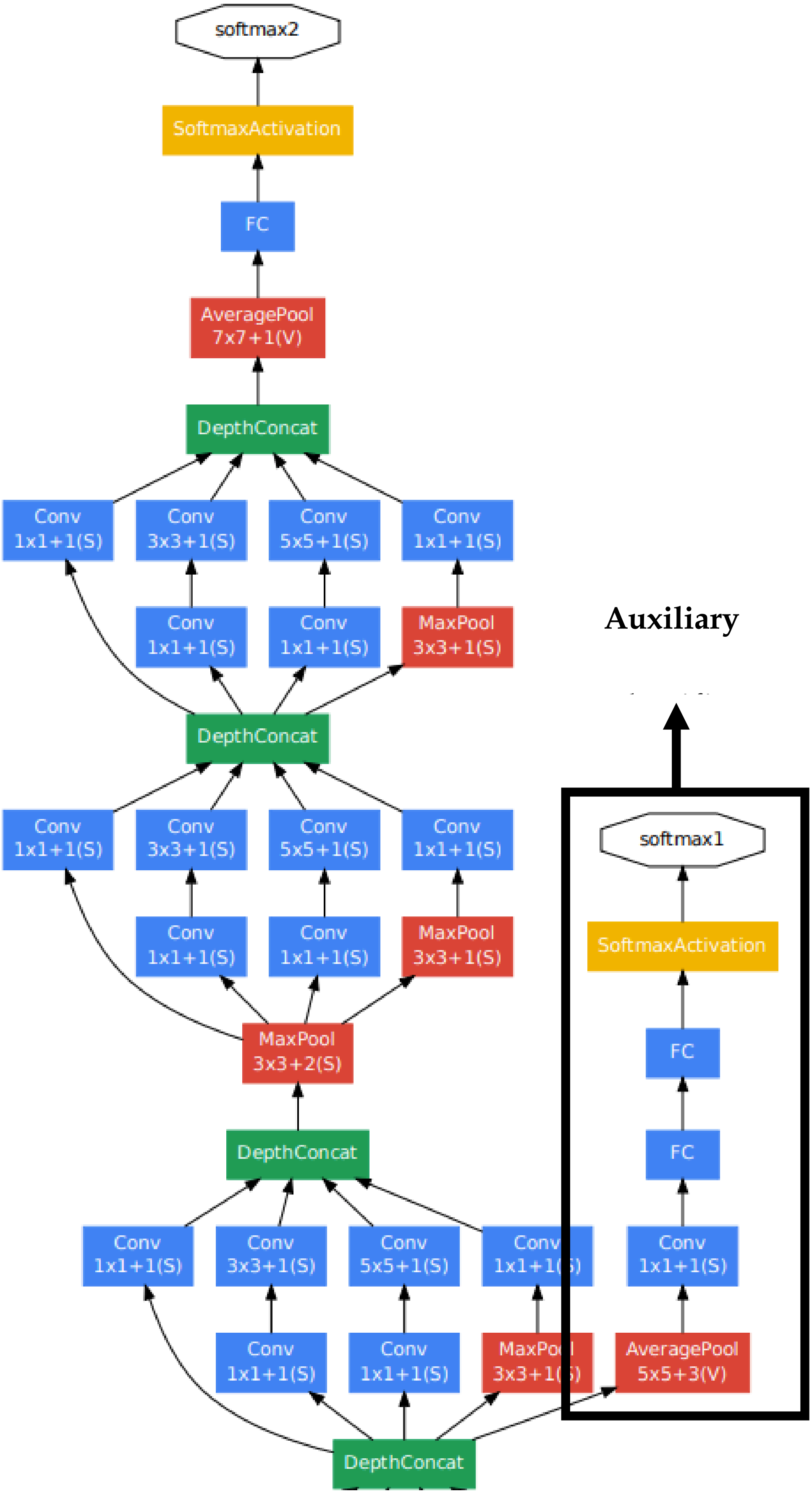
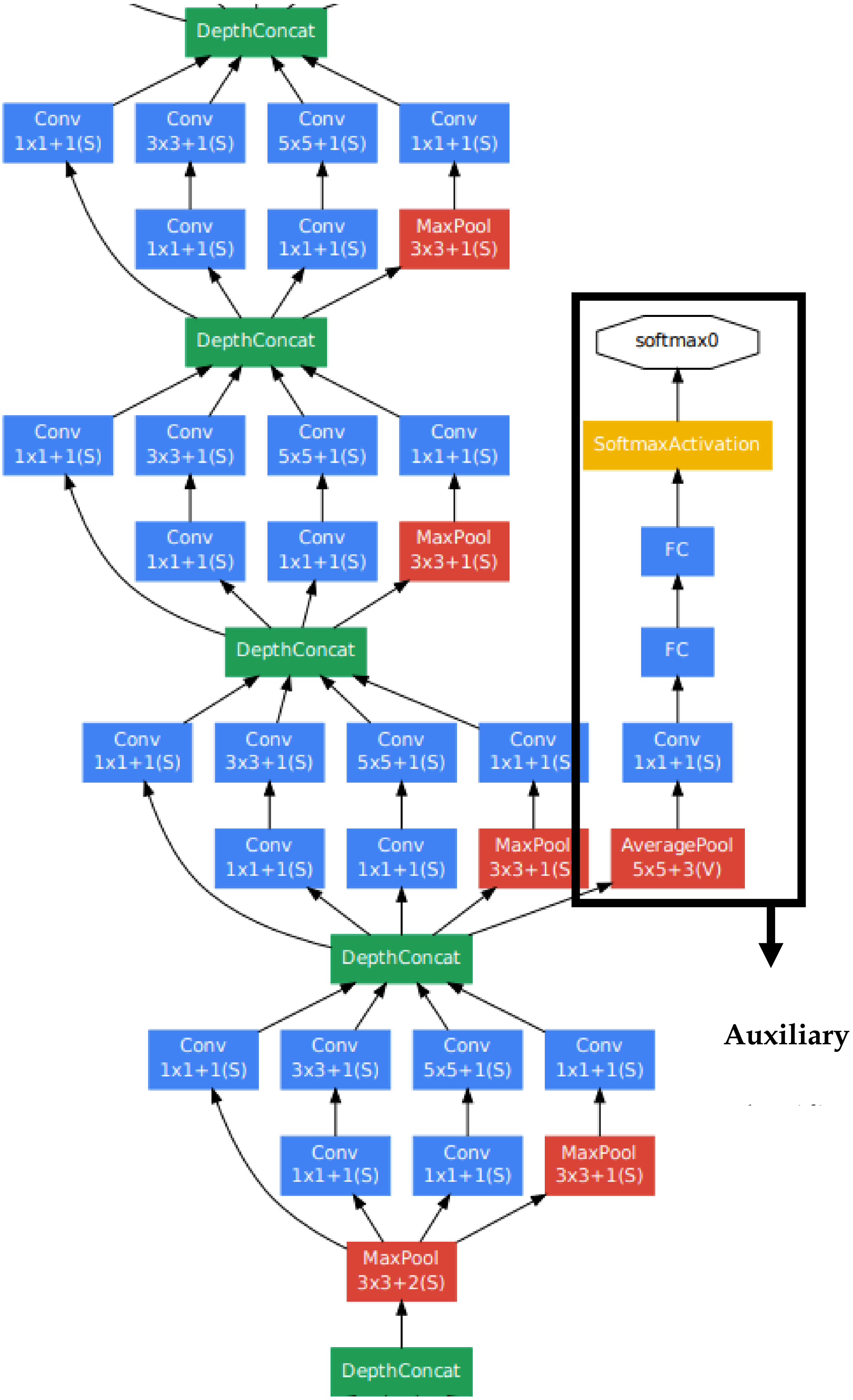


Figure 4. 3 Inception module with dimensionality reduction [103]

The GoogleNet architecture has 22 layers (or 27 layers including pooling layers) and there are nine inception modules in the network. Figure 4. 4 depicts the architecture of GoogleNet, spread over three pages. As shown in Figure 4. 4, there is an additional component, named the auxiliary classifier, which is used to avoid overfitting. Auxiliary classifiers are added to the intermediate layers of the architecture, the third and sixth inception modules.


softmax2
SoftmaxActivation
FC
AveragePool 7x7+1(V)
DepthConcat
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
Conv 1x1+1(S)
Conv 1x1+1(S)
Conv 1x1+1(S)
MaxPool 3x3+1(S)
DepthConcat
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
Conv 1x1+1(S)
Conv 1x1+1(S)
Conv 1x1+1(S)
MaxPool 3x3+1(S)
MaxPool 3x3+2(S)
DepthConcat
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
Conv 1x1+1(S)
Conv 1x1+1(S)
Conv 1x1+1(S)
MaxPool 3x3+1(S)
DepthConcat
Auxiliary
softmax1
SoftmaxActivation
FC
FC
Conv 1x1+1(S)
AveragePool 5x5+3(V)

4


DepthConcat
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
Conv 1x1+1(S)
Conv 1x1+1(S)
Conv 1x1+1(S)
MaxPool 3x3+1(S)
DepthConcat
softmax0
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
Conv 1x1+1(S)
SoftmaxActivation
Conv 1x1+1(S)
Conv 1x1+1(S)
MaxPool 3x3+1(S)
FC
DepthConcat
FC
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
Conv 1x1+1(S
Conv 1x1+1(S)
Conv 1x1+1(S)
Conv 1x1+1(S)
MaxPool 3x3+1(S
AveragePool 5x5+3(V)
DepthConcat
Auxiliary
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
Conv 1x1+1(S)
Conv 1x1+1(S)
Conv 1x1+1(S)
MaxPool 3x3+1(S)
MaxPool 3x3+2(S)
DepthConcat

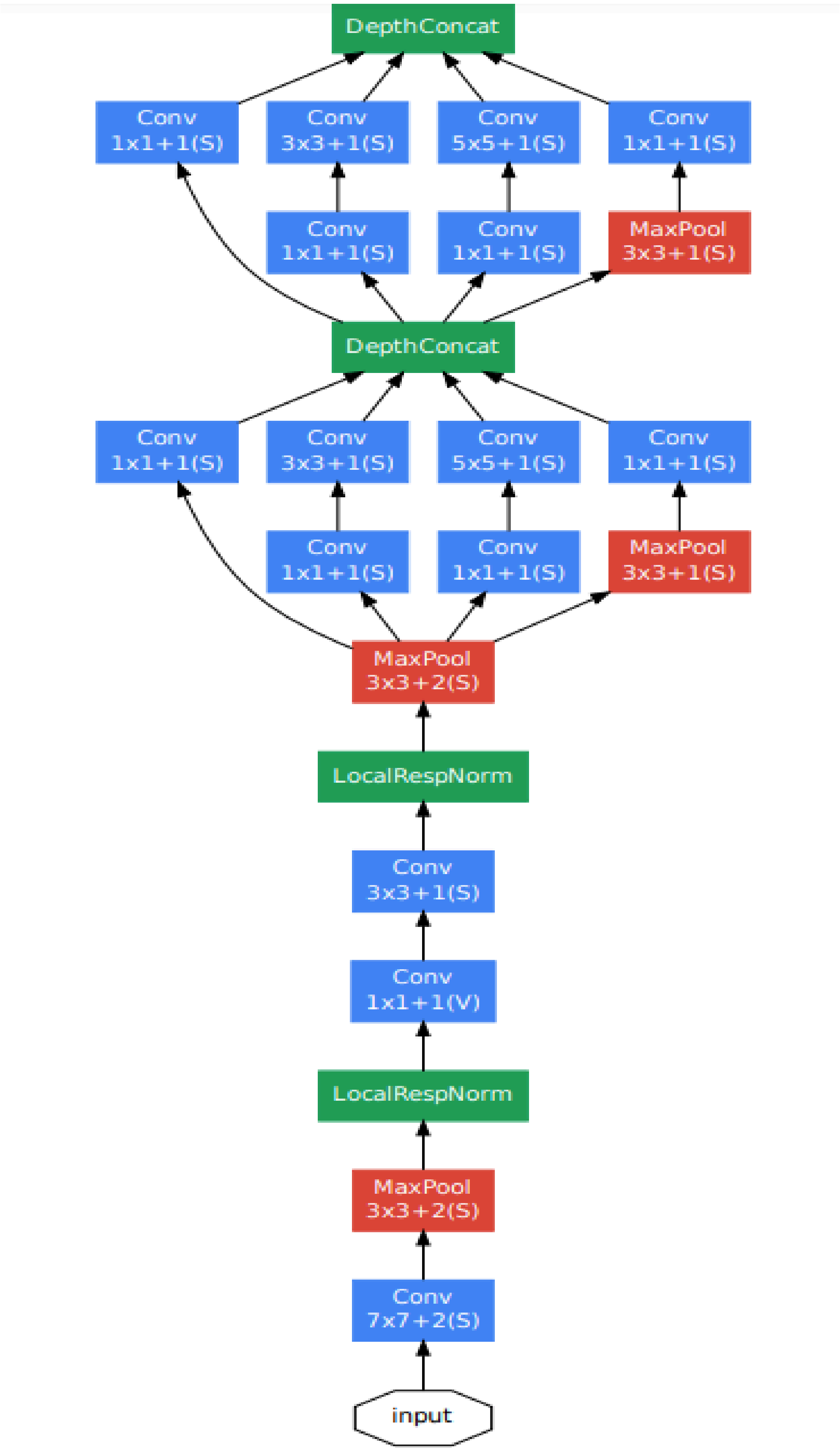


Figure 4. 4 GoogleNet architecture with all the bells and whistles [103]

4.3. Deep Residual Learning for Ear Biometrics

The deep residual learning method addresses the degradation problem by introducing a deep residual network [104]. This method allows some stacked layers to fit a residual mapping, instead of hoping these layers directly fit a desired underlying mapping. Formally, the desired underlying mapping is denoted as $\mathcal{H}(x)$, and the stacked nonlinear layers fit another mapping of $\mathcal{F}(x) := \mathcal{H}(x) - x$. The original mapping is recast into $\mathcal{F}(x) + x$. The formulation of $\mathcal{F}(x) + x$ can be realised by feedforward neural networks with skip connections, which are shown in Figure 4. 5. The building block of ResNet [104] is defined as:

$$y = \mathcal{F}(x, \{W_i\}) + x$$

where x, y are the input and output vectors of layers considered, and $\mathcal{F}(x, \{W_i\}) + x$ indicates the residual mapping to be learned.

Because of concerns about the training time, the building block is modified as a 'bottleneck' design [104]. For each residual function ($\mathcal{F}$), a stack of three layers instead of two layers is used, as shown in Figure 4. 6. The 50-layer ResNet contains three-layer bottleneck blocks instead of the two-layer blocks in the 34-layer network. The architecture of the ResNet is shown in Figure 4. 7.

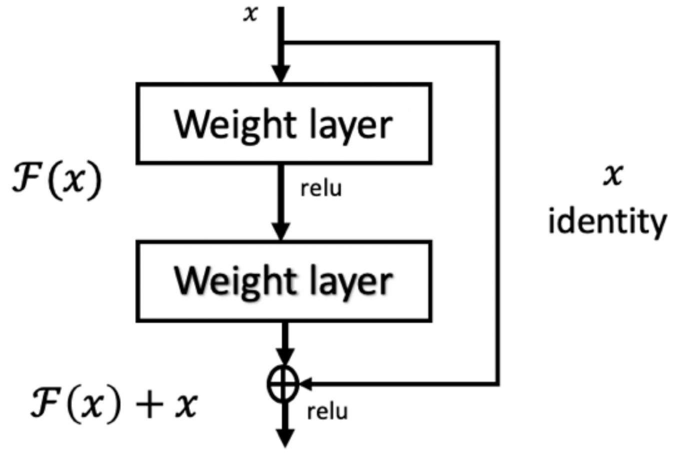


Figure 4. 5 Residual learning: a building block [104]

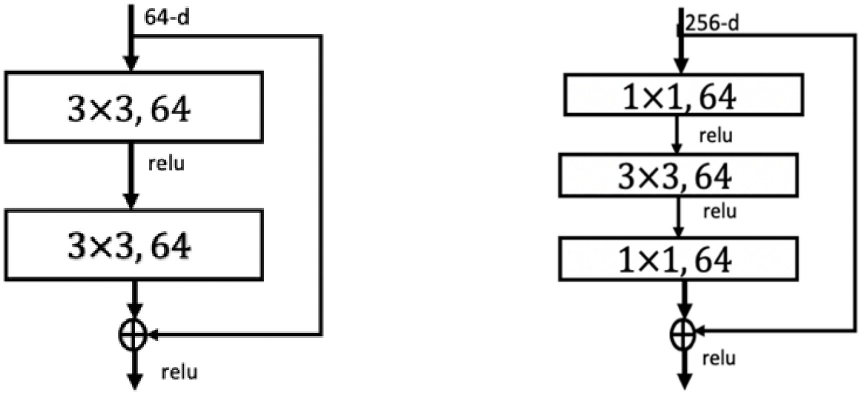


Figure 4. 6 A deeper residual function $\mathcal{F}$ for ImageNet [104] (left: building block for ResNet-34; right: 'bottleneck' building block for ResNet-50/101/152)

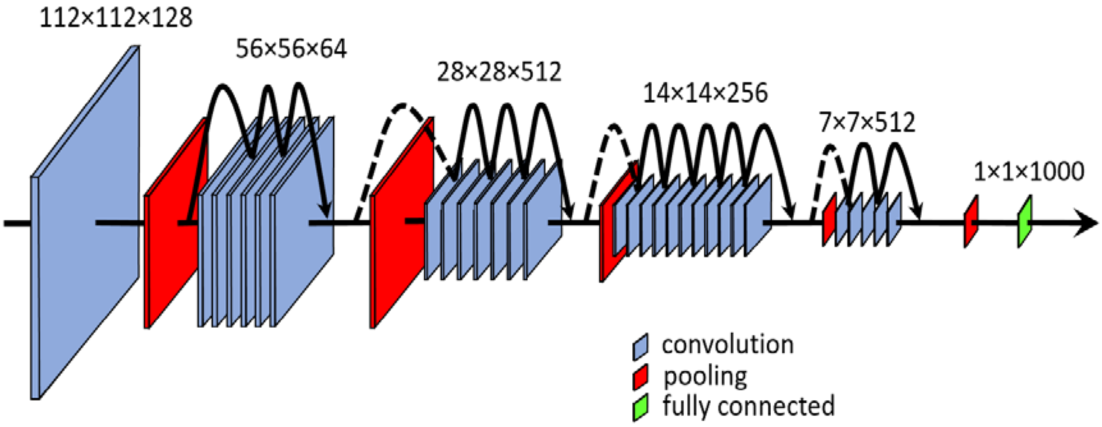


Figure 4. 7 ResNet architecture [98]

4.4. Fine Tuning Networks

Transfer learning is used to improve the performance of the network from one domain by transferring information from a related domain [105], especially for small datasets such as our ear datasets. The strategy of transfer learning is to pre-train a model based on a large labelled dataset. As we have a limited dataset for the training, we use VGG-16, GoogleNet and ResNet-50, and transfer the learning of the pre-trained models (VGG-16, GoogleNet, ResNet-50) to the ear datasets. For ear recognition, we replace the last fully connected layer and the classification layer in the pre-trained models with a set of layers that can classify 137 classes to recognise 137 subjects. For the case of gender classification, we replace those layers with layers that can classify the male and female classes into two groups. Finally, for ear symmetry experiments, we replace those layers with layers that can classify 100 classes to recognise 100 subjects.

4.5. Evaluating Visualisations

Grad-CAM [106] (Gradient-weighted Class Activation Mapping) uses the gradients of any target concept, say logits for 'female' or even a caption. They flow into the final convolution layer to produce a course localisation map highlighting the important regions in the image to predict the concept. The class discriminative localisation map Grad-CAM $L^c_{Grad-CAM} \in \mathbb{R}^{u \times v}$ of width u and height v for any class c, y^c (before the softmax), with respect to feature map A^k of a convolutional layer is $\frac{\partial y^c}{\partial A^k}$. The neuron importance weights α^c_k are:

$$\alpha^c_k = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}}$$

which are obtained by a global average pooling of the gradients derived from backpropagation. The weight α^c_k represents a partial linearisation of the deep network downstream from A, and captures the 'importance' of feature map k for a target class c. Z is the number of pixels in the feature map.

A weighted combination of forward activation maps is performed in [106], and this is followed by a ReLU to obtain the linear combination, written as:

$$L^c_{Grad-CAM} = ReLU\left(\sum\nolimits_k \alpha^c_k A^k\right)$$

Figure 4. 8 denotes the structure of Grad-Cam. It is given an image and a class of interest as input, which passes through the CNN part, and then calculates a raw score for the category. The gradient of desire class is set to one, and the other classes are set to zero. The signal is then backpropagated to the rectified convolutional feature maps of interest. They pointwise multiply the heatmap with guided backpropagation to obtain Guided Grad-CAM visualisations.

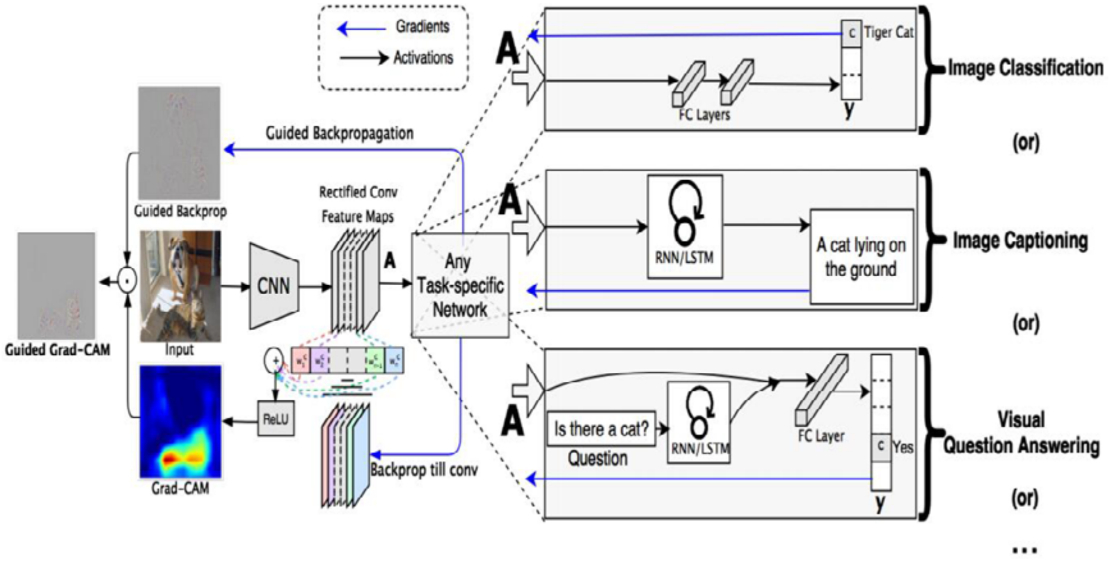


Figure 4. 8 Grad-Cam overview [106]

4.6. Conclusions

In this chapter, we have presented three different convolutional neural networks, fine tuning networks, and Grad-CAM. There are many different networks and models in deep learning. The main difference between networks consists of depth, width and the use of different layers. Therefore, the structure of the GoogleNet network is deeper and wider, while the architecture of the ResNet network is much deeper than, but not as wide as, the VGG network. VGG requires more time for training because it has more parameters. For GoogleNet, the convolutional layers switch to inception modules and two auxiliary classifiers are added to prevent overfitting. ResNet has four main stage blocks, and each stage block stacks upon several building blocks. We will present the fine-tuning networks for ear recognition, gender classification and ear symmetry, and analysis for the results of ear biometrics, in Chapter 6.

Chapter 5

Results and Analysis of the Model-Based Methodology

Many studies in ear recognition have been conducted, and multiple approaches have been used for ear recognition, including principal component analysis (PCA), local binary pattern (LBP) and deep learning, which were described in Chapter 2. In terms of occlusion by hair, local approaches, which are based on descriptions of local parts, can perform better than holistic methods. Such structural approaches are less susceptible to occlusion, and model-based approaches based on local features are less prone to occlusion than those based on global models. Holistic approaches require a database with images showing all possible occlusions, which is mitigated by the fact that the occlusion is largely at the rear and top of the ear. Local approaches based on the lower parts of the ear near the front are the least susceptible to occlusion by hair.

This study used models for ear recognition, gender classification, kinship verification and ear symmetry. All the approaches used ear recognition as a basic evaluation. This chapter presents the results and analyses of model-based approaches for ear biometrics. In addition, the only identification approach that considered the potency of different ear parts (Arbab-Zavar et al. [78]) used SIFT, and the model was constructed using a stochastic clustering method based on the SIFT keypoints of the training set. The recognition accuracy was 85.7%. The top ten most significant ear parts for recognition were shown with the most important being the inferior crus of the antihelix. Accuracies of different regions were used to analyse the contribution of different parts for the purposes of ear biometrics.

5.1. Geometric Model-Based Methodology

This section presents the separate geometric model-based results for human identification, gender classification and kinship verification obtained from ear images. Additionally, the performance of this model is analysed here.

5.1.1. Results of Ear Recognition

Chapter 3 described the geometric model used in this study. It contains nine points, and the distances between these points were extracted as geometric features. The ear model is presented in Figure 3. 2 and Figure 3. 3. Datasets I and II of the USTB datasets [109] were used to investigate ear recognition capability in the model. Additionally, as the images in dataset I were not rotated, unlike those in dataset II, an affine transformation was applied to those images in the second dataset. All results used kNN (k = 1) with a leave-one-out cross validation, which means one ear image for testing, the others for training.

Table 5. 1 Rank 1 recognition of USTB dataset I

Method	Euclidian distance	Manhattan distance
Geometric features	81.7%	82.0%
HOG	91.7%	90.0%
Geometric features + HOG	92.8%	95.6%

Table 5. 2 Rank 1 recognition of USTB dataset II (without affine transformation)

Method	Euclidian distance	Manhattan distance
Geometric features	53.5%	54.2%
HOG	62.1%	66.3%
Geometric features + HOG	62.5%	70.2%

Table 5. 3 Rank 1 recognition of USTB dataset II (with affine transform)

Method	Euclidian distance	Manhattan distance
Geometric features	56.1%	59.8%
HOG	59.9%	70.4%
Geometric features + HOG	64.2%	80.0%

Table 5. 1 presents the results for dataset I, which confirm that this study's model can match the contemporary performance with constrained datasets. The raw image results for dataset II are shown in Table 5. 2 and demonstrate that without affine transformation, recognition is much reduced. Clearly, the affine transformation is beneficial. Table 5. 3 presents the results for dataset II with viewpoint correction via an affine transformation. In all three tables, the first column is the description method, and the first row is the measurement distance used. As can be seen in Table 5. 1, the accuracy of HOG is slightly higher than for the geometric features. When the two feature sets are fused, the accuracy reaches as high as 95.6%. In Table 5. 2 and Table 5. 3, the accuracy is improved when the rotated image uses affine transformation. The accuracies of the geometric features are 54.2% and 59.8%, and the accuracies of HOG are 66.3% and 70.4%. The accuracies of the geometric features and HOG are 70.2% and 80.0%. Comparing the results in Table 5. 2 and Table 5. 3, the affine transformation can successfully correct the viewpoints, thus, it can work for the rotated ear images.

5.1.2. Results of Gender Classification

The results in Table 5. 1, Table 5. 2 and Table 5. 3 confirm that this geometric model is appropriate for use with ear recognition; therefore, it was applied for gender classification. We used SOTEAR dataset [21] for gender classification. It contained a total of 67 subjects, including 37 males and 30 females. The experiments in this section used k = 3 with a leave-one-out cross validation. Distance was measured using the Euclidian distance, Manhattan distance and Mahalanobis distance. The results of gender classification from the ear images are given in Table 5. 4.

Table 5. 4 Results of gender classification

Method	Geometric features	Geometric features + HOG
Euclidian distance	67.2%	62.9%
Manhattan distance	59.7%	68.7%
Mahalanobis distance	65.7%	58.2%

As can be seen in Table 5. 4, the correct gender classification of the geometric model is 67.2%, and the correct gender classification of geometric features and HOG is 68.7%. The accuracy of geometric features and HOG is slightly higher than geometric features alone, and the accuracy of this study's geometric model is higher than that of Yaman et al. [9], who also used geometric features (accuracy: 65%). Additionally, the present study automatically detected viewpoints, whereas [9] marked the points manually. Furthermore, the images in this dataset are unconstrained, whereas [9] used a controlled profile dataset.

5.1.3. Results of Kinship Verification

The results were applied to the geometric model for kinship verification from ear images. Little research has been conducted into kinship verification from facial images, and this is the first study of kinship verification from ear images. The experiments for kinship verification were performed on the database by using a kNN classifier with k = 1 using a leave-one-out cross validation, all the children's ear images for testing, and all the parents' ear images for training. The distance was measured using both Euclidian distance and Manhattan distance. The SOTEAR dataset [21] used for the kinship verification was collected by us and contains 134 images from 21 families (21 parents) and 25 children (16 sons and nine daughters). This is the first ear dataset contains kinship information. Each subject has a left and right image, and the ear images in the dataset are unconstrained. Table 5. 5 and Table 5. 6 depict the results of kinship recognition from ear images.

Table 5. 5 Results of kinship recognition (Euclidian distance)

Euclidian distance	Father	Mother
Son	40.6%	31.3%
Daughter	0%	33.3%
Son and daughter	20.0%	28.2%

Table 5. 6 Results of kinship recognition (Manhattan distance)

Manhattan distance	Father	Mother
Son	31.3%	25.0%
Daughter	0%	33.3%
Son and daughter	28.0%	28.0%

As observed in Table 5. 5 and Table 5. 6, the kinship classification accuracy between father and son is 40.6%, while the accuracy between mother and daughter is 33.3%. However, the correct kinship verification between father and daughter is zero. The results of the kinship verification reveal that a son is more similar to his father, whereas a daughter is more similar to her mother; therefore, kinship verification may be influenced by gender. In Table 5. 5 and Table 5. 6, the classification accuracies are determined using only geometric features.

Although the accuracy result for kinship verification is not very high, it is much better than the random rate of 4.7%. There are 21 families, so the random rate is 1/21 (4.7%). Thus, performing kinship verification from ear images is feasible. It is anticipated that this model for kinship recognition will be improved in the future. As more family ear data is required for analysis, the expansion of the dataset is ongoing.

5.1.4. Analysis of the Model

Figure 5. 1 shows the cumulative match characteristic (CMC) curve of the results. The horizontal axis indicates that the correct identity was among the top n (number of rank) results. The basic capability of the technique is shown by the result from the USTB dataset I (d1), which reaches 95.6% for the basic model-based approach. The inclusion of an affine transformation improves capability, but the recognition reduces to 80.0% (d2). Where an affine transformation is not included, the recognition drops to 70.2%, which indicates the successful operation of the affine transformation. At the present, all the ear recognition and gender classification results achieve 100% by rank 2, although at rank 1, identity and gender start at 80.0% and 68.7%, respectively. However, the father–son and mother–daughter kinship results start at 40.6% and 33.3%, respectively, and achieve 100% by rank 20. Clearly, these results are well above random performance and warrant further study.

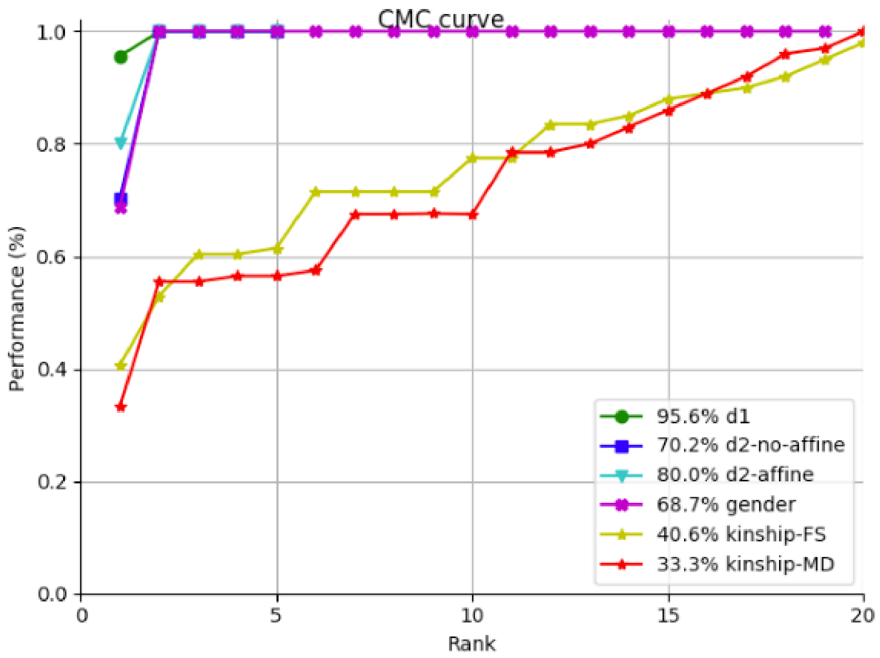
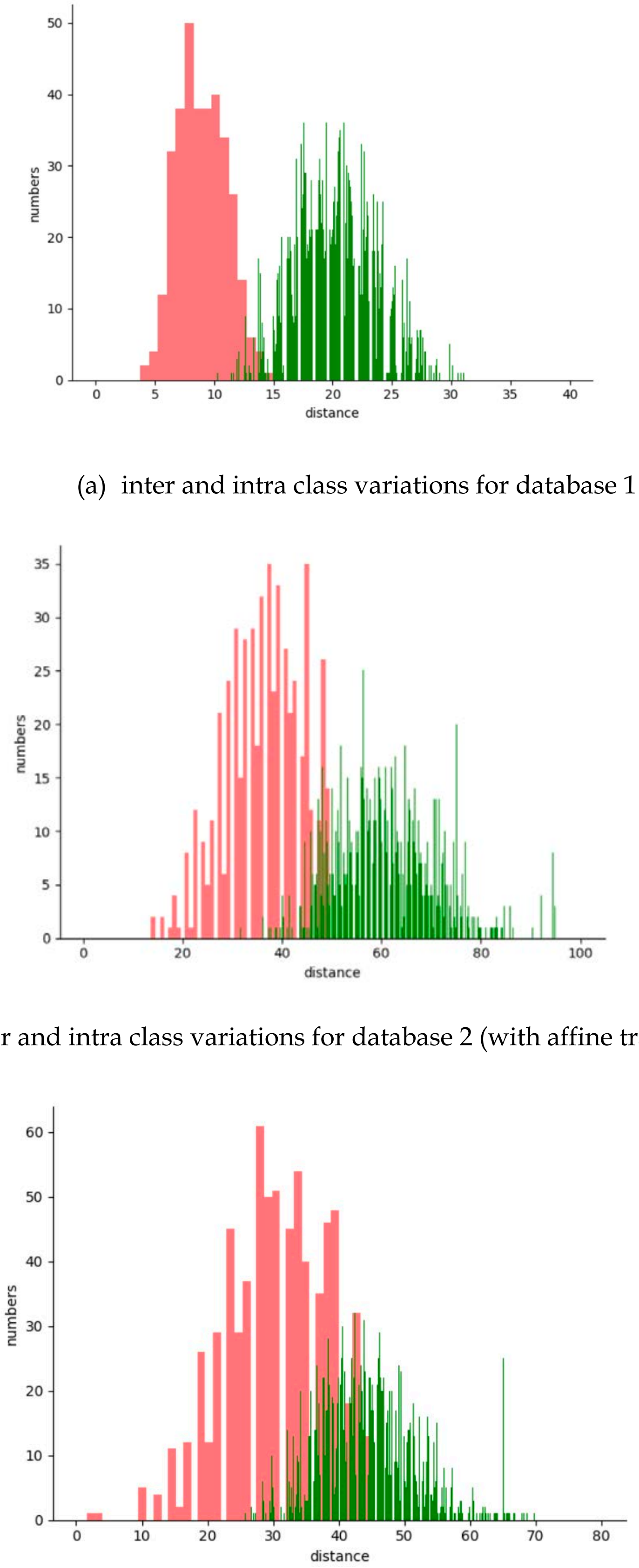


Figure 5. 1 CMC curve of results

Note: FS = father–son and MD = mother–daughter

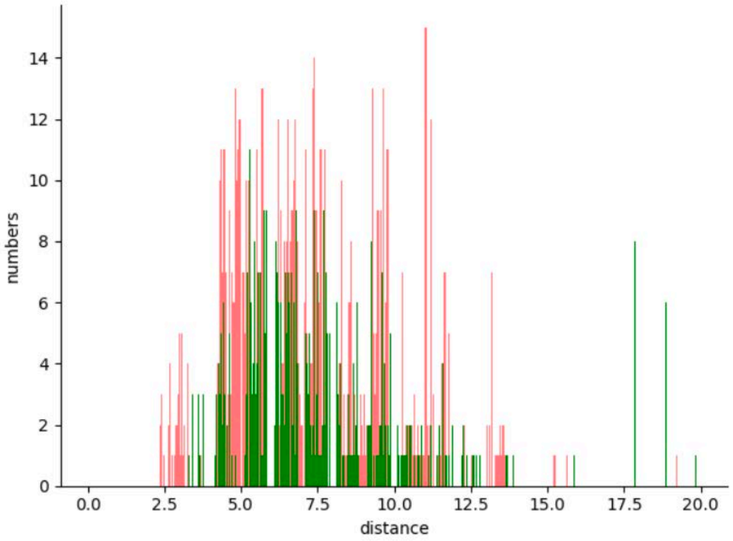
Figure 5. 2 illustrates the recognition capability and the inter-/intra-class variations for ear recognition and gender classification. The red bars show the distance between the same subject and client, while the green bars indicate the distance between different subjects and impostors. Figure 5. 2(a) shows the recognition capability of USTB dataset I; here, the presence of a small red bar, a large green bar and a small confusion region indicate that the base model for database 1 has a high recognition capability. Figure 5. 2(b) demonstrates the recognition capability of USTB dataset II with affine transformation. In this case, the confusion region is not large, which indicates the suitability of the base model for database II. However, in Figure 5. 2(c), the recognition capability of database II without affine transformation results in a larger difference between the confusion region than in Figure 5. 2(b). The disparity between the rotation database with and without affine transformation is clear. As affine transformation can compensate for rotation, the model performs better for the affine-transformed ear images. Figure 5. 2(d) shows the recognition capability for gender classification; the combined red and green bars indicate a poor recognition capability for this classification.



(a) inter and intra class variations for database 1

(b) inter and intra class variations for database 2 (with affine transformation)

(c) inter and intra class variations for database 2 (without affine transformation)



(d) inter and intra class variations for gender classification

Figure 5. 2 Histograms of distances

5.2. Model Improvement

As the results of gender classification accuracy based on the geometric model were not high, image pre-processing was considered to remove noise and hair from the ear images. A force field transform was applied for pre-processing, and HOG was used for feature extraction for gender classification from ear images.

First, the improved ear recognition model was used to evaluate the performance of the geometric model. A kNN classifier (k = 1) with a leave-one-out cross-validation strategy was used for ear identification. The Euclidian and Manhattan distances were used in a kNN classifier to evaluate the model. Pre-processing for images was applied, and the model was evaluated using USTB dataset I. The results for correct recognition are presented in Table 5. 7.

Table 5. 7 Ear recognition accuracies with image pre-processing

Method	Accuracy
Euclidian distance	94.5%
Manhattan distance	95.1%

Table 5. 7 shows that the model-based method achieves an accuracy of 95.1% for ear recognition, which confirms its good performance and its suitability for ear identification purposes.

However, as can be seen in Table 5. 4, the results were not very high; therefore, image pre-processing was applied before using HOG to extract the features for this dataset, and additional data was collected for inclusion in the dataset. These approaches were used for gender classification using the present study's own dataset of ear images; this contained 78 subjects (38 females and 40 males) with Euclidian, Manhattan and Mahalanobis distances for a kNN classifier with a value of k = 1 in a leave-one-out cross-validation strategy. Table 5. 8 presents the accuracies of the gender classification.

Table 5. 8 Gender classification accuracies for the model-based method with pre-processing

Method	Accuracy
Euclidian distance	72.9%
Manhattan distance	80.0%
Mahalanobis distance	82.9%

As Table 5. 8 demonstrates, this study's model-based technique achieved a gender classification accuracy of 82.9%, which is superior to the results given in Table 5. 4. The model has better performance with the force field transformed ear images. The following section analyses the different ear regions that contribute to gender classification.

5.3. Ear Symmetry

This subsection uses the model-based technique, which is based on ear structure, for matching a pair of ears. One of the SC face datasets for ear symmetry was used, which included 100 subjects, each subject contains three left side and three right side ear images; 300 left ear images for training and 300 right ear images for testing. As both ears were used, the right ear images were mirrored. Then, the kNN classifier was used with Euclidian and Manhattan distances for ear classification.

The quality of surveillance images is not good, thus, we add noise for ear images to evaluate the performance of the method. Zero mean Gaussian noise was added to the ear images to create noisy ear images with SNR (Signal-to-noise ratio) 22 dB (Figure 5. 3 shows the samples of noisy images).

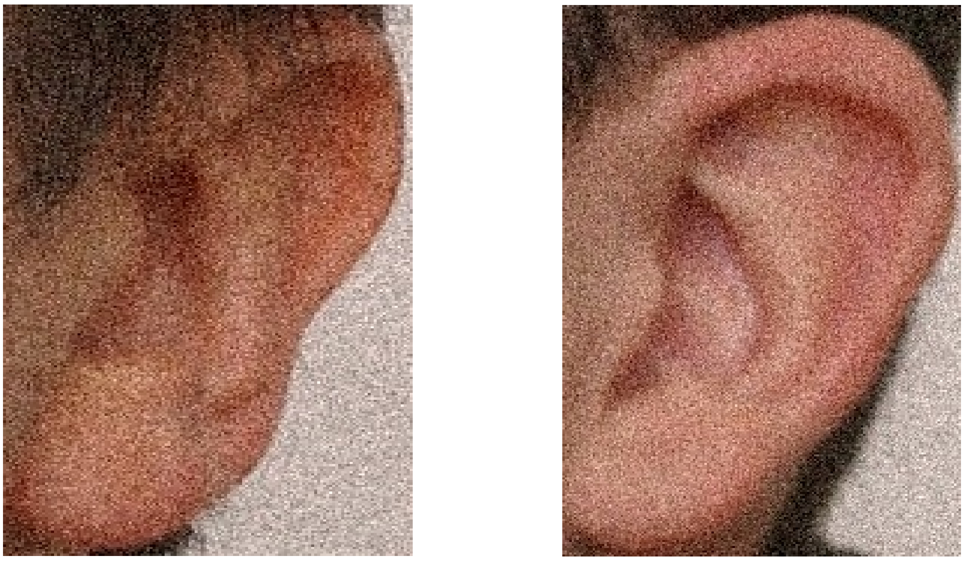


Figure 5. 3 Samples of noisy images

Table 5. 9 Ear symmetry accuracies for the model-based method (original images)

Distance	Accuracy
Manhattan distance	58.9%
Euclidian distance	60.7%

Table 5. 10 Ear symmetry accuracies for the model-based method (images with noise)

Distance	Accuracy
Manhattan distance	21.3%
Euclidian distance	23.1%

Table 5. 11 Ear symmetry accuracies for the model-based method (images (with no noise) after affine transformation)

Distance	Accuracy
Manhattan distance	66.9%
Euclidian distance	65.0%

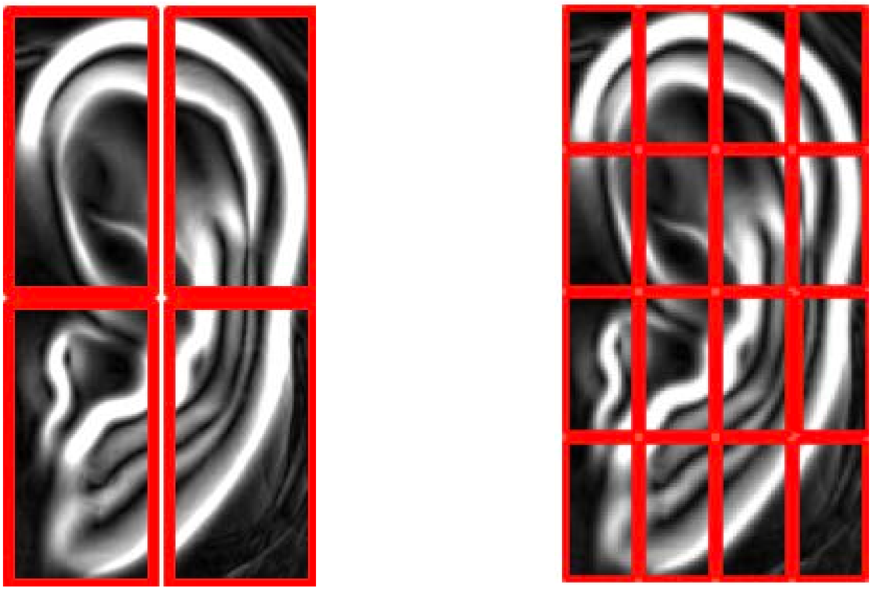
Table 5. 9, Table 5. 10 and Table 5. 11 present the results of this study's model-based technique and include the accuracies for the original images, noisy images and original images after affine transformation, respectively. The affine-transformed ear images demonstrated the highest accuracy of 66.9%. The original images produced a recognition rate of 60.7%, while the accuracy of the images with Gaussian noise was 23.1%, representing a reduction of 37.6%. As Figure 5. 3 shown, the outer ears' contours and structures of noisy images are not clear as the original ear images, thus, the model cannot

extract the ear features, such as triangular fossa, very clearly. Therefore, the model has poor robustness to noise.

5.4. Analysis of Different Regions for Ear Biometrics

The last sub-section confirmed the model's good performance for ear identification and gender classification. This sub-section investigates the contributions from different parts of the ears.

The ear images were divided into four regions (2×2) and 16 regions (4×4), as shown in Figure 5. 4. The sub-regions are referred to as $r_n(\text{eg.}\, r_1, r_2 \cdots)$, and their order is given in Figure 5. 5. We divide the ear images to four and 16 regions as shown in Figure 5. 4, and then calculate the HOG of each sub-regions separately.



Divided (2×2) regions

Divided (4×4) regions

Figure 5. 4 An example of a divided ear image

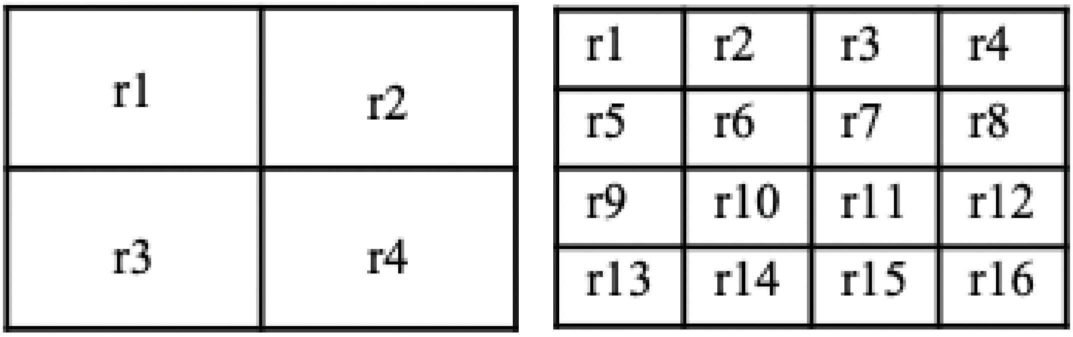


Figure 5. 5 The order of regions (left side: four regions, right side: 16 regions)

Each part of the ear was used separately for gender classification. Table 5. 12 shows the accuracies for four (2×2) parts of the ears, and Table 5. 13 presents the accuracies for 16 (4×4) ear parts.

Table 5. 12 and Table 5. 13 contain the accuracies of gender classification for each individual region, and the accuracies of the (2×2) sub-regions and (4×4) sub-regions are shown separately. r_n in Table 5. 12 and Table 5. 13 corresponds to r_n in Figure 5. 5.

Table 5. 12 Gender classification accuracies of different ear regions (four parts) using the model-based method

Ear region	k = 1	k = 3	k = 5	k = 7	k = 9
r1	62.0%	70.4%	**73.2%**	70.4%	67.6%
r2	53.5%	57.7%	60.5%	60.5%	57.7%
r3	47.8%	43.6%	33.8%	36.6%	35.2%
r4	52.1%	63.3%	60.5%	61.9%	60.5%

Table 5. 13 Gender classification accuracies of different ear regions (16 parts) using the model-based method

Ear region	k = 1	k = 3	k = 5	k = 7	k = 9
r1	63.4%	60.6%	52.1%	49.3%	50.7%
r2	53.5%	59.2%	59.2%	66.2%	69.0%
r3	35.2%	40.8%	39.4%	43.7%	42.3%
r4	**71.8%**	69.0%	64.8%	66.2%	66.2%
r5	61.9%	57.7%	53.5%	52.1%	53.5%
r6	42.3%	40.8%	49.3%	52.1%	47.9%
r7	50.7%	60.6%	56.3%	53.5%	49.3%

r8	50.7%	50.7%	50.7%	47.9%	47.9%
r9	42.3%	47.9%	56.3%	50.7%	49.3%
r10	57.7%	52.1%	49.3%	43.7%	42.3%
r11	67.6%	57.7%	59.2%	64.8%	62.0%
r12	52.1%	59.2%	60.6%	60.6%	63.4%
r13	43.7%	46.5%	45.1%	42.3%	38.0%
r14	50.7%	52.1%	50.7%	42.3%	45.1%
r15	49.3%	47.9%	49.3%	53.5%	59.2%
r16	52.1%	53.5%	52.1%	52.1%	56.3%

As shown in Table 5. 12 the r1 region (in a four-region setting) has the highest accuracy of 73.2%, which indicates it is the most important region for gender classification. The second-most important regions for gender classification are r2 and r4, while r3 is the least important part in this method. The accuracy for r3 is even lower than that of a random classifier. Table 5. 13 shows the classification accuracies for 16 ear regions. The highest accuracy of 71.8% is associated with region r4, while the r2 region has the second highest accuracy with 69.0%.

The results presented in Table 5. 12 and Table 5. 13 indicate that the upper helix is more important than the lower parts of the ears (such as the lower helix and the lobe) for gender classification purposes. Regions r4 and r2 in the 16-region-division setting are separate to r2 and r1 in the four-region-division setting. This indicates that the upper regions of the ears are more important than the lower parts.

After identifying the significant regions of the ear for gender classification (Table 5. 12 and Table 5. 13), the potential for confusion of the important ear parts in terms of gender classification was considered.

The ear images for ear recognition were divided into four and 16 sub-regions (as in Figure 5. 4), and each sub-region was used for ear identification. The results of each sub-region were compared for ear recognition. Table 5. 14 and Table 5. 15 show the accuracies of ear identification for each individual region. In Table 5. 14, the accuracy of r1 (in a four-region setting) reached 69.2%, whereas in Table 5. 15, r2 (in a 16-region setting) predicted correct recognition with 56.6% accuracy. The triangular fossa and the inferior crus of the antihelix are the most significant parts for ear recognition.

Table 5. 14 Ear recognition accuracies of different ear regions (four parts) using the model-based method

Ear region	Euclidian distance	Manhattan distance
r1	67.0%	**69.2%**
r2	53.8%	62.6%
r3	51.6%	51.6%
r4	37.9%	42.3%

Table 5. 15 Ear recognition accuracies of different ear regions (16 parts) using the model-based method

Ear region	Euclidian distance	Manhattan distance
r1	46.2%	46.2%
r2	51.6%	**56.6%**
r3	50.0%	53.8%
r4	29.1%	30.8%
r5	40.1%	40.7%
r6	48.9%	52.2%
r7	42.3%	51.6%
r8	32.4%	43.4%
r9	37.4%	41.7%
r10	41.2%	46.7%
r11	43.4%	47.3%
r12	35.2%	39.0%
r13	30.7%	40.1%
r14	33.6%	37.9%
r15	34.6%	37.9%
r16	24.7%	28.1%

The results presented in Table 5. 12, Table 5. 13, Table 5. 14 and Table 5. 15 were compared. When the ear images were divided into four sub-regions, the significant parts of the ears for gender classification and ear recognition were the same (r1). Therefore, the results of the 16 regions were compared. As the tables reveal, the most significant part for gender classification is the r4 region, whereas the most important part for ear recognition is the r2 region. Subdividing the ears into further sub-regions identifies distinct and significant parts for gender classification and ear recognition purposes. For gender classification, the outer region of the ears (r4) is more important than the other ear features, whereas the triangular fossa (r2) is the most significant part in terms of ear recognition.

In addition, we consider evaluating this model for ear symmetry. We also divided the images of ear symmetry experiment to four and 16 sub-regions (as in Figure 5. 4).Table 5. 16 and Table 5. 17 show the accuracies of ear symmetry for each individual region.

Table 5. 16 Ear symmetry accuracies of different ear regions (four parts) using the model-based method

Ear region	Euclidian distance	Manhattan distance
r1	22.5.%	20.7.%
r2	13.4%	14.4%
r3	8.4%	7.3%
r4	24.9%	**25.7%**

Table 5. 17 Ear symmetry accuracies of different ear regions (16 parts) using the model-based method

Ear region	Euclidian distance	Manhattan distance
r1	15.7%	14.1%
r2	13.9%	12.3%
r3	10.7%	12.3%
r4	9.7%	8.6%
r5	13.1%	13.4%
r6	14.1%	14.1%
r7	12.3%	11.3%
r8	7.6%	7.6%
r9	14.7%	13.4%
r10	11.8%	11.8%
r11	20.7%	**23.1%**
r12	8.6%	8.1%
r13	16.8%	17.8%
r14	11.5%	14.1%
r15	9.2%	8.9%
r16	7.1%	6.3%

In Table 5. 16, the accuracy of r4 (in a four-region setting) reached 25.7%, whereas in Table 5. 17, r11 (in a 16-region setting) predicted correct recognition with 23.1% accuracy. As

Figure 5. 4 and Figure 5. 5 shown, the r11 in 16 regions is a part of r4 in four regions, the anti-tragus is the most significant part for ear symmetry. However, the significant part of ear recognition is r1 in four regions, which is the triangular fossa; whereas, the accuracy for ear symmetry of r1 (triangular fossa region) is 22.5%, which is also important for ear symmetry.

5.5. Conclusions

This chapter presented the results of the model-based approaches for ear recognition, gender classification, kinship verification and ear symmetry. The geometric model has good performance for ear recognition but performs less well for gender classification. Although its accuracy is less than that of appearance-based results, it performs well among the model-based techniques, and it can handle affine transformations. The affine transform was used successfully for viewpoint correction. However, this study's gender classification system could benefit from the inclusion of additional features. Therefore, the model was improved using image pre-processing, which performs well in terms of gender classification. Then, the geometric model was applied for kinship verification. It is noted that in the kinship verification results, the ears of the mother appear to be more influential. The accuracy of the kinship verification indicates its suitability for ear biometrics. It is anticipated that more features will be added to improve the accuracy of the classification and provide a deeper understanding of how these results are related to basic ear structures. Additionally, a large family dataset is required for kinship verification.

Finally, this chapter presented a model-based technique for ear recognition with bilateral symmetry that produced an accuracy of 66.9%. This analysis shows that there is an implied similarity between the left and right ears, although the accuracy is not high enough. Therefore, the next chapter considers the use of deep learning for ear biometrics along with the contributions of different ear regions.

Chapter 6

Results and Analysis of the Deep Learning Approaches for Ear Biometrics

The development of the convolutional neural network (CNN) is receiving increasing attention. The increasingly good performance of convolutional neural networks (CNNs) suggests it is propitious to investigate their recognition performance further, here for specificity and for gender and symmetry. Recently, deep learning has been used for ear biometrics due to its strong performance, although it can be limited by a lack of training data. Transfer learning is an approach that requires a small amount of data. Therefore, we applied transfer learning to VGG-16 [101], GoogleNet [103] and ResNet-50 [104] networks to analyse human identification, gender classification and bilateral symmetry in ear images. Furthermore, there is no research on why the network can make a classification decision for ear biometrics. Selvaraju et al. [106] proposed a technique to visually explain the CNN-based models, which is called gradient-weighted class activation mapping (Grad-CAM). In this chapter, in addition to presenting the results of ear recognition, gender classification and ear symmetry based on deep learning approaches, we will apply Grad-CAM to our fine-tuned networks to analyse ear biometrics.

6.1. Ear Recognition

In this sub-section, we will deploy transfer learning of VGG-16, GoogleNet and ResNet-50 for ear recognition and the Equal Error Rate (EER) to evaluate models. Furthermore, we will consider the contributions of different regions in ears and use Grad-CAM to estimate the significance of each part.

6.1.1. Results and Analysis of Ear Recognition

USTB Database 1 [109] (180 images and 60 subjects with three images for each subject) and Database 2 [109] (308 images and 77 subjects with four images for each subject) are used for ear recognition. There are 351 images for training and 137 images for testing. We randomly selected one test ear image from each subject (137 test images from 137 subjects) and split the remaining images into training and validation datasets as follows: 70% data for training and 30% data for validation. We fine-tuned VGG-16, GoogleNet and ResNet-50 networks for ear recognition. Table 6. 1 shows the results.

Table 6. 1 Ear recognition Accuracies Based on Transfer Learning

Method	**Accuracy**
VGG-16	81.5%
GoogleNet	82.5%
ResNet-50	**92.9%**

As Table 6. 1 shows, the accuracies of ear recognition based on VGG-16 and GoogleNet networks are 81.5% and 82.5%, respectively, whereas the rate of correct ear recognition based on ResNet-50 is 92.9%. These mean that the fine-tuned ResNet-50 model has the best capacity for ear recognition. In previous studies [53][59][60], the highest accuracy was 100%. [60] used 250 training images for ten subjects. There is a small amount of data in this dataset, and it contains just ten subjects. However, when our model is trained on the rotated ear image datasets, we train 348 images for 137 subjects. Additionally, the datasets used here include rotated ear images, and the ear images are under weak illuminations. Our model appears to be immune to rotation and illumination.

We considered why these models can recognise humans, and which parts are important for identification by the models. Therefore, we employed Grad-CAM to analyse the parts of the ear that are significant for recognition. Figure 6. 1, Figure 6. 2 and Figure 6. 3 show the heatmaps for ear recognition. The red parts in the heatmaps are the most important parts, whereas the blue parts are less important.

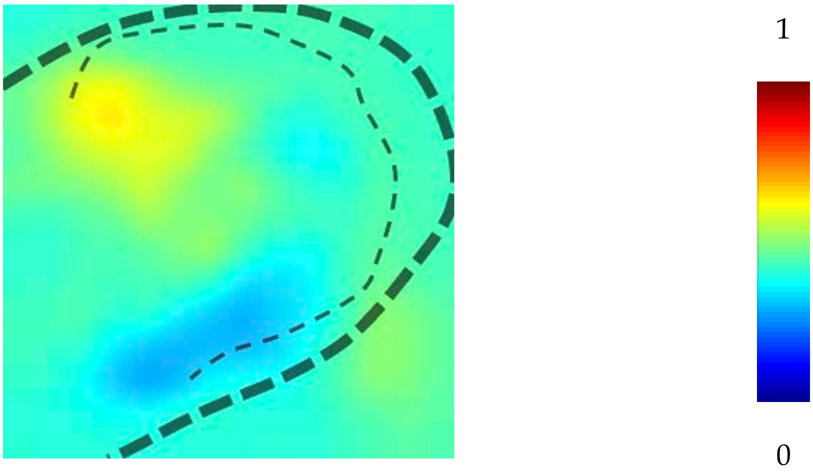


Figure 6. 1 Average heatmap for ear identification based on VGG-16

As Figure 6. 1 shows, the triangular fossa region of the ear is more significant for ear recognition based on VGG-16, whereas the antitragus is less important.

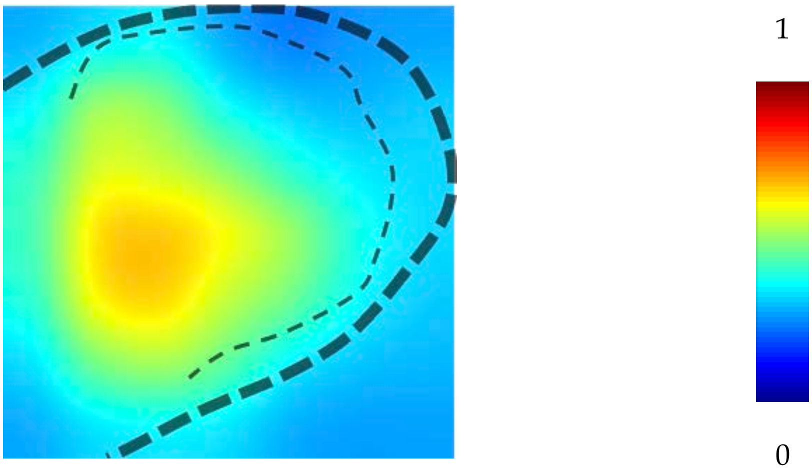


Figure 6. 2 Average heatmap for ear identification based on GoogleNet

Figure 6. 2 depicts the average heatmaps for ear identification computed by the GoogleNet network. The central parts of the ear are more significant for ear identification, and the helix is less important for ear recognition.

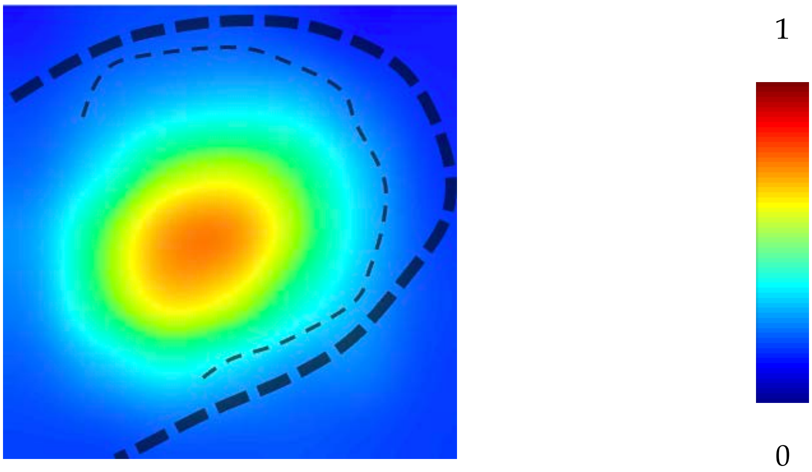


Figure 6. 3 Average heatmap for ear identification based on ResNet-50

As Figure 6. 3 shows, the central parts of the ear are more important, and the periotic regions are less important for ear recognition based on ResNet-50. When comparing the heatmaps based on these networks, the different regions of the ear show the different contributions. The divergence of the ear regions' contributions based on VGG-16 is slight, and the difference between the different regions based on GoogleNet is significant. Moreover, the contributions of the ear regions have a more significant difference based on ResNet-50. Therefore, the accuracy of GoogleNet is higher than that of VGG-16, and ResNet-50's accuracy is the highest. Furthermore, we considered which parts are more important for ear recognition out of the three networks. Figure 6. 4 shows the average heatmap of three different networks and highlights that the central parts are the most important for ear recognition out of the three networks.

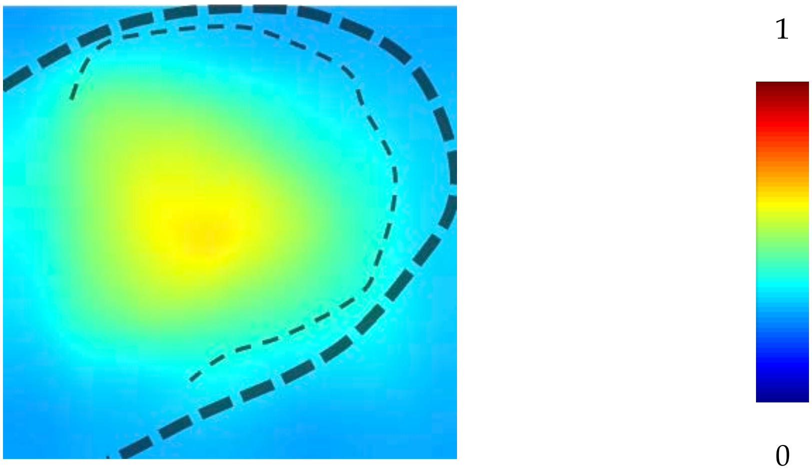


Figure 6. 4 Average heatmap for ear identification of the three networks

Comparing with the model technique (shown in Table 5. 14 and Table 5. 15), the important parts of model-based is triangular fossa, which is in the green/ yellow part of Figure 6. 4, thus, the important parts of model technique and deep learning are similar for ear recognition.

6.1.2. Ear Recognition Verification

We conducted verification experiments to evaluate the performance of our model. For the verification experiments, we randomly selected two ear images from each subject. We used the training model to represent each image using a vector and then calculated the cosine similarities as a distance metric between the two images.

If the stored ear image of subject A is represented by A_1 and another ear image for verification is represented by A_2, then the null hypothesis (H_0) and alternate hypothesis (H_1) are defined as follows:

Null hypothesis $\boldsymbol{H_0}$: A_1 and A_2 come from different subjects.

Alternate hypothesis $\boldsymbol{H_1}$: A_1 and A_2 come from the same subject.

If the value of cosine similarity is higher, then the system is more certain that $\boldsymbol{H_1}$ is correct. The system decides based on the following threshold: if the cosine similarity value is higher than or equal to the threshold, then $\boldsymbol{H_1}$ is correct. Otherwise, the cosine similarity value is lower than the threshold, and therefore, $\boldsymbol{H_0}$ is confirmed.

The false reject rate (FRR) is defined as rejecting the null hypothesis H_0 when it is true, and the false accept rate (FAR) is defined as failing to reject H_0 when it is false. The EER is the value defined as FRR=FAR. EER is an indicator of the performance of the model: the lower the value for EER, the higher the performance of the model.

Figure 6. 5, Figure 6. 6 and Figure 6. 7 show the recognition verification of ear images based on VGG-16, GoogleNet and ResNet-50. The values of the EER for ear recognition verification are presented in Table 6. 2.

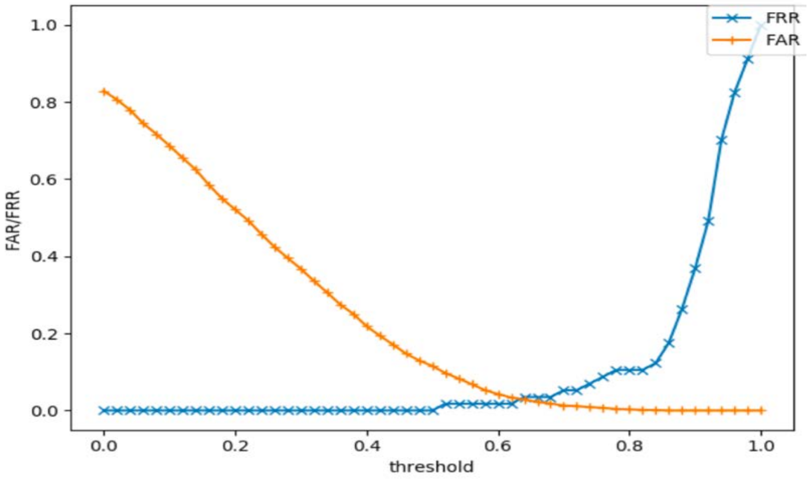


Figure 6. 5 Ear recognition verification of ear images (VGG-16)

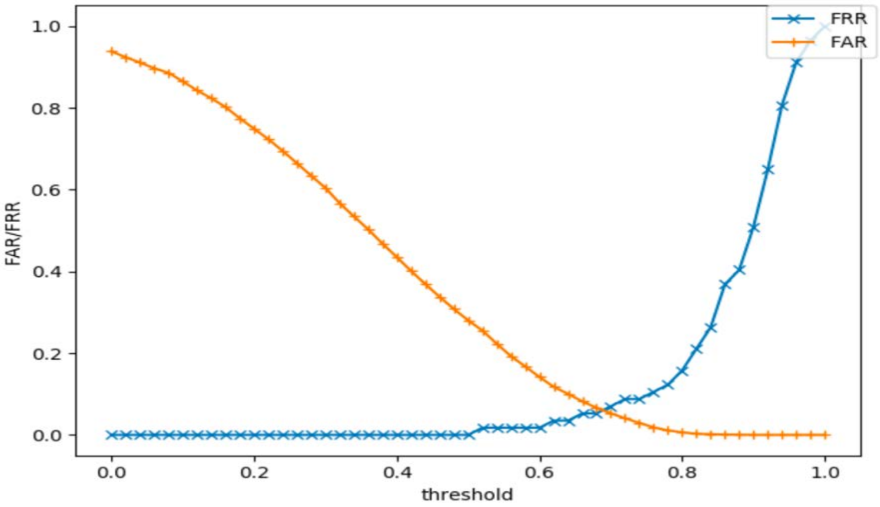


Figure 6. 6 Ear recognition verification of ear images (GoogleNet)

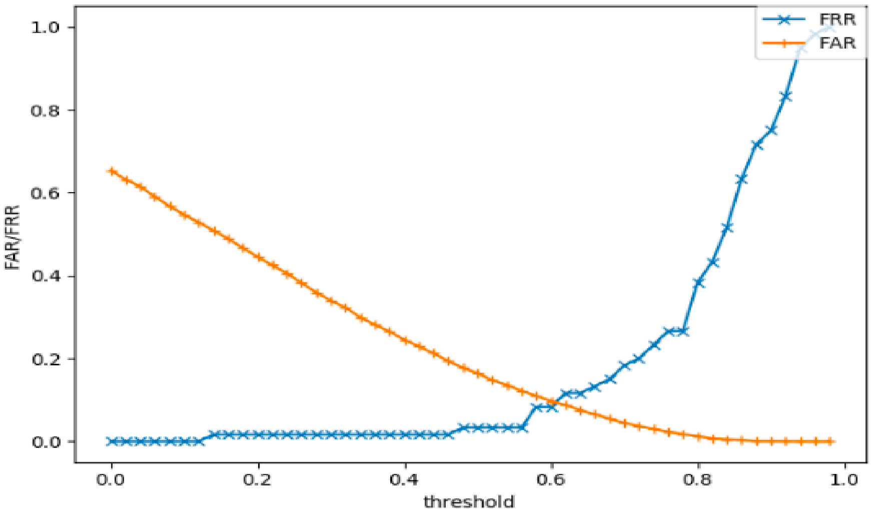


Figure 6. 7 Ear recognition verification of ear images (ResNet-50)

Table 6. 2 tabulates the EER for different networks. The EER for VGG-16, GoogleNet and ResNet-50 are 8.77%, 8.22% and 7.24%, respectively. The values of EER prove that the model based on the ResNet-50 network has the highest performance of these three networks.

Table 6. 2 Equal Error Rate of Different Networks (Ear Recognition)

Method	**EER**
VGG-16	8.77%
GoogleNet	8.22%
ResNet-50	**7.24%**

6.2. Gender Classification

As these fine-tuned networks perform well for ear recognition, we applied transfer learning to these three networks to conduct gender classification using the ear images. In addition, we used Grad-CAM to analyse the models. In this sub-section, we will present the results of gender classification from the ear images based on the three models and the analysis of these models.

6.2.1. Results and Analysis of Gender Classification

We used the USTB dataset III [109], SOTEAR dataset [21] and AWE dataset [110]. We selected several images from each dataset to create a merged dataset of ears, which would ensure a more balanced dataset between males and females. As a result, in our newly merged dataset, there are 90 female subjects and 100 male subjects. There are 800 images for training and validation (split as 70% data for training and 30% data for validation) and 110 images for testing. Additionally, we removed the background from each ear image to help us to focus on the ear and exclude the effects of the hair, head and neck regions. Table 6. 3 shows the results of gender classification from the ear images.

Table 6. 3 Gender Classification Accuracies for Deep Learning

Method	Accuracy
VGG-16	80.9%
GoogleNet	81.8%
ResNet-50	**90.9%**

As tabulated in Table 6. 3, the accuracies of fine-tuned VGG-16 and GoogleNet for gender classification are 80.9% and 81.8%, and ResNet-50 has the highest accuracy of the network based on transfer learning at 90.9%. As this accuracy is sufficiently high, it is possible to use this model to classify subjects into females and males. Although Yamen et al. [10] stated that the highest accuracy of gender classification is >99%, their experiments used controlled ear images, whereas our experiments for gender classification use datasets with unconstrained ear images. As ResNet-50 has a better performance than VGG-16 and GoogleNet, it is unclear why ResNet-50 has a higher accuracy than VGG-16 and GoogleNet. We used Grad-CAM to compute the heatmaps based on these three models shown in Figure 6. 8, Figure 6. 9 and Figure 6. 10 to determine which parts of the ears contribute to gender classification from ear images.

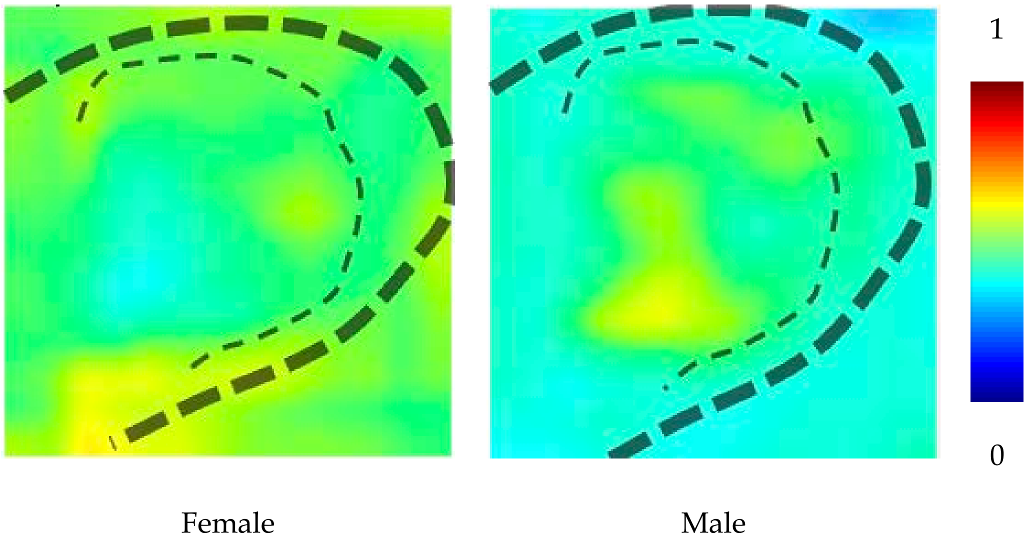
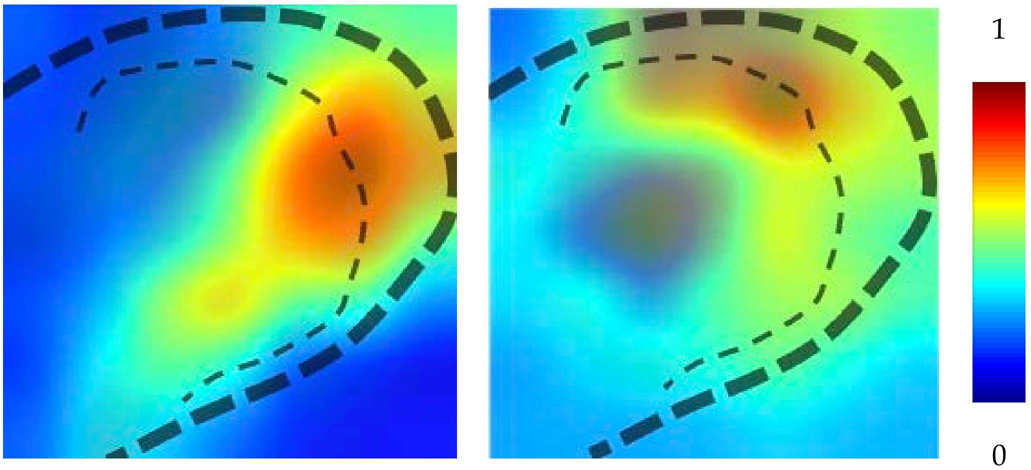


Figure 6. 8 Average heatmaps for different genders (VGG-16)

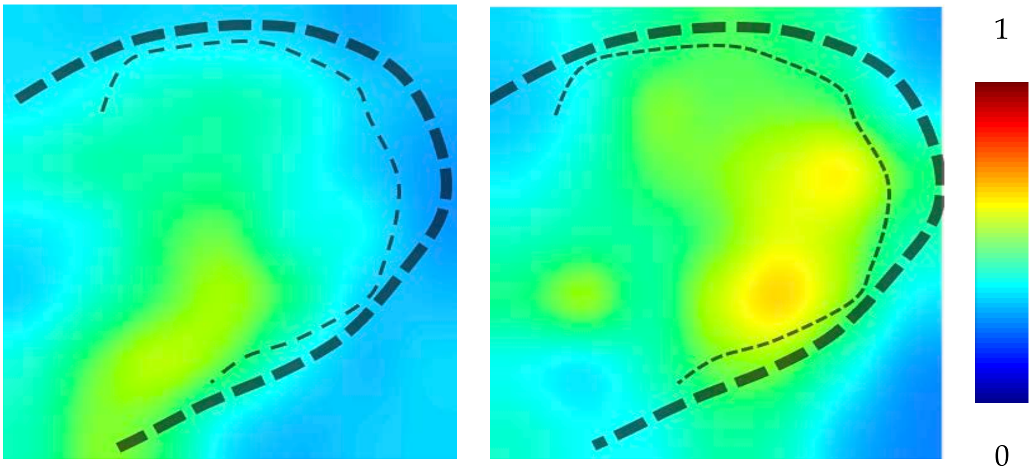
Figure 6. 8 shows the average heatmaps for different genders produced by the VGG-16 network. The middle parts, such as the tragus and antitragus, are more important for males, whereas the lower parts, especially the lobe, are more important for females. Compared to Figure 6. 1, the heatmaps of different genders differ to the heatmaps of ear recognition, as although the upper regions are more important for ear recognition, the lower regions are more significant for females and the middle parts are more important for males.



Female Male

Figure 6. 9 Average heatmaps for different genders (GoogleNet)

Figure 6. 9 depicts the average heatmaps for different genders, as computed by the GoogleNet network. The right-middle parts are more important for females, such as the middle helix, and the upper-middle parts are more significant for males, such as the upper helix and triangular fossa. Comparing Figure 6. 9 and Figure 6. 2 highlights that the heatmap of ear recognition based on GoogleNet shows that the central parts are very important, whereas the heatmap of females shows that the middle helixes are more significant, and the heatmap of males shows that the upper helixes are more significant.



Female Male

Figure 6. 10 Average heatmaps for different genders (ResNet-50)

Figure 6. 10 shows the average heatmaps for males and females, including a general outline of an ear to illustrate the important parts of the ear for gender classification. As depicted in Figure 6. 10, there is a clear difference between the genders; for males, the upper helix, triangular fossa, tragus and antitragus are more important than the other parts of the ear, whereas for females, the lobe plays a more significant role in ear recognition. However, Figure 6. 3 shows the heatmap of ear recognition based on ResNet-50, which shows that the central parts are more important for ear recognition.

Heatmaps show the different contributions of different parts. However, the average heatmaps of females and males based on VGG-16 (Figure 6. 8) and GoogleNet (Figure 6. 9) have overlapped regions, and the heatmaps of different genders based on ResNet-50 (Figure 6. 10) are more separated. Therefore, ResNet-50 performs better than VGG-16 and GoogleNet.

We considered which parts are different for females and males over the three networks. Figure 6. 11 shows the average heatmaps of three different networks.

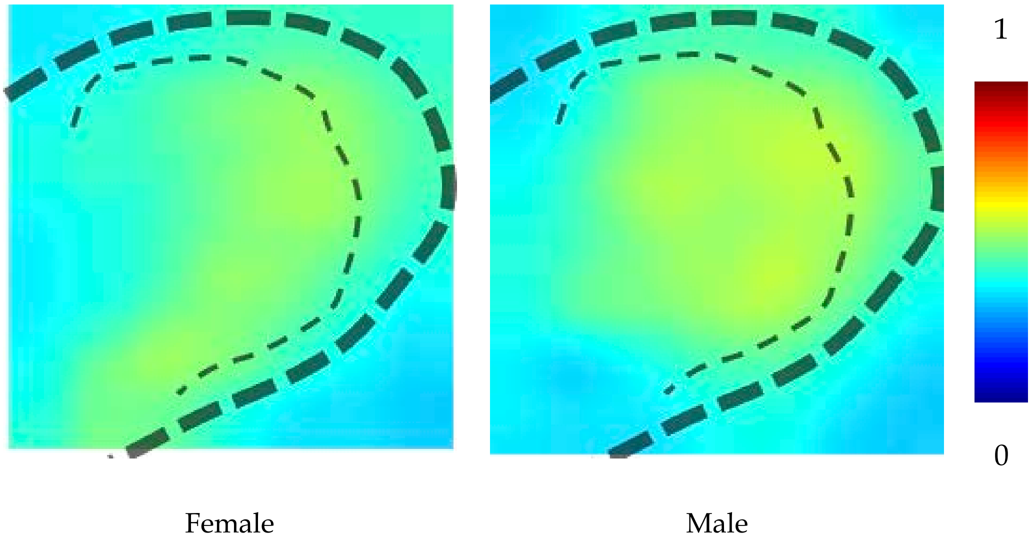


Figure 6. 11 Average gender heatmaps of three networks

As Figure 6. 11 shows, the lobe is the most significant difference between females and males based on the three different networks. This may be because women are more likely to wear earrings, which makes the ear lobe more important for female recognition. Moreover, females are more likely than males to have long hair, which increases the chance that the top of the ear is occluded, and therefore, not used in their recognition. Conversely, males are more likely to have short hair, and their recognition is based on the parts that are more easily observed, including the antitragus and antihelix. Some of the conclusions drawn for the heatmaps depicted in Figure 6. 8 differ from those obtained in our model-based technique (see Table 5. 12 and Table 5. 13). For example, the upper helix is important for gender classification, which is shown not only in our model-based technique but also by the difference between the heatmaps in Figure 6. 11. However, although the lobe did not play an important role in the model-based technique for gender classification, it is important for the convolutional network to classify genders.

It is interesting to compare identification with gender recognition. Figure 6. 4 and Figure 6. 11 show the average heatmap for ear identification and gender classification over three networks. The heatmaps for identification and gender classification are different, and some regions, such as the helix, strike a balance between a strong positive response for males and a strong negative response for females. As the heatmaps of different genders demonstrate, the helix and lobe are important for gender classification, whereas the central parts are more important for ear recognition.

6.2.2. Gender Classification Verification

Verification experiments of gender classification have been used to evaluate the performance of our models for gender classification.

If the stored ear image of a female subject A is represented by A_1, then the null hypothesis (H_0) and alternate hypothesis (H_1) are defined as follows:

Null hypothesis $\boldsymbol{H_0}$: A_1 is male.

Alternate hypothesis $\boldsymbol{H_1}$: A_1 is female.

If the value of predicted probability of the image having the label is higher, then the system is more certain that $\boldsymbol{H_1}$ is correct. The system decides based on the following threshold (T) : if the probability is higher than or equal to the threshold (T), then $\boldsymbol{H_1}$ is correct. Otherwise, the probability is lower than the threshold (T), and therefore, $\boldsymbol{H_0}$ is confirmed.

The FRR is defined as rejecting the null hypothesis $\boldsymbol{H_0}$ when it is true, and the FAR is defined as failing to reject $\boldsymbol{H_0}$ when it is false. The EER is the value defined as FRR=FAR. EER is an indicator of the performance of the model: the lower the value for EER, the higher the performance of the model.

Figure 6. 12, Figure 6. 13 and Figure 6. 14 show the gender classification verification of ear images. The values of the EER for gender classification verification are presented in Table 6. 4.

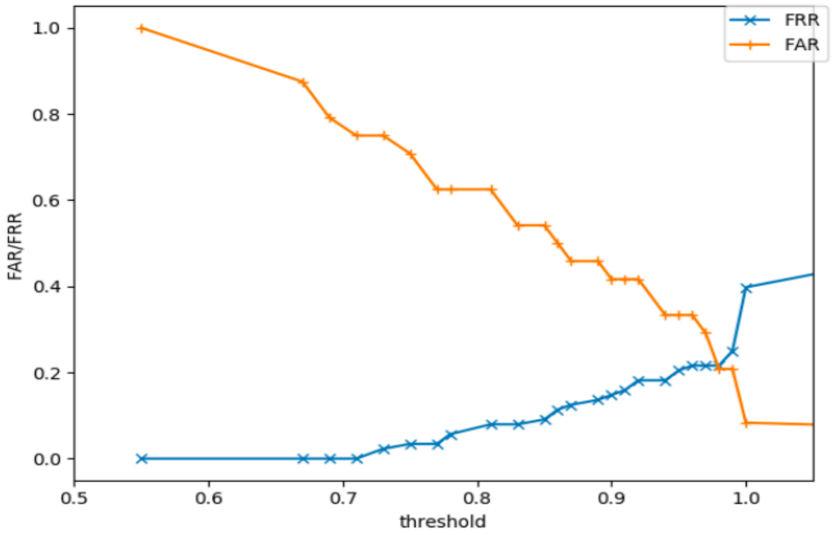


Figure 6. 12 Gender classification verification of ear images (VGG-16)

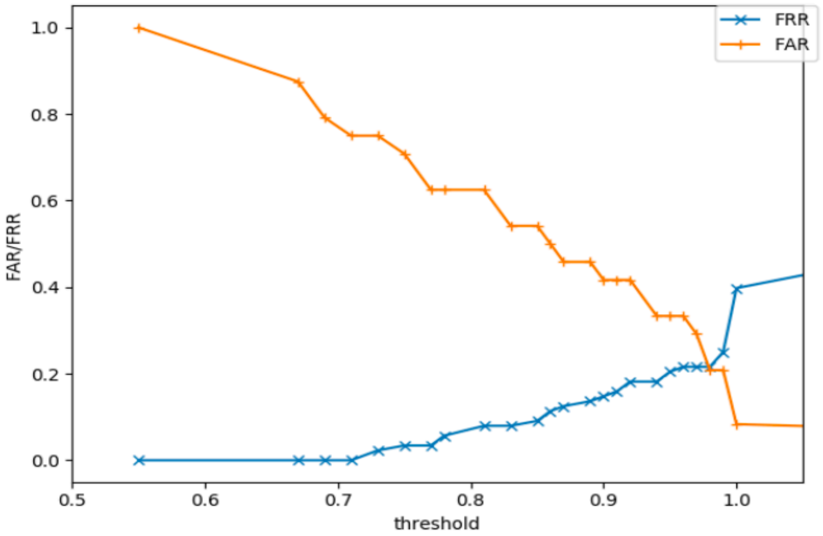


Figure 6. 13 Gender classification verification of ear images (GoogleNet)

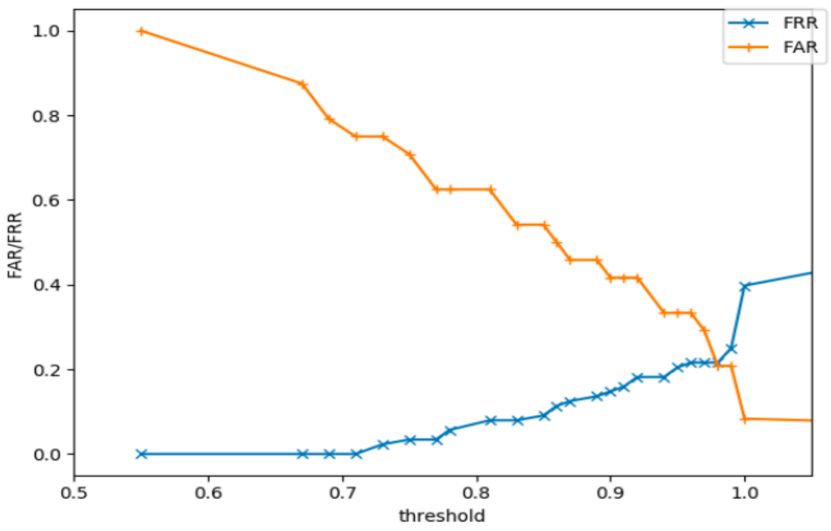


Figure 6. 14 Gender classification verification of ear images (ResNet-50)

Table 6. 4 tabulates the EER for different networks. The EER for VGG-16, GoogleNet and ResNet-50 is 21.59%, 28.68% and 13.62%, respectively. The values of the EER prove that the model based on the ResNet-50 network has the highest performance out of the three networks.

Table 6. 4 Equal Error Rate of Different Networks (Gender Classification)

Method	**EER**
VGG-16	21.59%
GoogleNet	28.68%
ResNet-50	**13.62%**

6.3. Ear Bilateral Symmetry

We will present the ear symmetry based on the model-based technique that was used in last chapter, where the highest accuracy was 66.9%. We will use the three fine-tuned networks (VGG-16, GoogleNet, and ResNet-50) for ear symmetry, and the Grad-CAM will be applied to explain which parts of the ear are more important for ear symmetry.

6.3.1. Results and Analysis of Ear Bilateral Symmetry

In the experiment described in this subsection, we transferred the pre-trained VGG-16, GoogleNet and ResNet-50 to the SC face datasets [17]. There are 100 subjects with 300 left ear images and 300 right ear images, and all the left ear images use for training and all the right ear images for testing. As both ears were used, we mirrored the right ear images for testing.

Table 6. 5 Two-Side Ear Recognition Accuracies of Three Networks

Method	**Accuracy**
VGG-16	66.9%
GoogleNet	62.1%
ResNet-50	**91.2%**

As observed in Table 6. 5, the accuracies of the VGG-16, GoogleNet, and ResNet-50 based on transfer learning are 66.9%, 62.1%, 91.2%, respectively. The performance of the ResNet-50 network is better than the other two networks. In our experiments, we noticed that the hair occlusion influences the correct prediction. Therefore, the hair occlusion presents a challenge for verifying ear symmetry.

However, as this dataset has rotated ear images, we considered applying affine transformation for this dataset. Additionally, we added noise to the ear images to evaluate the model. Zero mean Gaussian noise was also added to the ear images to result in noisy ear images with SNR 22 dB.

Table 6. 6 Two-Side Ear Recognition Accuracies for Deep Learning with Pre-processing Ear Images (VGG-16)

Images	**Accuracy**
Images with noise	60.7%
Affine transformed images	73.6%

Table 6. 6 depicts the accuracies of two-side ear recognitions for different types of data based on the VGG-16 network. Compared to the original images, the accuracy of the affine transformed images is 73.6%, which is an increase of 6.7%, whereas the accuracy of the noise image decreased by 6.2%.

Table 6. 7 Two-Side Ear Recognition Accuracies for Deep Learning with Pre-processing Ear Images (GoogleNet)

Images	**Accuracy**
Images with noise	60.4%
Affine transformed images	63.8%

As Table 6. 7 shows, the accuracy of the affine transformed images for two-side ear recognition based on the GoogleNet network is 63.8%, which is an increase of 2.6%, and the accuracy of the noise images is 60.4%, which is a decrease of 1.7%.

Table 6. 8 Two-Side Ear Recognition Accuracies for Deep Learning with Pre-processing Ear Images (ResNet-50)

Images	**Accuracy**
Images with noise	62.9%
Affine transformed images	93.1%

As observed in Table 6. 8, the accuracy of ResNet-50 based on transfer learning is 93.1%. The recognition rate of training original ear images is 91.2%, and the accuracy of training the ear images with Gaussian noise is 62.9%, which decreases by 28.3%. This shows that the presence of noise lowers the recognition rate, and the affine transformed ear images are responsible for the highest accuracy.

Comparing these three networks, the model of transfer learning based on GoogleNet is most robust for noisy ear images, which only decreases by 1.7%. The affine transformation is more effective for the transfer learning model based on the VGG-16 network.

It is interesting to compare deep learning approaches with model-based techniques. As presented in Table 5. 10, Table 5. 11 and Table 6. 8 the deep learning approach (fine-tuned ResNet-50) demonstrates the best performance. The results of training the original images indicate that deep learning is superior to the model-based method. Furthermore, these two methods also use ear images after affine transformation in the training images. As a result, the accuracy of deep learning increases by 1.9%, while that of the model-based method increases by 8%. As demonstrated in these experiments, affine transformation for the model-based method is more beneficial than deep learning. Furthermore, we have contaminated the ear images with zero mean Gaussian noise for training. Additive Gaussian noise causes the accuracies to decline, as the deep learning accuracy is lowered by 28.3%, and the model-based method decreases by 37.6%. Our results indicate that deep learning demonstrates a better performance than using a model in a noisy environment.

In addition, ResNet-50 has the best performance out of these three networks. Therefore, we applied ResNet-50 based on transfer learning to an in-the-wild ear image dataset (AWE dataset). We used 520 left ear images for training and 480 right ear images for testing, and the ear images have a severe rotation in the pitch and yaw axes. Additionally, the ear images in the dataset have severe occlusions. The recognition accuracy for the ear images without pre-processing is 20.2%. Although the accuracy is not high, it is still 19.2% higher than random chance. Compared to the SC face dataset, the accuracy has a sharp decrease of 70.3%; therefore, our results indicate that applying the system to the in-the-wild ear images presents a considerable challenge.

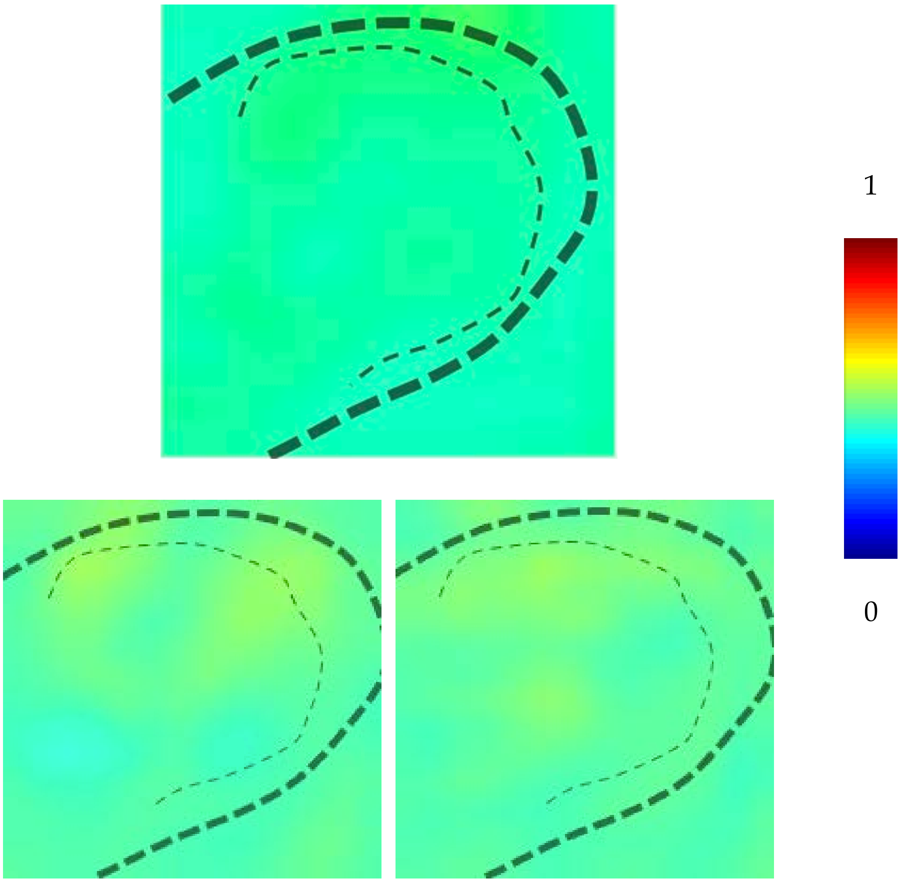


Figure 6. 15 Average heatmap for the recognition of a single ear from either side based on VGG-16 (top: original images; bottom left: affine transformed images; bottom right: images with added Gaussian noise)

We determined which region of the ear contributes the most to symmetry-based recognition rates using Grad-CAM to compute the heatmaps. We mirrored the right ears to the left ones and matched the right ears to the left ones. *Figure 6. 15*, *Figure 6. 16* and *Figure 6. 17* demonstrate how different parts of the ear contribute to ear symmetry based on fine-tuned VGG network, GoogleNet and ResNet.

As Figure 6. 15 shows, the upper regions of the ear images are more significant for ear symmetry based on VGG-16, and the heatmap of the ear images with the applied affine transformation differs to the other two heatmaps based on the VGG-16 network. Figure 6. 1 shows similar important parts (upper regions of ears) to Figure 6. 15 which means that although the heatmap of ear recognition is similar to ear symmetry, the heatmaps of different genders (Figure 6. 8) are different to Figure 6. 1 and Figure 6. 15.

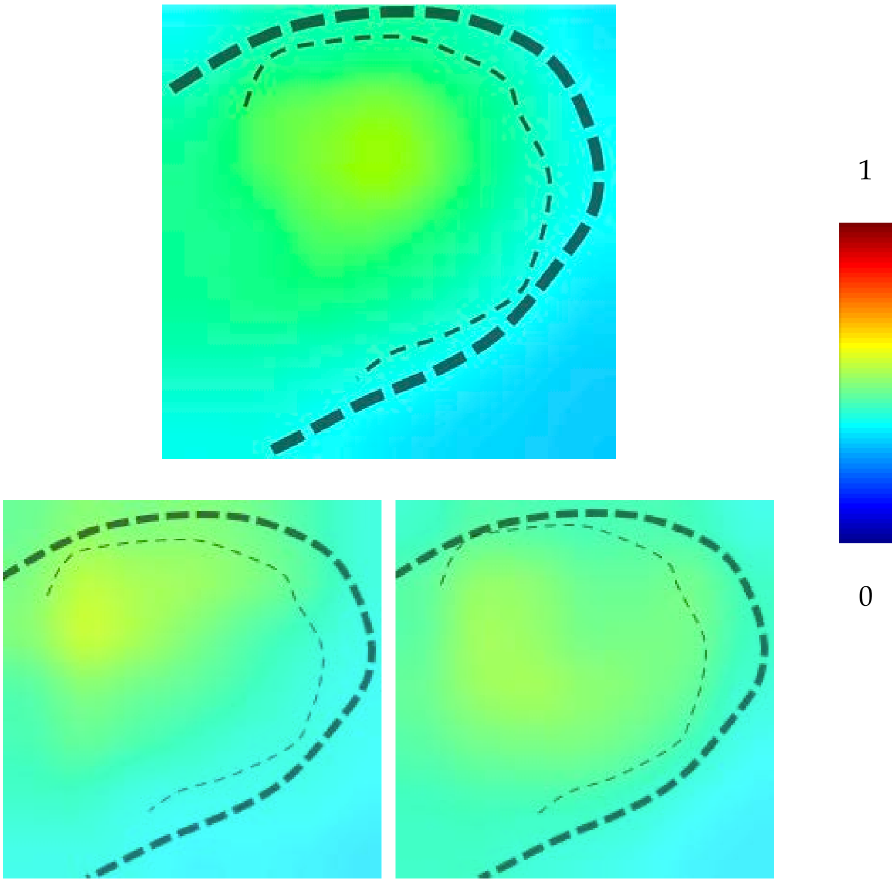


Figure 6. 16 Average heatmap for the recognition of a single ear from either side based on GoogleNet (top: original images; bottom left: affine transformed images; bottom right: images with added Gaussian noise)

As Figure 6. 16 shows, the middle regions of the original ear images are more important for ear symmetry based on GoogleNet, and the left-middle parts of the ear images are more significant in the heatmap of the ear images with applied affine transformation based on the GoogleNet network. The heatmap of ear recognition is similar to ear symmetry because Figure 6. 2 shows similar important parts (middle regions of ears) to Figure 6. 16. However, the heatmaps of different genders (Figure 6. 9) are different to Figure 6. 2 and Figure 6. 16.

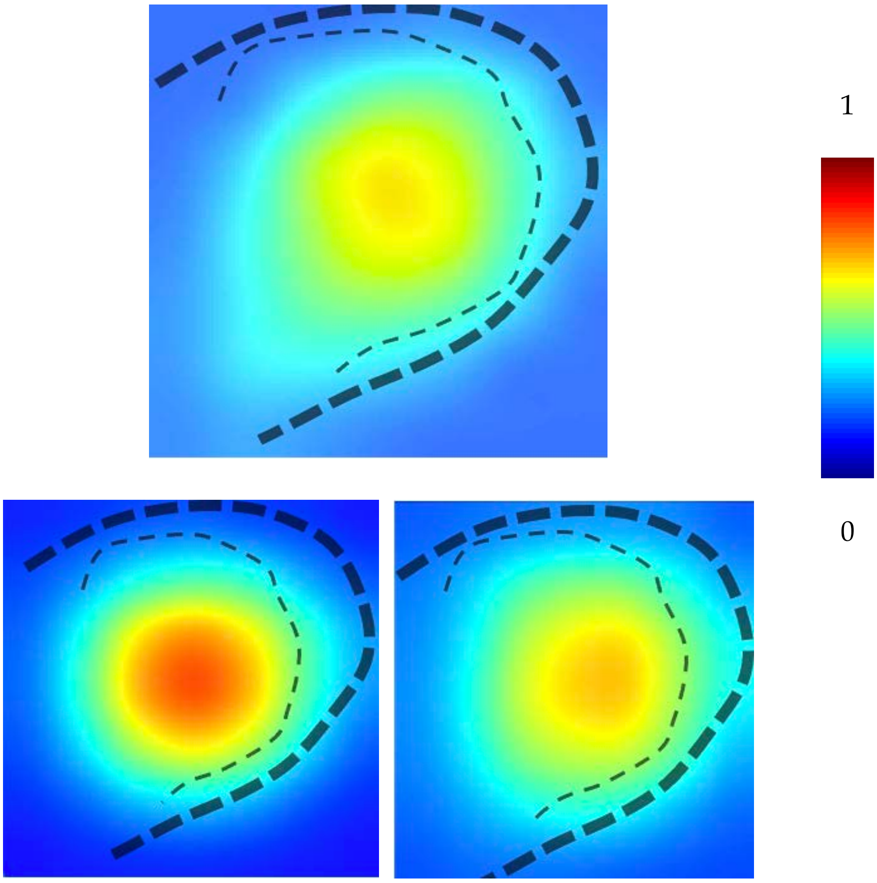


Figure 6. 17 Average heatmap for the recognition of a single ear from either side based on ResNet-50 (top: original images; bottom left: affine transformed images; bottom right: images with added Gaussian noise)

As Figure 6. 17 depicts, the central part of the ear is important for the recognition of single ears from either side, as well as for noisy images and affine transformed images. As Figure 6. 3 shows, the central parts are also important for ear recognition based on ResNet-50. However, the heatmaps of different genders (Figure 6. 10) based on ResNet-50 show the different contributions of ear parts and show that the lobe and right-middle parts play significant roles for gender classification.

As Table 6. 5 presents, the ResNet-50 network demonstrates the best performance. Comparing Figure 6. 15, Figure 6. 16 and Figure 6. 17 suggests that the heatmaps of ResNet-50 shows that the contributions of different parts of ears are very different, which is why ResNet-50 has the highest accuracy.

In addition, we calculated which part is more important for the three different networks using the original ear images. Figure 6. 18 shows the average heatmap of the original ear images over three different networks.

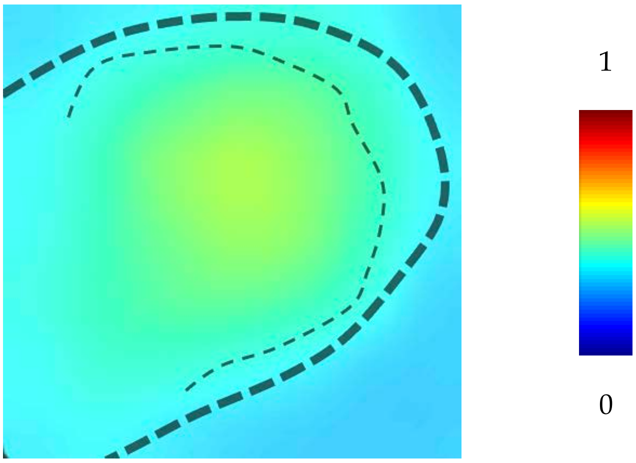


Figure 6. 18 Average heatmap for the recognition of a single ear from either side over three networks

As observed in Figure 6. 18, the heatmap shows that in the average heatmap generated by the three networks (VGG-16, GoogleNet and ResNet-50), the central regions are the most important parts for ear symmetry. We compared the recognition of a single ear from either side with one-sided ear recognition (Figure 6. 4) and found that they have similar heatmaps. Meanwhile, the important parts of model-based technique for ear symmetry are similar to the important parts of ear recognition based on model technique.

The heatmaps have shown the contributions of different regions of the ear. Therefore, we make an ellipse mask, mapping the central parts（red and yellow of heatmap) of the testing data in black while keeping the helixes (blue parts of heatmap), or we make only the central parts of the testing data visible and black out the rest of the image. Figure 6. 19 shows ear images with the mask mapping. Table 6. 9 presents the accuracy percentages of different parts of the ear.

Table 6. 9 Accuracies of different parts of ears (ear symmetry)

Images	**Accuracy**
Central parts	41.6%
Helixes parts	12.3%

As observed in Table 6. 9, the accuracy of the central parts of ear images is 41.6%, and the accuracy of the helixes parts is 12.3%. Thus, the performance of the red and yellow parts of heatmaps are much better than the blue. This has demonstrated that heatmaps are convinced.

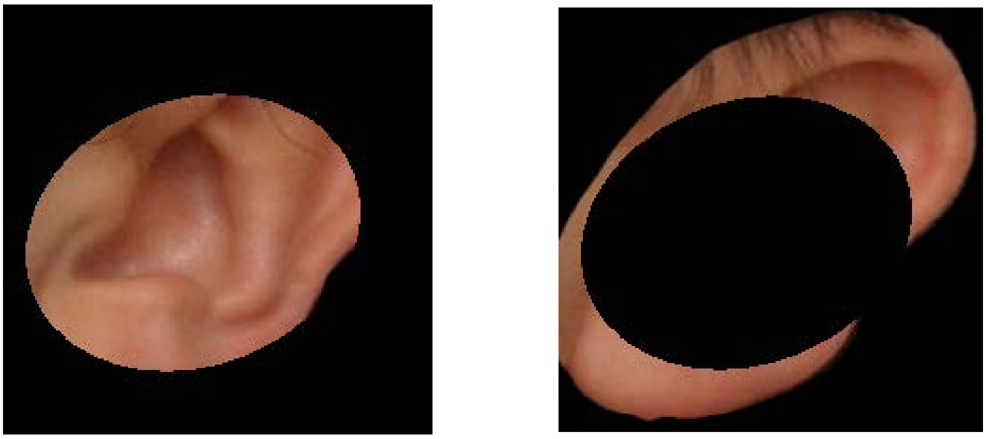
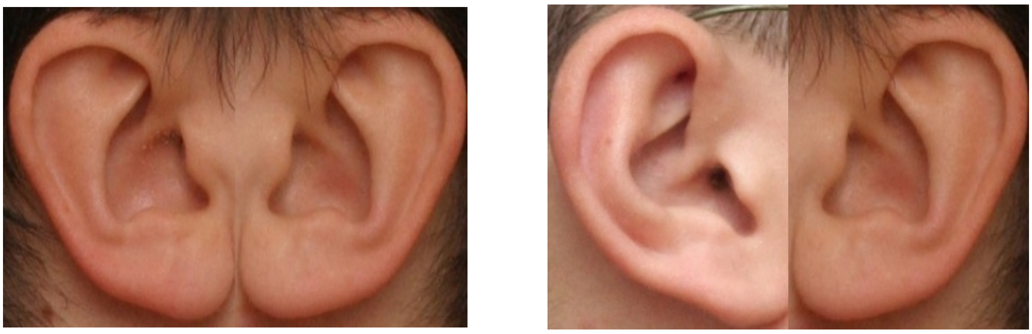


Figure 6. 19 Samples of mask mapped testing ear images

6.3.2. Combine Two-Side Ears to Test the Ear Symmetry

We considered whether the model can test the ear symmetry directly using an image combining a subject's left and right ear, as shown in Figure 6. 20. Then, we used Grad-CAM to calculate the heatmaps of combining ear images to explore which parts of ear are symmetry.



(a) Same subject (b) Different subjects

Figure 6. 20 Composite images of two sides

There are 300 combined two-side ear images from the same subjects (100 subjects) and 300 combined ear images from different subjects (100 different subjects are selected randomly). 150 ear images were from the same subject (50 subjects), 150 ear images were from different subjects for training and the others were for testing. We divided the testing set into five sub-sets, and each sub-set had 30 ear images from same subject and 30 ear images from different subjects. Because the ResNet-50 has the best performance for two-side ear recognition, we transferred the ResNet-50 learnings to this experiment. Table 6. 10 shows the results of the ear symmetry.

Table 6. 10 Ear Symmetry Accuracies for Deep Learning

Images	Accuracy
Original images	100%
Noisy images	100%
Affine transformed images	100%

Table 6. 10 shows the ear symmetry accuracies. The classification accuracies for the original images, affine transformed images and noisy images are 100%. As observed in Table 6. 10, the model can classify all the ear images into the correct groups and is not affected by noise. These accuracies, along with the recognition rates reported in Table 6. 5 and Table 6. 8, indicate that human ears are bilaterally symmetric. Figure 6. 21 shows the average heatmap of an ear image in this experiment.

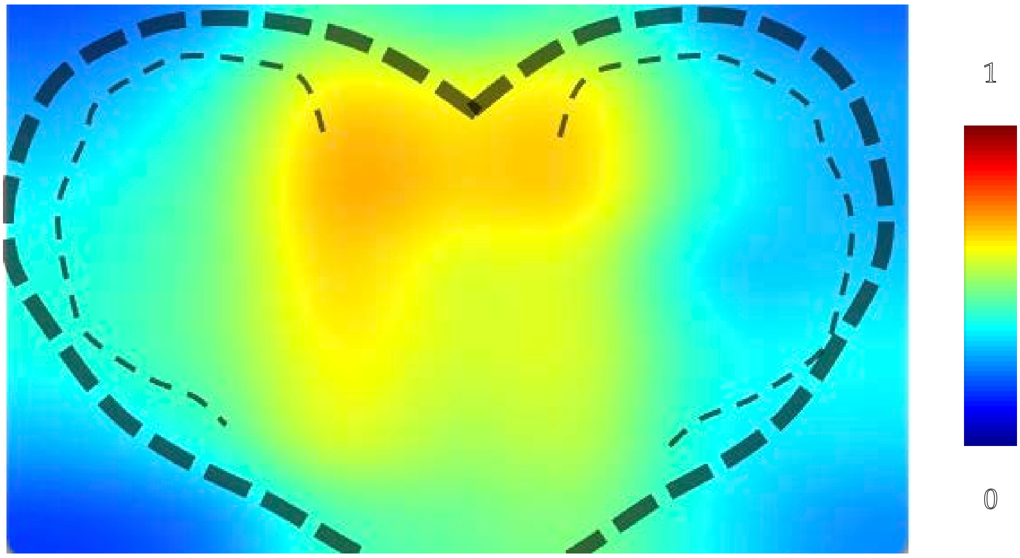


Figure 6. 21 Average heatmap for ear symmetry

As shown in Figure 6. 21, the middle parts are important for ear symmetry. If we compare the heatmap shown in Figure 6. 11 with the heatmaps of gender classification, although the lobe is significant for gender classification, it is not significant for ear symmetry. Therefore, the sections of the ear that dominate the gender analysis contribute less to ear symmetry. Comparing it with Figure 6. 17 shows which parts of the ear are not symmetric. As the heatmaps show, the helixes and lobes are not significant for ear symmetry, and the triangular fossa is the most important for ear symmetry.

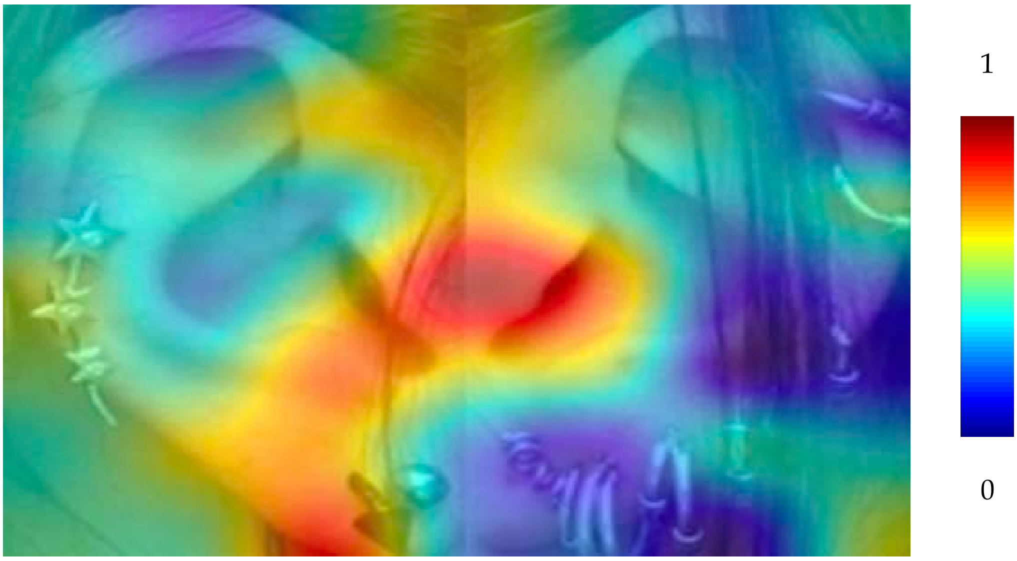


Figure 6. 22 Heatmap affected by earrings

In the next experiment, we consider the effect of wearing earrings on heatmaps. Earrings are usually worn on the helix, and the positions of the earrings differ among helices. Figure 6. 22 shows a heatmap of ear images derived from subjects wearing earrings. As shown in this figure, the red regions are concentrated on the central section, which suggests that the ear symmetry is less affected by earrings.

6.3.3. Ear Symmetry Verification

As deep learning is highly accurate for ear symmetry recognition, we conducted verification experiments to evaluate the performance of our model. For these experiments, we randomly selected one left and one right ear image from each subject. We used the training model to represent each image with a vector and calculated the cosine similarities of each image.

If the stored left ear biometric template of a subject A is represented by L_A and the right ear for verification is represented by R_A, the null hypothesis (h_0) and alternate hypothesis (h_1) are defined as follows:

Null hypothesis $\boldsymbol{h_0}$: The right ear R_A and left ear L_A come from different subjects.

Alternate hypothesis $\boldsymbol{h_1}$: The right ear R_A and left ear L_A come from the same subject.

If the cosine similarity score is higher, the system is more certain that $\boldsymbol{h_1}$ will be confirmed. The system's decision is regulated by the following threshold T: if the score is higher than or equal to the threshold T, $\boldsymbol{h_1}$ is confirmed. Otherwise, the score is lower than T, and the system infers that $\boldsymbol{h_0}$ is correct.

There are two incorrect conclusions. A type I error (FRR) is defined as rejecting the null hypothesis h_0 when it is true. A type II error (FAR) is defined as failing to reject h_0 when it is false. The EER is the value defined as FRR=FAR. EER is an indicator of the performance of the model: the lower the value for EER, the higher the performance of the model.

Figure 6. 23, Figure 6. 24 and Figure 6. 25present FRR and FAR regarding the threshold T for the original ear images of the three networks.

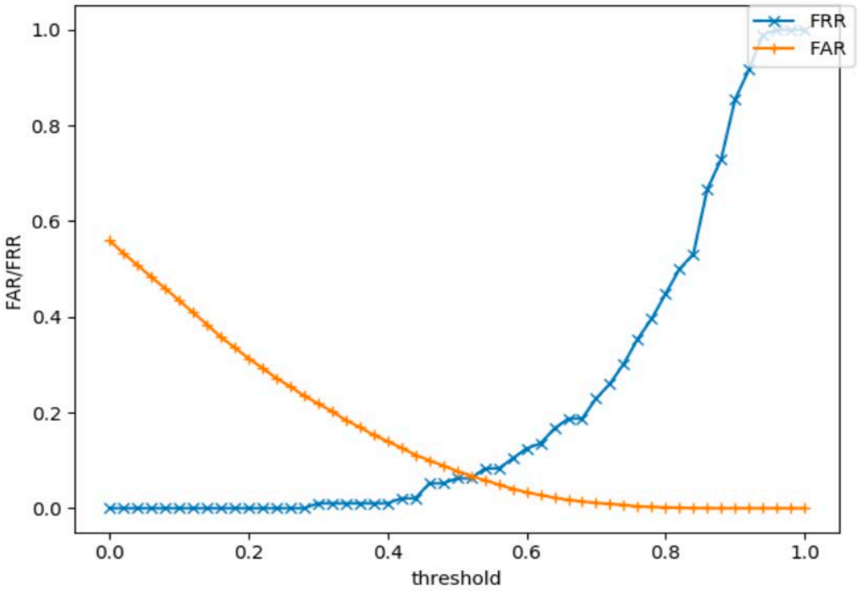


Figure 6. 23 Ear symmetry verification of the original ear images (VGG-16)

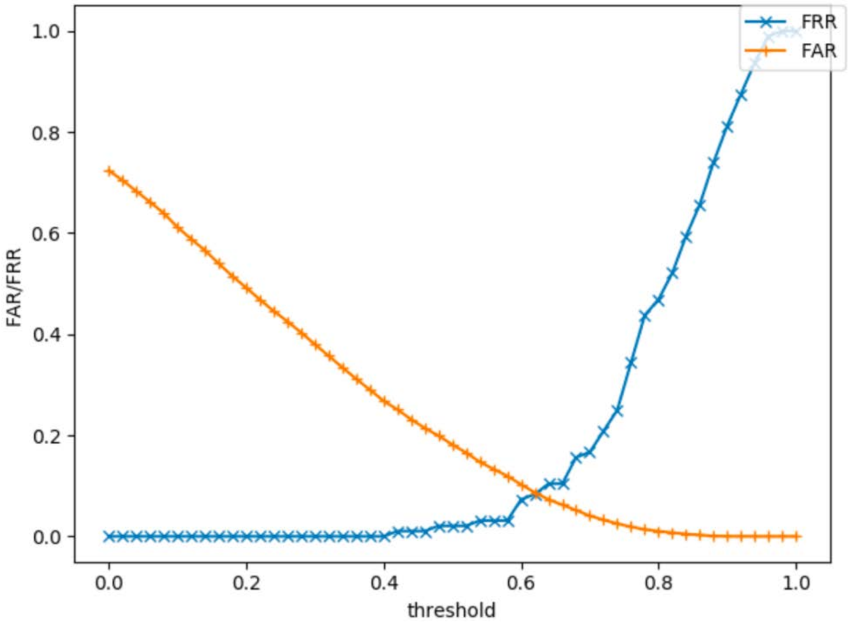


Figure 6. 24 Ear symmetry verification of the original ear images (GoogleNet)

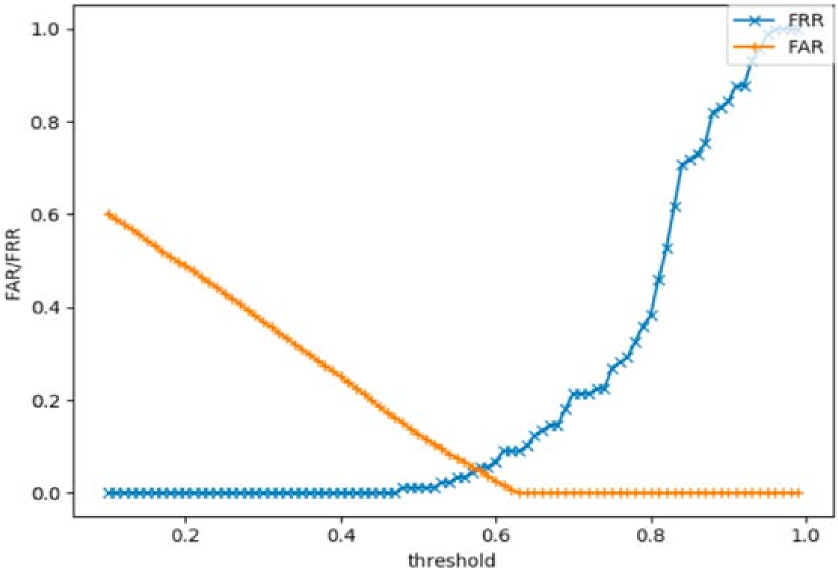


Figure 6. 25 Ear symmetry verification of the original ear images (ResNet-50)

Table 6. 11 Equal Error Rate of Two-Side Ear Recognition for Deep Learning

Method	**EER**
VGG-16	10.07%
GoogleNet	12.04%
ResNet-50	**3.57%**

Table 6. 11 tabulates the EER for the different network experiments. The EER for the original ear images based on VGG-16, GoogleNet and ResNet-50 is 10.07%, 12.04% and

3.57%, respectively. These results prove that the fine-tuned ResNet-50 network has the best performance out of the three networks.

In addition, we calculated the EER for different types of ear images using the fine-tuned ResNet-50 network. Figure 6. 26 indicates the ear verification of ear images with Gaussian noise, Figure 6. 27 indicates the ear verification of ear images via affine transformation, Figure 6. 25 shows that the model has a high level of performance and is sensitive for original ear images and Figure 6. 26 and Figure 6. 27 show that the model is affected by noise and affine transformation.

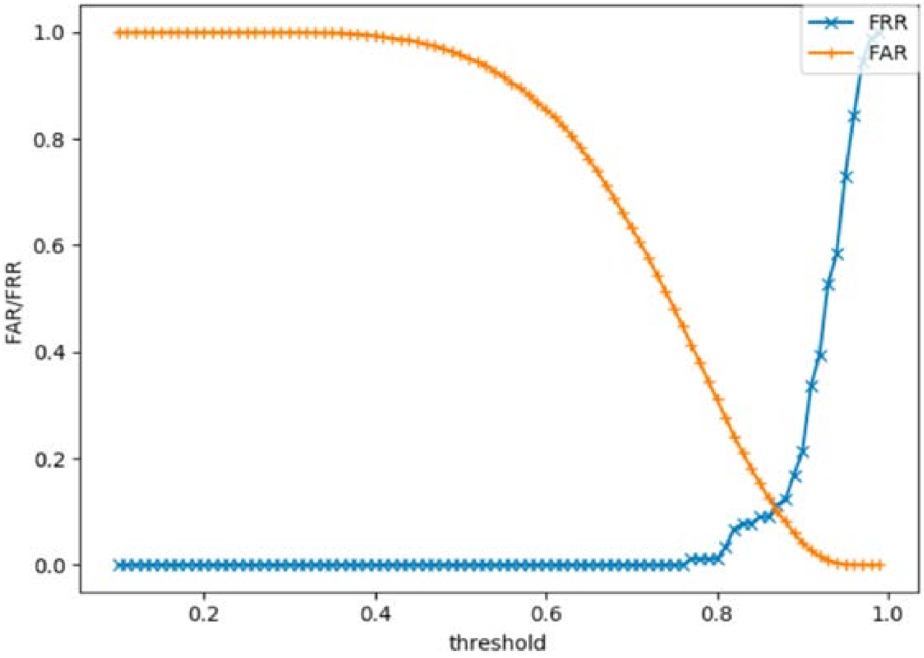


Figure 6. 26 Ear symmetry verification of ear images with Gaussian noise

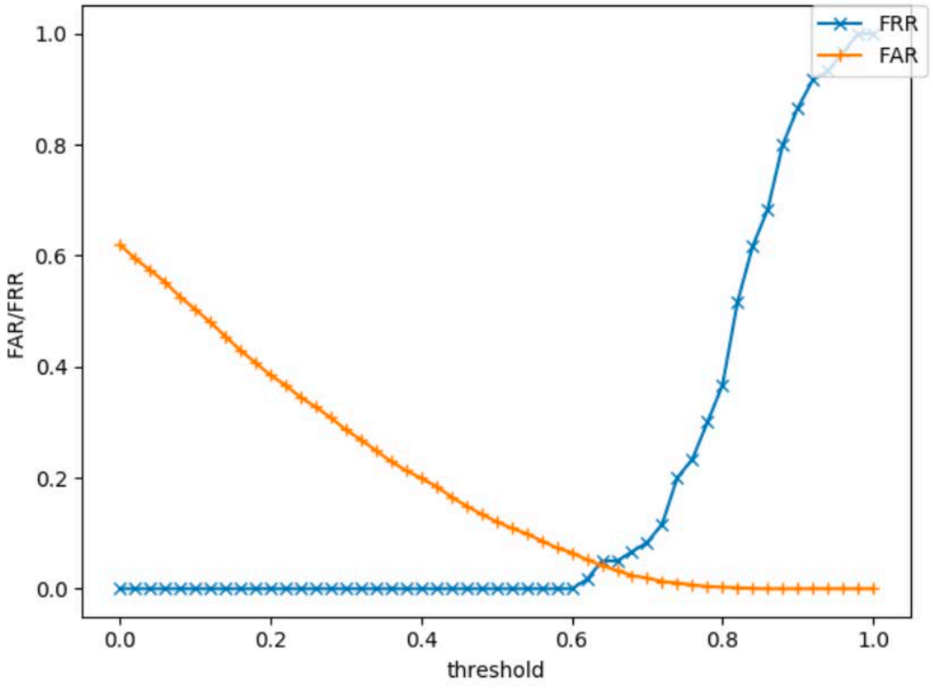


Figure 6. 27 Ear symmetry verification for affine transformed ears

Table 6. 12 represents the EER of different images. The EER images with affine transformation and images with Gaussian noise are 3.24% and 10.27%, respectively, and the model has a high level of performance.

Table 6. 12 Equal Error Rate of Pre-Processing Images Based on ResNet-50 (ear symmetry)

Images	EER
Noisy images	10.27%
Affine transformed images	**3.24%**

Figure 6. 28, Figure 6. 29 and Figure 6. 30 present the inter and intra class variations for fine-tuned ResNet-50 experiments of three different types of ear images. In these three figures, the green bars represent the distance histogram between the same subjects-client, and red bars show the distance between the different subjects-impostor. Figure 6. 28 shows the inter and intra class variations for the experiment where the original ear images without any pre-processing are used. In this figure, the green bar is small, the red bar is large and the confusion region (the overlap region between the green and red histograms) is the smallest. A small confusion region indicates a high recognition rate when the original ear images are used. Figure 6. 29 shows the inter and intra class variations for the affine transformed ear images, and the overlap region between the two histograms is smaller

than in Figure 6. 28. This smaller overlap region indicates that the recognition rate for the affine transformed ear images is expected to be higher than the original ear images.

Furthermore, Figure 6. 30 shows the inter and intra class variations for the ear images contaminated with the zero mean Gaussian noise. As observed in Figure 6. 30, the overlap region between the two histograms is larger than that in Figure 6. 28. Moreover, there is a disparity between the original images and affine transformed images or noisy images. The model has the best performance for the affine transformed ear images, as affine transformation can overcome rotation.

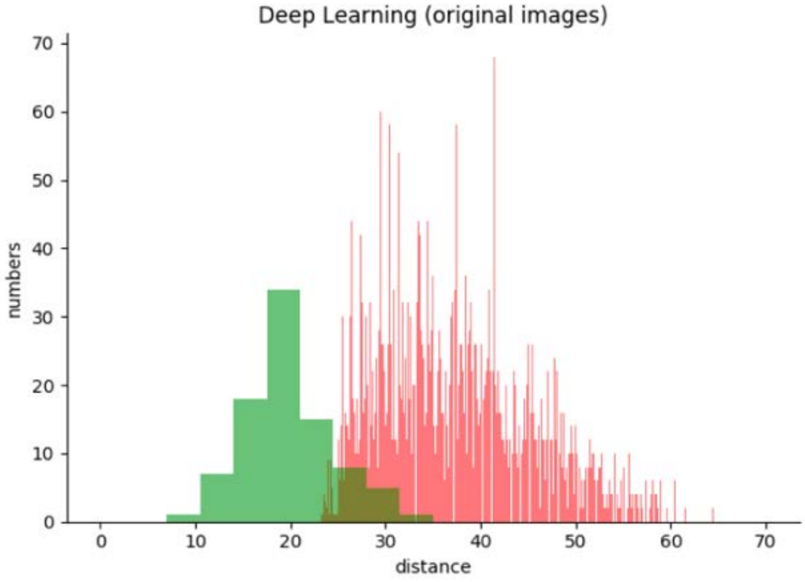


Figure 6. 28 Inter and intra class variations (original images)

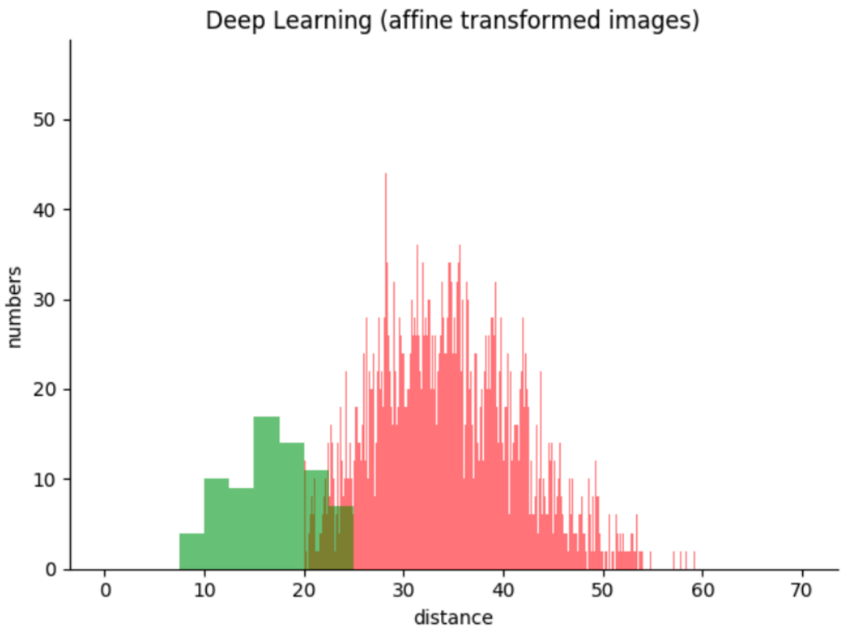


Figure 6. 29 Inter and intra class variations (affine transformed images)

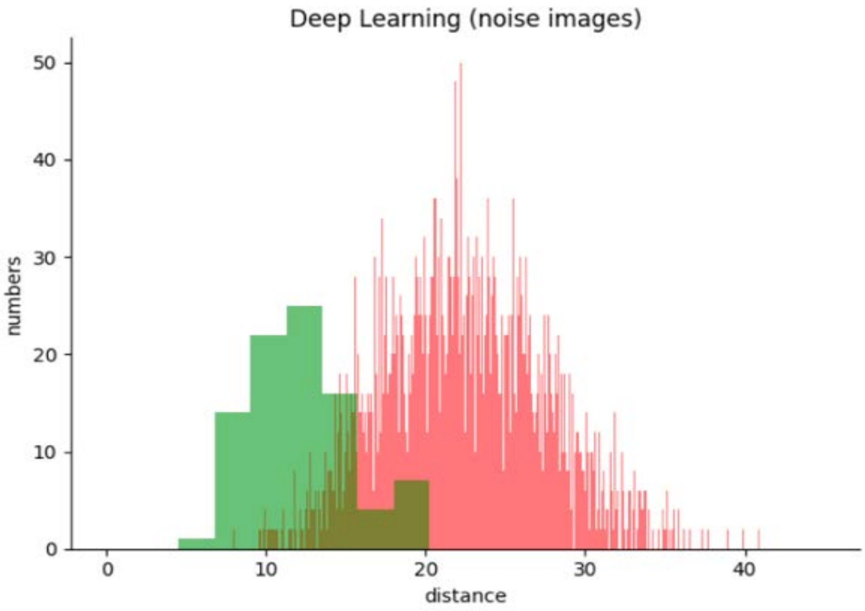


Figure 6. 30 Inter and intra class variations (noisy images)

6.4. Conclusion

In this chapter, we applied transfer learning to VGG-16, GoogleNet and ResNet-50 networks for human identification, gender classification and bilateral symmetry from ear images. The fine-tuned networks of ResNet-50 had the best performance out of these three networks. The performances for ear recognition, gender classification and ear recognition with bilateral symmetry were 92.9%, 90.9% and 93.1%, respectively. The deep learning approaches had better performances than the model-based technique, and the deep learning approaches were more robust for the noisy images. Additionally, we used deep learning for the paired ear recognition. Our results confirmed symmetry with a performance of 100%.

Additionally, we considered which parts of the ear contribute the most to ear recognition, gender classification and ear bilateral symmetry. The heatmaps presented in this chapter indicate that the central region is significant for ear recognition and ear bilateral symmetry. Furthermore, we compared the heatmap associated with ear recognition with that of gender classification. The lobe and upper helix are important for gender classification. However, the lobe does not play an important role in ear recognition and ear bilateral symmetry. Moreover, our analysis related to ear bilateral symmetry demonstrates that the recognition rate is insignificantly affected by jewellery.

Our experiments of ear symmetry were based on a controlled dataset. Although we used the in-the-wild dataset (AWE dataset), the accuracy was not good. Therefore, as the in-the-wild ear dataset presents a challenge for ear symmetry, we will consider using pre-processing approaches for in-the-wild data in future works to improve the accuracy of ear symmetry.

Chapter 7

Conclusions and Future Works

7.1. Conclusions

The ear as a biometric is no longer in the initial phase. There has been significant progress, not only in ear recognition but also in the soft traits that have been extracted from ear images. This thesis capitalises model-based and deep learning techniques for ear biometrics, including ear recognition, gender classification, kinship verification from ear images and ear bilateral symmetry.

The model used for human identification, gender classification and kinship verification from ear images is the first model-based approach for gender classification and kinship verification from ear images. The model is based on geometric features, which calculate the distance using the key points and apices of ear. The affine transformation was used successfully for viewpoint correction. Our model-based method produced promising results, with an identification accuracy of 95.6% for the non-rotated images. The affine transformation operates on the rotated images and increases performance.

In addition, our model-based algorithm was used for gender classification to produce a 67.2% classification accuracy. Although the accuracy was not as high as appearance-based results, it demonstrated a high performance among model-based techniques, as well as that it can handle affine transformations. Therefore, we could benefit from adding more features to our system for gender classification. Additionally, we can use HOG features to improve gender classification and kinship verification accuracies. We are the first to use ears for kinship verification, and the ear images of the SOTEAR dataset are unconstrained. There was a 40.6% accuracy for father and son, 33.3% accuracy for mother and daughter, 0% for father and daughter and 31.3% for mother and son. The kinship verification results show that the mothers' ears are more influential. The accuracy of kinship verification indicates that ear biometrics can contribute to kinship verification. We enhanced the performance of our model-based technique for gender classification by adding a pre-processing step for the model, applying the force field transform for pre-processing and using HOG for feature extraction. This model-based technique achieved a gender classification accuracy of 82.9%. Additionally, when we applied this model to ear symmetry, the accuracy achieved 66.9%.

Furthermore, the deep learning approaches were used for ear recognition, gender classification and ear symmetry, and the performances for ear recognition, gender classification, and ear recognition with bilateral symmetry were 92.9%, 90.9% and 93.1%, respectively. Our thesis used in-the-wild ear images for gender classification and ear symmetry. For gender classification, as the in-the-wild ear dataset (AWE dataset) was not balanced for gender classification, we merged it with the other datasets (USTB III and SOTEAR) to strike a balance between female and male, and the performance of gender classification was good. However, the accuracy for ear symmetry of in-the-wild data was not high. Additionally, we used deep learning for paired ear recognition to test the ear symmetry. We compared the three networks that we applied to transfer learning and used several pre-processing approaches to evaluate the performances of the networks. Deep learning approaches are more robust than the model-based technique of noisy ear images for ear bilateral symmetry. We are the first to use deep learning for ear recognition with bilateral symmetry, and our results confirm symmetry with a high performance.

We also considered which parts of the ear contribute the most to ear recognition, gender classification and ear bilateral symmetry. Although there is some research on ear recognition, gender classification and ear symmetry, there is no research on why the models or the deep learning approaches have these classifications. In the model-based approach, the highest accuracies of the divided ear images come from the upper parts of the ear, which means that these parts are more important for gender classification. For the deep learning approaches, we calculated the heatmaps of different networks and the average heatmaps over the three networks for ear recognition, gender classification and ear symmetry. The heatmaps, which are presented in Chapter 7, indicate that the central region is significant for ear recognition and ear bilateral symmetry, and the lobe and upper helix are important for gender classification. However, the lobe does not play an important role in ear recognition and ear bilateral symmetry.

In this thesis, through a set of experiments, we demonstrate that there is an implied similarity between the left and right ears. Therefore, ear recognition can be achieved regardless of which ear is used for analysis. This notion is important because human anatomy dictates that only one ear can be seen in a single image, as ears cannot be seen or analysed in images from a full-frontal view, such as passport style images, which rarely occur in surveillance. Therefore, our study indicates that the constraints on ear image acquisition are reduced. As such, this thesis establishes benchmarks for ear recognition by using ear bilateral symmetry and provides pointers for future applications and developments for ear symmetry, particularly as it is prudent for recognition to focus on the ear perimeter.

7.2. Future Works

Ear biometrics are improving, as some research discusses pose variation, rotation variation, illumination variation and occlusion. In addition, extracting soft traits from ear images is in the early stages of development. Although gender classification is being discussed by an increasing number of researchers, the performances of kinship verification and age classification need improvement. Other soft traits, such as ethnic background and genealogy, have not been developed. Additionally, gait biometrics have been used to find criminals, and the ear-prints can used as be evidence to accuse them. Moreover, ear features have been used to identify human remains. The Interpol DVI Form provides the following two options for ear features: ear lobes/pierced and distinctive features. The distinctive features include the different outer appearance of ears, such as Darwin's tubercle (a thickening on the helix at the junction of the upper and middle thirds [118].

The main problem is the need for larger family datasets of ear images, wherein a more accurate estimate of the performance of kinship verification can be obtained and potential associations between a family’s ears can be analysed. Additionally, we intend to find more features to improve classification accuracy and provide a deeper understanding of how these results are related to basic ear structures. The observation that the helix differs between female and male ears suggests that including the helix in our model-based system could conceivably improve the gender, and therefore, the kinship classification accuracy.

We aim to combine the model-based approach with deep learning to improve the ear recognition performance. The effect of ear occlusion by hair is another topic for future work. Another interesting topic for future work is to use deep learning attention networks for ear biometrics to avoid non-ear regions of ear images in ear recognition. Another issue for future work is to avoid using transfer learning during the training of the deep learning networks to produce heat maps with no effects of transfer learning, where non-ear regions are the least significant in heat maps. In such a scenario, the interpretation of heat maps will be clearer and more straightforward in the analysis of the deep learning methods for ear recognition, gender classification and ear symmetry. However, in such cases, a dataset with around 10,000 ear images for training would be required.

In addition, the in-the-wild ear images pose a challenge for ear biometrics, as although there is some research on ear recognition for in-the-wild ear images, the accuracies are not high. Our thesis used in-the-wild ear images for gender classification and ear symmetry, but the accuracy of ear symmetry was not high. We considered using different approaches to pre-process the ear images and remove the noise.

Although the ear occlusion by hair had been discussed for many years, and some studies has proposed some approaches to solve this problem. However, the accuracies of ear recognition for occlusion problem are not high, if the ear biometrics are deployed in the world, this problem must be solved.

Finally, as this approach is widely used to automatically authenticate individuals, fingerprint and face recognition has been applied to protect mobile phone security. Amazon's exposure of a patent shows that phones can use the camera to scan the user's ear. This enables mobile phones to automatically identify a user's ear. Based on this idea, the patent will enable users to answer a call without unlocking their phone, and the phone can automatically adjust the volume as the distance between the phone and ear changes [16].

References

[1] Bertillon, A., 1890. La photographie judiciaire: avec un appendice sur la classification et l'identification anthropométriques. Paris: Gauthier-Villars.

[2] Iannarelli, A.V., 1964. *Ear identification*. Paramont Publishing Company.

[3] Burge, M. and Burger, W., 1996. Ear biometrics. In *Biometrics* (pp. 273-285). Springer, Boston, MA.

[4] Hurley, D.J., Nixon, M.S. and Carter, J.N., 2000, September. A new force field transform for ear and face recognition. In *Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101)* (Vol. 1, pp. 25-28). IEEE.

[5] Chen, H. and Bhanu, B., 2005, January. Contour matching for 3D ear recognition. In *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1* (Vol. 1, pp. 123-128). IEEE.

[6] Shailaja, D. and Gupta, P., 2006, December. A simple geometric approach for ear recognition. In *9th International Conference on Information Technology (ICIT'06)* (pp. 164-167). IEEE.

[7] Meng, D., Nixon, M.S. and Mahmoodi, S., 2019, September. Gender and Kinship by Model-Based Ear Biometrics. In *2019 International Conference of the Biometrics Special Interest Group (BIOSIG)* (pp. 1-5). IEEE.

[8] Gnanasivam, P. and Muttan, S., 2013. Gender classification using ear biometrics. In *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)* (pp. 137-148). Springer, India.

[9] Yaman, D., Eyiokur, F.I., Sezgin, N. and Ekenel, H.K., 2018, June. Age and gender classification from ear images. In *2018 International Workshop on Biometrics and Forensics (IWBF)* (pp. 1-7). IEEE.

[10] Yaman, D., Irem Eyiokur, F. and Kemal Ekenel, H., 2019. Multimodal age and gender classification using ear and profile face images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 0-0).

[11] Sadler, T.W., 2006. Third to eight weeks: The embryonic period. *Langman's Medical Embryology, 10th edn. TW Sadler, ed. Lippincott Williams & Wilkins, Philadelphia*, pp.67-87.

[12] Davis, J., 1987. Surgical embryology. In *Aesthetic and reconstructive otoplasty* (pp. 93-125). Springer, New York, NY.

[13] Streeter, G.L., 1922. Development of the auricle in the human embryo. *Contrib Embryol*, *69*, p.111.

[14] Gray, H. and Standring, S., 2008. *Gray's anatomy: the anatomical basis of clinical practice*. Churchill Livingstone.

[15] Unar, J.A., Seng, W.C. and Abbasi, A., 2014. A review of biometric technology along with trends and prospects. *Pattern recognition*, *47*(8), pp.2673-2688.

[16] Wang, Z., Yang, J. and Zhu, Y., 2019. Review of ear biometrics. *Archives of Computational Methods in Engineering*, pp.1-32.

[17] https://www.scface.org/ Accessed: 2021-04-09

[18] Messer, K., Matas, J., Kittler, J., Luettin, J. and Maitre, G., 1999, March. XM2VTSDB: The extended M2VTS database. In Second international conference on audio and video-based biometric person authentication (Vol. 964, pp. 965-966).

[19] https://www.interpol.int/How-we-work/Forensics/Disaster-Victim-Identification-DVI Accessed: 2021-04-09

[20] Yan, P. and Bowyer, K.W., 2007. Biometric recognition using 3D ear shape. *IEEE Transactions on pattern analysis and machine intelligence*, *29*(8), pp.1297-1308.

[21] https://github.com/dm-vlc/dataset.git Accessed: 2021-04-09

[22] Abaza, A., Ross, A., Hebert, C., Harrison, M.A.F. and Nixon, M.S., 2013. A survey on ear biometrics. *ACM computing surveys (CSUR)*, *45*(2), pp.1-35.

[23] Islam, S.M., Bennamoun, M. and Davies, R., 2008, January. Fast and fully automatic ear detection using cascaded adaboost. In *2008 IEEE Workshop on Applications of Computer Vision* (pp. 1-6). IEEE.

[24] Yuan, L. and Zhang, F., 2009, July. Ear detection based on improved adaboost algorithm. In *2009 International Conference on Machine Learning and Cybernetics* (Vol. 4, pp. 2414-2417). IEEE.

[25] Said, E.H., Abaza, A. and Ammar, H., 2008, September. Ear segmentation in color facial images using mathematical morphology. In *2008 Biometrics Symposium* (pp. 29-34). IEEE.

[26] Arbab-Zavar, B. and Nixon, M.S., 2007, November. On shape-mediated enrolment in ear biometrics. In *International Symposium on Visual Computing* (pp. 549-558). Springer, Berlin, Heidelberg.

[27] Cummings, A.H., Nixon, M.S. and Carter, J.N., 2010, September. A novel ray analogy for enrolment of ear biometrics. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)* (pp. 1-6). IEEE.

[28] Ibrahim, M.I., Nixon, M.S. and Mahmoodi, S., 2010, November. Shaped wavelets for curvilinear structures for ear biometrics. In *International Symposium on Visual Computing* (pp. 499-508). Springer, Berlin, Heidelberg.

[29] Sarangi, P.P., Panda, M., Mishra, B.P. and Dehuri, S., 2017. An automated ear localization technique based on modified hausdorff distance. In *Proceedings of International Conference on Computer Vision and Image Processing* (pp. 229-240). Springer, Singapore.

[30] El-Naggar, S., Abaza, A. and Bourlai, T., 2018, August. Ear detection in the wild using faster R-CNN deep learning. *In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1124-1130). IEEE.

[31] Zhang, Y. and Mu, Z., 2017. Ear detection under uncontrolled conditions with multiple scale faster region-based convolutional neural networks. *Symmetry,* 9(4), p.53.

[32] Yuan, L. and Lu, F., 2018, July. Real-time ear detection based on embedded systems. *In 2018 International Conference on Machine Learning and Cybernetics (ICMLC)* (Vol. 1, pp. 115-120). IEEE.

[33] Emeršič, Ž., Gabriel, L.L., Štruc, V. and Peer, P., 2018. Convolutional encoder–decoder networks for pixel-wise ear detection and segmentation. *IET Biometrics*, *7*(3), pp.175-184.

[34] Ganapathi, I.I., Prakash, S., Dave, I.R. and Bakshi, S., 2020. Unconstrained ear detection using ensemble-based convolutional neural network model. *Concurrency and Computation: Practice and Experience*, 32(1), p.e5197.

[35] Bizjak, M., Peer, P. and Emeršič, Ž., 2019, May. Mask R-CNN for ear detection. *In 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 1624-1628). IEEE.

[36] Resmi, K.R. and Raju, G., 2019, March. A Novel Approach to Automatic Ear Detection Using Banana Wavelets and Circular Hough Transform. *In 2019 International Conference on Data Science and Communication (IconDSC)* (pp. 1-5). IEEE.

[37] Mursalin, M. and Islam, S.M.S., 2020, February. EpNet: A Deep Neural Network for Ear Detection in 3D Point Clouds. *In International Conference on Advanced Concepts for Intelligent Vision Systems* (pp. 15-26). Springer, Cham.

[38] Victor, B., Bowyer, K. and Sarkar, S., 2002, August. An evaluation of face and ear biometrics. *In Object recognition supported by user interaction for service robots* (Vol. 1, pp. 429-432). IEEE.

[39] Yan, P. and Bowyer, K., 2005, September. Empirical evaluation of advanced ear biometrics. *In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops* (pp. 41-41). IEEE.

[40] Oravec, M., Pavlovičová, J., Sopiak, D., Jirka, V., Loderer, M., Lehota, Ľ., Vodička, M., Fačkovec, M., Mihalik, M., Tomík, M. and Gerát, J., 2016, May. Mobile ear recognition application. *In 2016 International Conference on Systems, Signals and Image Processing (IWSSIP)* (pp. 1-4). IEEE.

[41] Zarachoff, M., Sheikh-Akbari, A. and Monekosso, D., 2018, October. Application of single image super-resolution in human ear recognition using eigenvalues. *In 2018 IEEE International Conference on Imaging Systems and Techniques (IST)* (pp. 1-6). IEEE.

[42] Jangilla, S., 2016, November. Ear recognition using bilinear Probabilistic Principal Component analysis and sparse classifier. *In 2016 IEEE Region 10 Conference (TENCON)* (pp. 979-983). IEEE.

[43] Banerjee, S. and Chatterjee, A., 2015. Ear Recognition Using Force Field Transform and Collaborative Representation-Based Classification with Single Training Sample Per Class. *In Intelligent Computing and Applications* (pp. 511-517). Springer, New Delhi.

[44] Arbab-Zavar, B., Nixon, M.S. and Hurley, D.J., 2007, September. On model-based analysis of ear biometrics. *In 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems* (pp. 1-5). IEEE.

[45] Guo, Y. and Xu, Z., 2008, October. Ear recognition using a new local matching approach. *In 2008 15th IEEE International Conference on Image Processing* (pp. 289-292). IEEE.

[46] Kumar, A. and Chan, T.S.T., 2013. Robust ear identification using sparse representation of local texture descriptors. *Pattern recognition*, 46(1), pp.73-85.

[47] Murukesh, C., Parivazhagan, A. and Thanushkodi, K., 2012. A novel ear recognition process using appearance shape model, fisher linear discriminant analysis and contourlet transform. *Procedia engineering*, 38, pp.771-778.

[48] Prakash, S. and Gupta, P., 2014. Human recognition using 3D ear images. *Neurocomputing*, 140, pp.317-325.

[49] Omara, I., Li, F., Zhang, H. and Zuo, W., 2016. A novel geometric feature extraction method for ear recognition. *Expert Systems with Applications*, 65, pp.127-135.

[50] Anwar, A.S., Ghany, K.K.A. and Elmahdy, H., 2015. Human ear recognition using geometrical features extraction. *Procedia Computer Science*, 65, pp.529-537.

[51] Ying, T., Debin, Z. and Baihuan, Z., 2014, May. Ear recognition based on weighted wavelet transform and DCT. *In The 26th Chinese Control and Decision Conference (2014 CCDC)* (pp. 4410-4414). IEEE.

[52] Wang, Y., Mu, Z.C. and Zeng, H., 2008, December. Block-based and multi-resolution methods for ear recognition using wavelet transform and uniform local binary patterns. *In 2008 19th International Conference on Pattern Recognition* (pp. 1-4). IEEE.

[53] Eyiokur, F.I., Yaman, D. and Ekenel, H.K., 2017. Domain adaptation for ear recognition using deep convolutional neural networks. *iet Biometrics*, 7(3), pp.199-206.

[54] Emeršič, Ž., Štepec, D., Štruc, V. and Peer, P., 2017, May. Training Convolutional Neural Networks with Limited Training Data for Ear Recognition in the Wild. *In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 987-994). IEEE.

[55] Chen, H. and Bhanu, B., 2007. Human ear recognition in 3D. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4), pp.718-737.

[56] Ganapathi, I.I., Ali, S.S. and Prakash, S., 2020. Geometric statistics-based descriptor for 3D ear recognition. *The Visual Computer*, 36(1), pp.161-173.

[57] Yan, P., Bowyer, K.W. and Chang, K.J., 2005, March. ICP-based approaches for 3D ear recognition. *In Biometric Technology for Human Identification II* (Vol. 5779, pp. 282-291). International Society for Optics and Photonics.

[58] Zhou, J., Cadavid, S. and Abdel-Mottaleb, M., 2011, June. A computationally efficient approach to 3d ear recognition employing local and holistic features. *In CVPR 2011 WORKSHOPS* (pp. 98-105). IEEE.

[59] Sinha, H., Manekar, R., Sinha, Y. and Ajmera, P.K., 2019. Convolutional neural network-based human identification using outer ear images. *In Soft computing for problem solving* (pp. 707-719). Springer, Singapore.

[60] Abd Almisreb, A., Jamil, N. and Din, N.M., 2018, March. Utilizing AlexNet deep transfer learning for ear recognition. *In 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)* (pp. 1-5). IEEE.

[61] Tian, Y., Dong, H. and Wang, L., 2020, July. Ear Recognition Based on Gabor-SIFT. *In International Conference on Artificial Intelligence and Security* (pp. 86-94). Springer, Cham.

[62] Chang, K., Bowyer, K.W., Sarkar, S. and Victor, B., 2003. Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Transactions on pattern analysis and machine intelligence,* 25(9), pp.1160-1165.

[63] Sajadi, S. and Fathi, A., 2020. Genetic algorithm based local and global spectral features extraction for ear recognition. *Expert Systems with Applications,* 159, p.113639.

[64] Khaldi, Y. and Benzaoui, A., 2020. A new framework for grayscale ear images recognition using generative adversarial networks under unconstrained conditions. *Evolving Systems,* pp.1-12.

[65] Omara, I., Hagag, A., Ma, G., Abd El-Samie, F.E. and Song, E., 2021. A novel approach for ear recognition: learning Mahalanobis distance features from deep CNNs. *Machine Vision and Applications,* 32(1), pp.1-14.

[66] Ying, T., Shining, W. and Wanxiang, L., 2018, June. Human ear recognition based on deep convolutional neural network. *In 2018 Chinese Control and Decision Conference (CCDC)* (pp. 1830-1835). IEEE.

[67] Emeršič, Ž., Štepec, D., Štruc, V., Peer, P., George, A., Ahmad, A., Omar, E., Boult, T.E., Safdaii, R., Zhou, Y. and Zafeiriou, S., 2017, October. The unconstrained ear recognition challenge. *In 2017 IEEE international joint conference on biometrics (IJCB)* (pp. 715-724). IEEE.

[68] Emeršič, Ž., SV, A.K., Harish, B.S., Gutfeter, W., Khiarak, J.N., Pacut, A., Hansley, E., Segundo, M.P., Sarkar, S., Park, H.J. and Nam, G.P., 2019, June. The unconstrained ear recognition challenge 2019. *In 2019 International Conference on Biometrics (ICB)* (pp. 1-15). IEEE.

[69] Alshazly, H., Linse, C., Barth, E. and Martinetz, T., 2020. Deep convolutional neural networks for unconstrained ear recognition. *IEEE Access*, 8, pp.170295-170310.

[70] Hansley, E.E., Segundo, M.P. and Sarkar, S., 2018. Employing fusion of learned and handcrafted features for unconstrained ear recognition. *IET Biometrics,* 7(3), pp.215-223.

[71] Zhou, Y. and Zaferiou, S., 2017, May. Deformable models of ears in-the-wild for alignment and recognition. *In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 626-633). IEEE.

[72] Radhika, K., Devika, K., Aswathi, T., Sreevidya, P., Sowmya, V. and Soman, K.P., 2020. Performance analysis of NASNet on unconstrained ear recognition. *In Nature Inspired Computing for Data Science* (pp. 57-82). Springer, Cham.

[73] Zhu, Q. and Mu, Z., 2020. PointNet++ and Three Layers of Features Fusion for Occlusion Three-Dimensional Ear Recognition Based on One Sample per Person. *Symmetry*, 12(1), p.78.

[74] Pflug, A. and Busch, C., 2012. Ear biometrics: a survey of detection, feature extraction and recognition methods. *IET biometrics,* 1(2), pp.114-129.

[75] Chen, H. and Bhanu, B., 2007. Human ear recognition in 3D. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 29(4), pp.718-737.

[76] Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision,* 60(2), pp.91-110.

[77] Bay, H., Tuytelaars, T. and Van Gool, L., 2006, May. Surf: Speeded up robust features. *In European conference on computer vision* (pp. 404-417). Springer, Berlin, Heidelberg.

[78] Arbab-Zavar, B. and Nixon, M.S., 2011. On guided model-based analysis for ear biometrics. *Computer Vision and Image Understanding,* 115(4), pp.487-502.

[79] Ojala, T., Pietikainen, M. and Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence,* 24(7), pp.971-987.

[80] Yuan, L. and chun Mu, Z., 2012. Ear recognition based on local information fusion. *Pattern Recognition Letters*, *33*(2), pp.182-190.

[81] Nosrati, M.S., Faez, K. and Faradji, F., 2007, November. Using 2D wavelet and principal component analysis for personal identification based on 2D ear structure. In *2007 International Conference on Intelligent and Advanced Systems* (pp. 616-620). IEEE.

[82] Taertulakarn, S., Pintavirooj, C., Tosranon, P. and Hamamoto, K., 2016, December. The preliminary investigation of ear recognition using hybrid technique. In *2016 9th Biomedical Engineering International Conference (BMEiCON)* (pp. 1-4). IEEE.

[83] Ganapathi, I.I., Prakash, S., Dave, I.R., Joshi, P., Ali, S.S. and Shrivastava, A.M., 2018. Ear recognition in 3D using 2D curvilinear features. *IET Biometrics*, *7*(6), pp.519-529.

[84] Ganapathi, I.I. and Prakash, S., 2018. 3D ear recognition using global and local features. *IET Biometrics*, *7*(3), pp.232-241.

[85] Khorsandi, R. and Abdel-Mottaleb, M., 2013, January. Gender classification using 2-D ear images and sparse representation. In *2013 IEEE Workshop on applications of computer vision (WACV)* (pp. 461-466). IEEE.

[86] Lei, J., Zhou, J. and Abdel-Mottaleb, M., 2013, June. Gender classification using automatically detected and aligned 3D ear range data. In *2013 International Conference on Biometrics (ICB)* (pp. 1-7). IEEE.

[87] Yaman, D., Eyiokur, F.I. and Ekenel, H.K., 2021. Multimodal soft biometrics: combining ear and face biometrics for age and gender classification. *Multimedia Tools and Applications*, pp.1-19.

[88] Nguyen-Quoc, H. and Hoang, V.T., 2020, December. Gender recognition based on ear images: a comparative experimental study. In *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 451-456). IEEE.

[89] Abaza, A. and Ross, A., 2010, September. Towards understanding the symmetry of human ears: A biometric perspective. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)* (pp. 1-7). IEEE.

[90] Toygar, Ö., Alqaralleh, E. and Afaneh, A., 2018. Symmetric ear and profile face fusion for identical twins and non-twins recognition. *Signal, Image and Video Processing*, *12*(6), pp.1157-1164.

[91] Hoogstrate, A.J., Van den Heuvel, H. and Huyben, E., 2001. Ear identification based on surveillance camera images. *Science & justice: journal of the Forensic Science Society*, *41*(3), pp.167-172.

[92] Lugt, C.V., 2001. Ear identification. *Bedrijifsinformatie's Gravenhage: Elsevier*.

[93] Van der Lugt, C., 1999. Report about the research on ears and earprints, Second International Ear Identification Course. *Zutphen (Netherlands)/Durham (UK): Dutch Police College/National Training Centre NTCSSCI*.

[94] Lugt, V.D.C., 1998. Ear identification–state of art. *Information Bulletin for Shoeprint/Toolmark Examiners*, *4*, pp.69-81.

[95] Meijerman, L., Nagelkerke, N.J.D., Van Basten, R., Van Der Lugt, C., De Conti, F., Drusini, A.G., Giacon, M., Sholl, S., Vanezis, P. and Maat, G.J.R., 2006. Inter-and intra-individual variation in applied force when listening at a surface, and resulting variation in earprints. *Medicine, science and the law*, *46*(2), pp.141-151.

[96] Meijerman, L., Sholl, S., De Conti, F., Giacon, M., van der Lugt, C., Drusini, A., Vanezis, P. and Maat, G., 2004. Exploratory study on classification and individualisation of earprints. *Forensic Science International*, *140*(1), pp.91-99.

[97] Alberink, I. and Ruifrok, A., 2007. Performance of the FearID earprint identification system. *Forensic science international*, *166*(2-3), pp.145-154.

[98] Nixon, M. and Aguado, A., 2019. *Feature extraction and image processing for computer vision*. Academic press.

[99] Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). Ieee.

[100] Song, Z., Zhou, S. and Guan, J., 2013. A novel image registration algorithm for remote sensing under affine transformation. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(8), pp.4895-4912.

[101] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[102] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, pp.1097-1105.

[103] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

[104] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[105] Pan, S.J. and Yang, Q., 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), pp.1345-1359.

[106] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

[107] Burge, M. and Burger, W., 2000, September. Ear biometrics in computer vision. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* (Vol. 2, pp. 822-826). IEEE.

[108] Carreira-Perpinan, M.A., 1995. Compression neural networks for feature extraction: Application to human recognition from ear images. *MSc thesis, Faculty of Informatics, Technical University of Madrid*.

[109] http://www1.ustb.edu.cn/resb/en/doc/Imagedb_123_intro_en.pdf Accessed: 2021-04-09

[110] http://awe.fri.uni-lj.si/ Accessed: 2021-04-09

[111] https://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Ear.htm Accessed: 2021-04-09

[112] http://robotics.csie.ncku.edu.tw/Databases/FaceDetect_PoseEstimate.htm Accessed: 2021-04-09

[113] https://www.nist.gov/srd/nist-special-database-18 Accessed: 2021-04-09

[114] https://www.nist.gov/itl/products-and-services/color-feret-database Accessed: 2021-04-09

[115] https://cvrl.nd.edu/projects/data/ Accessed: 2021-04-09

[116] Raposo, R., Hoyle, E., Peixinho, A. and Proença, H., 2011, April. Ubear: A dataset of ear images captured on-the-move in uncontrolled conditions. In *2011 IEEE workshop on computational intelligence in biometrics and identity management (CIBIM)* (pp. 84-90). IEEE.

[117] Hoang, V.T., 2019. EarVN1. 0: A new large-scale ear images dataset in the wild. *Data in brief*, *27*.

[118] https://en.wikipedia.org/wiki/Darwin%27s_tubercle Accessed: 2021-04-09

[119] Rutty, G.N., Abbas, A. and Crossling, D., 2005. Could earprint identification be computerised? An illustrated proof of concept paper. *International Journal of Legal Medicine*, *119*(6), pp.335-343.

[120] Zeng, H., Dong, J.Y., Mu, Z.C. and Guo, Y., 2010, October. Ear recognition based on 3D keypoint matching. In *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS* (pp. 1694-1697). IEEE.

[121] Kamboj, A., Rani, R., Nigam, A. and Jha, R.R., 2021. CED-Net: context-aware ear detection network for unconstrained images. *Pattern Analysis and Applications*, *24*(2), pp.779-800.

[122] Cintas, C., Delrieux, C., Navarro, P., Quinto-Sánchez, M., Pazos, B. and Gonzalez-José, R., 2019. Automatic Ear Detection and Segmentation over Partially Occluded Profile Face Images. *Journal of Computer Science and Technology*, *19*(01), pp.e08-e08.

[123] Prakash, S. and Gupta, P., 2012. An efficient ear localization technique. *Image and Vision Computing*, *30*(1), pp.38-50.

[124] Kumar, A. and Wu, C., 2012. Automated human identification using ear imaging. *Pattern Recognition*, *45*(3), pp.956-968.

[125] Tian, L. and Mu, Z., 2016, October. Ear recognition based on deep convolutional network. In *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 437-441). IEEE.

[126] Hassaballah, M., Alshazly, H.A. and Ali, A.A., 2019. Ear recognition using local binary patterns: A comparative experimental study. *Expert Systems with Applications*, *118*, pp.182-200.

[127] Kamboj, A., Rani, R. and Nigam, A., 2021. CG-ERNet: a lightweight Curvature Gabor filtering based ear recognition network for data scarce scenario. *Multimedia Tools and Applications*, pp.1-43.

[128] Hassin, A. and Abbood, D., 2021. Human–Ear Recognition Using Scale Invariant Feature Transform. *Artificial Intelligence & Robotics Development Journal*, pp.1-12.

[129] Alkababji, A.M. and Mohammed, O.H., 2021. Real time ear recognition using deep learning. *Telkomnika*, *19*(2), pp.523-530.

[130] Priyadharshini, R.A., Arivazhagan, S. and Arun, M., 2021. A deep learning approach for person identification using ear biometrics. *Applied Intelligence*, *51*(4), pp.2161-2172.

[131] https://en.wikipedia.org/wiki/International_Organization_for_Standardization Accessed: 2021-07-10

[132] https://www.iso.org/committee/45306.html Accessed: 2021-07-10

[133] https://www.iso.org/committee/313770.html Accessed: 2021-07-10