1)Paper Name: Deep Ear Biometrics for Gender Classification.

Abstract: Human gender classification based on biometric features is a major concern for computer vision due to its vast variety of applications. The human ear is popular among researchers as a soft biometric trait, because it is less affected by age or changing circumstances and is non-intrusive. In this study, we have developed a deep convolutional neural network (CNN) model for automatic gender classification using the samples of ear images. The performance is evaluated using four cuttingedge pre-trained CNN models. In terms of trainable parameters, the proposed technique requires significantly less computational complexity. The proposed model has achieved 93% accuracy on the EarVN1.0 ear dataset.
Keywords: Convolutional neural network (CNN) · Gender classification · Ear Biometric · Soft biometrics

Introduction: Gender identification has become a major concern due to its vast variety of applications, including social communication and connection, commercial visual supervision, banking transactions, illness prognosis, demographic data collection, artificial intelligence (AI)-based user interface for customization, consumer analysis for business growth, and many more [17, 35]. Biometric traits have proven their suitability in gender classification as they are non-intrusive, remain invariant with time, and are less influenced by emotions and circumstances. Several approaches have been proposed using various biometrics traits and their fusion. Biometrics can be broadly categorized into the conventional (aka hard) biometrics (e.g., face, hand, etc.) [5, 18] and soft biometrics (e.g., ear, gait, etc.) [26, 34]. In classifying human gender, based on the iris samples, Bansal and Sharma [2] have employed wavelet transforms and statistical methods. A technique to identify the gender of complete body photographs using part-based gender recognition is proposed in [7]. To determine the dialect and gender from parlance data, supervised machine learning techniques have been employed [13]. Tapia et al. [30] used the local binary patterns (LBP) and Histograms of Gaussian (HOG) descriptors for gender identification from iris images. Gait-based gender categorization methods are suggested by Yu et al. [36] and Li et al. [22]. Shan [28] has used real-world facial pictures and the LBP description to categorize gender. A novel patch-based LBP with adjustable weights is presented by Chen and Jeng [9] for the classification of gender on the person's face. Moreover, a human ear is an unique evidence among the existing biometric traits for gender identification. The anatomy of the human ear is regarded to be equally significant as compared to other biometric traits like the face, iris, and gait, for identifying an individual. Additionally, it is less affected by aging when compared to other biometric traits like the face [16] and gait [22]. Other biometric traits may change with age or changing circumstances, but the ear simply scales up. Numerous methods for ear identification have been proposed. Anwar et al. [1] use geometrical information from human ear images for identification. Various image descriptors have been studied including LPQ, BSIF, LBP, HOG, and surface descriptors in [27]. Gender classification from ear images has also been studied in [11, 12, 19, 34]. However, the effectiveness of ConvNet for gender identification is yet to be researched thoroughly. In this study, we suggest a basic yet successful CNN-based model that is computationally lightweight. For the experiment, we have considered the EarVN1.0 large-scale ear dataset and achieved significantly good accuracy.

2)Paper Name:Improved Ear Verification with Vision Transformers and Overlapping Patches.

Abstract—Ear recognition has emerged as a promising biometric modality due to the relative stability in appearance during adulthood. Although Vision Transformers (ViTs) have been widely used in image recognition tasks, their efficiency in ear recognition has been hampered by a lack of attention to overlapping patches, which is crucial for capturing intricate ear features. In this study, we evaluate ViT-Tiny (ViT-T), ViT-Small (ViT-S), ViT-Base (ViT-B) and ViT-Large (ViT-L) configurations on a diverse set of datasets (OPIB, AWE, WPUT, and EarVN1.0), using an overlapping patch selection strategy. Results demonstrate the critical importance of overlapping patches, yielding superior performance in 44 of 48 experiments in a structured study. Moreover, upon comparing the results of the overlapping patches with the non-overlapping configurations, the increase is significant, reaching up to 10% for the EarVN1.0 dataset. In terms of model performance, the ViT-T model consistently outperformed the ViT-S, ViT-B, and ViT-L models on the AWE, WPUT, and EarVN1.0 datasets. The highest scores were achieved in a configuration with a patch size of 28×28 and a stride of 14 pixels. This patch-stride configuration represents 25% of the normalized image area (112x112 pixels) for the patch size and 12.5% of the row or column size for the stride. This study confirms that transformer architectures with overlapping patch selection can serve as an efficient and high-performing option for ear-based biometric recognition tasks in verification scenarios.

INTRODUCTION: Ear recognition as a biometric modality has a long history, dating back many decades [13], and has seen attention as a possible alternative to traditional face recognition approaches. Ears have a relatively stable structural shape throughout an individual's adult lifetime [40], [45] and their images may be captured from a distance with little or no subject involvement, making ear recognition an appealing alternative to other modalities. Early studies on automatic ear recognition began in the mid-1990s, with researchers using geometrical analysis of ear contours and structures [14], [33]. More recent and advanced methods have displayed improved robustness to effects such as illumination and pose variations [25]. The accuracy of ear recognition systems has improved significantly in the last decade [38], [12], primarily driven by the advancements in deep learning methods. The large Unconstrained Ear Recognition Challenge (UERC) datasets [22], [19] have played a critical role in reigniting interest as they capture a variety of ear appearances encountered in real world scenarios—such as variances in pose, illumination, occlusion, and resolution. Recent advances in computer vision include the emergence of Vision Transformers (ViTs) [18] for recognition. These networks employ self-attention processes to identify both local and global patterns in image data. ViTs operate on a tokenized input consisting of image patches, allowing them to concentrate on broader contextual relationships than convolutional neural networks (CNNs), which are dependent on fixed receptive fields [31]. This patch-based approach makes ViTs particularly flexible for various image recognition applications, including ear recognition. The ViTEar work [19] fine-tuned pre-trained DINOv2 Vision Transformers [37] with datasets like UERC2023 [19] and EarVN1.0 [27] to obtain a rank-1 accuracy of 96.27% on the UERC2019 dataset [22]. Another research team [32] combined CNNs with ViTs to capitalize on their individual strengths, obtaining 99.36% and 91.25% accuracy on the Kaggle [26] and IITD-II [1] datasets, respectively. Furthermore, research [5] comparing ViT [18] and DeiT-based [42] models to CNNs found that transformer models outperform CNNs even without extensive data augmentation. Previous works exploring the use of ViTs for ear recognition [5], [32] trained ViTs on smaller datasets and

evaluated one or two datasets with fewer subjects in the test set due to computational constraints. In addition, the existing work not explore the use of overlapping patches [48], which can more effectively model features that extend beyond a patch boundary, leading to improved feature representation. Overlapping patches help capture both the small details and the larger, continuous structures of the human ear. Figure 2 illustrates the comparison between overlapping and nonoverlapping Patches. In our work, as shown schematically in Fig. 1, we train customized ViT models (obtained from the Inside Face repository) [11] using T, S, B, and L configurations on the UERC2023 dataset [19] (excluding the AWE component) and evaluate it on four datasets: OPIB [3], AWE [21], WPUT [24], and EarVN1.0 [27]. We also explore the impact of overlapping patches in ViTs. By employing a large dataset for training, we can reduce the risk of overfitting during training. Furthermore, testing on several different datasets gives a broader understanding of how the trained model reacts to variations in data distribution and subject demographics, revealing the model's generalizability to multiple settings.
The contributions of this work are:
1) A systematic study of the use of four transformer models trained on the same large and diverse data set.
2) The results of experiments with several patch-stride configurations, providing a comprehensive evaluation of the impact of different patch size and overlapping stride settings.
Our work establishes a robust benchmark for ear recognition in a verification scenario using ViTs, paving the way for future research and advancements in this domain.

3)Paper Name: Multimodal soft biometrics: combining ear and face biometrics for age and gender classification.

Abstract: In this paper, we present a multimodal, multitask deep convolutional neural network framework for age and gender classification. In the developed framework, we have employed two different biometric modalities: ear and profile face. We have explored three different fusion methods, namely, data, feature, and score fusion, to combine the information extracted from ear and profile face images. In the framework, we have utilized VGG-16 and ResNet-50 models with center loss to obtain more discriminative features. Moreover, we have performed two-stage fine-tuning to increase the representation capacity of the models.

To assess the performance of the proposed approach, we have conducted extensive experiments on the FERET, UND-F, and UND-J2 datasets. Experimental results indicate that ear and profile face images contain useful features to extract soft biometric traits. We have shown that when frontal face view of the subject is not available, use of ear and profile face images can be a good alternative for the soft biometric recognition systems. The presented multimodal system achieves very high age and gender classification accuracies, matching the ones obtained by using frontal face images. The multimodal approach has outperformed both the unimodal approaches and the previous state-of-the-art profile face image or ear image-based age and gender classification methods, significantly in both tasks.

Keywords: Multimodal learning · Multitask learning · Soft biometrics · Age estimation ·Gender classification · Convolutional neural networks

Introduction: Extracting soft biometric traits is an important research topic in biometrics [16, 17, 29, 33].It has been found that by utilizing soft biometric traits, subjects can be described better and this way the identification performance can be improved [16, 17, 24]. Two of the most widely used soft biometric traits are subjects' age and gender. Recently, with the convolutional neural networks (CNN) based approaches, superior results have been obtained for person identification, age estimation, and gender classification using ear and profile face images [22, 28, 35, 36]. There have been several works about extracting soft biometric traits, age and gender, using these modalities [2, 12, 14, 18, 21, 23, 25, 27, 30, 35]. In addition to the unimodal approaches, there has been one multimodal work that utilized both ear and profile face images to perform gender classification [39]. Most of these previous works have focused on gender classification. There is only one recent work [35] that has performed age classification from ear images, one that has performed age estimation from profile face images [5], and one multimodal approach [36]. Frontal face images have been used in many different biometric studies [3, 6, 8, 9, 22, 24, 28]. However, depending on the application, it may not be possible to always obtain frontal face images. For example, in a surveillance scenario, it would be very rare to acquire frontal faces of the subjects. In these situations, profile face and ear could be useful as alternative biometric modalities. Besides, ear biometrics has certain advantages, for instance, toleration to lighting changes compared to profile and frontal face images [10]. In this paper, extending our previous work

[36], we have presented a comprehensive study on age and gender classification using ear and profile face images as input biometric modalities. We have developed and investigated several multimodal and multitask deep CNN frameworks.We have explored three different fusion methods: data fusion —intensity fusion, spatial fusion, channel fusion—, feature fusion, and score-level fusion. In the study, we have employed two well-known deep CNN models, namely VGG-16 [32] and ResNet-50 [13]. To obtain more discriminative features, besides softmax loss, we have also benefited from center loss. In addition, we have performed domain adaptation via a two-stage fine tuning approach to increase the representation capability of the models. In this work, in addition to [36], we examined age regression which is one of the most important task in soft-biometric analysis. We also deeply analyzed the performances such as ± age classification accuracy, effect of context information over performance, effect of accessories, and etc. Besides, we trained a model using frontal face data for age and gender classification on the datasets that are used in this work in order to compare different biometric modalities. Please note that, profile face images can already contain ear region, when they are cropped using a large bounding box. However, when we enlarge the bounding box of the profile face images, then hair and background information are also included. In our experiments we have observed that including these irrelevant information degrades the performance. We have conducted extensive experiments on the UND-F, UND-J2, and FERET datasets [26, 37]. Experimental results have shown that profile face images contain a valuable source of information for age and gender classification. The proposed multimodal system has achieved very high age and gender classification accuracies. Moreover, we attained superior results compared to the state-of-the-art profile face image or ear image-based age and gender classification methods.

The contributions of the study can be summarized as follows:

– We have presented a multimodal, multitask deep convolutional neural network approach for age and gender classification.

– We have thoroughly explored several ways of benefiting from multimodal input for age and gender classification. We have investigated three different data fusion methods, as well as feature and score level fusion.

– We have adapted the utilized deep CNN models, namely VGG-16 and ResNet-50, to the
ear domain by using a two-stage fine-tuning approach. For this purpose, we have generated the extended version of theMulti-PIE ear dataset that was presented in

our previous work [35] and named it Multi-PIE extended-ear dataset. Moreover, we have employed center loss in combination with the softmax loss to learn more discriminative features.
– We have provided class activation maps to observe the CNN behaviour under different circumstances.
– We have conducted a comprehensive experimental analysis. We have used the UND-F, UND-J2, and FERET datasets for gender classification, and only the FERET dataset for age classification, since the UND-F and UND-J2 datasets do not contain age labels. We have achieved state-of-the-art age and gender classification results on these datasets.
– We have also performed age and gender classification experiments using the frontal face images of the same subjects in order to determine the usefulness of profile face and ear images as an alternative to the frontal face image.We have found that the accuracies achieved by the proposed multimodal approach are very close to the ones obtained by using the frontal face image as input. The remainder of the paper is organized as follows: In Section 2, we provide a brief overview of the previous works on the topic. In Section 3, we explain the used CNN architectures, proposed fusion methods, and two-stage fine-tuning strategy. In Section 4, we present the datasets, experimental setups, and experimental results. Finally, Section 5 concludes the paper.

4)Paper Name: Age and Gender Classification from Ear Images.

Abstract—In this paper, we present a detailed analysis on extracting soft biometric traits, age and gender, from ear images. Although there have been a few previous works on gender classification using ear images, to the best of our knowledge, this study is the first work on age classification from ear images. In the study, we have utilized both geometric features and appearance-based features for ear representation. The utilized geometric features are based on eight anthropometric landmarks and consist of 14 distance measurements and two area calculations. The appearance-based methods employ deep convolutional neural networks for representation and classification. The well-known convolutional neural network models, namely, AlexNet, VGG-16, GoogLeNet, and SqueezeNet have been adopted for the study. They have been fine-tuned on a large-scale ear dataset that has been built from the profile and close-to-profile face images in the Multi-PIE face dataset. This way, we have performed a domain adaptation. The updated models have been fine-tuned once more time on the small-scale target ear dataset, which contains only around 270 ear images for training. According to the experimental results, appearance-based methods have been found to be superior to the methods based on geometric features. We have achieved 94% accuracy for gender classification, whereas 52% accuracy has been obtained for age classification. These results indicate that ear images provide useful cues for age and gender classification, however, further work is required for age estimation.

Index Terms—Age and gender classification, deep learning, geometric features

INTRODUCTION: Ear biometrics has become a popular research topic [1].
A recent challenge, named as Unconstrained Ear Recognition Challenge [2] has shown the difficulties of performing person identification from ear images in the wild. To complement the identity related information from ear images, utilizing soft biometric traits, such as age and gender information can be auxiliary. For this purpose, in this paper, we have extensively investigated the tasks of age and gender classification from ear images. Biometric characteristics are expected to not change much over time, easy to obtain and unique for each individual [3]. Because of its several features, ear is an important modality in biometric studies and forensic science for identification. For example, compared to facial appearance, which is influenced by changes in facial expression, facial hair or makeup, ear appearance is relatively constant. Auricular is also a defining feature of the face [4]. Among the ear parts, earlobe is the most frequently used part in forensic cases. It is the only part of the ear that continues to grow and changes their shape [5]. Ear can be still visible in the whole or partly covered face in the captured images from security cameras, and can be used as an auxiliary information for identification. Also, when faces are viewed in profile, ear can be easily captured from video recordings or photos [6]. Although there have been many studies on using ear images for person identification [1], [6], the number of studies on extracting soft biometric traits, such as age and gender, from ear images is limited. To the best of our knowledge, this study is the first work on age classification from ear images. However, there have been a couple of previous work on using ear images for gender classification [7], [8], [9], [10]. In [7], the ear-hole is used as the reference point for the measurements. The Euclidean distances between the ear hole and seven features of ear, which are

identified from masked ear images, are calculated. They used an internal database, which has 342 samples, for the experiments. They have employed Bayes classifier, KNN classifier, and neural networks. The best performance is achieved by KNN with 90.42% classification accuracy. In [8], profile face images and ear images are used separately and are classified by support vector machines (SVM) with histogram intersection kernel. They performed score level fusion based on Bayesian analysis to improve the accuracy. The 2D images of UND biometrics dataset collection F [11] have been used for the experiments. Fusion leads to 97.65% accuracy, whereas face only performance is around 95.43% and ear only accuracy is around 91.78%. In [9], Gabor filters have been utilized to extract features and classification has been performed with extracted features based on dictionary learning. The dictionary has been built from training samples and used in the test phase to represent a test sample as a linear combination of the training data. UND biometric dataset collection J [11], which contains large appearance, pose, and illumination variability, has been used in the experiments. The best obtained accuracy reported in the paper is 89.49% has been achieved by using 128 features. In [10], gender classification is performed both on 2D and 3D ear images. 3D ears are automatically detected and aligned. The experiments were performed on UND dataset collections F and J2 [11]. Histogram of Indexed Shapes features were extracted and classified by SVM. The average performance of the system was 92.94%. In this paper, we present an extensive analysis on age and gender classification from ear images.We have explored use of both geometric features and appearance-based features for ear representation. Geometric features are based on eight landmarks determined on the ear. From these landmarks, to extract the features, we have calculated 14 different distances between them as well as performed two area calculations. To classify these extracted features, four different classifiers —logistic regression, random forests, support vector machines, neural networks— have been employed. The appearance-based methods are based on wellknown deep convolutional neural network (CNN) models, namely, AlexNet [12], VGG-16 [13], GoogLeNet [14], and SqueezeNet [15]. They have been fine-tuned twice, first on a large-scale ear dataset to provide domain adaptation, then on the small-scale target ear dataset. In the experiments, appearance-based methods have outperformed geometric feature-based methods. We have achieved 94% accuracy for gender classification, exceeding the attained accuracies in the previous studies. For age classification 52%

accuracy has been obtained. In summary, the contributions of the paper can be listed as below:
–We have explored geometric and appearance-based features for age and gender classification from ear images.

—For geometric features, we have used eight landmark points on the ear and derived 16 features from them.

–For appearance-based methods, we have utilized a large-scale ear dataset [16], which has been built from the profile and close-to-profile face images in the Multi-PIE face dataset [17]. This way, we have efficiently transferred and benefited from the wellknown CNN models for the problem at hand.

–We have achieved superior performance for gender classification compared to the previous work. We have presented the first work on age classification from ear images.
The remainder of the paper is organized as follows. In Section II, we explain the geometric features, the classifiers used with them, and the convolutional neural networks used for ear appearance representation and classification. In
Section III, we introduce the dataset and experimental setup, and present the obtained results. Finally, in Section IV, we conclude the paper and point the future research directions

5)Paper Name:Biometric Recognition of Infants using Fingerprint, Iris, and Ear Biometrics

ABSTRACT: Biometric recognition is often used for adults for a variety of purposes where an individual's identity must be ascertained. However, the biometric recognition of children is an unsolved challenge. Solving this challenge could protect children from identity theft and identity fraud, help in reuniting lost
children with their parents, improve border control systems in combatting child trafficking, and assist in electronic record-keeping systems. In order to begin the development of biometric recognition systems for children, researchers collected fingerprint, iris, and outer ear shape biometric information from infants. Each modality provides different challenges. Where possible, the performance of existing hardware and software that was developed for adults was assessed with infants. Where necessary, novel hardware or software was developed. For the

ear modality, existing hardware and software which have previously been applied to adults were applied to children. For the iris modality, existing hardware was used to acquire the images, while adjustments to the existing preprocessing algorithms were applied to cater for the localisation and segmentation of infant irises. For the fingerprint modality, novel hardware and image processing software were developed to acquire fingerprints from infants, and convert the images into a format which is backward compatible with existing international standards for minutiae extraction and comparison. The advantages and disadvantages of using each of these modalities during the rst year of life were compared, based on both qualitative assessments of usage, and quantitative assessments of performance. While there is no conclusively best modality, recommendations of usage for each modality were provided.
INDEX TERMS: Authentication, biometrics, ear recognition, fingerprint recognition, identification of persons, identification of infants, identity management systems, iris recognition

INTRODUCTION: Recognition of infants and minors precisely from birth is becoming ubiquitous. The choice of biometric modality to use for infants and minors has always been a bottleneck due to imaging devices and the uncooperative nature of infants. To mitigate these challenges a research project has been started with the aim of developing a prototype biometric recognition system to acquire biometric data from young children, and determine or verify the identities of these children from birth until they apply for their identification documents (which can be done at the age of 16 years in South Africa). To assess the performance of the existing and newly developed biometric acquisition and recognition systems for children and achieve the aim of the project, it is required to acquire biometric data from children and successfully compare this biometric data. The benets of developing such a system are manifold. The output of this research is meant to address issues of identity theft and fraud against children, help combat child trafficking, assist with reuniting small children who are lost with their parents, and improve healthcare management systems for children [1][5] The unique challenge that is posed is that existing technologies are not capable of acquiring biometric information from newborn infants and successfully matching it to the same individuals during growth and adulthood with accuracy and reliability, thus leaving children vulnerable to exploitation in various ways, such as identity theft and child trafficking. As rst step in solving this challenge, this paper addresses the acquisition of biometric information from children during the rst year of life. There has been some research into developing biometric recognition systems for children. However, there are still challenges to overcome in creating a complete biometric system for infants and minors. A review of several modalities was performed before reaching the decision to focus on the fingerprint, iris, and outer ear shape. These modalities were chosen after an assessment based on seven criteria of the desirability of biometric characteristics, namely: universality, uniqueness, permanence, collectability, performance, acceptability, and resistance to circumvention [6]. The analysis is summarised in Figure 1 and discussed in detail below. Face [7][11] and speech [12], [13] biometrics may work for older children but are ineffective for newborn babies and toddlers. Footprint crease patterns [14][16] are promising for newborns but become less user-friendly as people become older and start wearing shoes. There are also concerns regarding the hygiene of feet which may come into contact with biometric sensors. These concerns also translate to research into using friction ridge patterns of the feet [17]. However, friction ridge patterns of the fingers (also called fingerprints) [1], [3], [18][21] and palms (palm prints) [17], [22] have shown more promise. The main challenge to acquiring fingerprints is that conventional fingerprint scanners do not acquire fingerprints at a sufficiently high resolution to resolve the fingerprints of newborn infants, and the contact nature of conventional scanners may, at times, be incompatible with the soft, malleable skin of infants. One approach has been to use higher resolution contact-based scanners to increase the accuracy of using a single fingerprint [1], [3], [18][21]. Another approach has been to collect fingerprints from all 10 fingers using a conventional scanner and fuse the scores for higher reliability [11]. While this latter approach has resulted in a high level of accuracy for toddlers aged 18 months and older, it may be difficult and time consuming to collect all 10 fingerprints from babies. Furthermore, the reliability as children grow bigger and the reliability of this method below the age of 18 months remains an open question. In this paper, we have proposed to use a novel fingerprint scanner which is a contactless device that uses a higher resolution than the previously cited works [1], [3], [11], [18][21]. While the friction ridge patterns on palms are conceptually similar to fingerprints and may be ergonomically easier to capture from infants, palms present other challenges. Due to the much larger area, hardware costs and data transfer requirements would increase if the full area of the palm is acquired. Alternatively, if a sub-region of the palm is acquired, consistency in repeatedly acquiring the same region may prove challenging. Two other biometrics which have shown promise for young children are the outer ear shape [23][25] and the iris [11]. The advantage of the outer ear is that the collection of the biometric data is unobtrusive and hygienic since it is completely touchless. There is currently little research and commercial work done on ear recognition for children [26]. This includes work done by Tiwari et al. [23], [24], [27], [28] and Berra et al. [29] who attempted different recognition methods of newborns using ear images from hospitals. Kumar et al. [30] and Ntshangase et al. [31] evaluated the performance of recognition algorithms on ear recognition for children. There is still missing information in this eld that needs to be addressed, such as the effect of growth on ear recognition, more details are presented in a paper by Ntshangase and Mathekga [26]. However, a larger dataset and longitudinal studies are required to obtain more reliable information about the permanence of the shape of the ear and the performance of ear recognition for children. The iris is known to be effective for recognition from the age of 18 months and upwards. Daugman demonstrated in his pioneering iris recognition work that its recognition accuracy is seven times more than its major rival the fingerprint 32][36]. Even though Daugman reported high accuracy of the iris recognition system, no research was found where his works were extended to iris biometric recognition for infants and minors. The performance of iris image acquisition and recognition for children needs to be investigated. Preliminary findings of this research suggest that the variance in image quality between adult and infant iris images is minute [37]. In summary, in this paper, we report on the efforts towards developing and assessing three biometric recognition systems for infants using the fingerprint, iris, and outer ear shape biometric modalities. These systems have been developed independently, however, the long-term aim is to eventually fuse all three modalities in future work. This is done to improve the accuracy of each individual modality before they can be fused together. The rest of the paper is structured in the following manner. Section II discusses related work. Section III describes the approach taken for each biometric modality. Section IV provided the experimental results and a discussion thereof, as well as a discussion on the lessons learned in this endeavor. Section V concludes the discourse and suggests avenues for future work.