Person Identification by Models Trained Using Left and Right Ear Images Independently

K. R. Resmi[1](✉), G. Raju[2], Vijaya Padmanabha[3], and Joseph Mani[3]

[1] Santhigiri College, Kerala, India
resmykr@gmail.com
[2] Christ Deemed to be University, Bengaluru, India
[3] Modern College of Business and Science, Bowshar, Muscat, Sultanate of Oman
{Vijaya.Padmanabha,drjosephmani}@mcbs.edu.om

Abstract. The application of Deep Learning Techniques in biometrics has grown significantly during the last decade. The use of deep learning models in ear biometrics is restricted due to the lack of large ear datasets. Researchers employ transfer learning based on several pretrained models to overcome the limitations. For the unconstrained AWE ear dataset, traditional Machine Learning (ML) techniques and hand-crafted features fall short of providing a good recognition accuracy. This paper evaluates the influence of separating left and right ears and the effect of occlusion on the recognition accuracy in AWE dataset. The left and right ear of a person need not be identical. A study by separating the left and right ear into two different datasets is carried out with the pretrained ResNet50 based model. There is a remarkable increase in accuracy when the left and right ear images are independently considered. A new data augmentation technique, incorporating occlusion, is also proposed and experimented with the ResNet50 based model.

Keywords: Ear Recognition · Deep Learning · ResNet50 · Occlusion

1 Introduction

Authentication of an individual plays a critical role in applications like access control, security, surveillance, forensic etc. Personal authentication using Biometric is more prominent compared to conventional authentications like passwords, tokens, pins etc. Ear is an emerging physiological biometric trait that can be used to identify an individual. The most important feature of ear is its unique structure. Unlike face, ear features are not affected by mood, expression or aging. Feature extraction and selection plays a vital role in recognition accuracy using classical approaches. In recent years, deep convolutional neural network provides notable performance in tasks such as image classification [1, 2], medical image analysis [3], face recognition [4], object detection [5] and so on.

Ear recognition results using classical approaches based on handcrafted features gave low recognition accuracy using unconstrained ear datasets like AWE. Considerable improvement in recognition results on unconstrained datasets were reported using deep learning approaches based on CNN.CNN requires large dataset for training. The

N. Bacanin and H. Shaker (Eds.): ICIITB 2022, ACSR 104, pp. 281–288, 2023.
https://doi.org/10.2991/978-94-6463-110-4_20

major challenge in ear recognition using CNN is the limited ear dataset. To solve the problem of limited data, deep learning uses the technique called data augmentation and transfer learning. Data augmentation produce extensive training data by introducing more variations in input data. By using the concept of transfer learning, a network trained for one application can be reused for other applications with minimal time and resources.

In this paper, the focus is mainly to study the performance of deep CNN in ear recognition. Due to its computational efficiency, pretrained network ResNet50 [6] is adopted for feature extraction and classification. Present study proposed a classification method by separating the left and right ears for improving accuracy. A new data augmentation scheme by introducing occlusion is also proposed.

2 Related Works

Ear recognition are classified into geometrical, holistic, local and hybrid techniques based on handcrafted features [7]. The selection of the appropriate features and classification techniques is the key issue with handcrafted approaches. Ear recognition results using unconstrained ear datasets gave low recognition results using traditional feature extraction techniques. Ear recognition works using CNN reports high recognition results in unconstrained datasets. This section reviews some of the ear recognition models using CNN in unconstrained ear data set AWE.

Transfer learning using AlexNet, VGG-16, and SqueezeNet were presented in [8]. Selective learning using sqeezeNet, reports rank1 accuracy of 62% on combination of datasets AWE, CVLED And 500 images from internet. The same authors presented the unconstrained ear recognition challenge (UERC) in 2017 [9] and 2019 [10] respectively for personal identification in an uncontrolled environment. Several researchers participated in both the challenges and designed models using conventional features, deep learning methods, and a combination of both. Hansley et al. used a technique for ear recognition that combined hand-crafted characteristics with features from a CNN model created for face recognition [11]. On the AWE dataset, the combination of CNN features with HOG produced the greatest Rank1 recognition rate of 75.6%. Ear recognition using pretrained AlexNet, VGG 19, ResNet18, VGG16 and ResNet 50 is presented in [12]. Rank1 accuracy of 94.08%, 56.35%, 80.05% were reported on the CVLE, AWE and AWE + CVLE datasets. Ear recognition using Alexnet, ResNet50 and VGG16 were proposed in [13]. Rank1 accuracy of 63% was reported on AWE dataset using ResNet50 model. Ensemble based pretrained models was presented in [14]. Accuracy of 67.25%was reported on AWE dataset using average ensemble of pretrained networks.

The literature survey is limited to only works related to the AWE dataset. Several new unconstrained ear datasets with huge samples are available for experiments now a days.

3 CNN Pretrained ResNet50 Model

A powerful deep CNN called ResNet, also known as Residual Networks, is used in numerous computer vision tasks [6]. Deeper networks produce less accurate test results but learn more from the data. ResNet employs residual blocks for learning to address this issue [4]. ResNet's central concept is identity shorten connection by omitting one or more layers. ResNet comes in a variety of forms, and ResNet-50 is employed in this study.

4 Experiments

Tensorflow is used for the model building process in the Python environment (Keras). For the experiment, a computer with an Intel Core i5 processor and 4GB of RAM is used. The details of dataset and augmentation techniques used in the study are described below.

4.1 Data Set

Unconstrained dataset AWE [7] is used for experimentation. AWE contains 100 subjects and each subject contain tightly cropped ten images of varying size and quality. Both left and right ears of same subject are included in the database. The problem of over fitting in AWE is solved by using various data augmentation techniques. Figure 1 shows sample images from AWE dataset.

4.2 Data Augmentation

Data augmentation techniques generates vast data from limited data. More training data generated after data augmentation improves the efficiency of the model by lowering over fitting. A python library called "imgaug tool" is used to generate different augmented data [15]. Figure 2 shows augmented images generated from a sample image in AWE dataset using standard augmentation techniques.

To study the influence of occlusion in ear-recognition performance, one more data augmentation technique is incorporated by using controlled occlusion. By randomly

Fig. 1. Sample images from AWE Dataset

Fig. 2. Sample augmented images

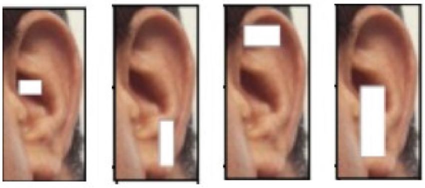

Fig. 3. Sample augmented images with different rectangular occlusion

Table 1. Rank-1 recognition accuracy of AWE on ResNet50

Rank1 accuracy (%) AWE on ResNet50		
Classification	Without Augmentation	Using Augmentation
Features (SVM)	36.25%	58.20%
Softmax	40.75%	63.00%

placing white rectangular blocks of varying sizes on the original image and allowing for a maximum of 30% occlusion, new augmented images are created. Figure 3 displays augmented images created with rectangular blocks and various occlusion percentages. A comparable investigation into the impact of occlusion on ear detection was conducted in [16].

5 Results and Discussion

The ResNet50 model receives the images from the AWE dataset as input. As a preprocessing step, all the images used as input for ResNet50 are resized to 224x224 pixels. Two methods are employed for classification [13]. The output of ResNet50's last convolutional layer is used in the first approach to extract features. In the second approach, 100 neurons are used in place of 1000 in ResNet50's final layer softmax. Rank-1 recognition accuracy generated by the two methods [13] on ResNet 50 model are shown in Table 1.

The results obtained using the proposed techniques in AWE are promising compared to the maximum accuracy of 49.6% reported in [7] using traditional features.

5.1 Occlusion in Ear Recognition

The problem of occlusion is common issue in real life ear recognition. Occlusions such as ear rings, hair, scarf/cap, spex, headphones etc. reduces the recognition accuracy. To study the effect of occlusion, a new data augmentation technique described in Sect. 4.2 is used. Occluded images were generated by adding rectangular block of various size at different positions in an ear image. In [16], a study is conducted to evaluate the effect of occlusion in ear detection accuracy. Occluded ear images are created by applying white rectangular blocks of 10% to 60% on top and left of ear images. Their study reported high detection rate up to 30% occlusion and the detection rate decreases above 30%

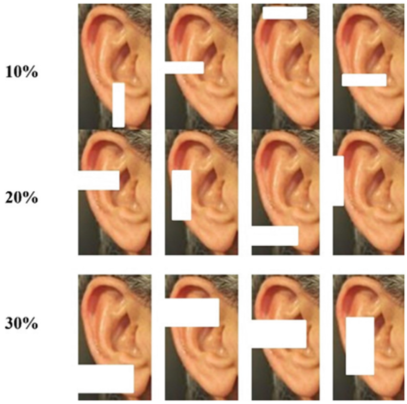

Fig. 4. Sample augmented images with different % of Occlusion

Table 2. Rank1 accuracy using different occlusion percentage

Rank-1 recognition accuracy [%]		
Epoch	Data Occlusion (%)	Accuracy
100	10	56.4%
	20	54.6%
	30	50.25%

occlusion. In the proposed method, augmented images are created using rectangular block size varying from 10% to 30% as shown in Fig. 4. For each sample in the dataset and for each occlusion percentage, 200 images with varying occlusion levels are created. Table 2 shows the result of recognition using different occlusion percentage. As occlusion increases, the accuracy using CNN found to decrease.

5.2 Experiments Considering Left and Right Ears Separately

In order to test whether there is an increase in the recognition performance if left and right ear images are considered separately, 480 left ear and 520 right ear images from AWE datasets are selected. ResNet50 is considered for experiments. The model is trained without data augmentation, augmentation by a factor of 10 times the original data and by a factor of 100 times the original data. Table 3 shows Rank-1 recognition accuracy.

As expected, the rank-1 recognition performance of ear recognition increases with data augmentation. There is a remarkable increase in the recognition accuracy when left, and right ear samples are separated into two subsets of data. Figures 5 and 6 depicts the performance. This opens up the possibility of a new recognition model by giving two different class labels for the same person's left and right ear.

Table 3. Rank-1 recognition accuracy Separating left and Right ears

Rank-1 recognition accuracy [%]				
	AWE Left ear database		AWE Right ear database	
Epoch	Data Augmentation	Accuracy	Data Augmentation	Accuracy
100	0	46.8%	0	49.3%
	10	73.33%	10	75.6%
	100	77.33%	100	80%

Fig. 5. Accuracy of the model on AWE left ear dataset with data augmentation of 100 time's original data

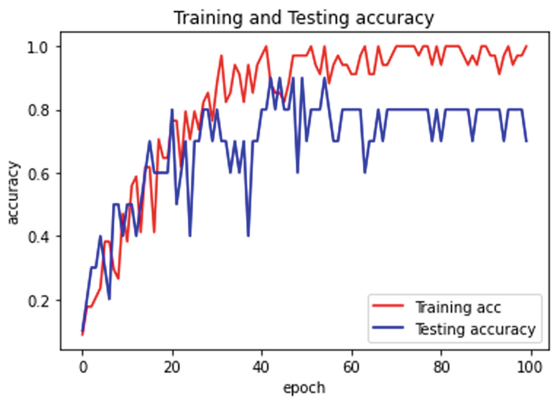

Fig. 6. Accuracy of the model on AWE Right ear dataset with data augmentation of 100 time's original data

6 Conclusion

This paper presents the usage of ResNet50 model to increase the ear recognition accuracy on unconstrained dataset AWE. To increase recognition accuracy experiments is conducted by separating the left and right ear. The effect of occlusion in ear recognition is also studied by creating an augmented database. A new data augmentation method, which introduces controlled occlusion, is used to analyze the variation in the recognition accuracy with the percentage of occlusion. The rank1 accuracy reported using ResNet50 on AWE is higher than the results using traditional features.

References

1. Krizhevsky, A., Sutskever, I.,Hinton, G. E.: ImageNet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems (2012). https://doi.org/10.1061/(ASCE)GT.1943-5606.0001284
2. Simonyan, K., Zisserman, A.:Very deep convolutional networks for large-scale image recognition, 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings (2015).
3. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., Sánchez, C. I.: A survey on deep learning in medical image analysis, In Medical Image Analysis (2017). https://doi.org/10.1016/j.media.2017.07.005
4. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)(2016). https://doi.org/10.1007/978-3-319-46478-7_31.
5. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence (2017). https://doi.org/10.1109/TPAMI.2016.2577031
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2016). https://doi.org/10.1109/CVPR.2016.90
7. Emeršič, Žiga, Vitomir Štruc, Peter Peer.: Ear Recognition: More than a Survey, Neurocomputing (2017). https://doi.org/10.1016/j.neucom.2016.08.139.
8. Emersic, Z., Stepec, D., Struc, V., Peer, P: Training convolutional neural networks with limited training data for ear recognition in the wild. In: Proceedings - 12th IEEE International Conference on Automatic Face and Gesture Recognition, FERA (2017). https://doi.org/10.1109/FG.2017.123
9. Emeršič, Ž., Štepec, D., Štruc, V., Peer, P., George, A., Ahmad, A., Omar, E., Boult, T. E., Safdaii, R., Zhou, Y., Zafeiriou, S., Yaman, D., Eyiokur, F. I., Ekenel, H. K.:The unconstrained ear recognition challenge, IEEE International Joint Conference on Biometrics, IJCB (2017). https://doi.org/10.1109/BTAS.2017.8272761
10. Emersic, Z., Kumar, A. S. V., Harish, B. S., Gutfeter, W., Khiarak, J. N., Pacut, A., Hansley, E., Segundo, M. P., Sarkar, S., Park, H. J., Nam, G. P., Kim, I. J., Sangodkar, S. G., Kacar, U., Kirci, M., Yuan, L., Yuan, J., Zhao, H., Lu, F., Štruc, V.: The Unconstrained Ear Recognition Challenge 2019, 2019 International Conference on Biometrics, ICB 2019, (2019). https://doi.org/10.1109/ICB45273.2019.8987337
11. Hansley, E.E., Segundo, M.P., Sarkar, S.: Employing fusion of learned and handcrafted features for unconstrained ear recognition. IET Biom. (2018). https://doi.org/10.1049/iet-bmt.2017.0210

12. Dodge, S., Mounsef, J., Karam, L.: Unconstrained ear recognition using deep neural networks. IET Biom. (2018). https://doi.org/10.1049/iet-bmt.2017.0208
13. Resmi, K.R., Raju, G. Ear Recognition Using Pretrained Convolutional Neural Networks. In: Singh, M., Tyagi, V., Gupta, P.K., Flusser, J., Ören, T., Sonawane, V.R. (eds) Advances in Computing and Data Sciences. ICACDS 2021. Communications in Computer and Information Science, vol 1440. Springer, Cham. (2021). https://doi.org/10.1007/978-3-030-81462-5_63
14. H. Alshazly, C. Linse, E. Barth, S. A. Idris and T. Martinetz, "Towards Explainable Ear Recognition Systems Using Deep Residual Networks," in IEEE Access, vol. 9, pp. 122254-122273, (2021), https://doi.org/10.1109/ACCESS.2021.3109441.
15. https://github.com/aleju/imgaug. Accessed 06 Oct 2022.
16. Ibrahim, M. I. S., Nixon, M. S., Mahmoodi, S.: Shaped wavelets for curvilinear structures for ear biometrics, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (2010).