JOIV : Int. J. Inform. Visualization, 7(4) - December 2023 2247-2254

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Comparative Analysis of VGG-16 and ResNet-50 for Occluded Ear Recognition

Hua-Chian Tey [a], Lee-Ying Chong [a,*], Siew-Chin Chong [a]

[a] *Faculty of Information Science and Technology, Multimedia University, Bukit Beruang, Melaka, 75450, Malaysia*

Corresponding author: [*]lychong@mmu.edu.my

***Abstract*— Occluded ear recognition is a challenging task in biometric systems due to the presence of occlusions that can hinder accurate identification. There is still a research gap in enhancing the robustness of deep learning to handle severities of occlusions with different datasets. This research focuses on developing a robust occluded ear recognition system by implementing fine-tuning techniques on three popular pre-trained deep learning models, Residual Neural Network (ResNet-50), Visual Geometry Group (VGG-16), and EfficientNet. The system is evaluated on two manually occluded ear datasets, which are the AMI ear dataset and the IITD ear dataset. The experiment results showed the fine-tuned ResNet-50 model performs better than the fine-tuned VGG-16 model. The results indicate that the model's ability to accurately predict the classes or labels decreases as more data is occluded. Higher occlusion rates lead to a loss of important information, making it more challenging for the model to distinguish between different patterns and make accurate predictions. According to the findings, the amount of occlusion influenced the identification accuracy and worsened as the occlusion became larger. In the future, ear recognition systems will likely continue to improve in accuracy and be adopted by a wider range of organizations and industries. They may also be integrated with other biometric technologies and used for personalization purposes. However, ethical considerations related to the use of ear recognition systems will also need to be addressed.**

***Keywords*— Occluded ear; fine-tune; ResNet-50; VGG-16; EfficientNet.**

Manuscript received 7 Dec. 2022; revised 29 Jul. 2023; accepted 2 Sep. 2023. Date of publication 31 Dec. 2023.

I. INTRODUCTION

Since the beginning of time, humans have been able to identify other people based on identifying features of their bodies, such as faces and voices. The first record of a biometric identifying system was made in the 1800s in Paris, France. Alphonse Bertillon designed a system that relied on precise body measures to categorize and evaluate criminals effectively. Biometrics measurements are categorized into two types, which are physiological and behavioral [1]. The physiological measurement is further grouped into morphological (fingerprints, face, ear, iris, and retina) and biological (DNA, blood, saliva, and urine). Behavioral measurements are voice, signature traits, keystrokes, and gestures. The widespread collection of biometric data by law enforcement and other government organizations throughout the globe has resulted in the creation of massive legacy databases such as databases of driver's licenses and immigration information.

It has been discovered that no two ears, not even those of identical twins, are equal [2]. Ear biometrics offers several advantages compared to other biometric modalities such as iris [3], [4], fingerprints, face, and retinal scans [5]. One advantage is that the ear is larger than the iris, fingerprint, making it easier to capture detailed images [6]. Furthermore, modern scanners allow ear scans to be conducted from a distance, enhancing convenience and efficiency. These factors make ear biometrics a promising option for automated human identification and verification systems.

Ear recognition technologies present multiple advantages, such as capturing ear images from afar, a feature beneficial for security and surveillance applications [7]. Additionally, the ear's morphology remains largely stable throughout a person's lifetime and is not influenced by facial expressions, making it a viable option for non-contact biometric identification [8], [9]. This technology is also resilient to emotional states and changes in facial expressions, contributing to its reliability as a biometric measure. Furthermore, ear recognition has broad applicability across various sectors, including forensics, surveillance, identity verification, and device unlocking [10], [11]. For those with hearing loss, associated wireless technology can enhance

speech recognition performance in noisy settings, offering better auditory experiences than those with normal hearing under similar conditions [12], [13].

This remarkable ability has inspired computer vision researchers and led to the exploration of similar techniques in occluded ear recognition. Occluded ear recognition refers to the challenging task of identifying individuals based on their ear images, even when the ears are partially covered or hidden behind objects like hair, hats, or accessories. Just like the uniqueness of the fingerprints, the ears possess distinct features that can serve as a reliable biometric trait for identification.

Unlike other biometric modalities, such as face recognition [14] or fingerprinting [15], ear recognition offers certain advantages. The ear is relatively stable throughout a person's life and remains largely unaffected by aging or minor physical changes. Furthermore, unlike facial features, the shape and structure of the ear are less prone to alteration due to facial expressions, making it a potentially robust and reliable biometric identifier [16]. Fig. 1 shows the ear structure of a person.

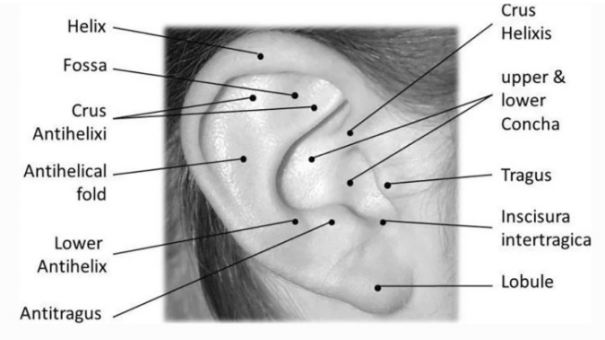

Fig. 1 Ear structure

However, recognizing occluded ears presents unique challenges for computer vision systems. The variability in ear shape, texture, and appearance, combined with the occlusion caused by various factors, poses significant hurdles. Traditional approaches rely solely on local feature extraction or template matching and struggle to handle these complexities effectively. To address these challenges, the paper aims to apply deep learning techniques such as ResNet-50 [17], VGG-16 [18], and EfficientNetB1 [19] for occluded ear recognition in two datasets which are AMI and IITD by implementing synthetic occlusion on ear images which has not been implemented before. By leveraging large-scale annotated datasets and pre-trained models, deep-learning models can learn intricate ear representations that are robust to occlusions and variations in appearance.

Besides that, it is noteworthy that artificial occlusion has primarily been studied and applied in the USTB dataset, with limited attention given to its application in the AMI and IITD ear datasets. This research expands the scope of occlusion analysis by investigating the effects of artificial occlusion on these datasets, which adds a novel dimension to the field. By addressing the gap in artificial occlusion analysis in the AMI and IITD ear datasets and highlighting the variations in results across different datasets, this research contributes to a deeper understanding of the challenges and opportunities in occluded ear recognition. It underscores the importance of dataset diversity and dataset-specific analysis in developing robust and reliable occluded ear recognition systems.

II. Material and Method

A. Impact of Covid-19 pandemic

The COVID-19 pandemic has posed significant challenges to the effectiveness of face recognition systems, particularly in security surveillance and attendance tracking, due to the widespread use of face masks as mandated by organizations such as the World Health Organization [20], [21]. To adapt, researchers have developed mask-aware recognition systems and real-time mask position monitoring using YOLO models, among other innovations [22], [23]. In India, facial recognition is even being proposed to support vaccination campaigns by identifying non-vaccinated individuals through Aadhaar-based systems. Educational organizations are also leveraging this technology for academic processes like online exams and class attendance [24], [25].

Despite these adaptations, ear recognition systems have emerged as a valuable supplementary technology. Ear recognition can effectively identify individuals even when face masks are worn, as the ear's structure remains visible and consistent over time. Moreover, ear recognition does not necessitate active participation from the subject, providing a robust alternative or supplement to face recognition, especially when other systems offer incomplete or inaccurate information [26].

The bulk of ear recognition studies, from the oldest to the most current, are based on built attributes such as local, holistic, and hybrids of the two. Researchers find statistical approaches to ear identification to be quite popular. In recent years, deep features have become increasingly important in ear recognition tasks because they include more concise and profound information than features generated by the machine learning approach. Deep neural networks, such as LeNet, AlexNet [27], ResNet, have been used for various biometric identification applications.

B. Ear Recognition System

Alshazly et al. [28] implemented Deep Residual Networks (ResNet) to distinguish between numerous databases of ear types. Multiple residual modules, in addition to typical convolution and pooling layers, are stacked to create a deep ResNet model. Ensemble approaches involve training multiple models to solve the same problem and then combining their predictions through techniques like averaging or voting. In this study, the authors created an ensemble of deep neural networks for ear identification, as shown in Fig. 2. They trained several individual networks with random starting points and then averaged their predictions. To initialize the network weights, the authors utilized pre-trained models on visual recognition tasks, explicitly training the fully connected layer on the ear identification task. Fine-tuning of all previously trained models was performed using a portion of each ear dataset.

Additionally, the authors used well-tuned ResNet models' second-to-last layer output as feature extractors for SVM classifiers. The performance of the generated models was extensively evaluated using ear images captured under both defined and uncontrolled imaging conditions from the AMI,

AMIC, WPUT, and AWE databases. The ensemble of networks achieved recognition accuracies of 98.57%, 97.85%, 81.89%, and 67.25% on the AMI, AMIC, WPUT, and AWE datasets, respectively, demonstrating the optimal performance.

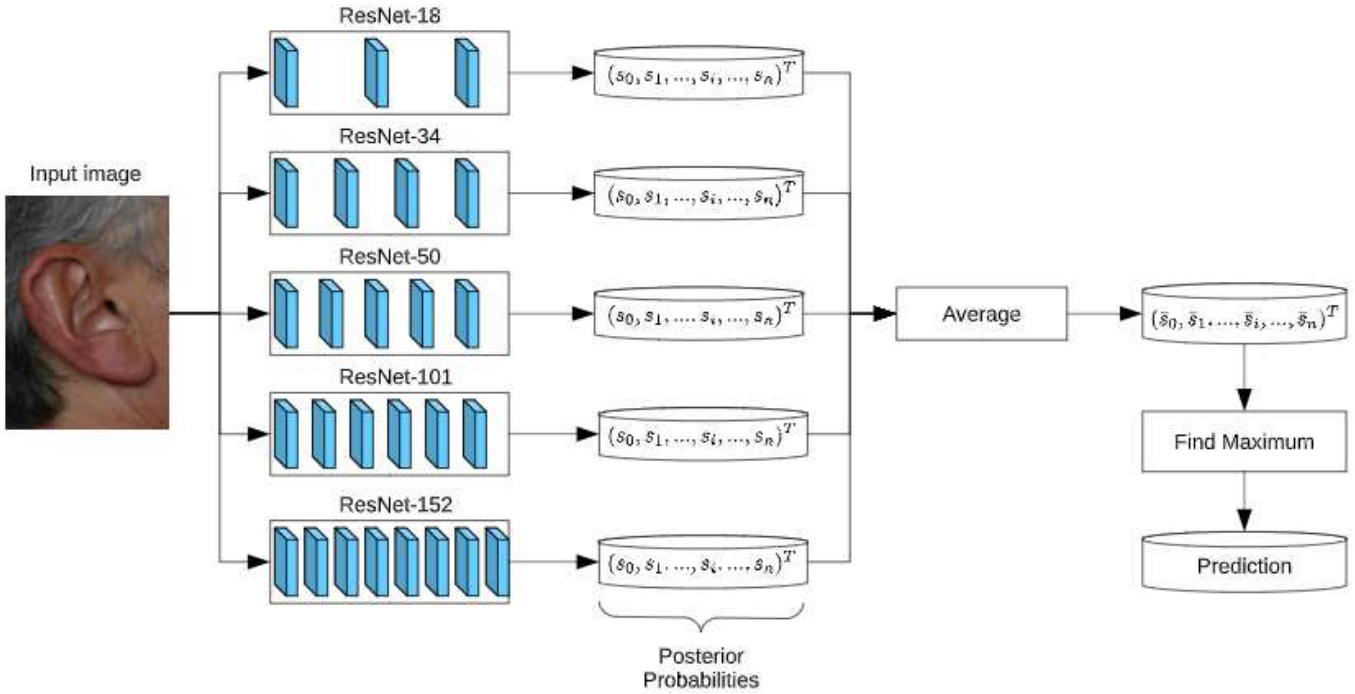

Fig. 2 Deep ResNet Ensemble Model

In their study, Sharkas et al. [29] utilized three different architectures, namely AlexNet, GoogleNet, and ResNet-50, for classifying ear images in an ear identification system. AlexNet consists of five convolutional layers and three fully connected layers. GoogleNet, on the other hand, employed a deep CNN with 22 layers, reducing the number of parameters from 60 million to 4 million. ResNet-50 addressed the issue by introducing skip connections that bypass one or more layers, enabling identity mapping by combining the outputs of these connections with the stacked levels. The authors examined two ear datasets: the AMI ear database and the IIT Delhi database. In the first scenario, both databases consisted of 100 classes, each containing seven images, resulting in a total of 700 images. The IIT Delhi database had 125 classes, with a minimum of three images per class, totaling 493 images. The results showed that ResNet-50 achieved the best performance on the AMI database, with an average mean accuracy of 94%. For the IIT Delhi segmented picture collection, both AlexNet and ResNet-50 performed similarly, with an accuracy of 62.86% for AlexNet and 59.29% for ResNet-50.

Kadhim Zaidan et al. developed an ear identification system based on contrast-limited adaptive histogram equalization (CLAHE) and convolutional neural networks (CNN). CLAHE is one of the most effective image-processing methods for enhancing picture contrast. CLAHE was created to address the shortcomings of adaptive histogram equalization, which tends to enhance noise in homogenous regions of the processed picture. CLAHE works on a small part of a picture instead of the whole picture to cut down on noise amplification. The suggested CNN technique has successfully obtained high accuracy in ear image classification, with an overall testing accuracy of 97.92% and a loss of 0.1254 across 45 epochs.

C. Occluded Ear Recognition

Wang et al. [30] classify occluded ear images using a Fisher Determination Dictionary Learning-based Sparse Representation Classifier (FDDL-based SRC). In SRC, signals may be represented with the fewest feasible atoms. The execution of sparse coding involves four main processes: input of dictionaries and feature extraction, solving sparse coding models, calculating residuals, and outputting classification results. The introduction of FDDL brings changes to the categorization process. Firstly, Adaptive Gamma Correction with Weighting Distribution (AGCWD) is applied to enhance the quality of the ear images. Secondly, a combination of DSIFT, Local Binary Pattern (LBP), and Histogram of Oriented Gradients (HoG) is utilized to extract features. Two feature selection methods based on robust sparse linear discriminant analysis (RSLDA) and inter-class sparsity-based discriminant least square regression (ICS-DLSR) are employed to improve the calculation speed. Eventually, the two sets of selected features are categorized using an FDDL-based Sparse Representation Classifier (SRC). At the decision level, the classification results from both sets are combined to generate high-precision outputs. The selected database is USTB 1, and it may be used as a dataset in an occlusion experiment with little to no outside influence. In the experiment, 120 pictures were used for training, 60 photos were used for testing, and varying amounts of random occlusions (10%, 20%, 30%, 40%, and 50%) were introduced to the test set. According to an occlusion rate of

10%, the outcomes are 93.33%, 83.33%, 83.33%, 63.89%, and 47.78%. The sample occluded ear is illustrated in Fig. 3.

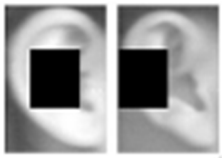

Fig. 3 Occluded images from USTB 1

Tian et al. [31] proposed a convolutional neural network that consists of an input layer, three convolutions, and two fully linked layers were suggested for a convolutional neural network. Each convolution layer was followed by a max-pooling (MP) layer, and the input layer was linked to the convolution layer. The final fully connected layer was a soft-max classifier with 79 categories. This classifier generated an output value representing the likelihood that an input fits into one of 79 categories. The author conducted an occlusion experiment using CNN on USTB III by splitting the ear pictures into random blocks with widths ranging from 5% to 50% of the original image's width. The obtained sample ears are exhibited in Fig. 4. The results of the proposed method are 98%, 96%, 95%, 88%, 74%, 60%, 58%, 41%, 30%, and 25% in the range of 5% occlusion rate until 50% occlusion rate at a 5% of increment.

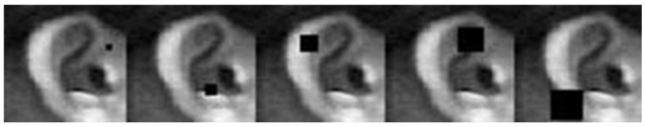

Fig. 4 Sample occluded ear images

Sultana et al. [32] presented a novel index-based rank fusion method for occluded ear recognition. This method grouped uncovered ear samples and assigned them a unique number. Each block's features were indexed and stored in a database of feature descriptors. The percentage of visible ear area was initially determined in the identifying process. If more than 60% of the ear canal was blocked, the sample was thrown out, and a reacquisition request may be made. The suggested technique collected features from the uncovered part and their related indices when the occlusion was less than 60%. The characteristics of the test and the enrolled samples were indexed, and nine corresponding scores were then derived. The author evaluated the synthetic occlusion USTB I methods and obtained 0.96%, 0.93%, 0.86%, 0.82%, 0.73%, and 0.62% correct recognition rate based on every 10% occlusion rate for USTB1 ear dataset.

D. Model Structure

Fine-tuning techniques on two popular deep learning models, VGG-16 and ResNet-50, are explored. Fine-tuning aims to utilize the learned representations from these pre-trained models and adapt them to the ear dataset relevant to occluded ear recognition. By leveraging the information captured by these models on general image features, better performance and faster convergence can be achieved.

During the fine-tuning process, certain layers of the pre-trained models are selectively frozen while allowing others to be trained. The proposed model has frozen 21 layers in VGG-16 and 143 layers in ResNet-50. This selective freezing helps to retain the learned representations in the earlier layers, which are more generic and transferable while allowing the later layers to be adapted to the specific dataset. This approach balances leveraging the pre-trained model's knowledge and tailoring it to the classification task. The fine-tuning of both VGG-16 and ResNet-50 models is implemented on the occluded dataset. By fine-tuning these models, the goal is to harness their powerful capabilities and achieve superior performance on the specific image classification task.

E. Datasets

Experiments involve two ear databases: the Mathematical Analysis of Images (AMI) dataset and the IIT Delhi (IITD) dataset [33]. AMI dataset contains information from 100 people between the ages of 19 and 65. Seven photographs of each person were taken, including six of their right ears and one of their left ears. The era images are in jpg format, with dimensions of 492 by 702, and a unique identification number is issued to each of the 700 subjects in the dataset. Some of the examples from AMI are shown in Fig. 5.

Fig. 5 Raw AMI ear samples

All images in the AMI ear dataset were manually cropped, rotated to the same angle, and resized to a dimension of 100 by 100. The dataset contains 700 images as the original dataset and is subjected to 107 classes. The cropped sample images are displayed in Fig. 6.

Fig. 6 Cropped and rotated AMI ear samples

The IITD dataset involves 222 people somewhere between 14 and 58 years old. The image was taken in jpeg format and has a 50-pixel wide and 180-pixel high resolution. This dataset contains the unprocessed photo data and the cropped ear pictures. The selected ear database was already automatically cropped and standardized and contained information from 222 people, totaling 793 different ear photographs. The examples from IITD ear dataset are shown in Fig. 7.

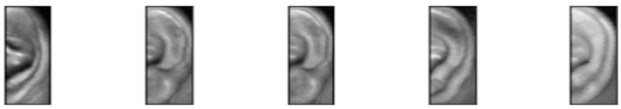

Fig. 7 Segmented IITD ear samples

A black occluded block is created to cover some portion of the ear randomly on both AMI and IITD datasets. Image is loaded from the dataset, and the range and increment for the occlusion block size are defined, varying from 5% to 50% with a 5% increment. Next, iterate over each image in the dataset, and for each image, iterate over the range of occlusion sizes. After calculation, the dimensions of the occlusion block are based on the image size and occlusion percentage. Using these dimensions, a black block is created as an occlusion of

the image. The black block is positioned at a random location on the original image. Finally, the modified image is saved with the occlusion block in a separate directory or appended to a new dataset. This process ensures the systematic creation of occluded versions of the original images with varying degrees of black block occlusion. The black occluded block with varying sizes on both AMI and IITD datasets is exhibited in Fig. 8 and Fig. 9.

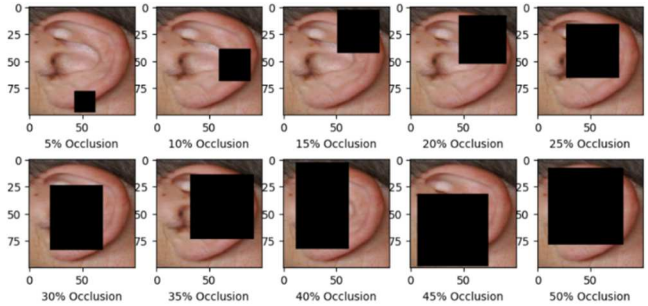

Fig. 8 Samples occluded images with different occlusion percentages on the AMI dataset

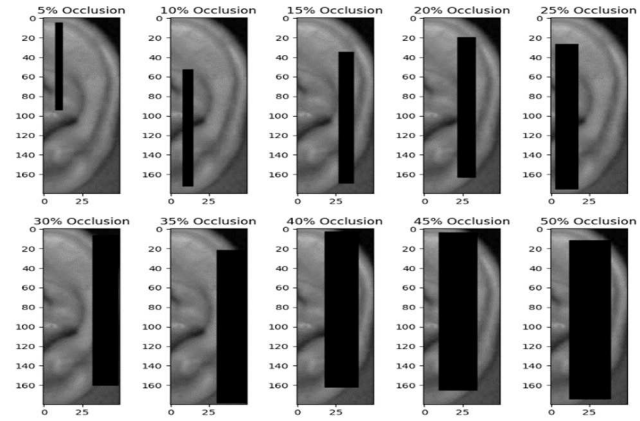

Fig. 9 Samples occluded images with different occlusion percentages on the IITD dataset

III. Result and Discussion

A. Evaluation Metric

The accuracy score measures the experimental result. Accuracy is a metric used to evaluate the performance of a classification model. It is calculated by dividing the number of correct predictions by the total number of predictions made. This metric indicates the proportion of instances that were correctly classified out of the total number of instances in the dataset. The accuracy value is typically expressed as a percentage ranging from 0% to 100%, where a higher accuracy value indicates a better classification model performance. The formula for accuracy is shown in Equation 1 below.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

TP represents true positives, TN refers to true negatives, FP is false positives, and FN represents false negatives.

B. Experiment Setup

The setup for occluded AMI and IITD is completed using Google Colab. Google Colab provides GPU T4, 25.5GB system RAM, and 15GB GPU RAM. Both datasets are separated into 70% of the training set and 30% of testing sets and trained with ResNet-50, VGG-16, and EfficientNetB1, respectively.

The VGG-16 model is trained on RMSprop optimizer with a learning rate from 0.00002 until 0.00001 when the error plateaus and freezes the first 21 layers of the model. The selected weight for the model is ImageNet and two fully connected layers with Rectified Linear Unit (ReLU) after the model, except IITD has only one fully connected layer. To avoid overfitting, a dropout layer with 0.5 value, a batch normalization layer on each fully connected layer, and early stopping are activated if the validation accuracy does not increase after 50 epochs. Fine-tune is set to 50 Epochs on AMI ear dataset and 100 Epochs on IITD ear dataset. After fine-tuning, the dataset fits the model for 500 epochs for each dataset.

For ResNet-50, the learning rate, optimizer, weight, fully connected layers, activation function, dropout, batch normalization, and early stopping are set as the same as VGG-16 for AMI except the optimizer is Adam, and the learning rate set from 0.00001 to 0.000001. The first 143 layers of the model are frozen for fine-tuning purposes. 50 Epochs of fine-tuning on AMI ear dataset and 100 epochs on IITD ear dataset will be trained for 500 epochs in ResNet-50.

The last one is EfficientNetB1; the model is trained on Adam optimizer with a learning rate from 0.0001 until 0.00001 when the error plateaus and freezes the first 200 layers of the model. The selected weight for the model is ImageNet and one fully connected layer ReLU. A dropout layer with 0.5 value and batch normalization layer on each fully connected layer and early stopping is activated if the validation accuracy does not increase after 50 epochs. Fine-tune is set to 50 epochs on both ear datasets. After fine-tuning, the dataset fits the model for 500 epochs for each dataset.

TABLE I
Experiment Setup on AMI and IITD

	Setup of AMI			**Setup of IITD**		
Model	VGG-16	ResNet-50	EfficientNetB1	VGG-16	ResNet-50	EfficientNetB1
Freeze Layer	21	143	200	21	143	200
Weight	ImageNet	ImageNet	ImageNet	ImageNet	ImageNet	ImageNet
Fully Connected Layer	2	2	1	1	2	1
Activation Function	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU
Fine-Tune	50 epochs	50 epochs	50 epochs	100 epochs	100 epochs	50 epochs
Optimizer	RMSprop	RMSprop	Adam	RMSprop	Adam	Adam
Learning Rate	0.00002 – 0.00001	0.00002 – 0.00001	0.0001-0.00001	0.00002 – 0.00001	0.00001 – 0.000001	0.0001-0.00001
Dropout	0.5	0.5	0.5	0.5	0.5	0.5
Batch Normalization	Yes	Yes	Yes	Yes	Yes	Yes
Early Stopping	50 epochs	50 epochs	50 epochs	50 epochs	50 epochs	50 epochs

C. Data Augmentation Setting

The occluded images are randomly rotated by up to 20 degrees, except the IITD ear dataset is rotated up to 10 degrees and randomly shifts images horizontally and vertically by up to 5% of the image width and height. The brightness range is set from 80% darker and 50% brighter. The images randomly zoom in and out of images by up to 5% and randomly flip images horizontally. The gaps in the image fill with the nearest pixel. The setup of data augmentation is illustrated in Table 2.

TABLE II
DATA AUGMENTATION SETTING

	AMI	IITD
Rotate Angle	20%	10%
Width Shift	5%	5%
Height Shift	5%	5%
Brightness	0.8-1.5	0.8-1.5
Zoom	5%	5%
Horizontal Flip	True	True
Gap Fill	Nearest Pixel	Nearest Pixel

D. Evaluation Results

The experiment involved evaluating the performance of ResNet-50, VGG-16, and EfficientNetB1 on two occluded ear recognition datasets, AMI and IITD, with varying levels of occlusion rates, as shown in Fig. 10, Fig. 11, and Fig. 12. The results revealed interesting trends in accuracy as the occlusion rate increased. The overall result for the occluded AMI ear dataset is better than the occluded IITD ear dataset for both models. It was discovered, during the process of experimental assessment on a dataset of occluded ear pictures, that the performance of the identification system changed depending on the degree to which the ear was concealed. The ear images, as illustrated in Fig. 8 and Fig. 9, show a black block that acts as occlusion with size increasing gradually, beginning with a coverage of 5% and rising by 5% increments up to 50%.

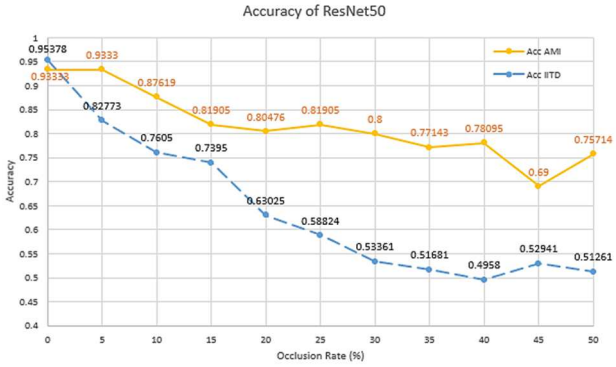

Fig. 10 Result of ResNet-50 for occluded ear from the AMI and IITD datasets

For Fig. 10, at an occlusion rate of 0%, where no occlusion is applied, the fine-tune ResNet-50 model achieves high accuracies for both occluded AMI (0.93333) and occluded IITD (0.95378). As the occlusion rate increases, the accuracy gradually declines for both datasets. At an occlusion rate of 5%, the model maintains relatively high accuracies for both datasets; the accuracy for AMI and IITD is at 0.93330 and 0.82773. However, as the occlusion rate increases, the accuracies decline more noticeably. When the occlusion rate reaches 25%, the model's accuracy decreases further. The results obtained for AMI and IITD are 0.81905 and 0.58824, respectively. This trend continues as the occlusion rate reaches 50%, where the accuracies further drop; the accuracy for AMI is 0.75714, and IITD only has 0.51261.

The fine tune VGG-16 model, as displayed in Fig. 11, when starting with an occlusion rate of 0%, where no occlusion is applied, the model achieves a high accuracy of 0.9905 for the occluded AMI ear dataset and 0.95378 for the occluded IITD ear dataset. As the occlusion rate increases, the accuracy score for both datasets drops. At an occlusion rate of 5%, the accuracy values decrease to 0.94762 for the AMI dataset and 0.80672 for the IITD dataset. As the occlusion rate increases to 10%, 15%, and 20%, the accuracy values decrease for both datasets. At an occlusion rate of 25%, the model achieves an accuracy of 0.76667 and 0.4538 for AMI and IITD ear dataset. As the occlusion rate increases to 30%, 35%, and 40%, the accuracy values remain relatively stable, but the IITD ear dataset is at lower levels than the AMI ear dataset. Continuing to higher occlusion rates of 45% and 50%, the model's accuracies decrease, with 0.69524 and 0.6619 for the AMI ear dataset and 0.4623 and 0.4244 for the IITD ear dataset.

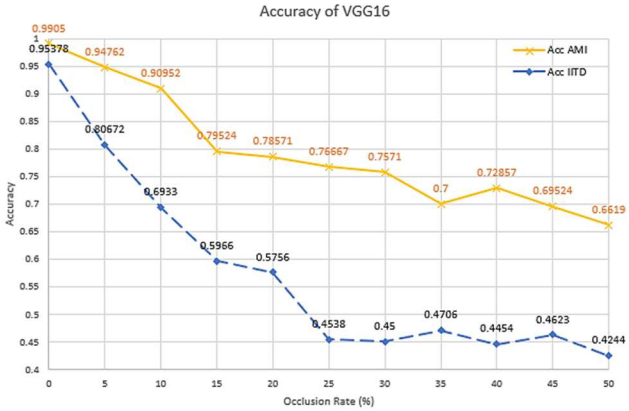

Fig. 11 Result of VGG-16 for occluded ear from the AMI and IITD datasets

For Fig. 12, the EfficientNetB1 model showcased strong performance across different occlusion rates on the AMI dataset. At 0% and 5% lower occlusion rates, the model achieved high accuracies of 0.9429 and 0.9238, respectively. As the occlusion rate increased to 10% and 15%, the model's accuracy declined slightly to 0.8619 and 0.8476, respectively. However, the model demonstrated a notable improvement in accuracy at an occlusion rate of 20%, reaching 0.8905. Beyond 20% occlusion, the accuracy experienced fluctuations, indicating a more challenging recognition task as the occlusion rate increased.

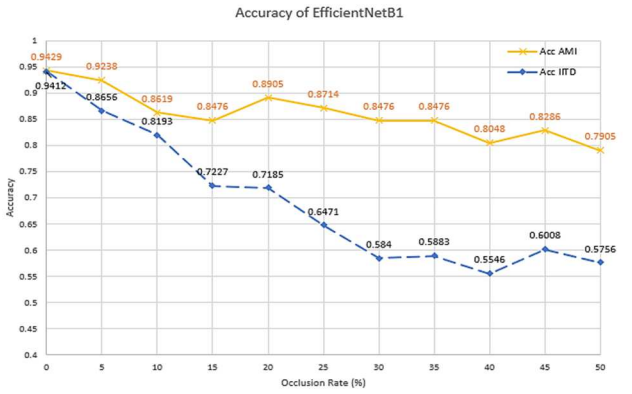

Fig. 12 Result of EfficientNetB1 for occluded ear from the AMI and IITD datasets

In contrast, the performance of the EfficientNetB1 model on the IITD dataset showed a different trend. The model exhibited relatively high accuracies at lower occlusion rates (0% to 15%) ranging from 0.9412 to 0.8193. However, as the occlusion rate increased to 20% and beyond, the model's accuracy experienced a significant drop. At 20% occlusion, the accuracy was 0.7185, and as the occlusion rate further increased, the accuracy continued to decline, reaching 0.5756 at a 50% occlusion rate.

The results indicate that the model's ability to accurately predict the classes or labels decreases as more data is occluded. Higher occlusion rates lead to a loss of important information, making it more challenging for the model to distinguish between different patterns and make accurate predictions. According to the findings, the amount of occlusion influenced the identification accuracy and worsened as the occlusion became larger. This demonstrates the difficulties that may be caused by occlusion in ear recognition systems, as well as the need to develop efficient methods for dealing with occlusion.

From the result for both models, the accuracy of the occluded IITD ear dataset is overall poorer than the occluded AMI ear dataset. This is because the IITD dataset is a grayscale image dataset, and it performs relatively poorly when using pre-trained models like VGG-16, ResNet-50, and EfficientNetB1 compared to the performance on color RGB images of the AMI dataset. As these models are trained on ImageNet, a large-scale color image dataset, these models have learned to extract features from color images, including patterns and structures that are indicative of the object classes. The color information is lost by converting the images to grayscale, negatively impacting the model's ability to distinguish between objects or classes.

Furthermore, while using a pre-trained model like VGG-16, ResNet-50, and EfficientNetB1, it is generally beneficial to have images with similar dimensions to the ones the model was trained on. In the case of VGG-16, it was trained on ImageNet with input images of size 224x224 pixels. The occluded AMI dataset in Fig. 8 with images of dimensions 100x100 is more likely to obtain a better result when using the pre-trained model. This is because the images in this dataset are closer to the expected input size of the model, and the model can process them without significant distortion or loss of information. The aspect ratio 1:1 in the 100x100 dataset also aligns better with the original aspect ratio used during training. On the other hand, the occluded IITD dataset in Fig. 9 with images of dimensions 180x50 has a different aspect ratio, which is not well-matched with the original input size of the pre-trained model. Resizing these images to fit the input size of the model would result in distortion and potential loss of information, which can negatively impact the model's performance.

From the experimental evaluation of occluded ear images from the two datasets, the recognition system's performance varied depending on the extent of occlusion. Occlusions such as black blocks were systematically applied to the images, starting from 5% coverage and increasing in increments of 5% up to 50%. The recognition accuracy is assessed at each level of occlusion to understand the system's robustness to occluded ear images.

IV. Conclusion

Recognizing an ear that is covered is a difficult but crucial part of biometric systems [34]. The procedure requires locating and identifying ear pictures when they are covered somehow. Experiment results demonstrated that as the level of occlusion increased, the recognition accuracy decreased. This highlights the challenges occlusion poses in ear recognition systems and the need for effective occlusion handling techniques. However, even with significant occlusion, the system showed promising performance, indicating its potential for real-world applications. It is worth noting that research into obstructed ear identification is still in its infancy, so there is room for improvement in the system's precision and dependability. Improving identification accuracy may require trying out new techniques for feature extraction, creating more advanced occlusion detection algorithms, or using deep learning methods. More work is needed to improve biometric systems and make them work reliably in the real world, where occlusion is a widespread problem.

In this regard, future research could explore more detailed and fine-grained models that can accurately identify and analyze the precise structure and parts of the ear. Fine-grained models can potentially improve recognition performance by capturing more discriminative information by considering the complex details of the ear's structure features, such as the helix, earlobe, tragus, and antitragus. Furthermore, the integration of advanced techniques, such as attention mechanisms or graph convolutional networks, could be explored to model the relationships and dependencies among different ear structures and parts effectively. This would enable the development of more sophisticated and context-aware recognition models, enhancing the accuracy and robustness of occluded ear recognition systems.

Acknowledgment

The work presented in this paper is supported by Multimedia University through the IR Fund 2021 (MMUI/210029).

References

[1] A. Kavipriya and A. Muthukumar, “Human Age Estimation based on Ear Biometrics using KNN,” 2019 IEEE International Conference on Clean Energy and Energy Efficient Electronics Circuit for Sustainable Development (INCCES), Dec. 2019, doi:10.1109/incces47820.2019.9167706.

[2] X. Xu, Y. Liu, C. Liu, and L. Lu, “A Feature Fusion Human Ear Recognition Method Based on Channel Features and Dynamic Convolution,” Symmetry, vol. 15, no. 7, p. 1454, Jul. 2023, doi:10.3390/sym15071454.

[3] V. Nazmdeh, S. Mortazavi, D. Tajeddin, H. Nazmdeh, and M. M. Asem, “Iris Recognition; From Classic to Modern Approaches,” 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Jan. 2019, doi: 10.1109/ccwc.2019.8666516.

[4] W.-H. Chuah, S.-C. Chong, and L.-Y. Chong, “The Assistance of Eye Blink Detection for Two- Factor Authentication,” Journal of Informatics and Web Engineering, vol. 2, no. 2, pp. 111–121, Sep. 2023, doi: 10.33093/jiwe.2023.2.2.8.

[5] L. F. Nakayama *et al.*, "Retinal Scans and Data Sharing: The Privacy and Scientific Development Equilibrium," *Mayo Clinic Proceedings: Digital Health*, vol. 1, no. 2, pp. 67–74, 2023, doi:10.1016/j.mcpdig.2023.02.003.

[6] L. Markićević, P. Peer, and Ž. Emeršič, "Improving Ear Recognition with Super-resolution," in *2023 30th International Conference on*

Systems, Signals and Image Processing (IWSSIP), 2023, pp. 1–5. doi:10.1109/IWSSIP58668.2023.10180250.

[7] M. M. Zarachoff, A. Sheikh-Akbari, and D. Monekosso, “Multi-band PCA based ear recognition technique,” Multimedia Tools and Applications, vol. 82, no. 2, pp. 2077–2099, Jun. 2022, doi:10.1007/s11042-022-12905-0.

[8] J. Jeyabharathi, S. Devi, B. Krishnan, R. Samuel, M. I. Anees, and R. Jegadeesan, “Human Ear Identification System Using Shape and structural feature based on SIFT and ANN Classifier,” 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), Mar. 2022, doi:10.1109/ic3iot53935.2022.9767893.

[9] R. Mehta, A. Sheikh-Akbari, and K. K. Singh, “A Noble Approach to 2D Ear Recognition System using Hybrid Transfer Learning,” 2023 12th Mediterranean Conference on Embedded Computing (MECO), Jun. 2023, doi: 10.1109/meco58584.2023.10154993.

[10] M. Sharkas, “Ear recognition with ensemble classifiers; A deep learning approach,” Multimedia Tools and Applications, vol. 81, no. 30, pp. 43919–43945, May 2022, doi: 10.1007/s11042-022-13252-w.

[11] R. Ahila Priyadharshini, S. Arivazhagan, and M. Arun, “A deep learning approach for person identification using ear biometrics,” Applied Intelligence, vol. 51, no. 4, pp. 2161–2172, Oct. 2020, doi:10.1007/s10489-020-01995-8.

[12] Y. Lei, J. Qian, D. Pan, and T. Xu, “Research on Small Sample Dynamic Human Ear Recognition Based on Deep Learning,” Sensors, vol. 22, no. 5, p. 1718, Feb. 2022, doi: 10.3390/s22051718.

[13] R. Mehta, A. Sheikh-Akbari, and K. K. Singh, “A Noble Approach to 2D Ear Recognition System using Hybrid Transfer Learning,” 2023 12th Mediterranean Conference on Embedded Computing (MECO), Jun. 2023, doi: 10.1109/meco58584.2023.10154993.

[14] D. Fitousi, N. Rotschild, C. Pnini, and O. Azizi, “Understanding the Impact of Face Masks on the Processing of Facial Identity, Emotion, Age, and Gender,” Frontiers in Psychology, vol. 12, Nov. 2021, doi:10.3389/fpsyg.2021.743793.

[15] D. Maltoni, D. Maio, A. K. Jain, and J. Feng, Handbook of Fingerprint Recognition. Springer International Publishing, 2022. doi:10.1007/978-3-030-83624-5.

[16] J. Patmanee, S. Kanprachar, and K. Chamnongthai, “Effects of Preprocessing in Person Identification Using Ear Features,” 2021 25th International Computer Science and Engineering Conference (ICSEC), Nov. 2021, doi: 10.1109/icsec53205.2021.9684602.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015, Accessed: Dec. 26, 2022. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[18] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks For Large-Scale Image Recognition," 2015, Accessed: May 29, 2023. [Online]. Available: http://www.robots.ox.ac.uk/

[19] M. Tan and Q. V Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, pp. 6105-6114, 2019, doi: 10.48550/arXiv.1905.11946.

[20] S. Mhadgut, “Masked Face Detection and Recognition System in Real Time using YOLOv3 to combat COVID-19,” 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Jul. 2021, doi:10.1109/icccnt51525.2021.9579525.

[21] J. S. Talahua, J. Buele, P. Calvopiña, and J. Varela-Aldás, “Facial Recognition System for People with and without Face Mask in Times of the COVID-19 Pandemic,” Sustainability, vol. 13, no. 12, p. 6900, Jun. 2021, doi: 10.3390/su13126900.

[22] S. Dharanesh and A. Rattani, “Post-COVID-19 Mask-Aware Face Recognition System,” 2021 IEEE International Symposium on Technologies for Homeland Security (HST), Nov. 2021, doi:10.1109/hst53381.2021.9619841.

[23] D. Min, S. Anandamurugan, K. Mohanasundaram, P. Pandiyan, R. Thangaraj, and V. K. Kaliappan, “Real-time face mask position recognition system using YOLO models for preventing COVID-19 disease spread in public places,” International Journal of Ad Hoc and Ubiquitous Computing, vol. 42, no. 2, p. 73, 2023, doi:10.1504/ijahuc.2023.10053539.

[24] A. Balmik, A. Kumar, and A. Nandy, “Efficient Face Recognition System for Education Sectors in COVID-19 Pandemic,” 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Jul. 2021, doi:10.1109/icccnt51525.2021.9579523.

[25] L. Yuningsih, G. A. Pradipta, D. P. Hostiadi, R. R. Huizen, and P. D. W. Ayu, “Ear Feature Extraction Methods - A Review,” 2022 4th International Conference on Cybernetics and Intelligent System (ICORIS), Oct. 2022, doi: 10.1109/icoris56080.2022.10031264.

[26] R. D. Balangue, C. D. M. Padilla, N. B. Linsangan, J. P. T. Cruz, R. A. Juanatas, and I. C. Juanatas, “Ear Recognition for Ear Biometrics Using Integrated Image Processing Techniques via Raspberry Pi,” 2022 IEEE 14th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Dec. 2022, doi:10.1109/hnicem57413.2022.10109479.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Accessed: Sep. 03, 2023. [Online]. Available: http://code.google.com/p/cuda-convnet/

[28] H. Alshazly, C. Linse, E. Barth, S. A. Idris, and T. Martinetz, “Towards Explainable Ear Recognition Systems Using Deep Residual Networks,” IEEE Access, vol. 9, pp. 122254–122273, 2021, doi:10.1109/access.2021.3109441.

[29] M. Sharkas, “Ear recognition with ensemble classifiers; A deep learning approach,” Multimedia Tools and Applications, vol. 81, no. 30, pp. 43919–43945, May 2022, doi: 10.1007/s11042-022-13252-w.

[30] Z. Wang, X. Gao, J. Yang, Q. Yan, and Y. Zhang, “Local feature fusion and SRC-based decision fusion for ear recognition,” Multimedia Systems, vol. 28, no. 3, pp. 1117–1134, Mar. 2022, doi:10.1007/s00530-022-00906-w.

[31] L. Tian and Z. Mu, “Ear recognition based on deep convolutional network,” 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Oct. 2016, doi: 10.1109/cisp-bmei.2016.7852751.

[32] M. Sultana, P. P. Paul, and M. Gavrilova, “A Novel Index-Based Rank Fusion Method for Occluded Ear Recognition,” 2015 International Conference on Cyberworlds (CW), Oct. 2015, doi:10.1109/cw.2015.30.

[33] A. Kumar and C. Wu, “Automated human identification using ear imaging,” Pattern Recognition, vol. 45, no. 3, pp. 956–968, Mar. 2012, doi: 10.1016/j.patcog.2011.06.005.

[34] S. Ramos-Cooper and G. Camara-Chavez, “Ear Recognition In The Wild with Convolutional Neural Networks,” 2021 XLVII Latin American Computing Conference (CLEI), Oct. 2021, doi:10.1109/clei53233.2021.9640083.