Unconstrained Ear Recognition through Domain Adaptive Deep Learning Models of Convolutional Neural Network

Marwin Alejo, Cris Paulo Hate

Abstract**:** ***Limited ear dataset yields to the adaption of domain adaptive deep learning or transfer learning in the development of ear biometric recognition. Ear recognition is a variation of biometrics that is becoming popular in various areas of research due to the advantages of ears towards human identity recognition. In this paper, handpicked CNN architectures: AlexNet, GoogLeNet, Inception-v3, Inception-ResNet-v2, ResNet-18, ResNet-50, SqueezeNet, ShuffleNet, and MobileNet-v2 are explored and compared for use in an unconstrained ear biometric recognition. 250 unconstrained ear images are collected and acquired from the web through web crawlers and are preprocessed with basic image processing methods including the use of contrast limited adaptive histogram equalization for ear image quality improvement. Each CNN architecture is analyzed structurally and are fine-tuned to satisfy the requirements of ear recognition. Earlier layers of CNN architectures are used as feature extractors. Last 2-3 layers of each CNN architectures are fine-tuned thus, are replaced with layers of the same kind for ear recognition models to classify 10 classes of ears instead of 1000. 80 percent of acquired unconstrained ear images is used for training and the remaining 20 percent is reserved for testing and validation. Results of each architectures are compared in terms of their training time, training and validation outputs as such learned features and losses, and test results in terms of above-95% accuracy confidence. Above all the used architectures, ResNet, AlexNet, and GoogleNet achieved an accuracy confidence of 97-100% and is best for use in unconstrained ear biometric recognition while ShuffleNet, despite of achieving approximately 90%, shows promising result for use in mobile version of unconstrained ear biometric recognition.***

Index Terms**:** ***ear recognition, domain adaptive deep learning, convolutional neural network, transfer learning.***

I. INTRODUCTION

Deep Learning is a universal method of machine learning that had been applied in several fields of researches including but not limited to image segmentation, object detection, classification, and biometric recognition [1, 2, 3, 4, 5]. Biometric recognition is a science and a technology (embedded systems) that deals with the identification of an individual by extracting unique physiological and/or behavioral characteristics from an individual's fingerprint, face, iris, retina, gait, and/or ear [6, 7, 8, 9]. Compared to the first five biometric modals stated, human ears had been proven to be more superior and advantageous for biometric recognition. The size and structure of ears are more static over ageing compared to other modal of biometry; human ears have robust and reliable features that could be extracted passively from distant measures; and human ears are universal, unique, permanent, and collectible [9, 10]. With these advantages, ear recognition had gained momentum of interests in multitudes of researches and applications as such computer vision and machine learning [11, 12, 13, 14].

Fusion of Computer Vision and Image Processing had been the most used methods in previous studies of ear biometrics [15]. Present studies suggest that the use of convolutional neural network shows promising results in ear recognition [16]. Deep Convolutional Neural Network or CNN is one of the deep learning algorithms and methods used in recent biometric recognition studies [16, 17]. It is a special type of deep neural network that works like the organization of an animal visual cortex. It is designed to automatically and adaptively extract and learn features from classes of images to perform tasks as such classification [18, 19]. Development of a CNN-based biometric recognition could be done by either modeling-from-scratch or transfer learning [20]. Modeling-from-scratch is a technique in developing a CNN-based biometric that heavily relies on the depth of available datasets through which features of classes will be extracted and learned for recognition. This method of CNN modeling heavily relies on the richness of used datasets and training time for learning new classes [21, 22]. However, this method posed a challenge to further explore CNN on ear biometrics due to limited collection of studies and ear datasets in both constrained and unconstrained environments [15, 23, 24, 25]. To widen recent solutions to this challenge, this paper exploits the study of ear recognition in an unconstrained environment with limited data through the fusion of computer vision and deep learning in the form of transfer learning.

Transfer Learning is another modeling technique that utilizes pre-trained CNN models by traversing earlier learned classes from one domain to classes of another domain [25]. In contrast with modeling-from-scratch, transfer learning learns new domains not by relying on the depth of its dataset but by recycling existing knowledge attained from previous training and hasten learning processes to be used for another specific learning task.

Revised Manuscript Received on 30 July 2019.

***** Correspondence Author

Marwin B. Alejo*, Department of Graduate Studies, Technological Institute of the Philippines, Quezon City, Philippines.

Cris Paulo G. Hate, Department of Graduate Studies, Technological Institute of the Philippines, Quezon City, Philippines.

The main goals of this paper can be summarized as follows:

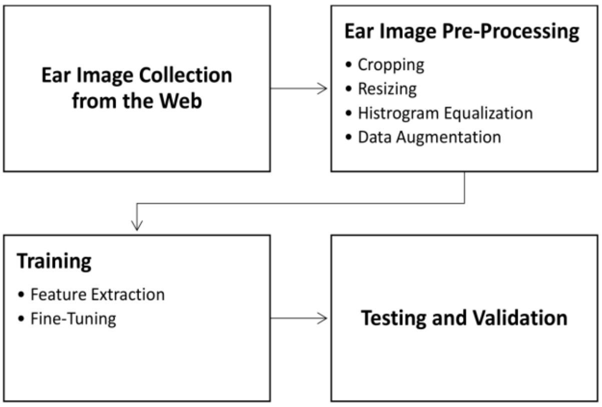
- Identify the best CNN model for an unconstrained ear biometric through domain adaptive deep learning methods.
- Analyze the concept of transfer learning on different deep convolutional neural network architectures for an unconstrained ear recognition.
- Compare each deep convolutional neural network architecture as used in ear recognition in the context of transfer learning and accuracy performance.

The rest of this paper is organized as follows: Section 2 discusses recent studies related to the focus of this paper including but not limited to biometric recognition with transfer learning. Section 3 of this paper provides a thorough discussions of the used methodologies and framework for an unconstrained ear biometric recognition including data preprocessing, data augmentation, pre-trained CNN architectures, and transfer learning. Section 4 of this paper discusses the experimental and validation results obtained after exploiting the used methodology. Lastly, Section 5 provides the conclusion of the paper and future application of this study.

II. REVIEW OF RELATED STUDIES AND LITERATURES

Transfer learning had been used in various recent biometric studies. Its application could be applied by either redirecting pre-trained weights for learning other knowledge domain and/or by fine-tuning network weights and learn new classes of domains from minimal number of datasets [26]. In the paper of [27], an AlexNet pre-trained model had been tailored with a handcrafted CNN architecture to extract features from fingernail plates and finger knuckles for the purpose of biometric authentication. The works of [28] proposes the use of a pre-trained CNN model for age range classification from an unconstrained face images due to the absence of large comprehensive unconstrained face dataset. In their paper, pre-trained CNN model is used as feature extractor from face images and applied fine-tuning to train their model for age classification. The contribution of [29] explores the utilization of transfer learning through AlexNet for finger-vein-based. Their work secured an accuracy level of 95% predictability. The paper of [30], although not related to biometry, presents the use of transfer learning on different CNN architectures for breast cancer detection. In their paper, GoogleNet, VGGNet, and ResNet are used and secured promising results. Likewise, this paper exploits several CNN architectures to explore the effectivity of different neural networks in an unconstrained ear recognition.

Although transfer learning had been widely applied to various biometric recognition, its application to ear recognition is very limited due to the scarcity in the collection of large ear data in either constrained or unconstrained settings [13, 15, 24]. To the knowledge of authors at the time of writing of this paper, there are only two papers that utilizes transfer learning for ear recognition. The works of [25] utilizes AlexNet transfer learning to recognize ears in controlled environments. Their paper presents a promising framework that their accuracy result achieved a rate of 100%. The paper of [31] focuses on ear recognition in an unconstrained environment while exploiting CNN architectures like AlexNet, VGGNet, ResNet and the datasets, AWE and CVLE from [13, 15]. Their paper achieved an accuracy rate of 88.75% to 99.69%. However, there had been no study that shows the effectivity of each available CNN architectures for unconstrained ear recognition through domain adaptive deep learning or transfer learning for benchmark purposes [25]. In addition to the above-stated solutions, this paper explores the effectivity of domain adaptive deep learning in an unconstrained ear biometric/recognition and compare each models through 9 different handpicked CNN architectures: AlexNet, GoogleNet, Inception-v3, Inception-ResNet-v2, ResNet-18, ResNet-50, SqueezeNet, ShuffleNet, and MobileNet-v2. Image processing of computer vision is also used in this paper to further enhance the quality and result of each architecture.


Ear Image Collection from the Web

Ear Image Pre-Processing
- Cropping
- Resizing
- Histrogram Equalization
- Data Augmentation

Training
- Feature Extraction
- Fine-Tuning

Testing and Validation

Figure 1. Ear Recognition Framework

Figure 2. Samples of Acquired and Preprocessed Ear Images from the Web

III. METHODOLOGY

In this section, ear image collection and data preprocessing are described. This is followed by the discussion of used CNN architectures. Latter subsections discuss the details of used method in each steps of the study framework as shown in figure 1.

A. Ear Image Collection

Ear image collection if the first step in the used framework for this study.

Inspired from the works of [13, 15], web crawler is used to collect ear images on the web randomly instead of using pre-defined ear databases. 25 ear images of 10 different personalities were extracted and collected from the internet in different positions, rotations, resolutions, scales, and distances, totaling to 250 ear images. Each ear image contains different look of occlusions. Extraction of features through this number of ear images is enough for transfer learning to commence in ear recognition [32]. Collected ear images are a typical 3-channel image composed of red, green, and blue.

B. Ear Image Preprocessing

Ear image preprocessing is the second step in the used ear recognition framework of the study. Each acquired ear image are manually cropped producing a set of ear images with different square dimensions. However, to satisfy the input requirement of the used CNN architectures, cropped images are resized to 277x277 square pixels. Also, to further enhance the quality of each ear images, Contrast Limited Adaptive Histogram Equalization (CLAHE) [33], a derivative of histogram equalization is applied. To avoid overfitting and memorizing the exact details of each image by CNN architectures, each ear images are automatically resized by 30 pixels horizontally and vertically in all directions. As preparation for transfer learning, 80% of the preprocessed ear images are allocated for the training while the remaining 20% is reserved for testing and validation of each CNN. There are no other methods applied aside from the processes stated above. Figure 2 shows sample of collected and preprocessed ear images.

C. Pre-Trained CNN Architecture for Feature Extraction

Third step of the study includes feature extraction and fine-tuning which are the core methods in training and modeling of the study. Subsection C discusses the used CNN architectures for modeling unconstrained ear recognition. Earlier layers of each architecture are used as feature extractor.

1) AlexNet

AlexNet is a CNN architecture proposed by Alex Krizhevesky and others that won the most difficult ImageNet challenge for visual object recognition in 2012 [34]. AlexNet is considered as the first state-of-the-art deep learning approach after it outperformed traditional computer vision methods in terms of accuracy and recognition rates [41]. It is the most well-studied CNN architecture due to its impact in most image classification tasks [42]. The architecture and network design of AlexNet is composed of 8 sequential layers with ~60 million parameters for a total of 25 layers. The basic structure of AlexNet architecture is composed of 5 convolutional layers with max-pooling, 2 fully connected layers, and a softmax layer. Softmax layer is the activation layer that connects the architecture to 1000 classes while earlier layers treated as feature extractors. AlexNet can induce 4096-dimensional feature vector from each image at the final layer which contain activations of hidden layers.

2) GoogLeNet

GoogLeNet is incarnated from the Inception architecture and is 22 layers deep composed of Inception modules for a total depth of 142 layers [36]. Although it is a deep architecture, its main goal is to learn with reduced parameters. It is a small CNN that is very close to human level performance with its exploitation of multiple layers of convolution or inception (LeNet-inspired) modules in parallel and bottleneck strategy to control overfitting and parameter explosion. On architectural-level performance, GoogLeNet uses small convolutions, batch normalization, and factorization and because it uses smaller convolutions, parameters are drastically reduced within the network. GoogLeNet’s parameter is ~4 million which is 15-times smaller than AlexNet.

3) Inception-v3

ReCeption or Inception-v3 is an improved version of GoogLeNet first used in 2016 ILSVRC Image Classification Challenge [37]. Instead of focusing on dimensionality reduction, factorization is introduced such that parameters will be reduced without decreasing the efficiency of network. Compared to GoogLeNet, ReCeption is 42 layers deep with a depth of 316 layers.

4) Inception-ResNet-v2

Inception-ResNet is a hybrid CNN model from the fusion of Inception-v3 and ResNet architectures, introduced by Szegedy and others in 2016 [43]. The main philosophy of this network is to go deeper with convolutions without sacrificing accuracy hence, ResNet’s philosophy is infused into Inception’s dimensionality reduction. Inception-ResNet is found to have the same performance as that of Inception-v4. On its architectural view, it is a 164 layers deep neural network composed of stem layers and residual blocks of inception modules totaling to a depth of 825 layers. Stem layers are a network of convolution layers found on the input layer of the architecture.

5) ResNet

Unlike the above-stated CNN architectures, ResNet architecture has the capacity to manage degradation of image classification accuracy on deeper layer of convolutions [39]. Neural networks traditionally learned through a stack of convolution layers, the deeper the depth, the deeper the model will learn. However, with the network depth increases, accuracy saturates and rapidly degrade. As a solution, ResNet introduces the use of residual blocks for learning. On architectural view, ResNet is composed of combined multiple sized convolution filters that can manage accuracy degradation and reduces training time. In this paper, ResNet-18 and ResNet-50 are used. ResNet-18 is a version of ResNet that has 2 layers of convolution in a residual block and is 18 layers deep for a total depth of 75 layers. ResNet-50 on the other hand depth 3 layers of convolution in a residual block and is 50 layers deep for a total depth of 177 layers.

6) SqueezeNet

SqueezeNet is a CNN architecture developed by Forrest Iandola and others in 2016. It dubbed as scaled-50 AlexNet as it can achieve AlexNet accuracy with 50x reduced parameters [35]. SqueezeNet is mainly composed of fire modules with compression strategies onto its layers. Fire module is a set of squeezed convolutional neural layers with parameters equivalent to 1x1 filters instead of the traditional 3x3.

Retrieval Number: B2865078219/19©BEIESP
DOI: 10.35940/ijrte.B2865.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Filter replacement and input channels limitation surrounds the core of SqueezeNet in reducing parameters with highest accuracy result.

Structurally, SqueezeNet begins with a standalone convolution layer followed by 8 fire modules and end with a final convolution layer for a total depth of 68 layers. Max-pooling with a stride of 2 are applied at the end of first and final convolution, and fourth and eight fire modules. Purposively, the design of the architecture is to increase the number of filters per fire module from beginning to end. Compared to AlexNet, SqueezeNet can classify accurately with only 1.24 million parameters.

Table 1. Fine-tuning Configuration of Pre-trained CNN Models

CNN Architecture	Layer #	Pre-Trained (1000 Classes Object Recognition)			Fine-Tuning (10 Classes for Ear Recognition)		
		Name	Type	Details	Name	Type	Details
AlexNet	23	fc8	fully connected	1000 fully connected	newFC	fully connected	10 fully connected
	24	prob	softmax	n/a	newProb	softmax	
	25	output	crossentropyex	n/a	newOutput	crossentropyex	
GoogLeNet	142	loss3_classifier	fully connected	1000 fully connected	newFC	fully connected	10 fully connected
	143	prob	softmax		newProb	softmax	
	144	output	crossentropyex		newOutput	crossentropyex	
Inception-v3	314	prediction	fully connected	1000 fully connected	newFC	fully connected	10 fully connected
	315	softmax	softmax		newProb	softmax	
	316	output	crossentropyex		newOutput	crossentropyex	
Inception-ResNet-v2	823	predicitons	fully connected	1000 fully connected	newFC	fully connected	10 fully connected
	824	predictions_softmax	softmax		newProb		
	825	classificationpredictions	crossentropyex		newOutput		
ResNet-18	70	fc1000	fully connected	1000 fully connected	newFC	fully connected	10 fully connected
	71	prob	softmax		newProb	softmax	
	72	classificationlayer_predictions	crossentropyex		newOutput	crossentropyex	
ResNet-50	175	fc1000	fully connected	1000 fully connected	newFC	fully connected	10 fully connected
	176	prob	softmax		newProb	softmax	
	177	classificationlayer_predictions	crossentropyex		newOutput		
SqueezeNet	67	prob	softmax	1x1x1000	newProb	softmax	1x1x10
	68	output	crossentropyex	n/a	newOutput	crossentropyex	
ShuffleNet	171	node_202	fully connected	1000 fully connected	newFC	fully connected	10 fully connected
	172	node_203	softmax		newProb	softmax	
	173	classificationlayer_node_203	crossentropyex		newOutput	crossentropyex	
MobileNet-v2	153	logits	fully connected	1000 fully connected	newFC	fully connected	10 fully connected
	154	logits_softmax	softmax		newProb	softmax	
	155	classificationlayer_logits	crossentropyex		newOutput	crossentropyex	

7) ShuffleNet

In 2018, [40] introduces ShuffleNet as an extremely efficient convolutional neural network for mobile devices. Shuffling of input image channels is the main method introduced in ShuffleNet. Compared to MobileNet architectures, ShuffleNet focuses on group convolution of image channels with point-wise group convolution than depth-wise group convolution. In this paper, this architecture is used to test its effectivity as base CNN model for unconstrained ear recognition despite that its main core is specifically designed for embedded devices.

8) MobileNet-v2

MobileNet is one of the recent (2018) CNN architecture developed by Sandler and others [38]. It is a light architecture that uses linear bottleneck depth-separable convolution with inverted residuals for image flattening in image classification tasks. Meaning, single convolution separately processes a color channels instead of three or more channels at once. The overall architecture of MobileNet is composed of 30 layers with stride 2 convolutional layers, depthwise layer, pointwise layer that doubles the number of channels, depthwise layer with stride 2, and another pointwise layer doubling the number of channels for a total depth of 155 layers. Although this architecture is designed to be used in embedded systems, its effectivity as base architecture for ear biometrics is tested and measured in this study.

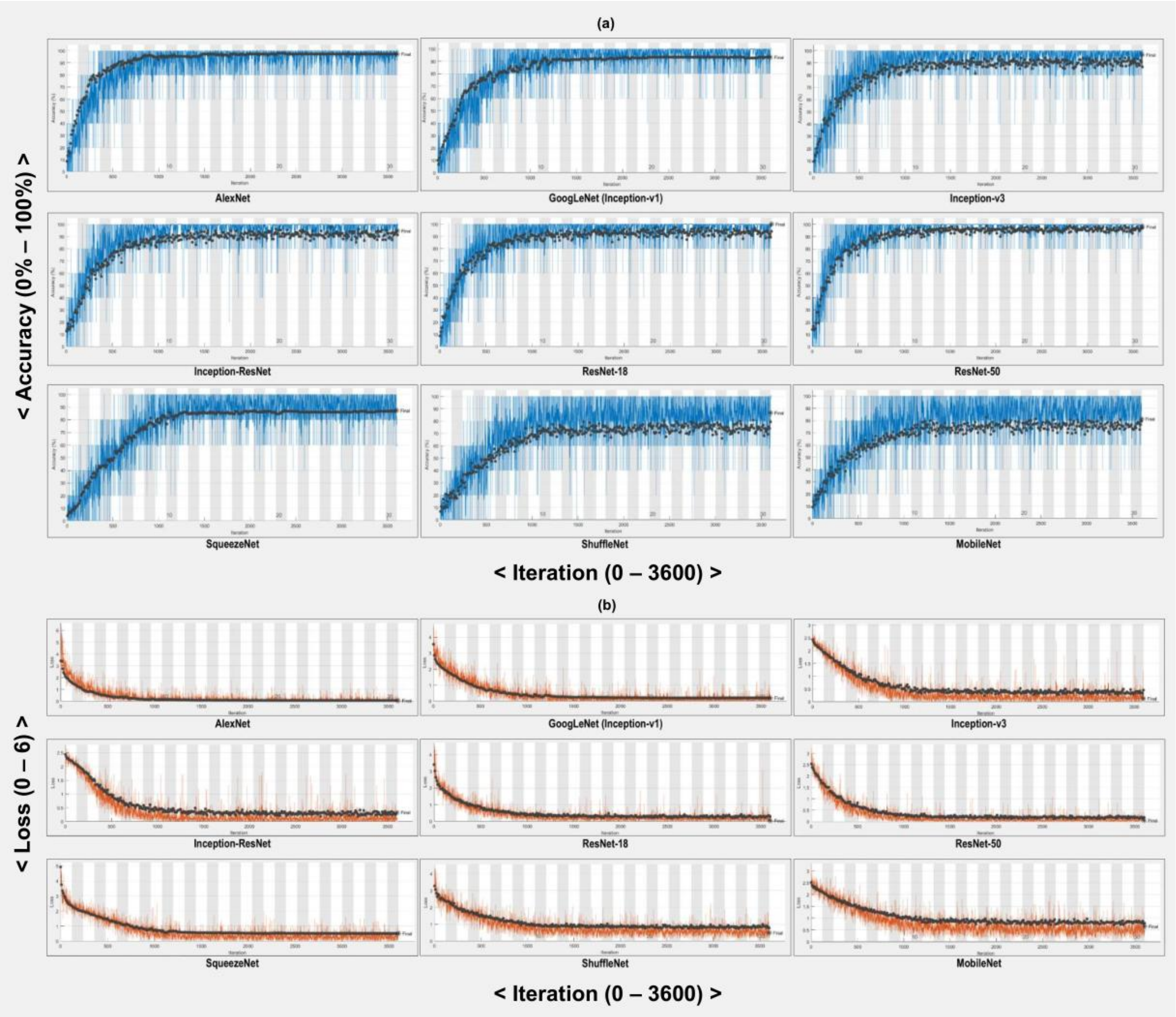


Figure 3. Training and Validation (a)Result and (b)Loss of used CNN Architectures. Series of black dots refer to validation line.

D. Fine-Tuning of Pre-Trained CNN Architectures

Subsection D discusses the final method used in the third step of the used framework of the study. Fine-tuning of the above-mentioned CNN models is the starting point of transfer learning or domain adaptive deep learning in this study. Earlier layers of each CNN architectures are used for feature extraction while the last 2-3 layers are for learning. To let each architecture, learn 10 classes of ears in unconstrained settings, the last 2-3 layers of each architectures are replaced with layers of the same kind but of different parameters. Table 1 shows the summarized and structured details of the layers replaced in each architecture. Table 1 also include the parameters applied on each altered layers of used architectures.

Each architecture is trained and fine-tuned with the same configuration. Each model is architected to learn 10 classes of ears in an unconstrained environment with a weight rate and neuron bias rate of 20 on both the fully connected layers and softmax or activation to further accelerate the process of learning on new layers. Fully connected, softmax, and classification layers are replaced in each architecture except for SqueezeNet as it is designed to learn with only convolutional layers and a softmax being present [35].

IV. RESULTS AND DISCUSSION

Table 3. Individual Test Results of each Unconstrained Ear Recognition Models

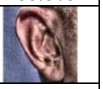
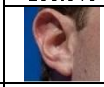
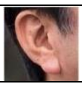
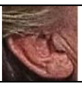
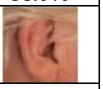
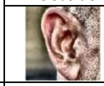
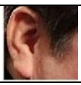
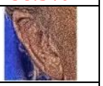
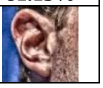
AlexNet	P6 100.0%	P0 100.0%	P0 100.0%	P5 100.0%	P1 100.0%	P2 100.0%	P8 100.0%	P1 100.0%	P5 100.0%	P4 100.0%
										
GoogLeNet	P5 100.0%	P0 100.0%	P8 100.0%	P3 86.9%	P2 91.2%	P7 100.0%	P4 92.8%	P1 99.9%	P6 100.0%	P7 100.0%
										
Inception-v3	P1 99.9%	P8 99.1%	P0 98.8%	P4 99.5%	P2 77.9%	P5 99.9%	P9 84.6%	P2 77.6%	P9 86.5%	P7 98.4%
										
Inception-ResNet	P9 48.8%	P7 98.7%	P5 99.9%	P1 99.7%	P9 90.5%	P3 99.0%	P6 96.1%	P7 98.4%	P3 92.9%	P2 98.0%
										
ResNet-18	P8 100.0%	P7 100.0%	P5 100.0%	P7 99.9%	P5 100.0%	P2 99.2%	P6 99.7%	P6 99.9%	P3 93.5%	P1 96.8%
ResNet-50	P5 100.0%	P2 100.0%	P4 73.5%	P1 24.8%	P1 100.0%	P9 100.0%	P0 100.0%	P5 97.0%	P9 98.4%	P7 95.0%
SqueezeNet	P4 100.0%	P0 100.0%	P9 99.9%	P1 99.6%	P6 100.0%	P3 98.5%	P5 99.9%	P6 64.9%	P9 62.3%	P3 98.3%
										
ShuffleNet	P0 99.8%	P2 99.5%	P6 98.7%	P0 100.0%	P3 56.3%	P1 99.2%	P4 32.7%	P3 92.4%	P5 51.6%	P8 65.6%
MobileNet-v2	P7 26.8%	P0 52.7%	P9 45.7%	P7 99.9%	P3 52.2%	P5 44.4%	P0 56.3%	P1 58.3%	P2 80.5%	P5 81.13%
										

Table 2. Summary of Confusion Matrices, Training and Validation Results, and Consumed Training Time of used CNN Architectures

CNN Architecture	15 ear image samples per person (P)										Summation of P0 to P9 (% Accuracy)	Total Training Time (min)
	P0	P1	P2	P3	P4	P5	P6	P7	P8	P9		
AlexNet	10.0	10.0	10.0	9.3	9.3	8.7	10.0	10.0	10.0	10.0	97.3	19.36
GoogLeNet	10.0	10.0	9.3	8.7	10.0	8.7	10.0	8.7	9.3	8.7	93.3	97.43
Inception-v3	10.0	10.0	10.0	8.7	9.3	9.3	10.0	9.3	10.0	10.0	96.7	269.7
Inception-ResNet	10.0	10.0	9.3	9.3	8.0	9.3	10.0	8.7	10.0	10.0	94.7	1078.37
ResNet-18	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	100.0	48.43
ResNet-50	10.0	10.0	10.0	10.0	10.0	7.3	10.0	9.3	10.0	10.0	96.7	159.29
SqueezeNet	8.7	9.3	8.0	9.3	7.3	8.7	9.3	10.0	8.7	8.0	87.3	40.28
ShuffleNet	9.3	9.3	8.7	6.0	9.3	8.0	10.0	10.0	9.3	10.0	86.7	163.35
MobileNet	10.0	8.7	8.7	3.3	8.0	8.0	9.3	8.7	9.3	7.3	81.3	155.37

Training, Testing and Validation of the study are commenced in the third and last step of the framework. 80% of total unconstrained ear images of different persona are trained with each fine-tuned CNN architecture. The training of these images is commenced with a single GPU natively built on top of a 7th generation i7 processor with a memory size of 12GB. Fixed configuration of training is set with a minimum batch-size of 5 and maximum epoch or training times of 30. This setup yields to 120 iterations per epoch for a total of 3600 maximum iterations. Maximum iteration is the set stabilization level at which validation accuracy is seen stabilized in all CNN architectures. 0.00001 is the set initial learning rate to each CNN architecture training such that new layers might catch-up to the pace of learning of existing layers. Due to the effectivity of 'adam' optimizer in achieving good results in most deep learning applications [44, 45], it is the chosen training function for the used CNN architectures in this study instead of using stochastic gradient descent momentum or SGDM. Shown in figure 3a are the training and validation results of used CNN architectures.

AlexNet, GoogLeNet, and SqueezeNet shows fine and smooth training and validation curves compared to the results of other CNN architectures. This observation of results is reflected to the results of training and validation loss as shown in figure 3b. To further generalize the performance and result of each used architecture in terms of accuracy, above-95.0% accuracy confidence must be met at each architecture's final validation result. This standard is adapted from the paper of [46]. However, only four of the used architectures achieved this standard namely AlexNet (97.3%), Inception-v3 (96.7), ResNet-18 (100.0%), and ResNet-50 (96.7). Recently developed CNN architectures and mobile architectures shows promising results despite of not achieving the 95% classification standard. AlexNet, GoogLeNet, and SqueezeNet shows smooth curve of training loss while ResNet-18 and ResNet-50 shows fine curve compared to the outcome of Inception-v3, Inception-ResNet, and recent CNN architectures. It is strongly believed that the occlusions included within the image area of unconstrained ears had contributed to these losses that degrade the performance of CNN architectures. Nevertheless, it is also believed that by increasing the maximum iteration instead of using the set 3600 might enhance the performance of Inception-ResNet, SqueezeNet, ShuffleNet, and MobileNet as basis for ear recognition in the context of transfer learning.

The remaining 20% of the dataset is used for testing. Shown in table 2 is the summarized result of the confusion matrices of each CNN architectures. Top 4 CNN architectures that achieved above 95.0% classification accuracy shows promising results in classifying ears with consideration of their training time. ResNet-18 had been trained for 48.43 minutes to be able to fully recognize ears in unconstrained environment with an accuracy confidence of 100.0%. AlexNet is trained the fastest with an accuracy confidence of 97.3%, while ResNet-50 (159.29 minutes), and Inception-v3 (269.7 minutes) are trained longer due to the depth of their layers and achieved an accuracy confidence of 96.7% on final iteration. Training time of other CNN architectures are shown in table 2. Table 3 shows random test results of each CNN architectures with 10 samples each to further validate the results of used CNN architectures.

Actual validation results shown in table 3 states that AlexNet, GoogLeNet, Inception-v3, ResNet, and ShuffleNet can classify ears in unconstrained environment. GoogLeNet, despite of achieving 93.3 accuracy confidence and finished training for ~98 minutes performed well on actual validation. Inception-ResNet-v2 and SqueezeNet shows promising classification results however is not able to classify ears correctly. MobileNet-v2 appears to have difficulty in performing learning and classification tasks like ear recognition making it yields below 95% accuracy confidence and misclassifications, however, is believed that might be used if trained with longer maximum iterations.

Overall, results above states that network structures of AlexNet, ResNets, Inception-v3, and ShuffleNet architectures are best to be used as basis in unconstrained ear recognition and/or biometrics. GoogLeNet posed above promising results and might be used as modeling CNN for the defined task.

V. CONCLUSION AND FUTURE SCOPE

This paper explores the concept of domain adaptive deep learning or transfer learning in unconstrained ear biometric recognition through computer vision and transfer learning using 9 handpicked CNN models: AlexNet, GoogLeNet, Inception-v3, Inception-ResNet, ResNet-18, ResNet-50, SqueezeNet, ShuffleNet, and MobileNet-v2. Each layers of CNN architectures are analyzed structurally such that context of transfer learning will be used in the forms of feature extraction and fine-tuning. 80% of acquired ear images from the web is used as training data while the remaining 20% is used for testing and validation. Performance results of each generated unconstrained ear recognition models are compared in context of their training time, learned features and losses, and accuracy confidence in recognizing ear images in the wild. It is identified that ResNet-18, AlexNet, Inception-v3, GoogLeNet, and ShuffleNet are the best CNN models for an unconstrained ear biometric recognition in context of domain adaptive deep learning or transfer learning.

For future studies, further exploration of other CNN architectures developed later than MobileNet-v2 and/or handcrafted CNN architectures for ear recognition should be performed under a standard training configuration. Also, adaptation of the methods used in this study to other biometric studies should be attempted to provide better benchmarking and references towards the development of different advances in biometric engineering to the context of deep learning.

ACKNOWLEDGMENT

The authors of this paper would like to acknowledge both the faculty and staff of both the Department of Computer Engineering and the Department of Graduate Studies of Technological Institute of the Philippines Quezon City for their support and guidance to the development of this study.

REFERENCES

1. T. Nagasawa, H. Tabuchi, H. Masumoto, H. Enno, M. Niki, H. Oshugi and Y. Mitamura, "Accuracy of deep learning, a machine learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting idiopathic macular holes," PeerJ, vol. 6, no. e5696, pp. 1-10, 2018.
2. S. Ruder, "An overview of gradient descent optimization algorithms," CoRR, vol. abs/1609.04747, 2016.
3. D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," CoRR, vol. abs/1412.6980, 2014.
4. C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," ICLR 2016 Workshop, 2016.
5. A. Khan, A. Sohail, U. Zahoora and A. S. Qureshi, "A Survey of the Recent Architectures of Deep Convolutional Neural Networks," arXiv, 2019.
6. M. Alom, T. Taha, C. Yakopcic, S. Westberg, M. Hasan, B. Esesn, A. Awwal and V. Asari, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," ArXiv, 2018.

I. Zhang, X. Zhou, M. Lin and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," in 2018 IEEE/CVF Conference on Computer Vision and Pattern

Recognition, Salt Lake City, UT, USA, 2018
7. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016.
8. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018.
9. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
10. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Vanchoucke and A. Rabinovich, "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
11. F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size," CoRR, vol. abs/1602.07360, p. 1602.07360, 2016.
12. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet Classification with Deep COnvolutional Neural Networks," Advances in Neural Information Processing Systems, pp. 1097-1105, 2012.
13. Y. Chang, C. Jung, P. Ke, H. Song and J. Hwang, "Automatic Contrast-Limited Adaptive Histogram Equalization with Dual Gamma Correction," IEEE Access, vol. 6, pp. 11782 - 11792, 2018.
14. Y. H. C. Q. Z. L. Y. L. C. L. Dengyu Xiao, "Transfer learning with convolutional neural networks for small sample size problem in machinery fault diagnosis," Journal of Mechanical Engineering Science, 2019.
15. S. Dodge, J. Mounsef and L. Karam, "Unconstrained ear recognition using deep neural networks," IET Biometrics, vol. 7, no. 3, pp. 207-214, 2018.
16. S. Khan, N. Islam, Z. Jan, I. U. Din and J. J. P. C. Rodrigues, "A Novel Deep Learning based Framework for the Detection and Classification of Breast Cancer Using Transfer Learning," Pattern Recognition Letters, vol. 125, no. 1 July 2019, pp. 1-6, 2019.
17. S. Fairuz, M. H. Habaebi and E. M. A. Elsheikh, "Finger Vein Identification Based on Transfer Learning of AlexNet," in 2018 7th International Conference on Computer and Communication Engineering (ICCCE), Kuala Lumpur, Malaysia, 2018.
18. A. Mallouh, Z. Qawaqneh and B. D.Barkana, "Utilizing CNNs and transfer learning of pre-trained models for age range classification from unconstrained face images," Image and Vision Computing, vol. 88, no. August 2019, pp. 41-51, 2019.
19. S. H. Choudhury, A. Kumar and S. H. Laskar, "Biometric Authentication through Unification of Finger Dorsal Biometric Traits," Information Sciences, vol. 497, no. September 2019, pp. 202-218, 2019.
20. M. Talo, U. B. Baloglu, O. Yildirim and U. R. Acharya, "Application of deep transfer learning for automated brain abnormality classification using MR images," Cognitive Systems Research, vol. 54, no. May 2019, pp. 176-188, 2019.
21. A. Almisreb, N. Jamil and N. M. Din, "Utilizing AlexNet Deep Transfer Learning for Ear Recognition," in 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), Kota Kinabalu, Malaysia, 2018.
22. Ž. Emeršič, D. Štepec, V. Štruc and P. Peer, "Training Convolutional Neural Networks with Limited Training Data for Ear Recognition in the Wild," in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 2017
23. Y. Zhang, Z. Mu, L. Yuan and C. Yu, "Ear verification under uncontrolled conditions with convolutional neural networks," IET Biometrics, vol. 7, no. 3, pp. 185-198, 2018.
24. S. Khalid, T. Khalil and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in 2014 Science and Information Conference, London, UK, 2014.
25. T.Banzato, G.B.Cherubini, M.Atzori and A.Zotti, "Development of a deep convolutional neural network to predict grading of canine meningiomas from magnetic resonance images," The Veterinary Journal, vol. 235, no. May 2018, pp. 90-92, 2018.
26. R. Yamashita, M. Nishio, R. K. G. Do and K. Tagashi, "Convolutional neural networks: an overview and application in radiology," Insights into Imaging, vol. 9, no. 4, pp. 611-629, 2018.
27. J. Zbontar and Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches," The Journal of Machine Learning Research, vol. 17, no. 1, pp. 2287-2318, 2016.
28. H. Hofbauer, E. Jalilian and A. Uhl, "Exploiting superior CNN-based iris segmentation for better recognition accuracy," Pattern Recognition Letters, vol. 120, no. 1 April 2019, pp. 17-23, 2019.
29. Y. Zhang, Y. Huang, L. Wang and S. Yu, "A comprehensive study on gait biometrics using a joint CNN-based method," Pattern Recognition, vol. 93, no. September 2019, pp. 228-236, 2019.
30. P. L. Galdamez, W. Raveane and A. G. Arrieta, "A brief review of the ear recognition process using deep neural networks," Journal of Applied Logic, vol. 24 Part A, no. November 2017, pp. 62-72, 2017.
31. Z. Emersic, V. Struc and P. Peer, "Ear recognition: More than a survey," Neurocomputing, vol. 255, pp. 26-29, 13 September 2017.
32. H. Chen and B. Bhanu, "Human Ear Recognition in 3D," in IEEE Transactions on Pattern Analysis and Machine Intelligence, Washington, DC, USA, 2007.
33. Z. Emersic, D. Stepec, V. Struc, P. Peer, A. George, A. Ahmad, E. Omar, T. Boult, R. Safdaii, Y. Zhou, S. Zafeiriou, D. Yaman, Y. Eyiokur and H. Ekenel, "The unconstrained ear recognition challenge," in 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 2017.
34. D. Hurley, B. Arbab-Zavar and M. Nixon, "The Ear as Biometric," in 15th European Signal Processing Conference (EUSIPCO 2007, Poznan, Poland, 2007.
35. S. Anwar, K. K. A.Ghany and H. Elmahdy, "Human Ear Recognition Using Geometrical Features Extraction," in International Conference on Communication, Management and Information Technology (ICCMIT 2015), 2015.
36. K. Delac and M. Grgic, "A survey of biometric recognition methods," in Proceedings. Elmar-2004. 46th International Symposium on Electronics in Marine, Zadar, Croatia, Croatia, 2004.
37. Jain, A. Ross and S. Prabhakar, "An Introduction to Biometric Recognition," IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, pp. 4-20, 2004.
38. Ross and A. Jain, "Biometrics, Overview," in Encyclopedia of Biometrics, Boston, MA, Springer, 2009, p. 8.
39. Tariq, M. A. Anjum and M. U. Akram, "Personal Identification Using Computerized Human Ear Recognition System," in 20 II International Conference on Computer Science and Network Technology, Harbin, China, 2011.
40. J. M. White, Security Risk Assessment: Managing Physical and Operational Security, Butterworth-Heinemann, 2014.
41. D.-X. Zhou, "Universality of deep convolutional neural networks," Applied and Computational Harmonic Analysis, no. June 2019, pp. 1-13, 2019.
42. M. Singh, R. Singh and A. Ross, "A comprehensive overview of biometric fusion," Information Fusion, vol. 52, no. December 2019, pp. 187-205, 2019.
43. Elmahmudi and H. Ugail, "Deep face recognition using imperfect facial data," Future Generation Computer Systems, vol. 99, no. October 2019, pp. 213-225, 2019.
44. Z. Zhaoxiang, S. Shiguang, F. Yi and S. Ling, "Deep Learning for Pattern Recognition," Pattern Recognition Letters, vol. 119, no. March 2019, pp. 1-2, 2019.
45. D. Rong, L. Xie and Y. Ying, "Computer vision detection of foreign objects in walnuts using deep learning," Computers and Electronics in Agriculture, vol. 162, pp. 1001-1010, 2019.

AUTHORS PROFILE

Marwin B. Alejo earned his Bachelor of Science in Computer Engineering at Technological Institute of the Philippines Quezon City in 2015 and is currently completing his Master of Engineering at Technological Institute of the Philippines. He had been a Computer Engineering faculty at Technological Institute of the Philippines and is currently a faculty at National University Manila while being engaged into various computer vision and deep learning researches. Before going to academe, he had been part of the industry for research works in remote sensing, data science, machine learning, and embedded systems. Partaking on community growth and involvement, he had also provided speakership and researches in the fields of data science, computer vision, machine learning, artificial intelligence, and emerging engineering technologies.

Cris Paulo G. Hate received his Bachelor of Science and Master of Engineering in Computer Engineering at Technological Institute of the Philippines in 2014 and 2016. Currently, he is taking his Doctor of Engineering at the same institution. He is a faculty of Technological Institute of the Philippines Quezon City under the Department of Computer Engineering, College of Engineering and Architecture and the Graduate Studies Department. He also authored multiple articles in journals and conference proceedings. He is also a Cisco Certified Network Associate in Routing and Switching Technologies. His research interests include embedded system, robotics, internet-of-things, algorithm analysis, artificial intelligence, and emerging technologies.

Retrieval Number: B2865078219/19©BEIESP
DOI: 10.35940/ijrte.B2865.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication