Proc. 1st International Conference on
Machine Learning and Data Engineering (iCMLDE2017)
20-22 Nov 2017, Sydney, Australia
ISBN: 978-0-6480147-3-7

Automatic Ear Detection using Deep Learning

MD Moniruzzaman[1] and Syed Mohammed Shamsul Islam[2]

[1]Postgraduate Student, School of Science, Edith Cowan University, Joondalup, WA 6027, Australia
[2]Lecturer, School of Science, Edith Cowan University, Joondalup, WA 6027, Australia

Corresponding author's E-mail: syed.islam@ecu.edu.au

Abstract

Ear based biometric identification can be the solution for instance such as surveillance where other biometric traits are simply very hard to access. Although many semi-automatic approaches have been made to detect ear and use it for human recognition purpose, most of them are based on feature extraction or shallow machine learning approaches. Very few approaches those used deep neural network architectures are either having less hidden layers, or a combination of deep neural network and feature extraction classifiers or already trained complex deep convolutional network. In this research approach, a deep but simple raw convolutional neural network have been used to detect ear from an ear and non-ear environment which is the initial part of ear based biometric implication. Data-Augmentation have also used and a comparative analysis have been done for original data set and Augmented dataset. Using this deep but from scratch architecture trained by 792 original images we achieved promising output which show larger data could achieve higher accuracy.

Keywords: *Ear detection, Deep Learning, Convolutional Network, Data augmentation*.

1. INTRODUCTION

During the course of last decade, researchers are focusing actively on biometrics research due to the growing need for automatic authentication of human individuals. In biometric identification, ear based authentication and human recognition proved itself a novel research field as suggested by Zhang and Mu (2017). The potential, credentials and possibilities of using human ear for recognition and human identification was proposed by Bertillon (1890). According to Jain et al (2007) and Omara et al (2017), to satisfy the traditional personal authentication properties, a biometric trait should be unique, universal, permanent and easily collectible. The uniqueness of the helix part of human ear has been proven by Iannarelli (1989). Not only helix, anti-helix, which is a parallel part of helix with a distinctive hairpin-bend shape also proved prominent by Hurley et al (2007). Human facial expression doesn't have any effect on ear, with age the change is negligible and the background is always predictable as ear is fixed firmly on the side of human head. Collection of ear image does not create a sense of anxiety and no hygiene issue raises as the collection process is touch-less. Moreover, as the size of the ear is significantly larger than other biometric traits like finger print, iris, retina, it proves more accessible while the incident is associated with criminal act. Present Artificial Neural Network based ear detection and recognition approaches can be broadly categorized in to 3D point cloud based approaches or 2D image data approaches. This research work deals with only 2D images.

One of the earliest approaches of using Neural Network based Ear recognition was experimented by Galdámez et al (2014). In their approach they used Speeded Up Robust Features (SURF) and Fisher Linear Discriminant Analysis (LDA) as the input of their two neural networks. On their later approach Galdámez et al. (2016) used deep Convolutional neural networks (CNN) based ear recognition system to identify a person by the ear image. Tian and Mu (2016) designed a neural network having three convolutional layer followed by a fully connected layer and a soft-max classifier. They used University of Science and Technology, Beijing (USTB-III) dataset. Emeršič et al. (2017) tried to train convolutional neural networks with small training data. They implemented aggressive data augmentation with selective model learning technique for limited training data. For the solution of lack of training data problem for deep learning based technique Omara et al. (2017) proposed a technique called pairwise SVM. Their neural networked was based on VGG-M Net. Along with pairwise SVM, they also used Principal Component Analysis (PCA) so that the dimension can be reduced before classification. They used USTB-I and USTB-II dataset for their experiment and Accuracy was more than 98%. Cintas et al (2017) proposed a new approach based on deep learning and Geometric Morphometrics for feature extraction and ear detection in the form of landmarks. They trained three different convolutional neural networks with a set of manually landmarked examples. They used a dataset (CANDELA, Consortium for the Analysis of the Diversity and Evolution of Latin Americans) of 2735 manually landmarked ear images. Zhang and Mu (2017) in their recent approach, proposed an efficient method based on Multiple Scale Faster Region-Based Convolutional Networks (Faster R-CNN) for ear detection from 2D images. For this experiment they created an ear dataset consisting 100 images collected from internet and are captured from real world situation and named it WebEar dataset. Along with this dataset they also experimented on UBEAR and UNDJ-2 (University of Notre Dame J2) dataset. All the above approaches are summarised into Table 1.

Table 1. Summary of Deep Neural Network based ear detection and recognition approaches

Author	Technique Used	Number of Image	Accuracy
Galdámez et al (2014)	SURF and LDA as input of Neural Network	300	97% for SURF and 93% for LDA
Galdámez et al (2016)	CNN Architecture with sliding window optimisation	52800	99.02%
Tian and Mu (2016)	3 Layer CNN	79 Subjects	98.27%
Emeršič et al (2017)	Aggressive data augmentation and Selective model learning	2304	>95%
Omara et al. (2017)	Pairwise SVM & PCA	488	>98%
Cintas et al (2017)	Deep Learning with Geometric Morphometric	2735	
Zhang and Mu (2017)	Faster R-CNN	10995	>98%

The very first step of ear based recognition system is automatic ear detection. The main intention of the project was to experiment and judge the possibility of using a raw deep convolutional neural network architecture based algorithm to detect ear in the small 2D data sets available for ear and maximising efficiency and reducing time to ensure practical use of the system. The outcome of the project will set the base for a fully automated, efficient and real time ear based human authentication system in the future work.

The organisation of the rest of the paper is as follows. The elaboration of the approached research design and methodology is discussed in section 2. Experimental results and discussion is provided at section 3. Finally conclusion is drawn and indication of future research is provided in section 4.

2. RESEARCH DESIGN AND METHODOLOGY

Three separate attempts have been made to detect ear from an ear and no-ear environment in this research approach using Deep Convolutional Neural Network. In this Section the architecture of the Deep Neural Network have been have been discussed along with the description of the dataset used and the Data Augmentation technique applied as well as training and testing methods.

2.1. Dataset

To create a perfect ear and non-ear environment for ear detection purpose, total five classes of images were selected for experiment one and dataset-1 was created. The five image categories involved digital photos of cat, dog, ear, horse and human. Cat, dog, ear and human every category have 202 images collected from github (https://github.com/anujshah1003/own_data_cnn_implementation_keras) and for ear 182 images were used from University of Science and Technology Beijing- I (USTB-I) [Omara et al (2017)]. So the whole dataset includes 990 images. For Second attempt the image class have been reduced from five to three that includes cat, ear and horse. With the previous 202 images for cat and horse, another 300 images were added from Canadian Institute for Advanced Research (CIFAR-10) database for each of the category to create dataset -2. And for ear 309 images were added from USTB-II and 11 photos were added from USTB-III. So for second approach 1506 images were used from 3 classes, each having 502 images. For third attempt, the size of the data set was 10574, including 3542 ear images, 3514 cat images and 3518 horse images. This dataset-3 is made up with all the images from dataset-2 and further data augmentation.

2.2. Data Augmentation

It is commonly accepted that the more data can be used during the Deep Neural Network training phase, the more effective it becomes for validation phase [Wang and Perez (2017)]. As for first and second attempts, the size of the data were fairly small, Data augmentation technique was used to increase the size of the total data more than ten thousands. The augmentation procedures that were performed in this experiment is listed below.

1. Rotation of up to 40° in both direction
2. Width sift from 0 to 20% range
3. Height shift from 0 to 20% range
4. Shearing from 0 to 20% range
5. Image Zoom up to 20%
6. Horizontal flipping

Figure 1. Sample Ear images after Augmentation

2.3. Network Architecture

To keep the consistency in the research, for all the attempts same architecture have been used. The architecture of the Deep Neural Network used in this project (https://github.com/anujshah1003/own_data_cnn_implementation_keras) is shown in the Table-3 below.

Table 3. Layers of Deep Neural Network Architecture Used

Layer Index	Type	Filter Size	Output Size
L1	Convolution2D	3 × 3	32 × 126 × 126
L2	ReLU		32 × 126 × 126
L3	Convolution2D	3 × 3	32 × 126 × 126
L4	ReLU		32 × 126 × 126
L5	MaxPooling2D	2 × 2	32 × 63 × 63
L6	Dropout		32 × 63 × 63
L7	Convolution2D	3 × 3	64 × 61 × 61
L8	ReLU		64 × 61 × 61
L9	MaxPooling2D	2 × 2	64 × 30 × 30
L10	Dropout		64 × 30 × 30
L11	Flatten		57600 × 1
L12	Dense		64 × 1
L13	ReLU		64 × 1
L14	Dropout		64 × 1
L15	Dense		5 × 1
L16	Softmax		5 × 1

2.4. Training and Testing

For all the three instances Datasets were divided into training set and test set. For dataset-1 training set contains 792 images and 198 test images. Dataset-2 was divided into 1204 training images and 302 test images, and dataset-3 was divided into 8459 training and 2115 test set. For all three attempts, input images were converted to 128 × 128 grey scale images. Training was done using Rectified Linear Unit (ReLU) and Softmax Activation function. Training batch size was set 16 as the training was done on a GPU less computation system. Number of epoch was 20 for all the attempts to keep the validation result trustworthy. The whole experiment was done using Theano backend Keras library on a python 2.7 based programming language. Open CV was used for image processing. The experimental setup used a computation machine having 2 GHz Intel Core i5 processor, 8GB RAM and an Intel 540 Iris Graphics card. So the whole system is really cost effective and Table 4 indicates the very less execution time.

3. RESULT AND DISCUSSION

From Table-4 we can see that the training accuracy for dataset one 86.11% but the validation accuracy is 56.56% which is fairly low. The reason behind the low validation accuracy is the very small size of the data. The size of the training data was only 792. When the data was increased to 1204 for training at dataset 2 and the number of class was decreased to three from five, the validation accuracy increased to 81.79%. So it is fairly obvious that the increase of main data increased the accuracy for the architecture.

Table 4. Training and Validation accuracy with loss for all the 3 datasets

Data Set	Execution Time Per Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
Dataset- 1	153 sec	86.11 %	56.56%	0.3920	1.3328
Dataset-2	235 sec	90.12%	81.79%	0.2616	0.3933
Dataset-3	1621 sec	87.21%	77.45%	0.3668	0.5812

Dataset-3 includes the augmented images of dataset-2. The training size of this third attempt was 8459 which is much larger than that of dataset-1 and 2. But the validation accuracy is 77.45% which is slightly lower than that of attempt-2. This results indicated that the augmentation procedure (shearing, flipping, zooming, rotation etc.) chosen for this experiment actually did not performed up to expectation. Comparison for the training and validation accuracy is presented in Figure-2.

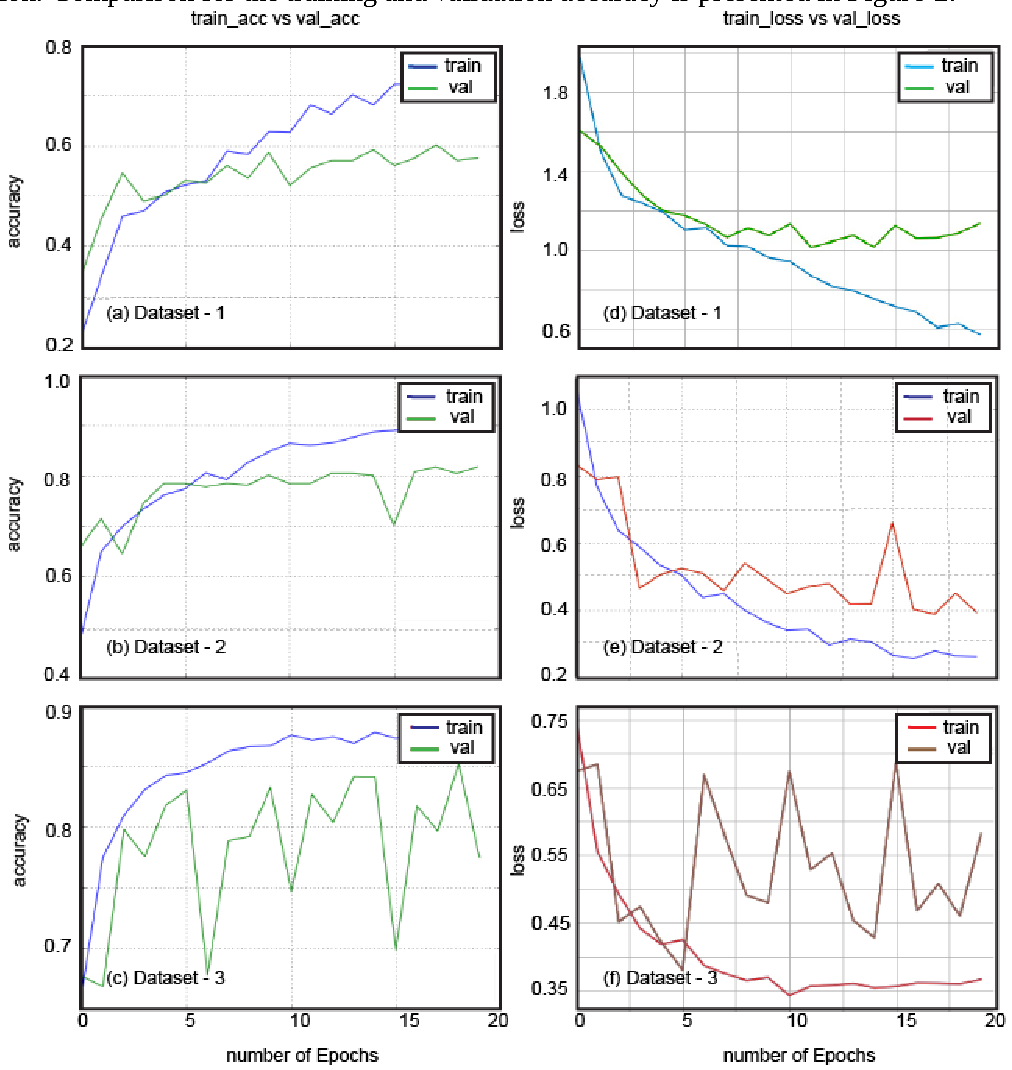

Figure 2. Train and Validation Accuracy with loss for dataset-1, 2 and 3 consecutively

For Neural Network loss is measured as negative log-likelihood which is not a percentage. The training loss and validation loss for all the attempts are presented in table-4 and Figure-2, which also indicate the correctness of the accuracy data. For attempt -2 for dataset-2 the network performed better than other two attempts.

4. CONCLUSION AND FUTURE WORK

In this research project deep learning architecture based ear detection was attempted in an ear and non-ear environment. The proposed architecture is capable of detecting ear from a mixed environment and the highest validation accuracy of 81.79 % achieved for attempt two with dataset-2 which indicates the success of the research project. The result of the experiment is considerable as the time period of the research project was very short (only 12 weeks) and experiment with deep neural network architecture takes a significant amount of work and time especially the network selection, data augmentation and training phase. The accuracy of the network would be much higher simply if the size of the original image data were larger. So the same experiment will be done after collection of more digital ear images from different data sources. This project tested the accuracy of a single deep neural network. Several other networks will be used in the future work so that a comparative analysis can be done and a more robust architecture can be selected. Regularisation, Network initialisation, batch normalisation and some other type of data augmentation technique will be used in future work which may help to increase the accuracy of the ear detection. Moreover, this experiment deals with the detection of ear images from a mixed image dataset, but the future research will be focused on the biometric implication of ears i.e., to detect a person from ear image which will address a real life problem.

ACKNOWLEDGMENTS

Authors would like to acknowledge that this work is done with the support of Australian Government Research Training Program Scholarship and Edith Cowan University ECR grant. All the ear data were collected from University of Technology, Beijing (USTB I, II, III). Non-ear data were collected from Canadian Institute for Advanced Research (CIFAR-10) and data available in Github.

REFERENCES

Bertillon A (1890). La photographie judiciaire, avec un appendice sur la classification et l'identification anthropometriques, Gauthier-Villars, Paris.

Cintas C et al (2017). Automatic ear detection and feature extraction using Geometric Morphometrics and convolutional neural networks, IET Biometrics, 6(3), 211-223.

Deepak R, Nayak AV, Manikantan K (2016). Ear detection using active contour model, IEEE Emerging Trends in Engineering, Technology and Science, 1-7.

Emeršič Ž, Štepec D, Štruc V, Peer P (2017). Training Convolutional Neural Networks with Limited Training Data for Ear Recognition in the Wild, IEEE International Conference on Automatic Face & Gesture Recognition, Washington, DC, 987-994.

Galdámez PL, Arrieta MAG, Ramón MR (2014). Ear Recognition with Neural Networks Based on Fisher and Surf Algorithms, International Conference on Hybrid Artificial Intelligence Systems, Springer, 254-265.

Galdámez PL, Raveane W, Arrieta AG (2016). A brief review of the ear recognition process using deep neural networks, Journal of Applied Logic, 62-70.

Ganesh MR, Krishna R, Manikantan K, Ramachandran S (2014). Entropy based Binary Particle Swarm Optimization and classification for ear detection, Engineering Applications of Artificial Intelligence, 27, 115-128.

Hurley DJ, Arbab-Zavar, B, Nixon, MS (2007). The ear as a biometric, 15th European Signal Processing Conference (EUSIPCO), Poznan, Poland, 25-29.

Iannarelli AV (1989). Ear identification, Paramount Publishing Company.

Jain A, Flynn P, Ross AA (2007). Handbook of biometrics, Springer Science & Business Media, 131-150 pp.

Omara I, Wu X, Zhang H, Du Y, Zuo W (2017). Learning pairwise SVM on deep features for ear recognition, IEEE International Conference on Computer and Information Science, Wuhan, 341-346.

Tian L, Mu Z (2016). Ear recognition based on deep convolutional network, International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, Datong, 437-441.

Wang J, Perez L (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning, Stanford University.
Yan P, Bowyer K (2005). Empirical evaluation of advanced ear biometrics, IEEE Computer Vision and Pattern Recognition-Workshops, 41.
Yan P, Bowyer KW (2007). Biometric recognition using 3D ear shape, IEEE Transactions on pattern analysis and machine intelligence, 29(8), 1297-1308.
Zhang Y, Mu Z (2017). Ear Detection under Uncontrolled Conditions with Multiple Scale Faster Region-Based Convolutional Neural Networks, Symmetry, 9(4), 53.