Ear Recognition Development

Assignment #3

Image Based Biometrics 2020/21, Faculty of Computer and Information Science, University of Ljubljana

Dimitar Stefanov

Abstract—**This paper presents a deep learning model for ear recognition developed for the purpose of completing Assignment #3 in the subject Image-based Biometry. The model builds upon on the already existing convolutional neural network (later in the paper referred to as CNN) VGG16 researched as part of [1]. This neural network has been extensively tested and has achieved satisfactory results, hence our decision to utilize it in our recognition problem as well. Unfortunately, due to a number of reasons, varying from the fact that VGG16 has not been constructed to classify ear images solely, to the fact that the dataset we used contains images which are rather difficult to classify (additional information on this remark can be found in [2]), we were unable to replicate the results VGG16 usually achieves in simpler classification tasks.**

I. INTRODUCTION

To begin with, we would like to provide a brief description of the dataset used for this recognition task. It consists of 1000 cropped ear images of 100 subjects collected via various sources on the Internet. This data collection named AWE Dataset which can be obtained at [3] was also the test dataset for the First Unconstrained Ear Recognition Challenge (UERC) in 2017, and as such can be considered a representative test of a model's quality. On this set of images, aside from the standard classification of persons to whom the images belong, ethnicity and gender can be analyzed as well, because along with the images, there are annotation files which contain this additional information. Therefore, we opted to perform a binary classification - male and female persons.

II. METHODOLOGY

VGG16 is a convolutional neural network architecture that takes advantage of convolution layers of 3x3 filter and maxpool layer of 2x2 filter. This arrangement of convolution and max pool layers remains consistent throughout the whole architecture. At the end, there are 2 fully connected layers followed by a softmax layer for the output. And, this last layer is the segment of the neural network we modified for the purposes of our classification. The implementation of VGG16 in Keras, the library we made use of, in its final layer classifies into a 1000 classes by default. However, the number of classes in our problem is 2, so we replaced the softmax layer with another dense layer classifying into 2 different classes.

In the machine learning community, it has been well established that for the cross-validation of a model 5-fold or 10-fold cross-validation suffices. Also, in the AWE Toolbox a 5-fold cross validation is performed, so we followed this procedure as well. In addition, we split the AWE Dataset into 70 % training data and 30 % test data, and judging by the performance of our model which will be discussed in the following sections of the paper, we believe the correct choice was made.

III. RESULTS

As we have already mentioned in the previous section, the results after the cross-validation of our model were rather satisfactory. A more detailed representation of them can be found in the following Table 1:

Table I

SUMMARIZATION OF THE 5-FOLD CROSS-VALIDATION PERFORMED ON OUR MODEL THROUGH THE COMPUTATION OF THE METRICS: ACCURACY AND AUC (AREA UNDER THE CURVE).

	validation accuracy	validation AUC
fold #1	0.95000	0.95000
fold #2	0.90000	0.90071
fold #3	0.95000	0.95071
fold #4	0.83571	0.83571
fold #5	0.91429	0.91367

Table 1 shows that the validation accuracy of our model was in all cases except fot the fourth fold above 90 %. This kind of percentages can be considered acceptable, but the value we should look more carefully at, is the validation Area Under the Curve. As we can observe, the validation AUC is also in most cases in the rank from 90 % to 100 %. Such scores mean that our neural network completes the classification task outstandingly.

Lastly, we would like to mention the accuracy our model achieved on the test set. It was 91 %, again a satisfactory value, which comes as no surprise bearing in the scores received during the cross-validation.

IV. CONCLUSION

All in all, it can be said that our fine-tuning of the VGG16 neural network turned out to be adequate for the binary classification task we chose to tackle.

REFERENCES

[1] Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556, 2015.

[2] Žiga Emeršič, Aruna Kumar S. V., B. S. Harish, Weronika Gutfeter, Jalil Nourmohammadi Khiarak, Andrzej Pacut, Earnest Hansley, Mauricio Pamplona Segundo, Sudeep Sarkar, Hyeonjung Park, Gi Pyo Nam, Ig-Jae Kim, Sagar G. Sangodkar, Ümit Kaçar, Murvet Kirci, Li Yuan, Jishou Yuan, Haonan Zhao, Fei Lu, Junying Mao, Xiaoshuang Zhang, Dogucan Yaman, Fevziye Irem Eyiokur, Kadir Bulut Özler, Hazım Kemal Ekenel, Debbrota Paul Chowdhury, Sambit Bakshi, Pankaj K. Sa, Banshidhar Majhi, Peter Peer, Vitomir Štruc. The Unconstrained Ear Recognition Challenge 2019 - ArXiv Version With Appendix. arXiv:1903.04143, 2019.

[3] Computer Vision Laboratory, Faculty of Computer and Information Science, University of Ljubljana. Annotated Web Ears Dataset - AWE Dataset. http://awe.fri.uni-lj.si/, 2017.

*The URL of my Github repository containing the code written to obtain the above presented results is the following: https://github.com/dstefanov46/IBB_Assignment_3.