Transfer learning for ear recognition

Assignment #3

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

Jana Řežábková

***Abstract*—The ears biometric modality for person recognition is an active area of research, yet not as popular as more traditional modalities such as fingerprints. Its advantages over them include the possibility to obtain ear images from distance and without cooperation. However, the size of available datasets is limited which makes it challenging to apply deep learning approaches that have achieved state-of-the-art results on most image based tasks. In this report, we present closed-set results of deep learning ear classification on the 100 person AWE-W dataset. We combat the limited number of data per person with image augmentations and transfer learning. We reach 44.40 % rank-1 accuracy and 72.00 % rank-5 accuracy on holdout test dataset. Further, we show that augmentations are essential for the success of our model.**

I. Introduction

Deep learning, namely convolutional neural networks (CNNs), have proven to be the state-of-the-art technique in myriad of image based tasks. Transfer learning in particular has been integral for tasks with small available dataset. Models pretrained on ImageNet [1] jumpstart the training process by extracting general image features and can be finetuned to specific domain.

Ear recognition research is relatively newer among other biometric modalities. The first attemp on using CNNs for closed-set ear recognition has been made in [2]. Since then paper comprising the full biometric pipeline has been published [3] and the recognition problem extended as open-set problem [4].

In this paper we perform a closed-set recognition (similarly to [2]) on the AWE-W dataset [5] using transfer learning without any additional subject information.

II. Methodology

We present a simple CNN for the recognition task. We resort to transfer learning, because the dataset is small. We build our classification head on top of ImageNet pretrained EfficientNet-B0. We use the smallest of EfficientNet models [6] as our computing resources are limited. To combat the limited dataset size and limit overfitting we artificially enrich it using image augmentations. These include random horizontal flips, HSV augmentation, grayscale conversion and shifts. We then validate the importance of augmentations by training a model without them.

III. Experiments

The AWE-W dataset for ear recognition consists of 1000 ear images cropped out of headshot photos of people. In total it comprises of 100 different subjects each having exactly 10 ear images. The dataset is split into train set of 750 images and test set of 250 images. Each image is differently sized and comes with a brief 'json' annotation that we do not use.

First we rescale the input images to 256×128. We choose to have the height dimension larger to reflect the ear shape.

For model training we further split the training data to obtain a validation set (25 % of train data). However, since our dataset is already limited we perform the split again before every epoch thus effectively training on the full train set.

The EfficientNet backbone is frozen during training. We append a classification head of two blocks each with convolutional layer, max pooling layer and batch normalization layer. Since we are dealing with a closed-set problem, we finish with a softmax dense layer to predict distribution over 100 classes. The final architecture does not include additional non-linearities apart from the max pooling layers.

We use Adam optimizer with learning rate of 0.01 for first 25 epochs then decrease it tenfold. The batch size is set to 32 and we train over 100 epochs to loss convergence.

The model has total of 4.4M parameters, 4.0M of them belonging to the EfficientNet-B0.

To evaluate model performance we compute the Rank-1 and Rank-5 accuracies, Cumulative Match-score (CMC) curve and Area Under the CMC Curve (AUCMC).

IV. Results and Discussion

Table I summarizes performance metrics for both the augmented and non-augmented models. Clearly, image augmentations play an important role in preventing over-fitting. We have observed that the non-augmented model quickly reaches near perfect results, but as demonstrated in Table I does not generalize to test data. Image augmentations improve the Rank-1 accuracy by 11 %. Still, Rank-1 accuracy drops by half from training and more regularization would probably be desirable as demonstrated in [2].

Table I
Performance metrics on hold-out test set. Note comparison to training accuracy.

Augmentations	Rank-1	Rank-5	AUCMC	Rank-1 Train
Yes	44.40	72.00	93.65	89.91
No	33.20	58.80	89.17	99.72

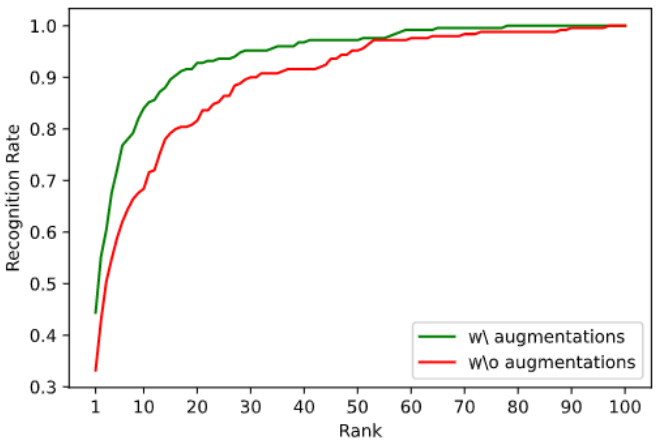

Figure 1. CMC curve

The CMC curve depicted in Figure 1 confirms that image augmentations improve results across all ranks. The AUCMC

score increases by over 4 % with augmentations. Our results are comparable to [2] with similar number of augmentations. However, they see an additional increase up to 62 % Rank-1 accuracy with 10 times more augmentations.

Figure 2 shows the distribution of misclassification rate in test data on model with augmentations using the Rank-1 and Rank-5 metrics. We can observe that 28 out 99 subjects (one subject is missing in test set) could not be classified at all under Rank-1. On the other hand 19 people were classified in all their test images. This suggests that some people are hard and some easy to classify for the model.

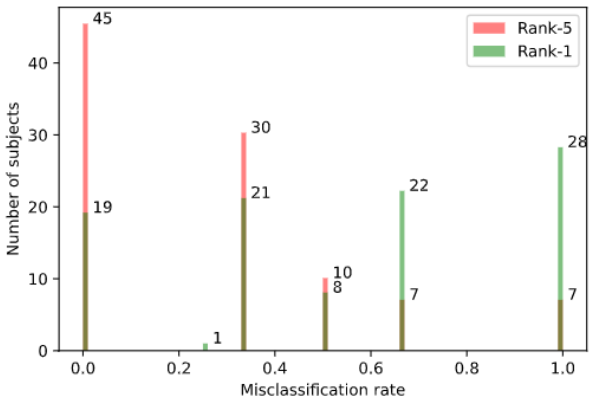

Figure 2. Histogram showing the distribution of subject misclassification rate by model with image augmentations.

V. Conclusion

We successfuly created a transfer learning based classification model for ear recognition. We showed that image augmentations are integral to success on limited size dataset. Our results are comparable with results achieved in similar settings albeit on an extended version of the AWE-W dataset [2].

We have not investigated the effect of specific augmentations on the generalization performance, which could serve as an extension of our work.

We have chosen the easier closed-set setting in this paper. A natural step forward is to remove the softmax layer and view the convolutional output as feature descriptor of given ear image. Descriptors are then classified based on distance metrics. This so called open-set protocol does not require retraining of models when new subjects are added to database and is under active research of the ear biometrics community [7].

All code with additional result plots is available in GitHub repository `https://github.com/janarez/AWE-W-segmentation/`.

References

[1] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. Li, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*. IEEE Computer Society, 2009, pp. 248–255. [Online]. Available: https://doi.org/10.1109/CVPR.2009.5206848

[2] Ž. Emeršič, D. Štepec, V. Štruc, and P. Peer, "Training convolutional neural networks with limited training data for ear recognition in the wild," *CoRR*, vol. abs/1711.09952, 2017. [Online]. Available: http://arxiv.org/abs/1711.09952

[3] Ž. Emeršič, J. Križaj, V. Štruc, and P. Peer, "Deep ear recognition pipeline," in *Recent Advances in Computer Vision - Theories and Applications*, ser. Studies in Computational Intelligence, M. Hassaballah and K. M. Hosny, Eds. Springer, 2019, vol. 804, pp. 333–362. [Online]. Available: https://doi.org/10.1007/978-3-030-03000-1_14

[4] Ž. Emeršič, D. Štepec, V. Štruc, P. Peer, A. George, A. Ahmad, E. Omar, T. E. Boult, R. Safdaii, Y. Zhou *et al.*, "The unconstrained ear recognition challenge," in *2017 IEEE International Joint Conference on Biometrics, IJCB 2017, Denver, CO, USA, October 1-4, 2017*. IEEE, 2017, pp. 715–724. [Online]. Available: https://doi.org/10.1109/BTAS.2017.8272761

[5] Ž. Emeršič, V. Štruc, and P. Peer, "Ear recognition: More than a survey," *Neurocomputing*, vol. 255, pp. 26–39, 2017. [Online]. Available: https://doi.org/10.1016/j.neucom.2016.08.139

[6] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *CoRR*, vol. abs/1905.11946, 2019. [Online]. Available: http://arxiv.org/abs/1905.11946

[7] Ž. Emeršič, A. K. SV, B. Harish, W. Gutfeter, J. Khiarak, A. Pacut, E. Hansley, M. P. Segundo, S. Sarkar, H. Park *et al.*, "The unconstrained ear recognition challenge 2019," in *2019 International Conference on Biometrics, ICB 2019, Crete, Greece, June 4-7, 2019*. IEEE, 2019, pp. 1–15. [Online]. Available: https://doi.org/10.1109/ICB45273.2019.8987337