EXPLORING THE EFFICACY OF EAR IMAGES FOR BIOMETRICS IDENTIFICATION

Sakshamdeep Singh & Ashutosh Rastogi

Introduction

Our project examines the use of ear pictures as reliable biometric identifiers considering the security risks associated with current procedures. We want to extract critical identification data from ear pictures using cutting-edge deep learning architectures. We want to know if ear pictures provide a trustworthy, non-invasive substitute for traditional biometric identification methods.

Methods

Our technique employs a streamlined two-step process that utilizes video feed as the input for reliable ear-based identification.

1) Detection: The first step involves processing the video feed using the YOLOv8s model. We leverage the model to locate and extract the ROIs containing the ear from the input image.
2) Recognition: In the second phase, the extracted ear region is fed into the YOLOv8m-cls model trained on the EarVN1.0 dataset. This model has been optimized to accurately recognize and identify ears.

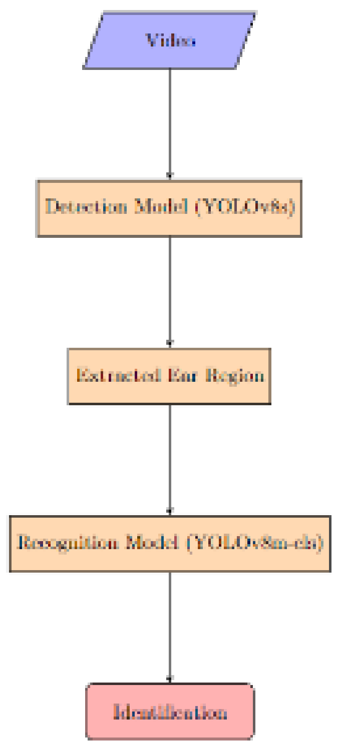

Figure 1: Experimental Pipeline

Dataset

A custom dataset of 69 images of 5 subjects was created for the detection task. The dataset was annotated using the LabelImg tool. For recognition, the EarVN1.0 dataset of 28,412 images of 164 Asian individuals was used, along extracted ROIs from the detection dataset. Data Augmentation techniques were applied including horizontal flip, random rotation, and Gaussian blur to enhance dataset diversity and model generalization.

Figure 2: Detection Task Training Samples

Figure 3: Recognition Task Training Samples

Evaluation Metrics

Evaluation metrics such as precision, recall, mAP, top-1 accuracy, and top-5 accuracy assess the model's performance in terms of prediction accuracy, detection capability, and classification accuracy. These metrics provide valuable insights into the model's effectiveness and reliability in various tasks.

Results

Initially, we used a small dataset of 29 images with 2 subjects for detection training, but the performance was unsatisfactory. To improve detection recall, we increased the dataset size, resulting in notable improvements. Our best recognition model achieved 83.3% accuracy using the YOLOv8m-cls model for testing, which had a smaller size but still performed well. However, during workflow execution, we observed excellent detection but declining performance in recognition tasks. This indicates the challenge of transferring performance from static images to dynamic video scenarios..

Architecture	Total Parameters (M)	Input Size	Epochs	Train Acc	Val Acc (Top-1) (%)	Top-5(%)	Training Time (hh:mm:ss)
ResNeXt101 (SotA)	87.1	(224, 224)	441	-	**93.45**	-	-
ResNeXt50	25.3	(32, 64)	50	99.22	53.44	-	0:39:44
ResNeXt50	25.3	(64, 128)	25	96.36	71.47	-	0:33:43
ResNeXt50	25.3	(128, 256)	25	97.00	73.12	-	1:20:55
ResNeXt101	88.1	(64, 128)	25	97.29	65.54	-	1:20:42
ResNeXt50	25.3	(128, 256)	10	91.88	80.64	-	1:31:39
YOLOv8m-cls	17.0	(224, 224)	10	-	83.00	94.10	0:23:24
YOLOv8x-cls	57.4	(224, 224)	10	-	**83.30**	94.40	0:52:48

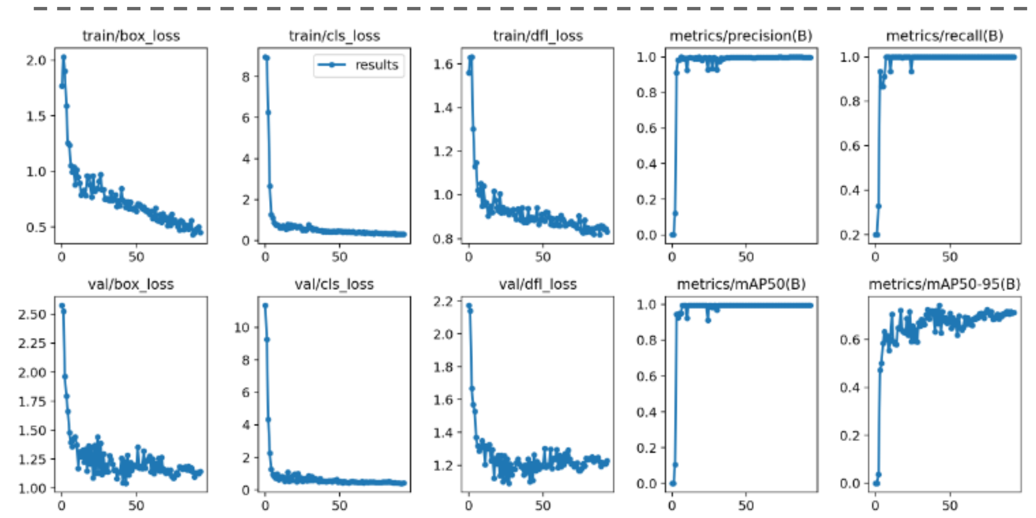

Figure 4: Detection Training Graphs

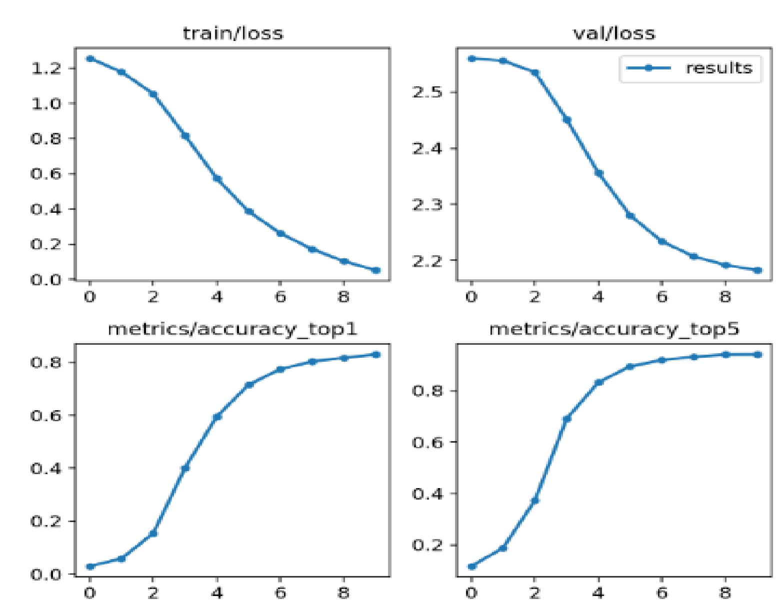

Figure 6: Recognition Training Graphs

Comparison with SotA

This study introduces the YOLOv8s model for ear detection, outperforming custom CNN models. The recognition task achieved a high accuracy of 93.4% with the ResNeXt101 model on the EarVN1.0 dataset. However, on our modified dataset with 169 subjects, ResNeXt-based models reached a maximum accuracy of 80.64%. To improve performance, we employed the YOLOv8x-cls model, resulting in an accuracy of 83.30% after applying data augmentation techniques.

Future Work

Our models perform well in detection tasks but need improvement in recognition tasks. Future research aims to enhance recognition performance by extending the model's capabilities from static images to real-time video datasets. Open set identification techniques will be explored to accurately identify individuals not seen during training. Additionally, investigating multi-modal systems that combine different biometric modalities like side profile, face, iris, and ear is of interest.

References

1. Xie, S. et al. (2017) Aggregated residual transformations for deep neural networks, arXiv.org. Available at: https://arxiv.org/abs/1611.05431
2. H. Alshazly, C. Linse, E. Barth and T. Martinetz, "Deep Convolutional Neural Networks for Unconstrained Ear Recognition," in IEEE Access, vol. 8, pp. 170295-170310, 2020, doi: 10.1109/ACCESS.2020.3024116.
3. Raveane, W.; Galdámez, P.L.; González Arrieta, M.A. Ear Detection and Localization with Convolutional Neural Networks in Natural Images and Videos. Processes 2019, 7, 457. https://doi.org/10.3390/pr7070457
4. hila Priyadharshini, R., Arivazhagan, S. & Arun, M. A deep learning approach for person identification using ear biometrics. Appl Intell 51, 2161–2172 (2021). https://doi.org/10.1007/s10489-020-01995-8

University at Buffalo The State University of New York

Department of Computer Science and Engineering
School of Engineering and Applied Sciences