EXPLORING THE EFFICACY OF EAR IMAGES FOR BIOMETRICS IDENTIFICATION

Sakshamdeep Singh
50475857
ssingh86@buffalo.edu

1 Abstract

Trustworthy biometric identification systems play a crucial role in security and identification procedures. While various biometric identifiers have been proposed and deployed, the potential of ear pictures remains underutilized in this field. This study aims to explore the effectiveness of ear pictures in biometric identification by employing state-of-the-art machine learning and deep learning techniques. Specifically, we focus on utilizing Convolutional Neural Networks (CNNs) to develop a reliable ear image-based biometric identification system. Additionally, we contribute to the existing knowledge by conducting a thorough analysis of current ear detection methods.

Our approach begins with capturing video and using it as the input. The video feed is then processed using the detection model, which extracts the region containing the ear. Subsequently, the extracted ear region is utilized as input for a trained CNN model recognition model to perform identification tasks.

Through this research, we aim to demonstrate the potential of ear pictures as a valuable biometric identifier. The proposed workflow, integrating advanced machine learning models and detection methods, offers a promising direction for building robust ear image-based identification systems.

2 Introduction

Automatic person identification using biometric traits has gained significant attention in the research community. The human ear, with its specific and unique features, offers potential for reliable identification, similar to other biometrics like face, iris, and fingerprints. This paper focuses on exploring the feasibility and effectiveness of using ear images for person identification, considering the factors of temporal consistency, ease of acquisition, and individual uniqueness.

In the context of the COVID-19 pandemic, face identification systems often fail due to mask-wearing scenarios. The ear provides a reliable alternative for passive person identification, as it does not rely on user cooperation. Additionally, ear images can be obtained in a contact-less and non-intrusive manner, making them suitable for various environments.

By leveraging the temporal consistency, ease of acquisition, and uniqueness of ear features, we can develop robust person identification systems. This research aims to contribute to the advancement of automatic identification by exploring the potential of ear images as a valuable biometric modality.

Through methodologies, experimental results, and analysis, we provide insights to guide the development and implementation of efficient and accurate ear-based identification systems, furthering the field of biometrics.

3 Related Work

The investigation of ear pictures as a biometric identifier has garnered significant attention, with several noteworthy studies conducted in this field.

One notable study, "Ear Detection and Localization with Convolutional Neural Networks in Natural Images and Videos" by Raveane, W., Galdámez, P.L., and González Arrieta, M.A. (2019), proposed a method that

utilizes convolutional neural networks (CNNs) to detect and localize ears in real-world photos and videos. This research showcases the application of machine learning techniques in effectively managing ear imaging data, particularly in the preprocessing stages of ear-based biometric identification systems.

In terms of recognition, a study by Ramar Priyadharshini, Selvaraj A., and Madakannu A. titled "A deep learning approach for person identification using ear biometrics" developed a custom 6-layer CNN model for person identification using ear biometrics. They achieved a recognition rate of 97.36% on the constrained ear dataset called AMI.

Another study by Hammam Alshazly, Christoph Linse, Erhardt Barth, and Thomas Martinetz (2020) focused on evaluating deep learning models for ear recognition tasks using the unconstrained EarVN1.0 dataset. Their best-performing model, RexNeXt101, achieved an accuracy of 93.45%.

These research efforts collectively emphasize the potential of deep learning methods for both ear detection and identification tasks. Building upon these foundational findings, our study aims to further explore the effectiveness of CNNs in utilizing ear pictures as a reliable biometric identifier.

4 Proposed Algorithm

4.1 Algorithm

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, a real-time object identification system known for its speed and accuracy. We leverage YOLOv8s to locate and extract the region of interest (ROI) containing the ear from the input image.
2. **Recognition**: In the second phase of the algorithm, 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.

By leveraging these two steps, our approach enables efficient and effective ear detection and recognition, facilitating the development of robust ear-based identification systems.

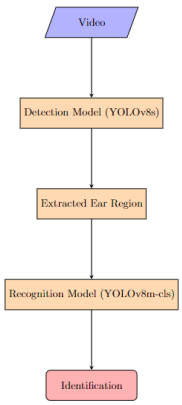

Figure 1: Experimental Pipeline

4.2 Dataset

4.2.1 Detection Dataset

To facilitate the detection task, a custom dataset consisting of 69 images was created, involving 5 subjects. These images were captured using the camera of an iPhone 12 from different distances and orientations. The dataset was annotated using the labelImg tool, and the annotations were stored in the YOLO format.

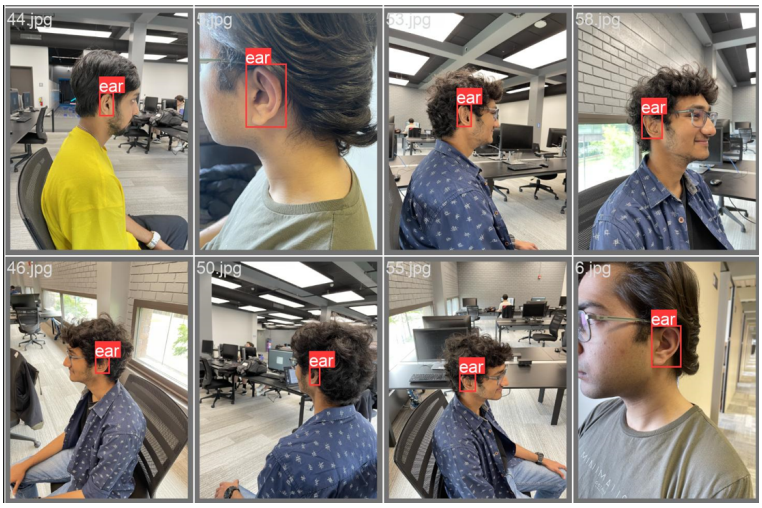

Figure 2: Detection Task Training Samples

4.2.2 Recognition Dataset

In the recognition task, we utilized the EarVN1.0 dataset, which consists of 28,412 color pictures of 164 Asian individuals (98 males and 66 females). These images were collected in an unrestricted environment, presenting challenges for our model to perform effectively in diverse real-world scenarios. In addition, we incorporated the extracted Region of Interest (ROIs) from the detection dataset for the closed-set recognition task.

To enhance the diversity of the training dataset, data augmentation techniques were employed. These techniques included randomly applying horizontal flips with a probability of 0.5, introducing random rotations within the range of -15 to 15 degrees, and applying Gaussian blur with a probability of 0.5. By incorporating these augmentations, the training dataset became more varied and robust, enabling the model to generalize better to different real-world scenarios.

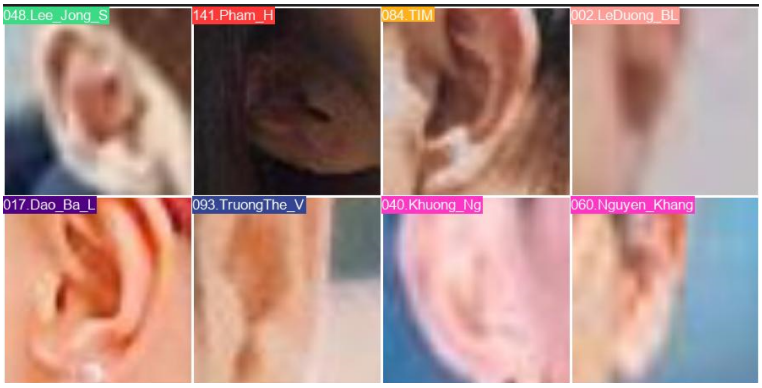

Figure 3: Recognition Task Training Samples

5 Software and Hardware Requirements

The project implementation relied on the following hardware and software specifications:

1. **Processor**: 12th Gen Intel(R) Core(TM) i7-12700H, with a base clock speed of 2.30 GHz.
2. **Memory**: 16.0 GB of installed RAM, with 15.7 GB usable.
3. **Operating System**: 64-bit version with x64-based processor.
4. **Graphics**: NVIDIA GeForce RTX 3060 GPU, with 6GB dedicated video memory.

6 Results

6.1 Evaluation Metrics

- **Precision**: Precision measures the accuracy of the model's positive predictions. It is the ratio of the number of true positive detections to the total number of detections made by the model. In other words, precision is the fraction of positive detections that are actually positive.
- **Recall**: Recall measures the model's ability to detect all positive instances. It is the ratio of the number of true positive detections to the total number of actual positive instances in the test set. In other words, recall is the fraction of actual positive instances that are correctly detected by the model.
- **mAP**: mAP is a summary metric that measures the average precision across different recall levels. It is calculated by computing the precision at different recall levels, and then taking the mean over those precision values. mAP is typically reported at different intersection over union (IoU) thresholds, such as IoU=0.5 and IoU=0.5:0.95.
- **Top-1 Accuracy**: Top-1 accuracy is a measure of the model's accuracy in correctly predicting the single most probable class label. It represents the percentage of test samples for which the model's top prediction matches the ground truth label.
- **Top-5 Accuracy**: Top-5 accuracy evaluates the model's performance by considering whether the ground truth label is within the top five predicted class labels. It indicates the percentage of test samples where the correct label is present within the model's top five predictions.

These evaluation metrics provide insights into the performance of the model in terms of precision, recall, average precision, as well as the accuracy of top-1 and top-5 predictions. They help assess the effectiveness and reliability of the model in various detection and classification tasks.

6.2 Detection Results

The YOLOv8s model provided by Ultralytics APIs was utilized for training purposes. The dataset used for training and validation was divided in an 80:20 ratio. The model consisted of 225 layers and approximately 11 million trainable parameters. To optimize the training process, the SGD optimizer with a learning rate of 0.01 was employed.

6.2.1 Training Results

Class	Images	Instances	Box(Precision	Recall	mAP50	mAP50-95)
all	52	52	0.999	1	0.995	0.856

6.2.2 Validation Results

Class	Images	Instances	Box(Precision	Recall	mAP50	mAP50-95)
all	15	15	0.996	1	0.995	0.742

6.3 Recognition Results

The preliminary outcomes of our suggested ear biometric identification technology are encouraging. With 132,863,336 trainable parameters, we used the VGG-11 model to obtain training accuracy of 63.9% and validation accuracy of 63.3%. These outcomes show an acceptable degree of performance, indicating that ear pictures can be utilized for biometric identification after all.

However, given that recent research indicates that the ResNeXt50 model might provide greater accuracy, we intend to evaluate it in order to investigate additional enhancements to the functionality of our system. In particular, it has been demonstrated that ear image categorization using data augmentation and the ResNeXt50 model can reach up to 93% accuracy for a 70:30 train-test split and 90% accuracy for a 60:40 split.

We later conducted training using different architectures, including ResNeXt50, ResNext101, YOLOv8m-cls, and YOLOv8x-cls. For our experiments, we utilized an 80:20 data split. The state-of-the-art model achieved

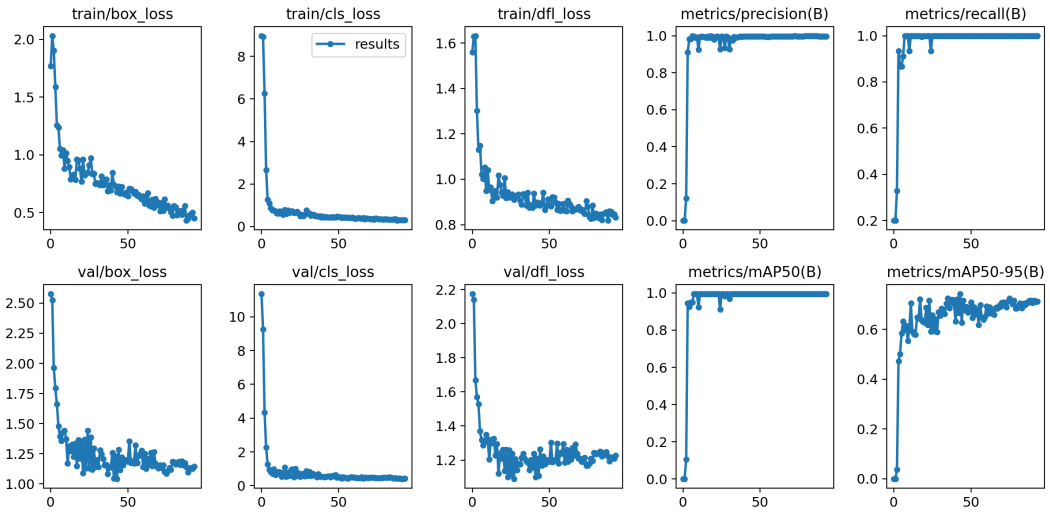

Figure 4: Detection Training Graphs

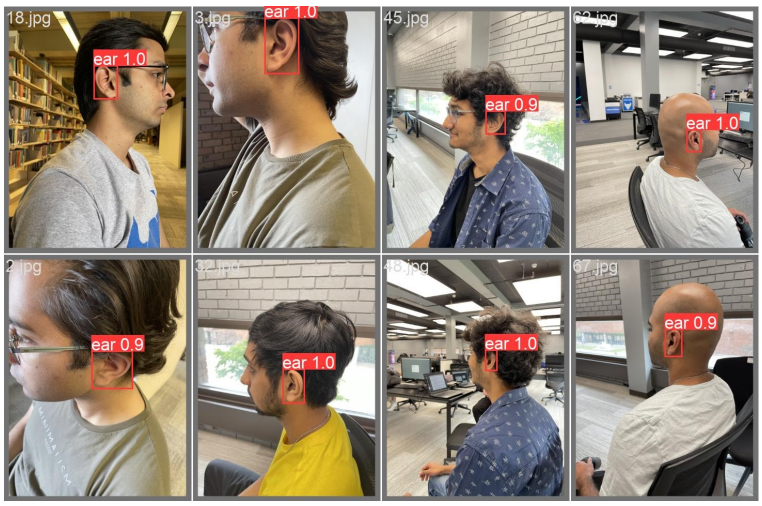

Figure 5: Detected ear regions on the validation data

an impressive rank-1 accuracy of 93.45%, while our modified dataset allowed us to achieve 83% accuracy using YOLOv8m-cls model.

The YOLOv8m-cls model provided by Ultralytics APIs was utilized for training purposes. The model consisted of 141 layers and approximately 15 million trainable parameters. To optimize the training process, the SGD optimizer with a learning rate of 0.01 and weight decay of 0.0005 was employed.

6.3.1 Training Results

6.3.2 Validation Results

Class	Top1-Accuracy (%)	Top5-Accuracy (%)
all	83.00	94.1%

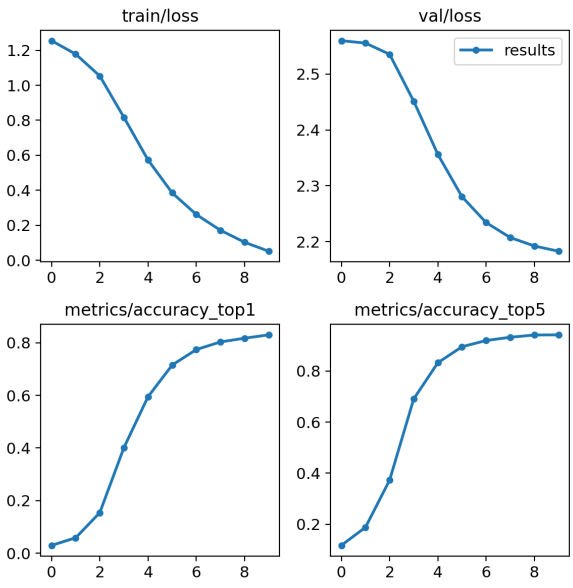

Figure 6: Recognition Training Graphs

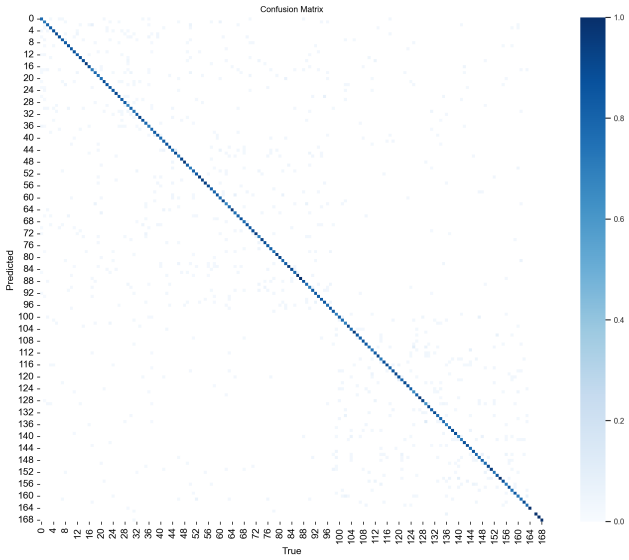

Figure 7: Recognition Confusion Matrix

6.4 Comparison with SotA

This study marks the inaugural utilization of the YOLOv8s model for ear detection, surpassing the performance of custom CNN models by a significant margin. In the recognition task, the ResNeXt101 model achieved the highest performance on the EarVN1.0 dataset, attaining an accuracy of 93.4%. However, on our modified dataset with 169 subjects (instead of 165), our ResNeXt-based models achieved a maximum accuracy of 80.64%. To enhance performance further, we employed the YOLOv8x-cls model, which yielded an improved accuracy of 83.30%, as demonstrated in the comparative graphs presented below. The state-of-the-art (SotA) model incorporated data augmentation techniques, and we also applied data augmentation in our last three models.

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

6.5 Interpretation

During the training phase for detection, we initially employed a dataset comprising 2 subjects with 29 images, but the performance obtained was below expectations. To improve the detection recall performance and enhance the confidence of ear region detection, we increased the dataset size. This led to notable improvements in performance.

Regarding the recognition task, our best model achieved an accuracy of 83.3%. For testing purposes, we utilized the YOLOv8m-cls model instead of the YOLOv8x-cls model due to its significantly smaller size, while still maintaining a satisfactory level of identification performance.

However, when executing our workflow, we observed excellent detection results, but the performance in recognition tasks showed a decline. This indicates that the good performance achieved on the image dataset could not be readily generalized to video feed inputs, highlighting the challenges of transferring performance from static images to dynamic video scenarios.

7 Future Work

Our models and workflow demonstrate satisfactory performance in detection tasks but fall short in achieving optimal results for recognition tasks.

- **Recognition Performance:** Concerning the performance of recognition, our upcoming research endeavors seek to enhance our model's capabilities, aiming to either match or surpass the performance of the State of the Art feature extractor.
- **Complete the Architecture:** Implement the remaining component of our biometric system, encompassing the template generator, enrollment process, and matching functionality.
- **Open Set Identification:** Additionally, we plan to explore the utilization of open set identification techniques, allowing our system to accurately identify individuals who were not seen during the training phase. This approach will bring a new level of adaptability and flexibility to our biometric identification system.
- **Multi-modal systems:** Furthermore, we are interested in investigating the potential of multi-modal systems, where different biometric modalities such as side profile, face, iris, and ear are combined.

8 Conclusion

The study leveraged Convolutional Neural Networks (CNNs) and the YOLOv8 object detection model, alongside other advanced deep learning techniques, to explore the viability of using ear pictures as a reliable biometric identifier.

While still in its early stages, our approach has made notable advancements in the field of biometric identification. This includes incorporating real-time video feed input and utilizing a custom dataset specifically designed for detection training and evaluation. By employing established evaluation metrics such as precision, recall, top-1, top-5, we effectively compared the performance of our system with other biometric identification systems.

Acknowledging the scope for improvement, our ongoing research aims to enhance recognition precision, implement multi-ear detection, diversify the dataset, and improve detection performance. Additionally, we are considering the exploration of open set identification techniques.

Despite the challenges and obstacles associated with developing a robust ear-based biometric identification system, our preliminary findings and future research plans highlight the potential effectiveness and reliability of ear pictures as a trustworthy biometric identifier.

9 Bibliography

[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] Ahila 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

[5] Kamboj, A., Rani, R. & Nigam, A. A comprehensive survey and deep learning-based approach for human recognition using ear biometric. Vis Comput 38, 2383–2416 (2022). https://doi.org/10.1007/s00371-021-02119-0

[6] Truong Hoang, Vinh (2019), “EarVN1.0”, Mendeley Data, V3, doi: 10.17632/yws3v3mwx3.3