EXPLORING THE EFFICACY OF EAR IMAGES FOR BIOMETRICS IDENTIFICATION

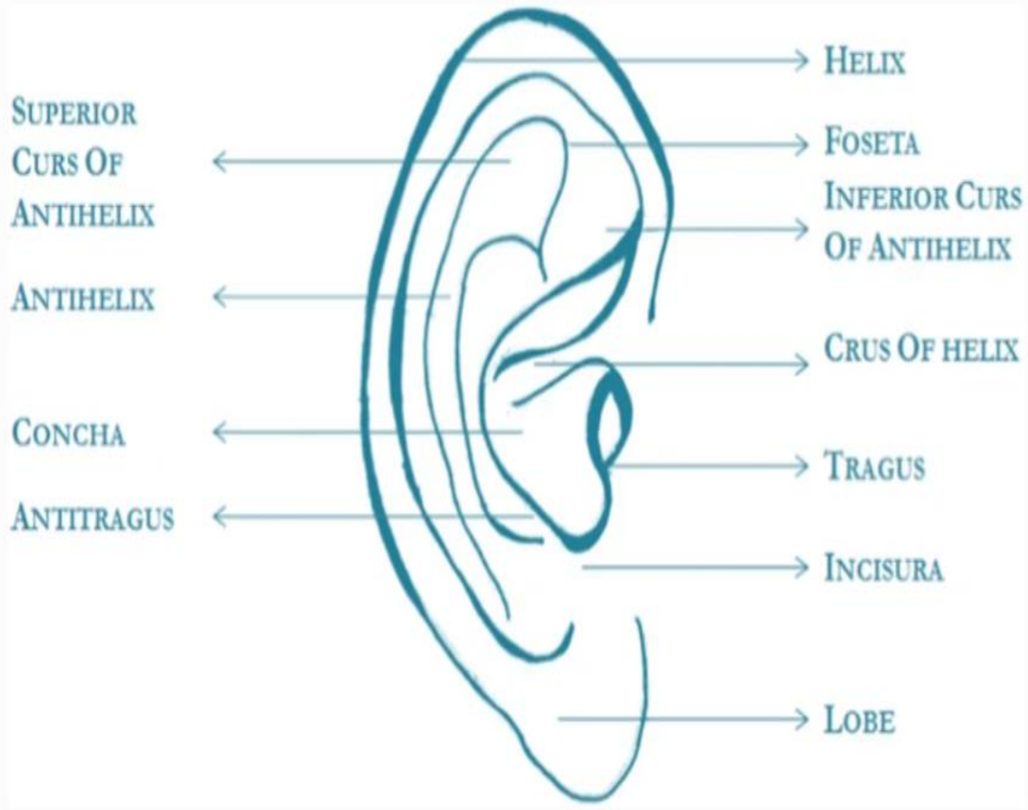
Sakshamdeep Singh – 50475857

Ashutosh Rastogi - 50486793

University at Buffalo The State University of New York

Research Problem

- Automatic person identification from ear images is an active field of research within the biometric community.
- Like other biometrics such as face, iris and fingerprints, ear also has a large amount of **specific and unique features** that allow for person identification.
- Three important factors for person identification are **temporal consistency**, ease of **acquisition**, and **uniqueness** to everyone.

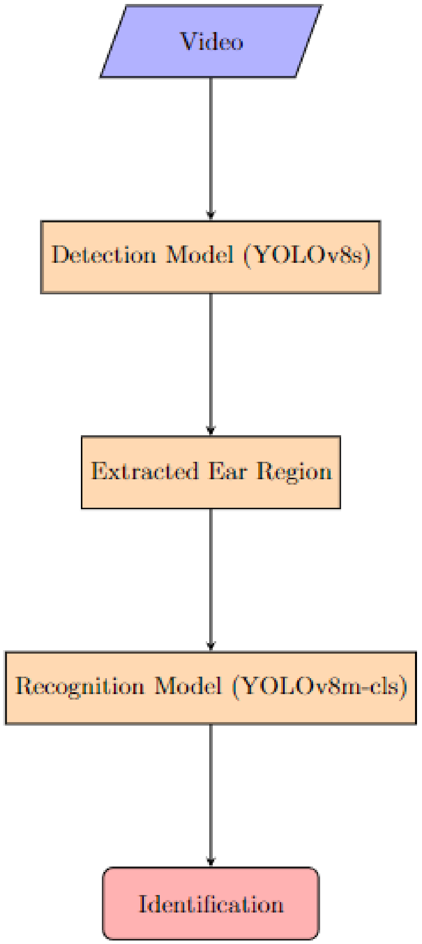


Approach

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.



Novelty

- Evaluation of the performance of ear biometric system using real time video feed input.
- Using state of the art YOLOv8 (You Only Look Once) model for ear detection. YOLO popular object detection algorithm that can detect objects in real-time is the YOLO model. This method is quicker and more effective than conventional object detection techniques.
- Small custom dataset was created of annotated ear images for detection training and evaluation.
- Using state of the art YOLOv8m-cls model for subject identification. Among the various deep learning models tried, best performance was achieved using YOLOv8m-cls model.

Related Works

- 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
- 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

Table 1. Summary of the available ear databases in literature.

Datasets	Country	Number of peoples	Number of images	Image size
IIT Delhi-I [1]	India	121	471	272×204
USTB Ear [3]	China	77	308	varied
AWE [2]	Slovenia	100	1000	varied
AWE extend [8]	Slovenia	346	4104	varied
AMI [6]	Spain	106	700	492×702
WPUT [5]	Poland	501	2071	varied
UERC [7]	Slovenia	3706	11,804	varied
EarVN1.0	Vietnam	164	28,412	varied and low resolution

Detection Dataset

Custom dataset for detection

- Initial Dataset
  - 2 subjects
  - 31 image
- Improved Dataset
  - 5 subjects
  - 69 images

Figure 2: Detection Task Training Samples

Recognition Dataset

EarVN1.0: A new large-scale ear images dataset in the wild

- 164 Asian Individuals (98 males, 66 females)
- 28,412 Color Images
- Unconstrained conditions

Figure 3: Recognition Task Training Samples

Evaluation Protocol

Some common metrics we will be using in our biometric identification systems include:

- **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.
- **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.
- **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.
- **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.

These metrics used to evaluate the performance of an ear biometric identification system and to compare its performance with other biometric identification systems.

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.

Training Results

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

Validation Results

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

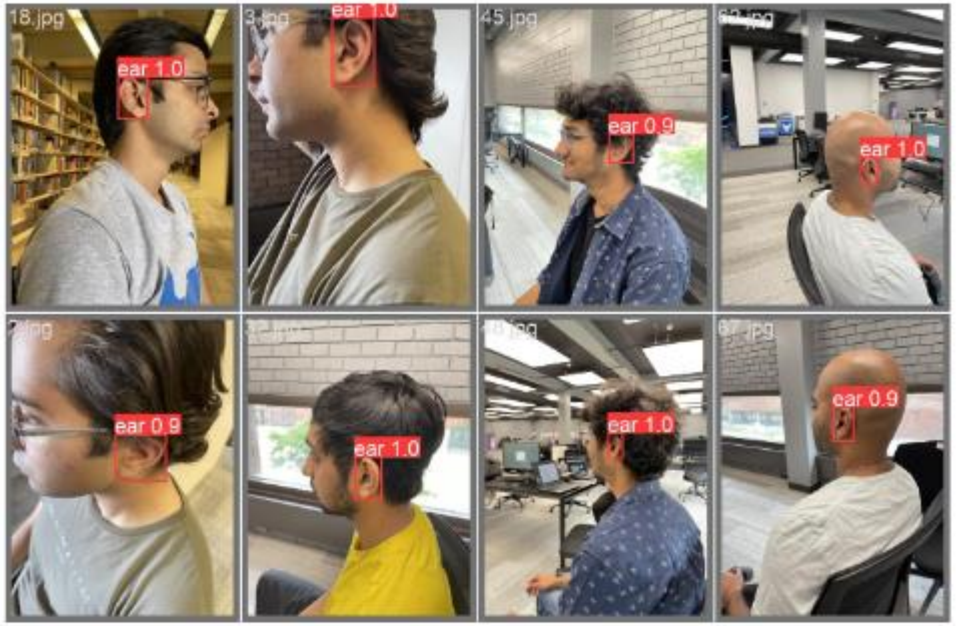


Figure 5: Detected ear regions on the validation data

Recognition Results

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

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

Future Work and Improvements

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

- **Recognition Performance:** Regarding recognition performance, our future research aims to extend the capabilities of our model, which is currently trained on static images, to effectively handle real-time video datasets. This generalization will enhance the model's ability to accurately recognize individuals in dynamic scenarios.
- **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.

References

- 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
- 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
- Mursalin, M.; Ahmed, M.; Haskell-Dowland, P. Biometric Security: A Novel Ear Recognition Approach Using a 3D Morphable Ear Model. *Sensors* **2022**, *22*, 8988. https://doi.org/10.3390/s22228988
- Truong Hoang, Vinh (2020), “EarVN1.0”, Mendeley Data, V4, doi: 10.17632/yws3v3mwx3.4
- Xie, S. et al. (2017) Aggregated residual transformations for deep neural networks, arXiv.org. Available at: https://arxiv.org/abs/1611.05431
- 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
- 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

Thank You