A Novel Embedding Architecture and Score Level Fusion Scheme for Occluded Image Acquisition in Ear Biometrics System

Presented by:
Archishman Biswas
Authors:
Archishman Biswas, Goutham A P, Saket Pateriya, Divyang S. Jadav, Satish Mulleti and Vikram M. Gadre

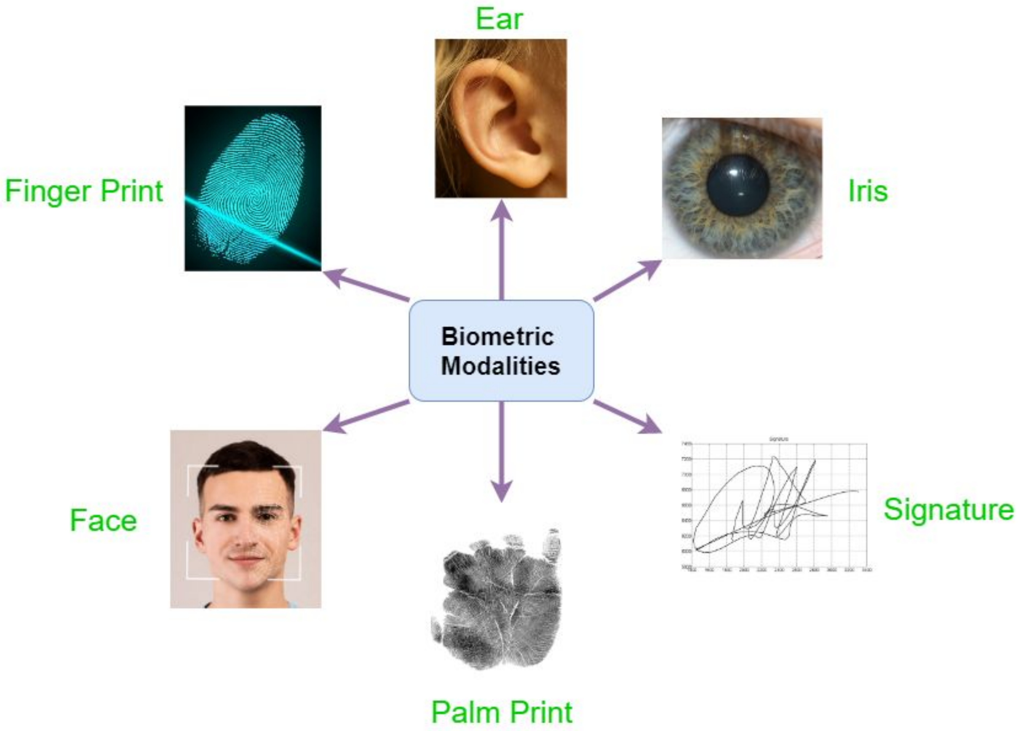
Department of Electrical Engineering, Indian Institute of Technology Bombay, India 400076

1

Introduction to Ear Biometrics

Biometrics
(Bio=life; Metrics=measurement)

- Biometric is the science of determining an individual's **unique identity** based on one or more **distinguishing physical or behavioural characteristics.**



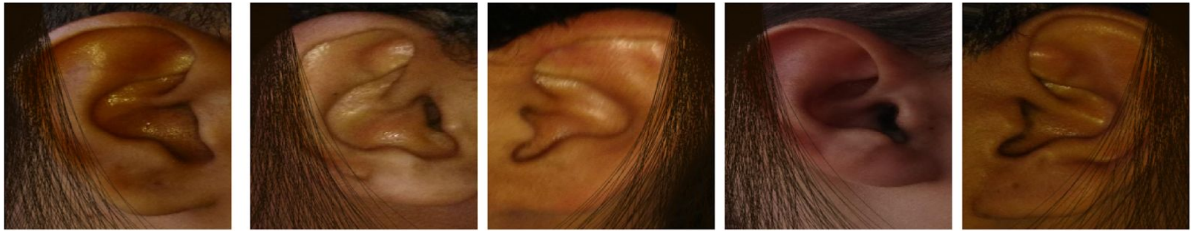
Ear Biometrics

- ❑ Past studies have shown that ears are **unique** for each individual.
- ❑ Ear image acquisition process is **contactless** and **non-intrusive**.
- ❑ Ear is **consistent** and doesn't change with expressions.

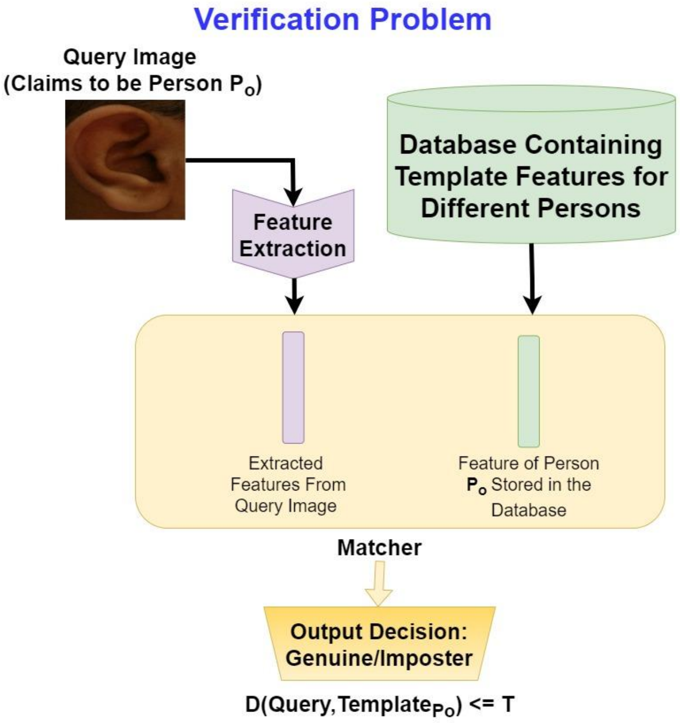
We use our own IITB-dataset for studying ear based biometrics

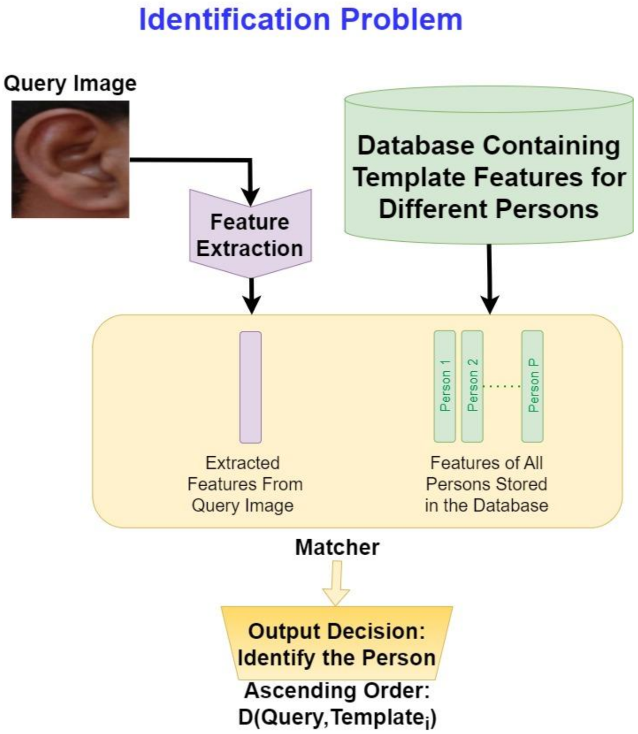
- ❑ 1000 ear images of 100 subjects, 5 images of each side ear, all of size (224,224,3)

- ❑ We have also studied the effect of occlusions due to hair, for this, we have added artificial random occlusions (covering 35% of the image on average)



Verification and Identification Problems in Biometrics

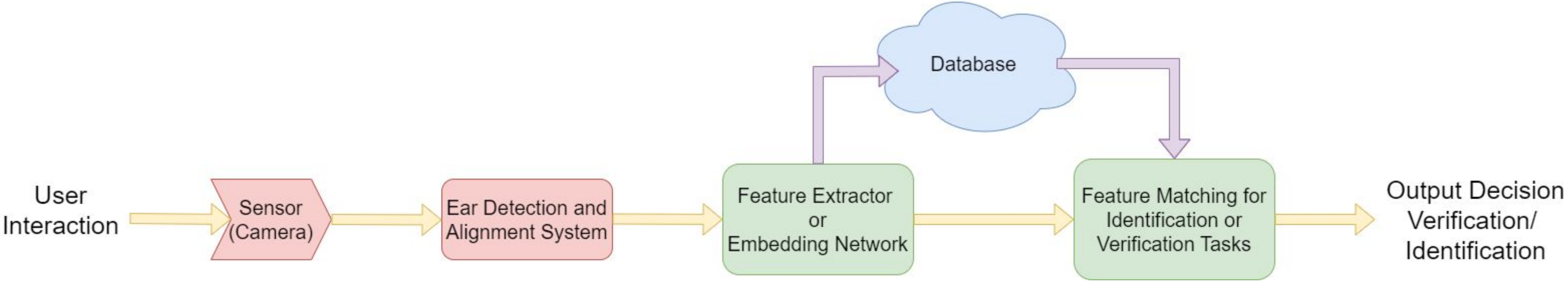




2

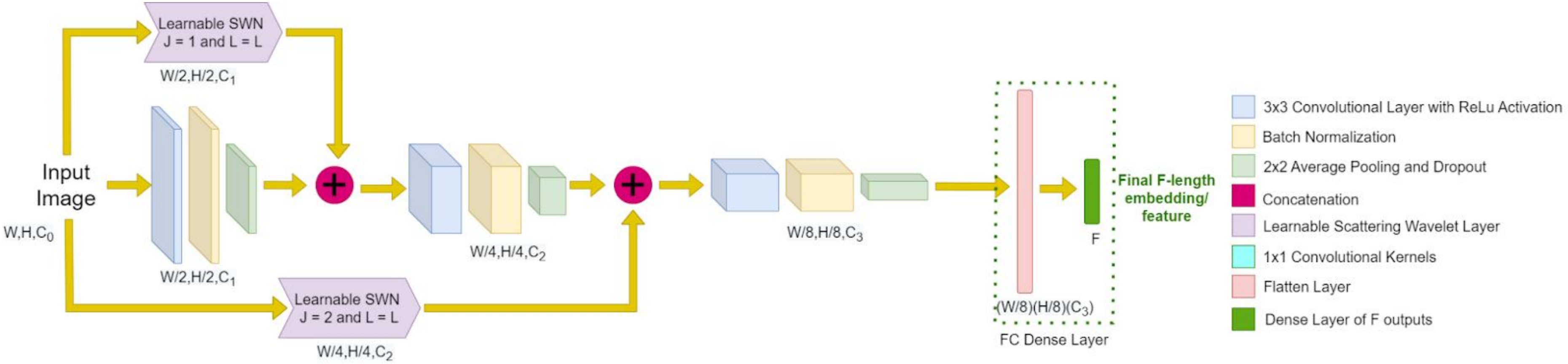
Embedding Network and Siamese Training

Pipeline of Ear Biometrics System



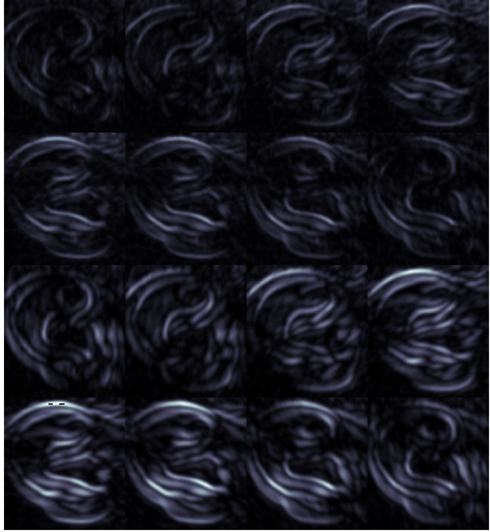
- The initial **detection and alignment system** is an important preprocessing step that gets rid of many of the irrelevant background information
- Our proposed innovations are only in the **Embedding Network** and **Feature Matching** systems/blocks

Embedding Network Used



- ❑ The architecture of the embedding network consists of conventional CNNs in the central path and **learnable Scattering Wavelet Network (SWN)** in the parallel paths
- ❑ The parallel learnable SWN paths pickup information directly from the input image and inserts them into further downstream parts of the embedding network

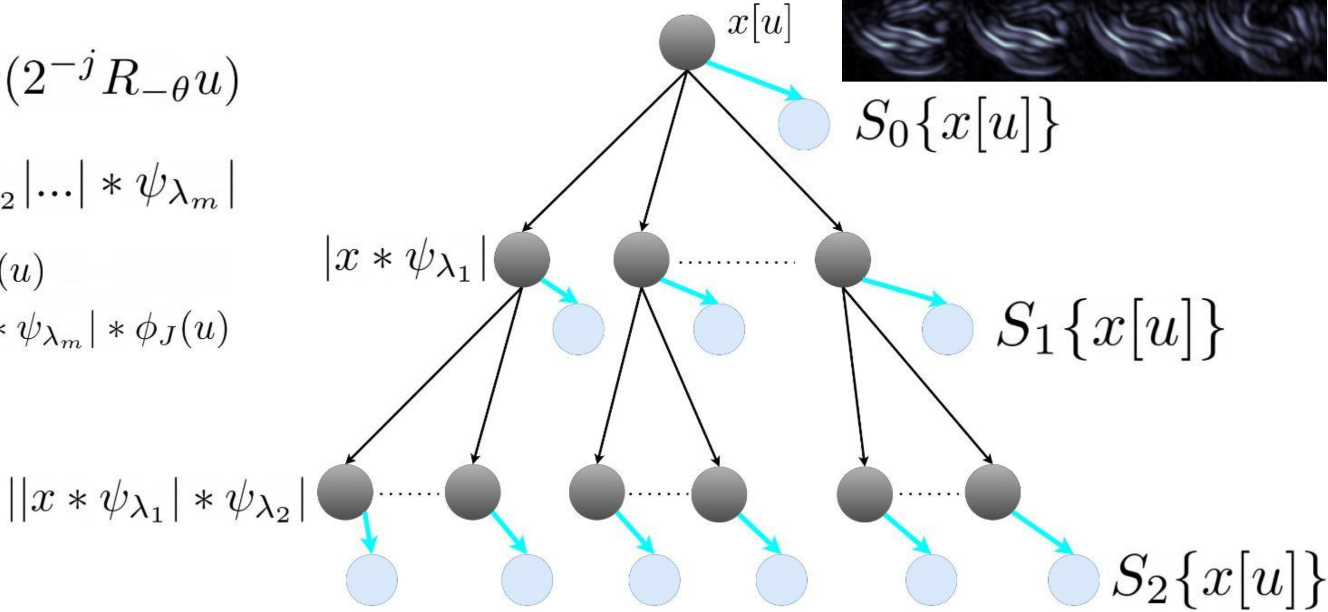
Scattering Wavelet Network



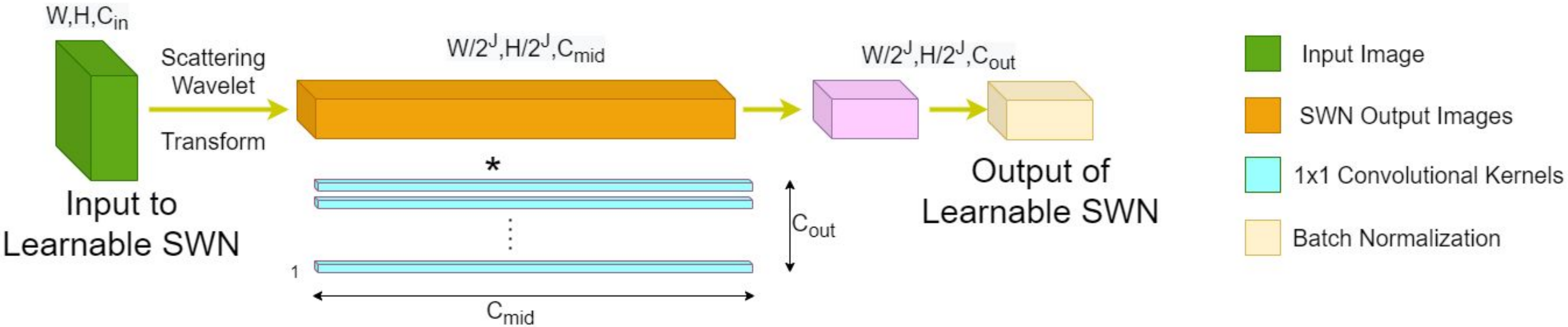
$$\psi_\lambda(u) \triangleq \psi_{j,\theta}(u) = 2^{-j}\psi(2^{-j}R_{-\theta}u)$$

$$U[p]x = ||...|x * \psi_{\lambda_1}| * \psi_{\lambda_2}|...| * \psi_{\lambda_m}|$$

$$\begin{aligned} S[p]x &= U[p]x * \phi_J(u) \\ &= ||...|x * \psi_{\lambda_1}| * \psi_{\lambda_2}|...| * \psi_{\lambda_m}| * \phi_J(u) \end{aligned}$$



Learnable Scattering Wavelet Network

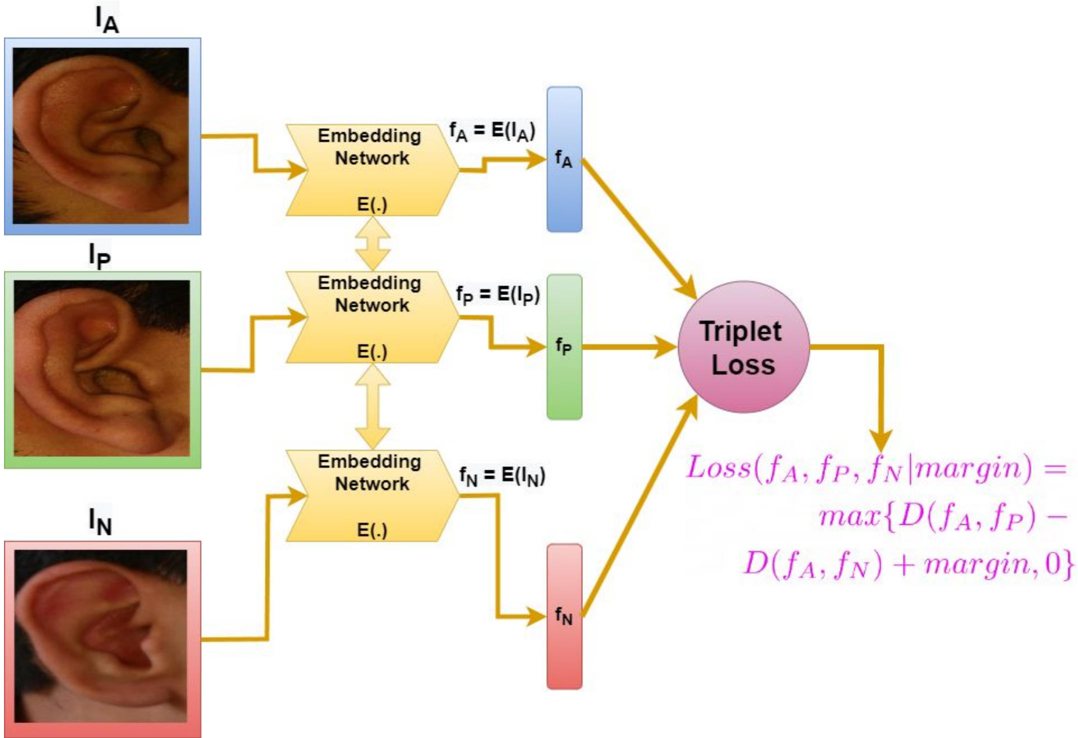


- ❑ Number of output channels (denoted by $\mathbf{C_{mid}}$) of SWN is given by:

$$C_{mid} = C_{in} \times \left(JL + \frac{J(J-1)}{2}L^2 + \ldots + \binom{J}{m}L^m\right)$$ {0th layer (low-pass) is omitted}

- ❑ Then, convolution with 1x1 kernel is carried out to learn the linear combination across the $\mathbf{C_{mid}}$ channels and finally get $\mathbf{C_{out}}$ channels.

Siamese Training Framework

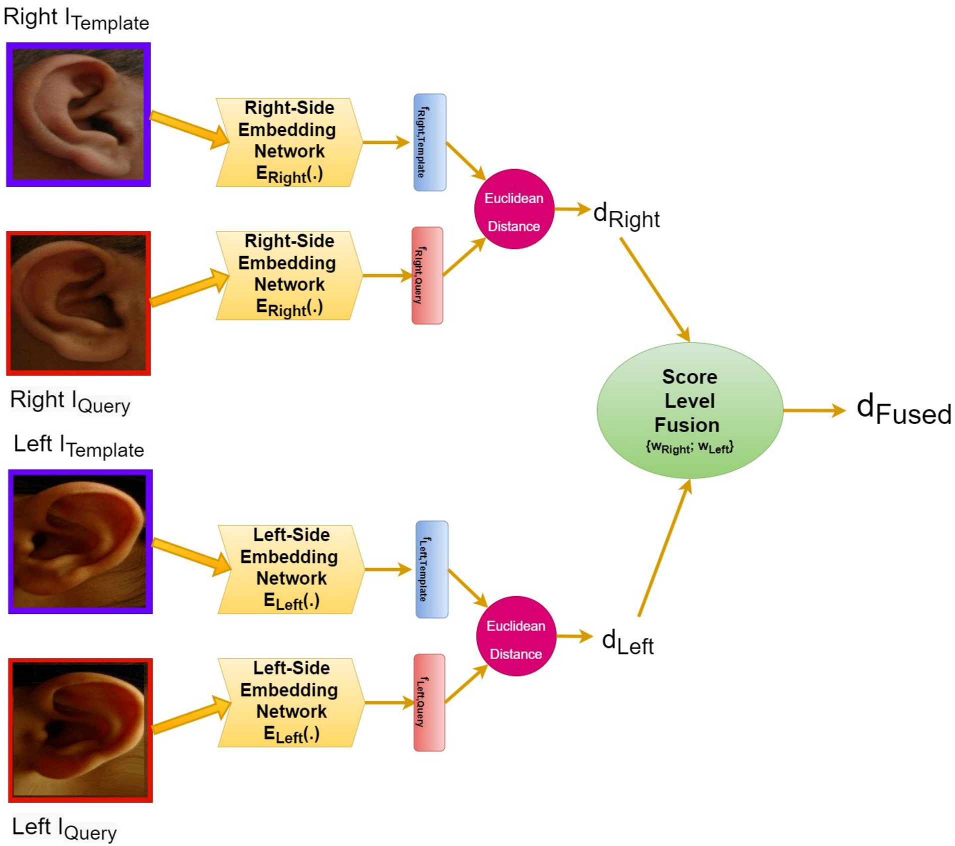


- The siamese training of the network is carried out by providing three images: Anchor(A), Positive(P), and Negative(N) together
- The distance D(,) here is taken as the Euclidean distance

Score Level Fusion: E_{Right} and E_{Left}

- For fusing information from both the right side and left side of the image, we use a score level fusion scheme

- Hence, we only need a trained left embedding network and a trained right embedding network, no separate training for the fused decision is required

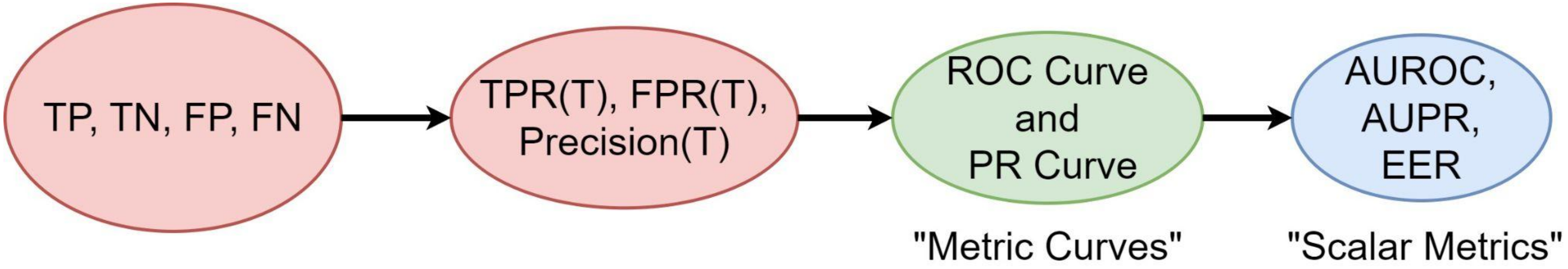


3

Evaluation Metrics for Biometrics System

For Verification Problem

$$\text{True Positive}: I_{query} \text{ actually belonged to person } P_n : D(f_{template}, f_{query}) \leq T$$
$$\text{True Negative}: I_{query} \text{ actually didn't belonged to person } P_n : D(f_{template}, f_{query}) > T$$
$$\text{False Positive}: I_{query} \text{ actually didn't belonged to person } P_n : D(f_{template}, f_{query}) \leq T$$
$$\text{Flase Negative}: I_{query} \text{ actually belonged to person } P_n : D(f_{template}, f_{query}) > T$$



For Identification Problem

- ❑ In identification mode the system returns ranking for the query image, I_{Query} of a person P_i. The system is working correctly in Rank-R sense if:

$$\boldsymbol{r_i \leq R}, \textbf{ for query image of } \boldsymbol{i^{th}} \textbf{ person.}$$

$$\text{Rank-R Accuracy} = \frac{\text{Number of test cases declared correct in Rank-R sense}}{\text{Total number of query images}}$$

- ❑ Using above results the plot of Rank-R Accuracy versus Rank-R called Cumulative Match Curve (***CMC***).

4

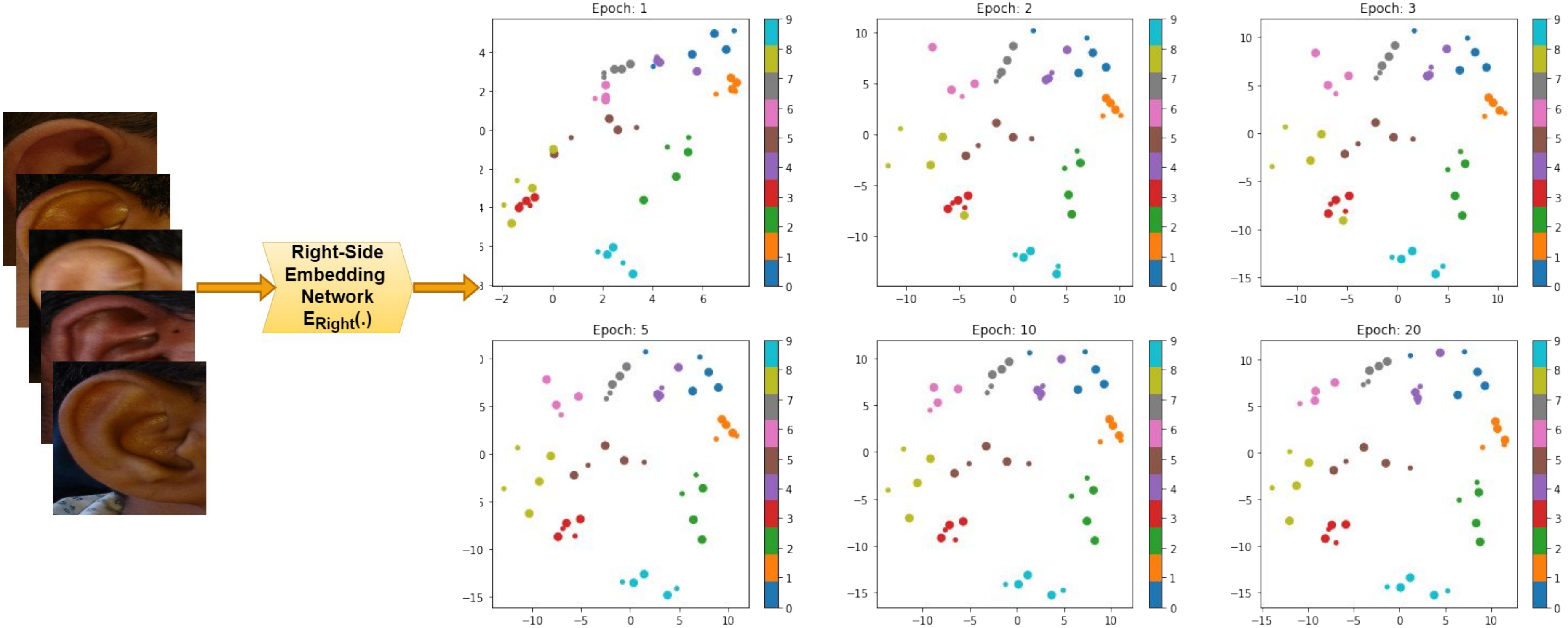
Experimental Results

Set Hyperparameters and Type of Experimental Settings

- ❑ Train the embedding network with margin = 1 using Adam optimizer in Keras library with learning rate of 1e-4 and β parameters set to default.
- ❑ Run the training for 20 epochs with a batch size of 100. Embedding network for each of the ear side is trained for 5 independent trials.
- ❑ The AUPR, EER and Rank-1 metrics are reported with mean and SD across the 5 trials in the paper for different experimental settings
- ❑ Experimental settings: "Closed" and "Open" sets; "Clean" and "Occluded" image acquisition; "Clean" and "Augmented" training of model; 3 different embedding network: no parallel paths, L = 4, L = 8

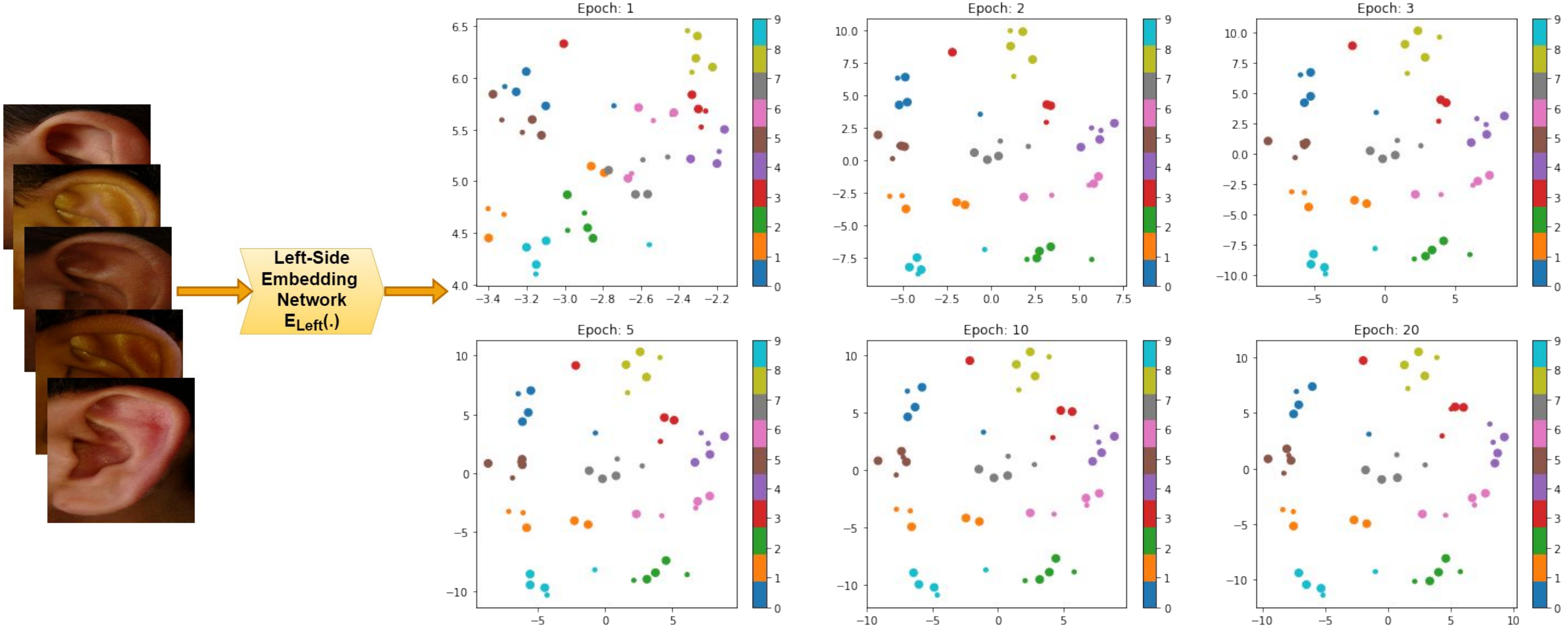
Visualizing the Extracted Features (Right Embeddings)

- The figure shows the extracted features (2D projection of 128D vector) from **10 person's right ear images**. Train: Larger dots, Test: Smaller dots.



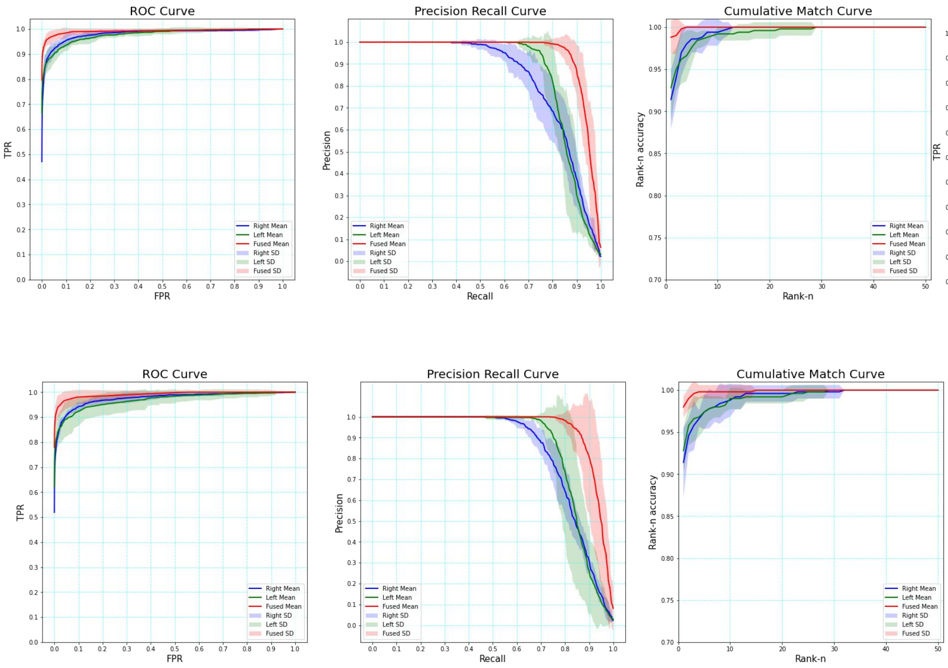
Visualizing the Extracted Features (Left Embeddings)

- The figure shows the extracted features (2D projection of 128D vector) from **10 person's left ear images**. Train: Larger dots, Test: Smaller dots.

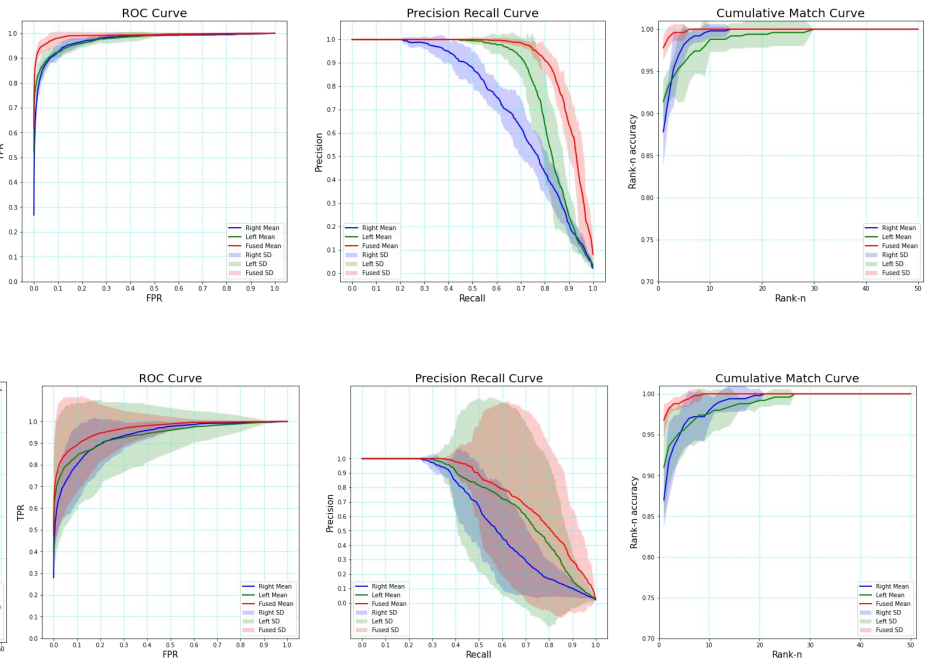


Comparing Results for L = 4 and No Parallel Paths (No Data Augmentation)

- ❑ Top Row: For no parallel Paths
- ❑ Bottom Row: With L = 4



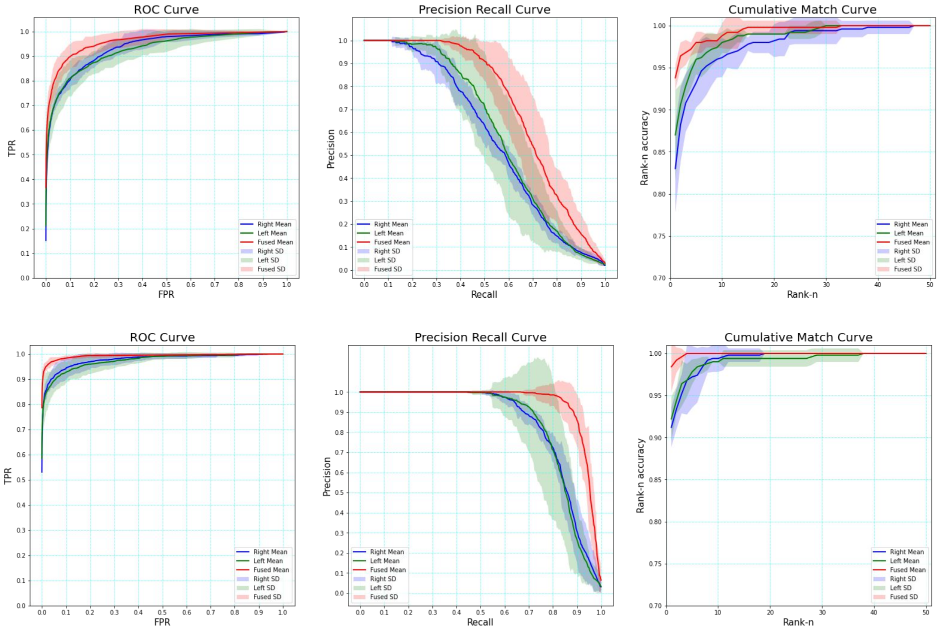
Test Data is Clean



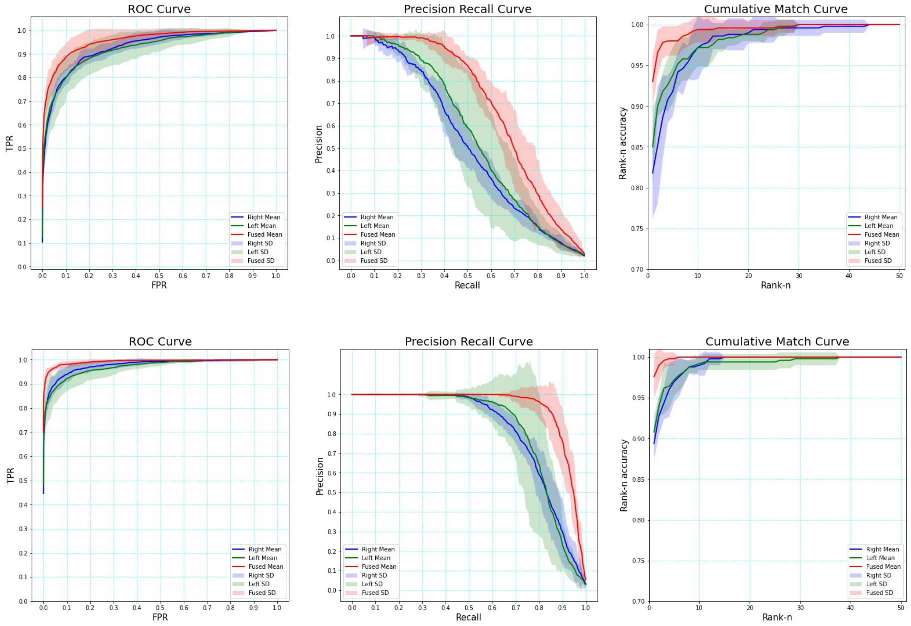
Test Data is Occluded

Comparing Results for L = 4 and No Parallel Paths (With Data Augmentation)

- Top Row: For no parallel Paths
- Bottom Row: With L = 4



Test Data is Clean



Test Data is Occluded

Numerical Results

Test set with no occlusions

	Clean Training			Augmented Training		
	AUPR	EER	Rank-1 Acc.	AUPR	EER	Rank-1 Acc.
No Parallel	**94.23, 0.69**	**3.53, 0.34**	**98.8, 0.40**	76.88, 3.19	8.64, 1.17	97.20, 1.16
Parallel L = 4	93.94, 1.35	3.47, 0.65	98.20, 1.16	93.58, 1.28	3.62, 0.63	**99.40, 0.49**
Parallel L = 8	93.50, 1.34	3.70, 0.51	97.60, 1.02	**95.05, 0.82**	**3.08, 0.76**	98.60, 0.80

Numerical Results

Test set with occlusions

	Clean Training			Augmented Training		
	AUPR	EER	Rank-1 Acc.	AUPR	EER	Rank-1 Acc.
No Parallel	**89.73, 1.24**	**5.09, 0.37**	**98.80, 0.40**	72.67, 2.42	9.56, 1.11	95.40, 1.50
Parallel L = 4	69.34, 14.33	13.59, 8.35	96.4, 1.35	92.38, 1.54	4.24, 0.76	**99.00, 0.63**
Parallel L = 8	84.79, 3.94	5.85, 1.42	97.20, 1.17	**94.01, 0.89**	**3.54, 0.74**	98.00, 0.89

Size of embedding networks used and the corresponding average training time required per step for a batch-size of 100

Embedding Architecture	# trainable params	training time per step
CNN + Parallel SWN with L = 8	562343	374.5 ms/step
CNN + Parallel SWN with L = 4	560603	287.5 ms/step
Baseline with only CNN	560858	80.3 ms/step

- There is very low increment in the number of trainable parameters as the learnable SWN blocks adds very little amount to the number of parameters.

- The training time increases for parallel paths as backpropagation has to pass through multiple paths.

5

Future Scope

Conclusion and Future Scope

- The results show an **agility of a system to recognize occluded and non-occluded images using a common framework**. We believe that it will perform equally with other biometric modalities.

- The score-level fusion technique used in the research can also be used for **multimodal biometric recognition systems**.

- As a part of future work, we will explore the **utility of other types of wavelets** and **advanced information fusion** methods.

References

References

1. P. Birajadar, M. Haria, S. G. Sangodkar, and V. Gadre, “Unconstrained ear recognition using deep scattering wavelet network,” in 2019 IEEE Bombay Section Signature Conference (IBSSC), pp. 1–6, 2019.
2. G. Koch, R. Zemel, R. Salakhutdinov, et al., “Siamese neural networks for one-shot image recognition,” in ICML deep learning workshop, vol. 2, p. 0, Lille, 2015.
3. H. Mirzaalian, T. K. Lee, and G. Hamarneh, “Hair enhancement in dermoscopic images using dual-channel quaternion tubularness filters and mrf-based multilabel optimization,” IEEE Transactions on Image Processing, vol. 23, no. 12, pp. 5486–5496, 2014.
4. J. Bruna and S. Mallat, “Invariant scattering convolution networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1872–1886, 2013.

References

5. S. G. Sangodkar, N. Suthar, A. Palde, P. Birajadar, and V. M. Gadre, “Ear recognition for multimedia security,” in Advanced Security Solutions for Multimedia, IOP Publishing, Sept. 2021.
6. F. Cotter and N. Kingsbury, “A learnable scatternet: Locally invariant convolutional layers,” in 2019 IEEE International Conference on Image Processing (ICIP), pp. 350–354, 2019.
7. M. Andreux, T. Angles, G. Exarchakis, R. Leonarduzzi, G. Rochette, L. Thiry, J. Zarka, S. Mallat, J. Andén, E. Belilovsky, J. Bruna, V. Lostanlen, M. Chaudhary, M. J. Hirn, E. Oyallon, S. Zhang, C. Cella, and M. Eickenberg, “Kymatio: Scattering transforms in python,” Journal of Machine Learning Research, vol. 21, no. 60, pp. 1–6, 2020.

References

8. E. E. Hansley, M. P. Segundo, and S. Sarkar, “Employing fusion of learned and handcrafted features for unconstrained ear recognition,” IET Biometrics, vol. 7, no. 3, pp. 215–223, 2018.
9. J. Bromley, I. Guyon, Y. LeCun, E. S¨ackinger, and R. Shah, “Signature verification using a” siamese” time delay neural network,” Advances in neural information processing systems, vol. 6, 1993.
10. F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.
11. D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

Thank You

Questions?