Mobile based Human Identification using Forehead Creases: Application and Assessment under COVID-19 Masked Face Scenarios

Rohit Bharadwaj[1] Gaurav Jaswal[2] Aditya Nigam[3]
Kamlesh Tiwari[1]

[1]Dept. of CSIS, Birla Institute of Technology and Science Pilani

[2]Dept. of EE, Indian Institute of Technology Delhi

[3]SCEE, Indian Institute of Technology Mandi

IEEE Winter Conference on Applications of Computer Vision

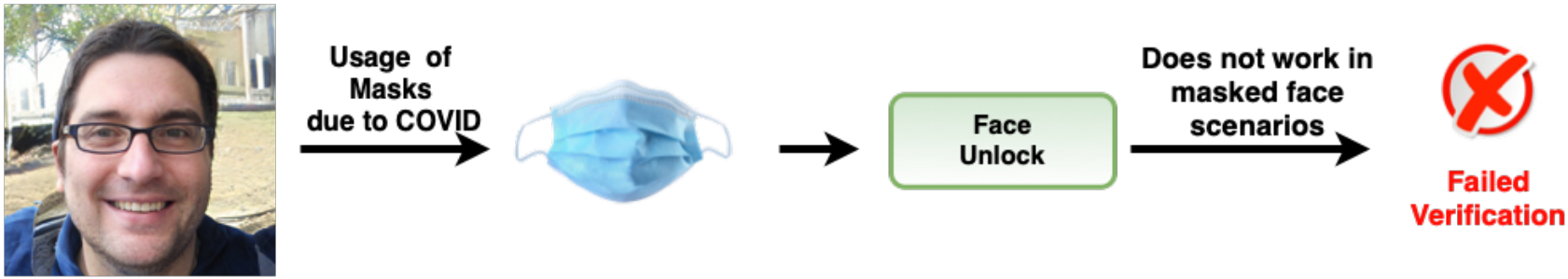
Background

Existing biometrics have undergone serious challenges due to COVID-19.

Background

Existing biometrics have undergone serious challenges due to COVID-19.

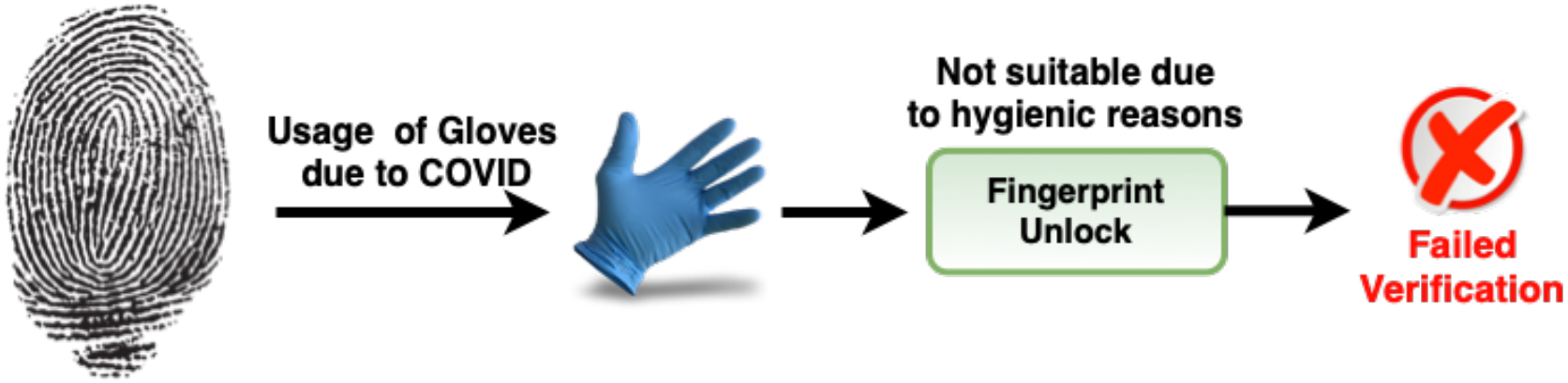
- Face Authentication fails due to masks.



Background

Existing biometrics have undergone serious challenges due to COVID-19.

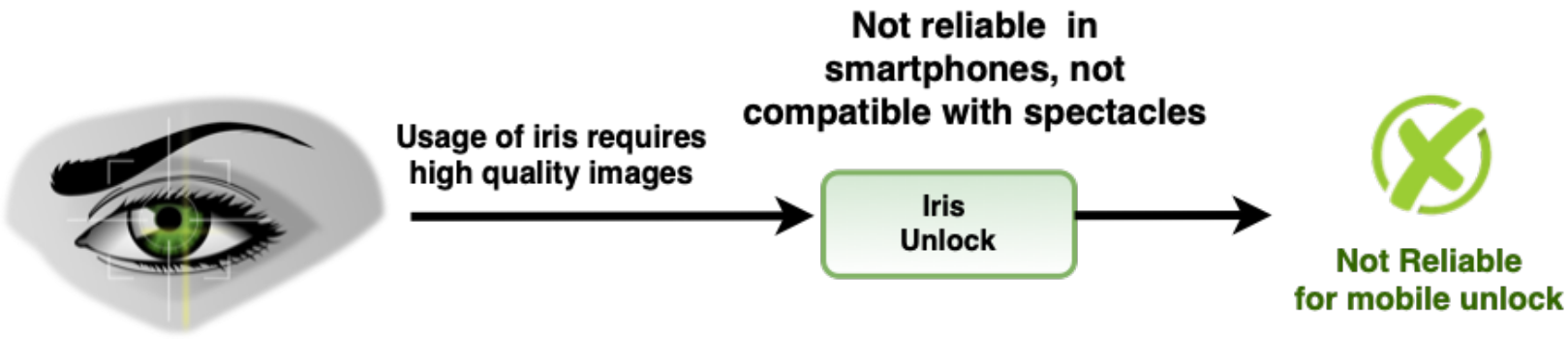
- Face Authentication fails due to masks.
- Fingerprint scanners increases the possibility of contamination.



Background

Existing biometrics have undergone serious challenges due to COVID-19.

- Face Authentication fails due to masks.
- Fingerprint scanners increases the possibility of contamination.
- Iris Authentication requires high-quality images; not generally used.



Alternate Biometrics?

- Voice?:

Alternate Biometrics?

- Voice?: Recent studies have shown the consequences of face coverings upon the quality of speech.[1]

[1] David Haws and Xiaodong Cui. "Cyclegan bandwidth extension acoustic modeling for automatic speech recognition". In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 6780–6784.

Alternate Biometrics?

- Voice?: Recent studies have shown the consequences of face coverings upon the quality of speech.
- Periocular?:

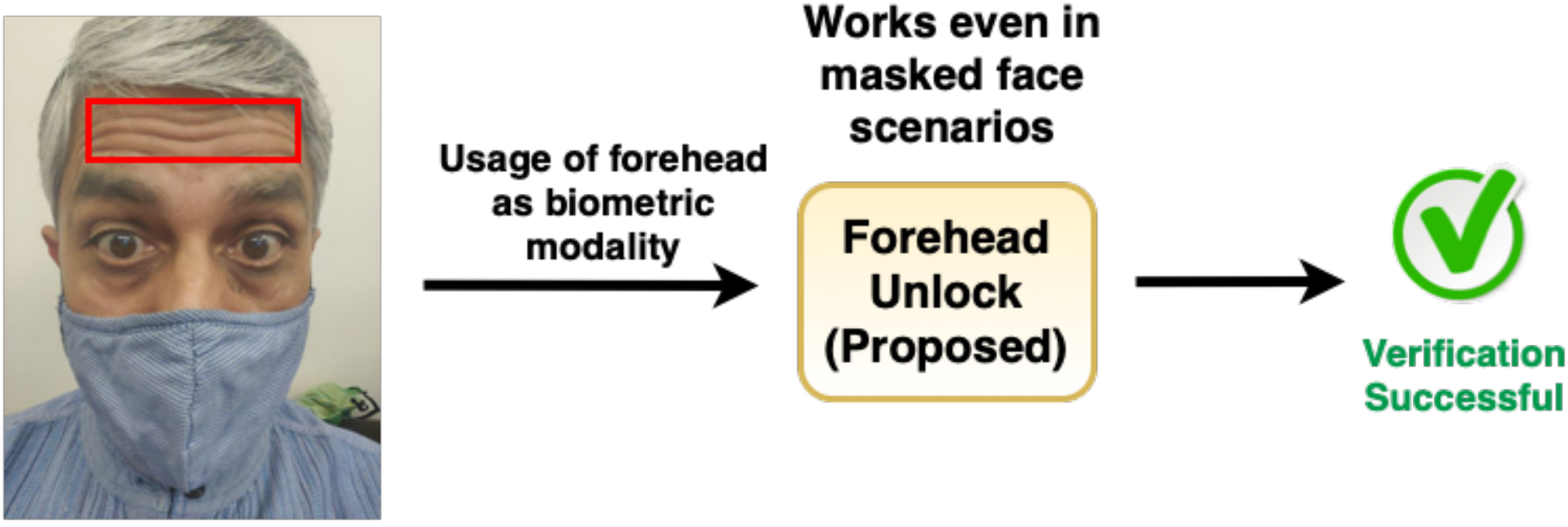
Alternate Biometrics?

- Voice?: Recent studies have shown the consequences of face coverings upon the quality of speech.
- Periocular?: Face mask and low-quality images might present an obstruction to reliable segmentation of periocular region.[1]

[1] Juan Tapia et al. “Selfie Periocular Verification using an Efficient Super-Resolution Approach”. In: *arXiv preprint arXiv:2102.08449* (2021).

Forehead Creases

Our work investigates forehead creases (under given facial expressions) as a biometric trait.



Related Work

- Forehead Crease patterns are a significantly less explored biometric modality in the literature so far.
- The first work in forehead recognition was presented by Manit et al. (2019)[2]

[2] Jirapong Manit et al. “Human forehead recognition: a novel biometric modality based on near-infrared laser backscattering feature image using deep transfer learning”. In: *IET Biometrics* 9.1 (2019), pp. 31–37.

Related Work

- Forehead Crease patterns are a significantly less explored biometric modality in the literature so far.
- The first work in forehead recognition was presented by Manit et al. (2019)[2]
  1. Purpose not helpful for smartphone-based applications.

[2] Jirapong Manit et al. “Human forehead recognition: a novel biometric modality based on near-infrared laser backscattering feature image using deep transfer learning”. In: *IET Biometrics* 9.1 (2019), pp. 31–37.

Related Work

- Forehead Crease patterns are a significantly less explored biometric modality in the literature so far.
- The first work in forehead recognition was presented by Manit et al. (2019)[2]
  1. Purpose not helpful for smartphone-based applications.
  2. Their system acquired forehead images using a complex imaging system based on near-infrared laser scanning.

[2] Jirapong Manit et al. “Human forehead recognition: a novel biometric modality based on near-infrared laser backscattering feature image using deep transfer learning”. In: *IET Biometrics* 9.1 (2019), pp. 31–37.

Related Work

- Forehead Crease patterns are a significantly less explored biometric modality in the literature so far.
- The first work in forehead recognition was presented by Manit et al. (2019)[2]
  1. Purpose not helpful for smartphone-based applications.
  2. Their system acquired forehead images using a complex imaging system based on near-infrared laser scanning.
  3. Not suitable in COVID-19 applications due to hygienic concerns, sensor surface noise, and user inconvenience.

[2] Jirapong Manit et al. "Human forehead recognition: a novel biometric modality based on near-infrared laser backscattering feature image using deep transfer learning". In: *IET Biometrics* 9.1 (2019), pp. 31–37.

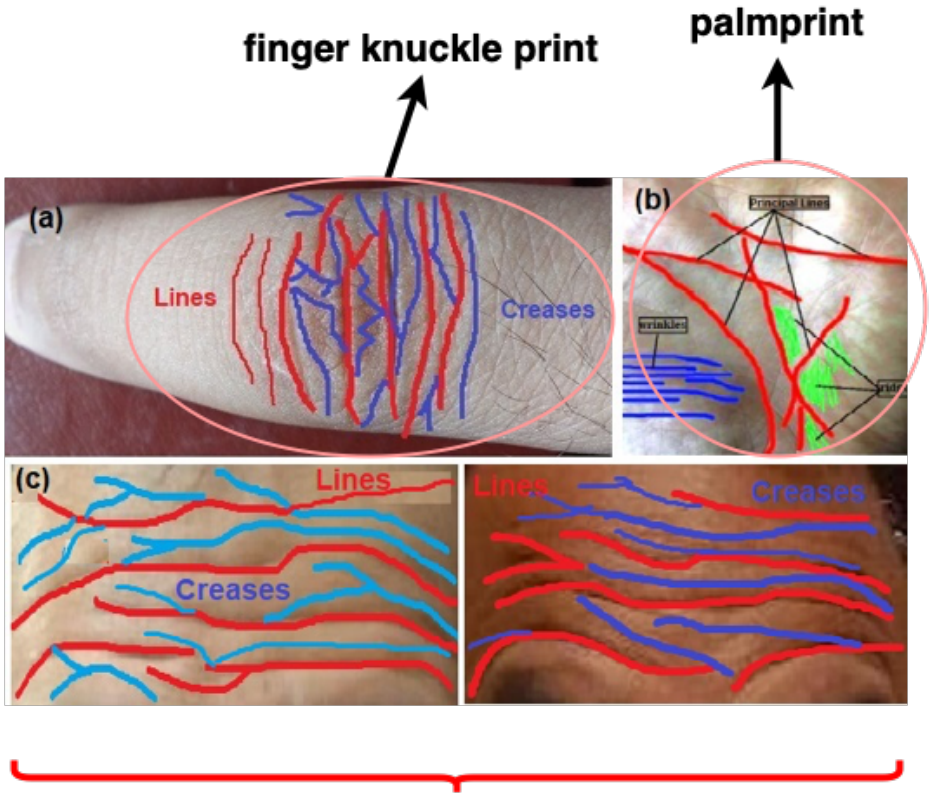
Our Work

In forehead images, creases are observed to be distinct and permanent for each individual due to the differences in frontal bone morphology and thickness of the skin tissues.[3]

[3] Jirapong Manit, Achim Schweikard, and Floris Ernst. "Deep convolutional neural network approach for forehead tissue thickness estimation". In: *Current Directions in Biomedical Engineering* 3.2 (2017), pp. 103–107.

Our Work

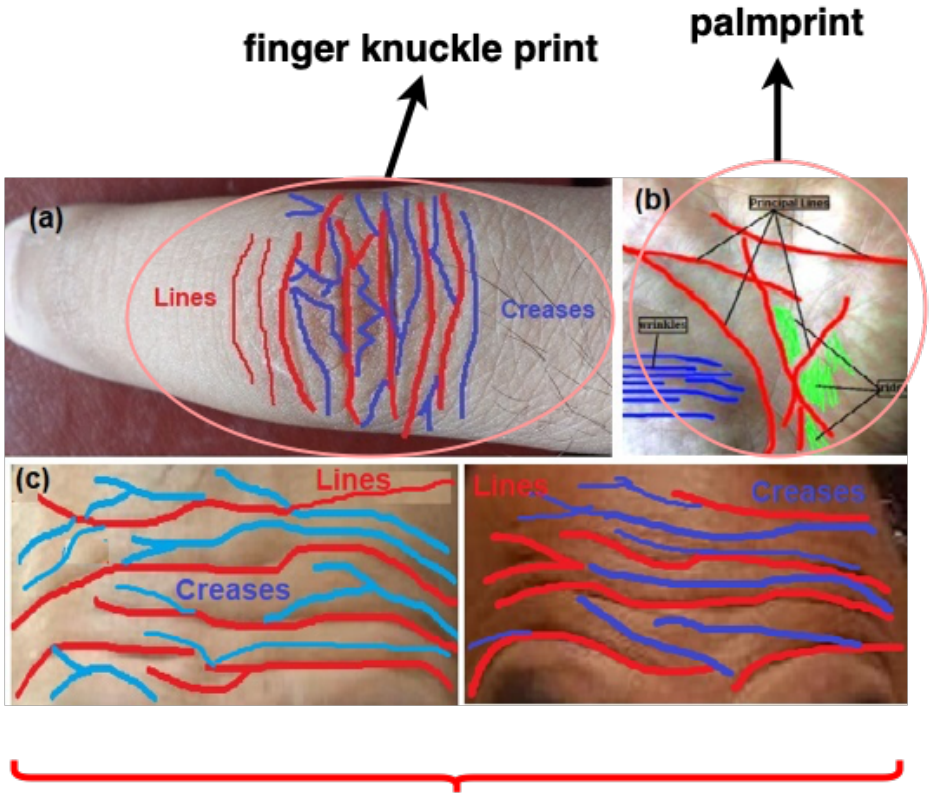
We assume and hypothesize that forehead creases have similar image properties to finger knuckle print and palmprint.



Forehead Creases show similar line and crease-patterns as finger knuckle print and palmprint.

Our Work

We assume and hypothesize that forehead creases have similar image properties to finger knuckle print and palmprint.



Forehead Creases show similar line and crease-patterns as finger knuckle print and palmprint.

However, the uniqueness of these creases is insufficient to establish the personal identity in non-cooperative conditions. Thus, a full potential from the forehead biometrics is yet to be realized.

Contributions

- First smartphone-based biometric recognition approach with forehead images taken under masked face scenarios.
- We incorporate LMCL loss, that is highly optimized for extracting discriminative features for forehead recognition.
- Incorporate a dual attention mechanism that considers spatial attention to learn semantic regularities and channel attention to compute correlation between all channels independently.
- Devised an intelligent data acquisition strategy using our android application.
- Developed forehead-image database consisting of 4,964 cropped forehead ROI's from 247 subjects.

Complete Pipeline

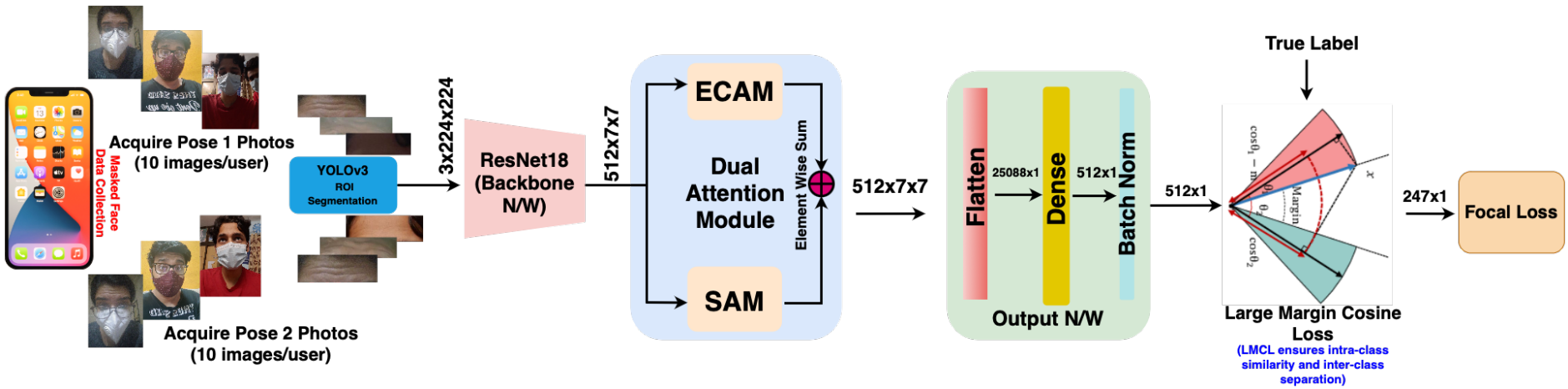
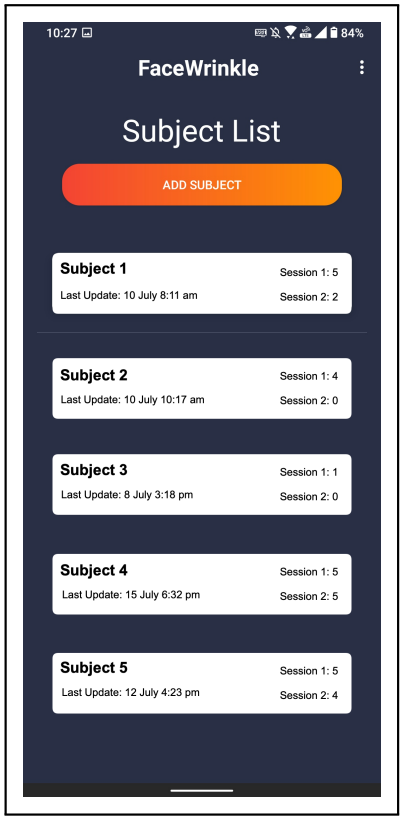


Figure: Block diagram of the proposed forehead recognition network assisted with dual attention network. The network performs discriminative feature learning in cosine space and thus has richer embedding representation.

Forehead Image Acquisition



App UI

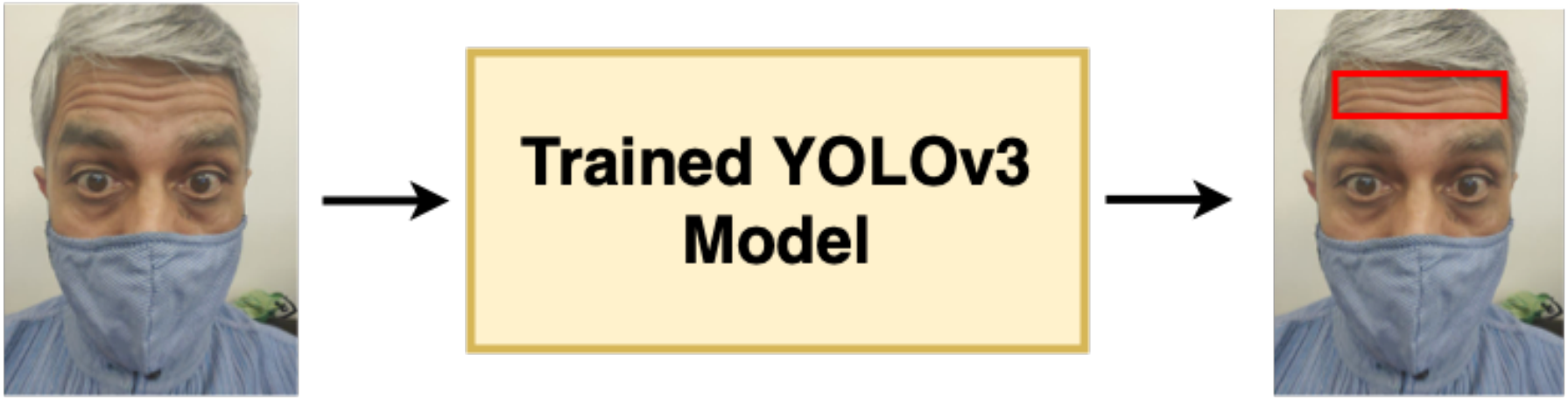
Pose-1

Pose-2

- The data is collected remotely using our android application. We consider the above two poses to account for variance due to distance.
- The data is taken in two sessions, wherein a subject must provide five photographs of the two poses in each session

Extracting the Region of Interest

- After obtaining the forehead images, we deployed pre-trained YOLOv3 model (which initially was trained on Imagenet) on around 600 raw forehead images from 30 subjects.
- Using the trained weights from the YOLOv3 model, we obtained bounding boxes over our entire dataset. Thus we obtain cropped ROI's from our initial face-selfie images.



BITS-IITMandi-ForeheadCreases Images Database

Figure: Sample Images from the BITS-IITMandi-ForeheadCreases Images Database

- This dataset consists of **4,964** forehead-images from **247** subjects, with each user giving roughly 20 samples.
- This is the first and the largest dataset on forehead photos, which is being made public to the research community.
- Accessed from: http://ktiwari.in/projects/foreheadcreases/

Backbone Network

- The feature map $E^{(i)} \in \mathbb{R}^{C' \times H' \times W'}$ obtained from the backbone network f can be given by:

$$E^{(i)} = f(x^{(i)}; \theta^{(1)}) \tag{1}$$

where $x^{(i)} \in \mathbb{R}^{C \times H \times W}$ is the input ROI image.
$\theta^{(1)}$ = backbone network parameters, C = number of input channels, H = height, and W = width of the image. Similarly, C', H', W' denotes the feature maps' number of channels, height, and width, respectively.

- We have considered ResNet-18 as our backbone network f, and the feature map $E^{(i)}$ is extracted from the **conv5_x** layer of ResNet-18.

Dual Attention Module

- Forehead ROI's have a lot of background in the form of eyebrows, acne, and hair which can lead to distorted features and reduce the discriminatory power.
- To ensure backbone network is not confused by unimportant features, we employ a dual attention module (DAM) mechanism to jointly model the inter-dependencies of spatial and channel level forehead features.
- Spatial Attention Module (SAM): The output feature map $E^{(i)}$ is passed through SAM to emphasize the spatial dependencies.

$$F_1^{(i)} = E^{(i)} \bigotimes A_{SAM}(E^{(i)}; \theta^{(2)}) + E^{(i)} \tag{2}$$

- Channel Attention Module (CAM): The output feature map $E^{(i)}$ is passed through CAM to capture channel-wise relationship of features.

$$F_2^{(i)} = E^{(i)} \bigotimes A_{CAM}(E^{(i)}; \theta^{(3)}) \tag{3}$$

Feature Fusion

To fuse the attention outputs from CAM and SAM, we have considered the following two scenarios.

1. Concatenate the outputs obtained from the two attention modules.

$$O^{(i)} = F_1^{(i)} \| F_2^{(i)} \tag{4}$$

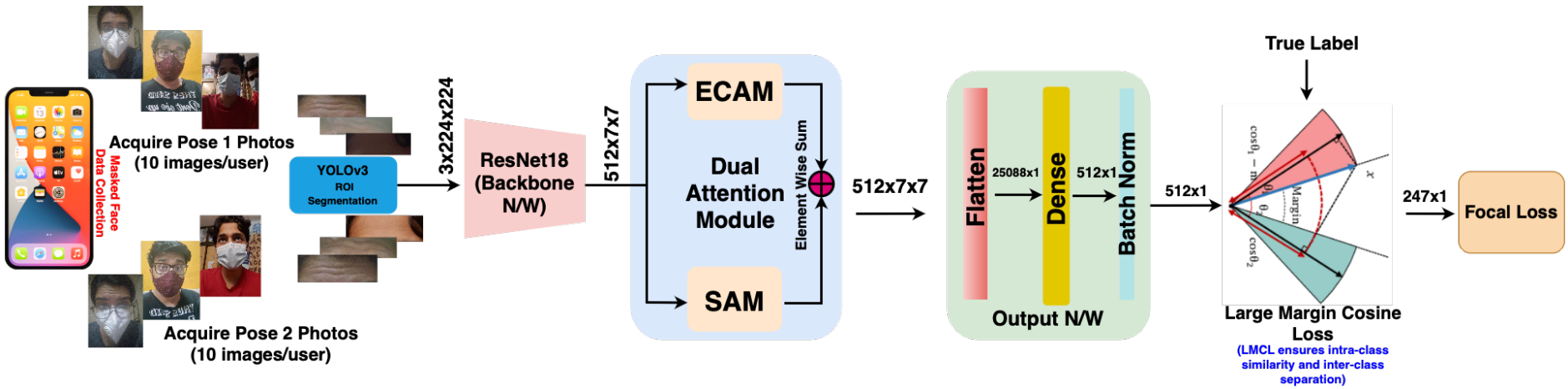
where $O^{(i)} \in \mathbb{R}^{2C' \times H' \times W'}$ is the feature map obtained after concatenating the two attention module outputs, and $\|$ denotes the concatenation operator.

2. Element-wise sum of the two attention outputs. Mathematically,

$$O_j^{(i)} = {F_1}_j^{(i)} + {F_2}_j^{(i)} \tag{5}$$

where, ${F_1}_j^{(i)}$ is the j-th element of $F_1^{(i)}$, ${F_2}_j^{(i)}$ is the j-th element of $F_2^{(i)}$, $O_j^{(i)} \in \mathbb{R}^{C' \times H' \times W'}$ is the j-th element of output feature map.

Final Embedding



The fused output O is then flattened to 1-dimension, passed into a fully-connected layer followed by a batch-normalization layer to obtain the final output feature vector $I^{(i)} \in \mathbb{R}^{512}$, which represents the embedding of the input forehead RoI image $x^{(i)}$.

Loss Function

- Large Margin Cosine Loss (LMCL)[3] is utilized for training the entire network in an end-to-end fashion.
- To have less intra-class variation and more inter-class variations, $m \geq 0$ is set as enforced margin.
- LMCL loss function can be formulated as:

$$L = \frac{1}{N} \sum_{i} -\log \frac{e^{s(\cos(\theta_{y_i}, i) - m)}}{e^{s(\cos(\theta_{y_i}, i) - m)} + \sum_{j \neq y_i} e^{s \cos(\theta_j, i)}} \tag{6}$$

where, $\cos(\theta_j, i) = W_j^T O^{(i)}$ and W_j is the weight vector of the j-th class.

[3] Hao Wang et al. "Cosface: Large margin cosine loss for deep face recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 5265–5274.

Network Training

- Our dataset was divided roughly equally into train and test sets.
- We have set LMCL parameters $m = 0.35$, $s = 300$, and used feature embedding dimension as 512.
- The output obtained from LMCL is a vector of class probabilities. To facilitate backpropagation, output from LMCL is further passed to the focal loss module with $\gamma = 2$.
- The model weights are updated using adam optimizer, initial learning rate is set to $lr = 3 \times 10^{-4}$ and it is decayed by $\gamma = 0.1$ at every 20 epochs. We trained the network for 100 epochs with L2 weight penalty of $\lambda = 5 \times 10^{-4}$.

Experiments

- Images in train set were used as gallery, while the images in test set were used as query to perform matching.
- Each query image is matched with each of the gallery images.
- A score between two embeddings is termed as genuine matching if they are obtained from the same subjects; otherwise, it is an imposter matching.
- We perform the following two kinds of matching experiments.
  1. **Closed Set Matching:** Consider all 247 subjects and divide entire dataset roughly equally into gallery and query. In our work we have **2,462** images in gallery and **2,502** images in the query. Thus, we obtain **22,768** genuine matchings and **5,565,663** imposter matchings.
  2. **Open Set Matching:** We exclude 30% subjects (i.e. 75) from the entire dataset and the model is then trained only on **172** subjects. This trained model is then used to generate matching scores on the remaining **unseen 75 subjects**. We have **744** images in gallery and **742** images in query, and we obtain **7,386** genuine and **544,662** imposter matchings.

Data Augmentation

- For data augmentation, we sampled 5 points randomly from each of 45×45, 47×47, 49×49, and 51×51 neighborhoods around the center of the original bounding box of each ROI.
- We generated new ROI's using these sampled points as the center of the new bounding boxes (while keeping the height and width of the boxes unchanged).
- Thus, the dataset is augmented by **20 times**.

Data Augmentation

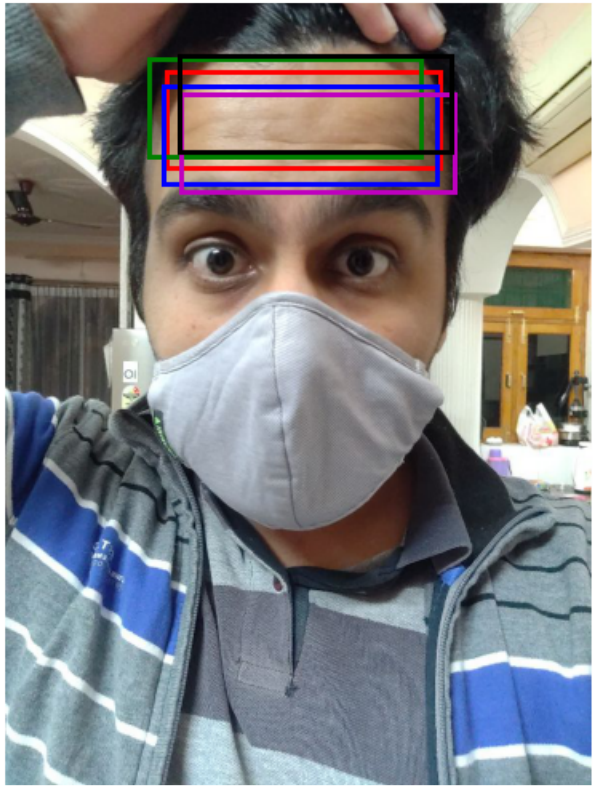


Figure: Augmentation process shown for a single subject. Red Bounding box is the original bounding box and the other boxes are obtained by following the augmentation process on 45×45 neighbourhood.

Results

- Augmented Data is used **only** to train the model, and only the training set data was augmented. The gallery and query image sets are unchanged and only the model is varied.
- **Model V1** refers to our proposed model trained on non-augmented dataset.
- **Model V2** refers to our proposed model trained on augmented dataset.

	Subject Count	Matching Strategy	CRR (%)	EER (%)	DI
Model V1	75	Open Set	96.08	17.50	1.44
Model V2	**75**	**Open Set**	**97.84**	**12.40**	**1.66**
Model V1	247	Closed Set	97.22	4.46	1.91
Model V2	**247**	**Closed Set**	**99.08**	**0.44**	**2.79**

Table: Closed Set and Open Set System performance on forehead dataset.

Ablation Study

Model	Scenario	Subjects/ Poses	CRR %	EER %
V1 (No attention)	closed set	247	97.64	4.18
V2 (No attention)	closed set	247	99.36	0.38
V1 (No attention)	open set	75	96.92	17.06
V2 (No attention)	**open set**	**75**	**97.44**	**12.98**
V1 (Fused by Element Wise Sum)	open set	75	96.08	17.50
V2 (Fused by Element Wise Sum)	**open set**	**75**	**97.84**	**12.40**
V1 Fused by Concatenation	open set	75	95.52	17.45
V2 Fused by Concatenation	open set	75	97.71	13.22
V1 (SAM)	open set	75	93.43	19.19
V2 (SAM)	**open set**	**75**	**96.90**	**12.95**
V1 (SAM)	closed set	247	97.00	4.88
V2 (SAM)	closed set	247	99.32	0.39
V1 (CAM)	**closed set**	**247**	**98.10**	**3.02**
V2 (CAM)	closed set	247	99.48	0.90
V1 (Fused by Concatenation)	closed set	247	96.96	4.47
V2 (Fused by Concatenation)	closed set	247	99.24	0.48
V1 (Fused by Element Wise Sum)	closed set	247	97.22	4.46
V2 (Fused by Element Wise Sum)	**closed set**	**247**	**99.08**	**0.44**

Comparative Analysis

Approach	Dataset Statistics	Testing Protocol	CRR %	EER %
ResNet18 (Work by[4])	30 Subjects, 270 images	Validation Accuracy	94.47	NA
ResNet101 (Work by[4])	30 Subjects, 270 images	Validation Accuracy	98.55	NA
MLP Mixer[5] (Our Work)	247 Subjects, 4964 images	Standard Matching	92.31	4.08
ArcFace Loss[6] (Our Work)	247 Subjects, 4964 images	Standard Matching	97.47	4.54
Proposed (element sum fusion)	247 Subjects, 4964 images	Standard Matching	97.22	4.46
Proposed (concat fusion)	247 Subjects, 4964 images	Standard Matching	96.96	4.47
Proposed (SAM only)	247 Subjects, 4964 images	Standard Matching	97.00	4.88
Proposed (CAM only)	247 Subjects, 4964 images	Standard Matching	98.10	3.02
Proposed (element sum fusion) *	247 Subjects, 4964 images	Standard Matching	99.08	0.44
Proposed (concat fusion) *	247 Subjects, 4964 images	Standard Matching	99.24	0.48
Proposed (SAM only) *	247 Subjects, 4964 images	Standard Matching	99.32	0.39
Proposed (CAM only) *	247 Subjects, 4964 images	Standard Matching	99.48	0.90

Table: Comparison with other methods and settings. (* denotes model trained using augmented dataset.)

[4] Jirapong Manit et al. “Human forehead recognition: a novel biometric modality based on near-infrared laser backscattering feature image using deep transfer learning”. In: *IET Biometrics* 9.1 (2019), pp. 31–37.

[5] Ilya Tolstikhin et al. “Mlp-mixer: An all-mlp architecture for vision”. In: *arXiv preprint arXiv:2105.01601* (2021).

[6] Jiankang Deng et al. “Arcface: Additive angular margin loss for deep face recognition”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 4690–4699.

Closed Set ROC Plot

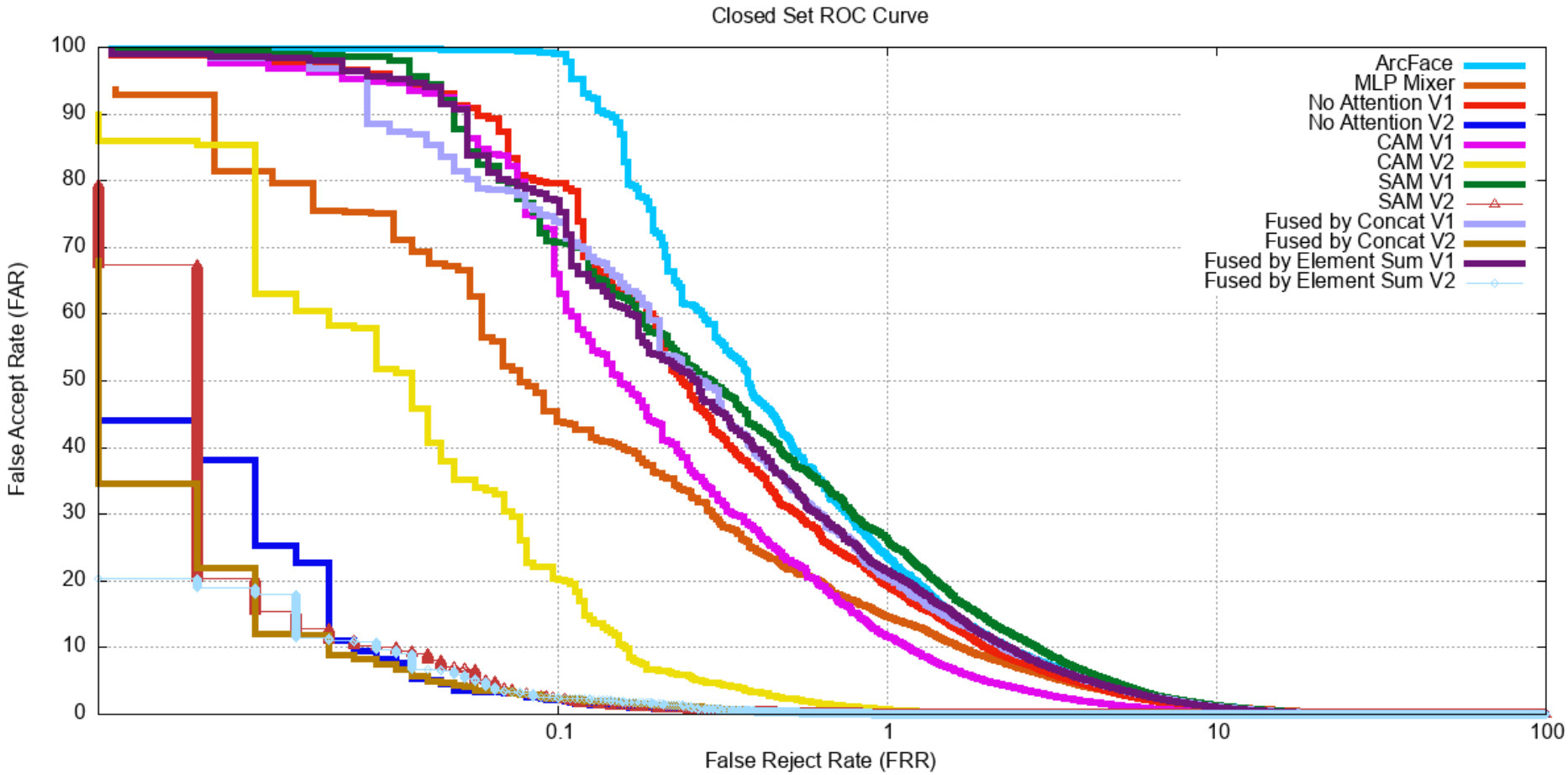


Figure: ROC curves for the closed set experiments. The x-axis is shown in log scale.

Open Set ROC Plot

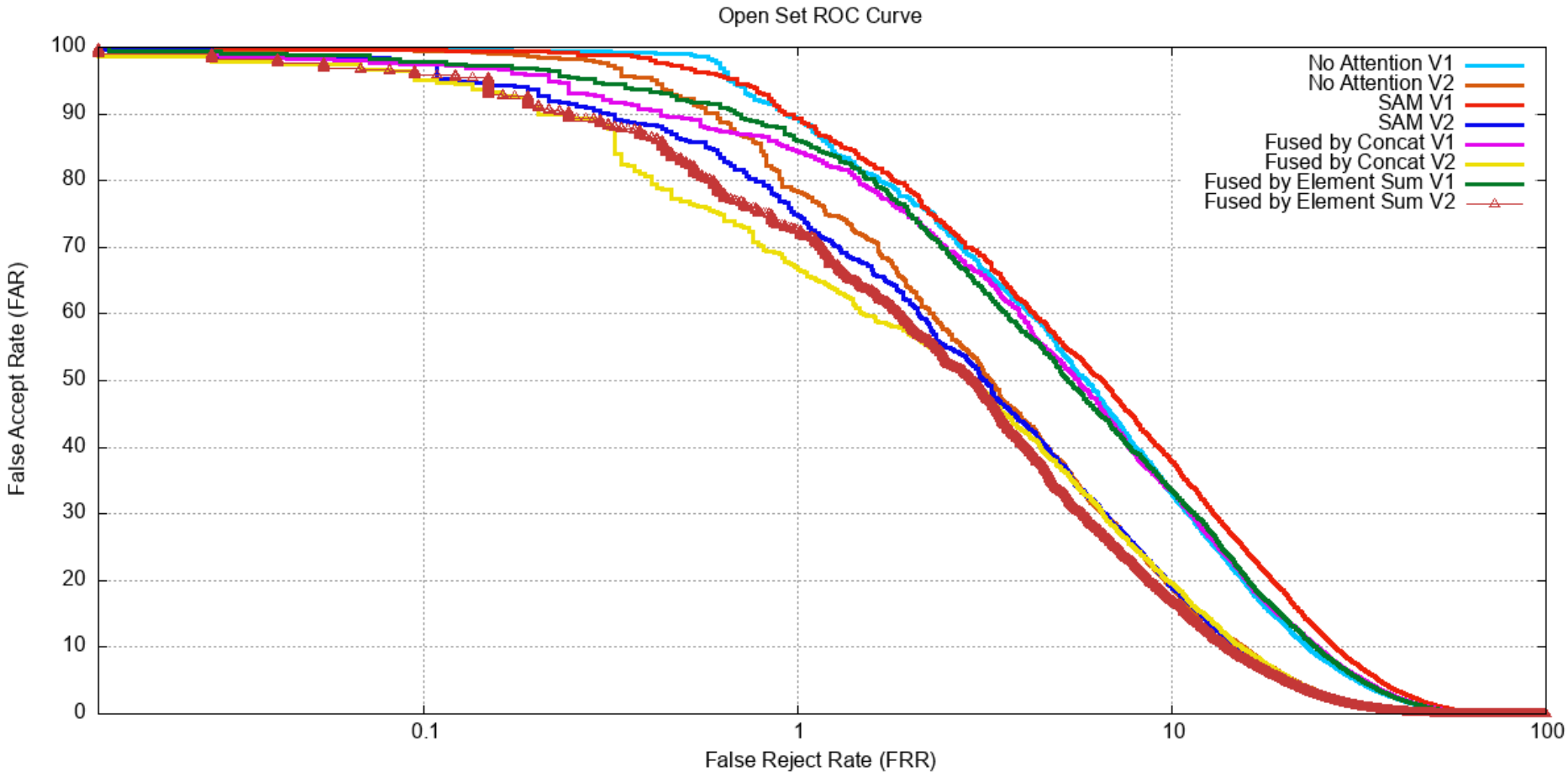


Figure: ROC curves for the open set experiments. The x-axis is shown in log scale.

Conclusion

- This work investigates the usefulness of the forehead creases, under surprised facial expressions, as a biometric modality for smartphone-based user recognition
- We device touch-less image acquisition using a mobile application and present a generalized deep learning framework with an attention-guided mechanism that further regularizes metric learning for better inter-class variability.
- The system is evaluated on a masked face dataset acquired from **247 subjects** that contain **4,964 selfie images**.
- Our proposed network reports high performance results in the closed and open set matching protocols: CRR: 99.08%, EER: 0.44% on the closed set, and CRR: 97.84%, EER: 12.40% on open set experiments.
- These outperforming results are comparable to the performance of iris, finger knuckle, and palmprint biometrics. Thus, we validate our assumption for considering forehead creases as a biometric modality to improve masked face scenarios.

Database and Code

We are making the database and the code used (network and android data-collection app) public, so as to facilitate further research in this direction. The database and the code can be accessed by visiting:

1. **Database:** http://ktiwari.in/projects/foreheadcreases/
2. **Code:** https://github.com/rohit901/ForeheadCreases

Thank you!!