
Journal of Informatics and Web Engineering

Vol. 3 No. 3 (October 2024)

eISSN: 2821-370X

Conditional Deployable Biometrics: Matching Periocular and Face in Various Settings

Jihyeon Kim¹, Tiong-Sik Ng², Andrew Beng Jin Teoh^{3*}

^{1,2,3}Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, South Korea

*corresponding author: (bjteoh@yonsei.ac.kr; ORCID: 0000-0001-5063-9484)

Abstract - In this paper, we introduce the concept of Conditional Deployable Biometrics (CDB), designed to deliver consistent performance across various biometric matching scenarios, including intra-modal, multimodal, and cross-modal applications. The CDB framework provides a versatile and deployable biometric authentication system that ensures reliable matching regardless of the biometric modality being used. To realize this framework, we have developed CDB-Net, a specialized deep neural network tailored for handling both periocular and face biometric modalities. CDB-Net is engineered to handle the unique challenges associated with these different modalities while maintaining high accuracy and robustness. Our extensive experimentation with CDB-Net across five diverse and challenging in-the-wild datasets illustrates its effectiveness in adhering to the CDB paradigm. These datasets encompass a wide range of real-world conditions, further validating the model's capability to manage variations and complexities inherent in biometric data. The results confirm that CDB-Net not only meets but exceeds expectations in terms of performance, demonstrating its potential for practical deployment in various biometric authentication scenarios.

Keywords— Biometrics, Face, Periocular, Deep Learning, Matching

Received: 30 July 2024; Accepted: 18 September 2024; Published: 16 October 2024

This is an open access article under the [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.



1. INTRODUCTION

Biometrics can be easily associated with a user's identity, in which it pertains to how it is unique to each person [Error! Reference source not found.]. In other words, considering this uniqueness, it is said that biometrics can be a considerable solution to second-factor identification components such as passwords. Face biometrics is a popular biometric modality, given that it is non-cooperative and is easy to obtain. However, there may be instances where face biometrics may underperform. For example, occlusions on the face may occur, such as the usage of surgical mask, make-up, and plastic surgery. In such a case, periocular biometrics, which is the ocular area around the eye region, is a good alternative Error! Reference source not found.].

Though not as well studied as face, periocular biometrics have been relatively popular since the advent of machine learning, beginning with different hand-crafted techniques such as the Local Binary Pattern (LBP) Error! Reference source not found.]. However, given its lack of representation power, periocular biometrics have a

relatively low performance in comparison with face. Since the introduction of the Convolutional Neural Networks (CNN) **Error! Reference source not found.**, deep learning has been a popular technique used for different classification techniques, which naturally includes periocular recognition.

Conditional multimodal biometrics, has garnered attention in terms of boosting the performance of periocular biometrics. In **[Error! Reference source not found.]**, the concept of conditional multimodal biometrics comprising a shared-parameter deep neural network alongside a specially designed loss function was proposed for periocular and face images. In said work, the matching of intra-modal and multimodal was thoroughly studied. In the case of the former, intra-modal matching refers to the matching between similar biometric modalities, e.g., periocular vs. periocular, or face vs. face, while the latter refers to the fused matching between several biometric modalities, e.g., fused periocular and face vs. fused periocular and face.

In this paper, we take a step further upon improving the conditional multimodal biometrics by proposing the notion of Conditional Deployable Biometrics (CDB), whereby we not only study the intra-modal matching and multimodal matching of periocular and face, but we also include the non-trivial study of cross-modal matching. Our proposed CDB regimen enables flexible deployment for biometric recognition systems, accepting either modality that are trained during the enrollment phase (i.e., feature extraction and storage) and the query phase (i.e., feature extraction and perform matching for authentication). Figure 1 illustrates the enrollment and query phases after the training with several biometric modalities are done. Particularly, suppose that periocular and face are trained, the input biometrics during the enrollment and query phases can be consisted of either face, periocular, or a combination of both modalities.

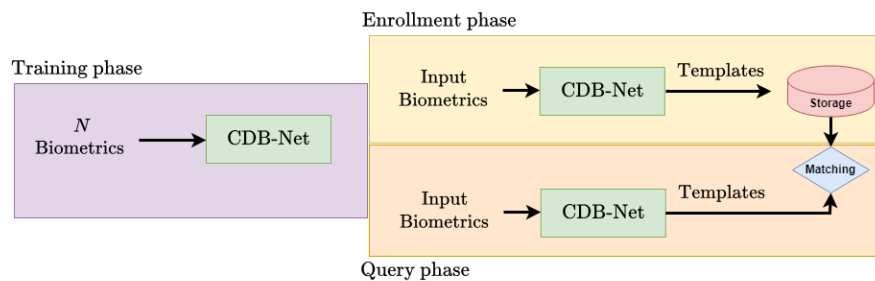


Figure 1. Training, Enrollment, And Query Phases Of CDB

In this paper, we realize the CDB by proposing the CDB-Net, a deep neural network that comprises a CNN architecture and a specifically designed loss function that enables matching of periocular and face modalities that are deployable in multiple settings. The architecture of the CDB-Net is built upon the lightweight MobileFaceNet **Error! Reference source not found.** architecture, whereby we modify the network by appending a CDB block designed specifically for periocular and face to further boost its performance. In addition, our proposed CDB loss enables cross-modal matching between periocular and face, in which though the former is a subset image of the latter, there lies a modality gap between both due to the different image sizes and the different salient areas.

In summary, our contribution is as the following:

1. We propose the CDB notion, whereby we enable flexible deployment of biometric authentication systems in different settings.
2. We realize the CDB regimen by proposing the CDB-Net, a deep neural network consisting of a Convolutional Neural Network architecture and the CDB loss, specifically tailored for periocular and face.
3. The CDB-Net is benchmarked on five in-the-wild datasets, in which these datasets are obtained in an unconstrained environment regardless of the angle, pose, or lighting, deemed to be more challenging than datasets that are obtained in a controlled setting. We demonstrate the effectiveness of the CDB-Net on these datasets by identification and verification matching protocols.

This paper is organized as follows. Section 2 presents the latest literature reviews on periocular and face biometrics, while also studying the concept of conditional multimodal biometrics and several works related to cross-modal

matching. Section 3 then presents the research methodology of our work, introducing in detail the CDB-Net architecture and loss functions. Section 4 then demonstrates the benchmarking of the CDB-Net via experiments conducted, leading to the conclusion of our research in Section 5.

2. LITERATURE REVIEW

2.1 Face and Periocular Biometrics

Face recognition has become an essential part of biometrics recognition ever since the advent of the Yale Face Database [Error! Reference source not found.] due to its non-cooperative nature and its ease of performing authentication through biometric systems. Though it is well studied, the performance of face biometrics has been underwhelming due to its large intra-class variations [Error! Reference source not found.]. With the introduction of deep neural networks such as the Convolutional Neural Network [Error! Reference source not found.] alongside angular margin losses such as the ArcFace [Error! Reference source not found.] and CosFace [Error! Reference source not found.], the research on face recognition was significantly improved geared towards its accuracy improvement, particularly for then-challenging in-the-wild datasets such as the Labelled Faces in the Wild (LFW) [Error! Reference source not found.]. In particular, the LFW was used as a standard benchmark due to the dataset being obtained from an unconstrained environment, whereby there are no restrictions in pose, angles, or lighting settings. As of current state-of-the-art works, the LFW dataset has managed to reach a plateau in performance, leading to variants of it such as the Cross-Pose LFW (CPLFW) [Error! Reference source not found.] to be used as a more challenging benchmark.

Periocular recognition on the other hand, has also garnered early attention with the advent of machine learning techniques. To be specific, early works on periocular recognition involve hand-crafted techniques such as the Local Binary Pattern (LBP) [Error! Reference source not found.] to extract its features and perform matching. However, due to a lack of representation power in periocular compared with face, its performance similarly has also been underwhelming, even moreso than face. As a result, periocular biometrics is typically paired with other biometric modalities such as face [Error! Reference source not found.], or iris [Error! Reference source not found.]. In [Error! Reference source not found.], the proposed MDLN relies on the LBP extracted left and right periocular images to boost its performance. However, the accuracy performance was benchmarked on a dataset obtained in a controlled environment. It is not until [Error! Reference source not found.] that in-the-wild periocular dataset obtained a good performance using a knowledge distillation technique, distilling knowledge from the face teacher network to the periocular student network.

Though our work benchmarks the CDB-Net on intra-modal matching of face and periocular recognition, we place an equal emphasis also on matching in the multimodal and cross-modal settings. For a thorough survey on face and periocular advancements, we refer the readers to the survey paper in [Error! Reference source not found.], [Error! Reference source not found.].

2.2 Conditional Multimodal Biometrics and Cross-Modal Matching

In [Error! Reference source not found.], the Conditional Multimodal Biometrics (CMB) notion was proposed to boost the performance of periocular via leveraging its learning with face, such that a shared parameter network accepting both inputs is able to perform mutual learning of both modalities. This leads to the performance boost of both modalities, in which the benchmark was carried out on in-the-wild datasets, outperforming the knowledge distillation technique in [Error! Reference source not found.]. Though that is the case, the study on cross-modal matching between periocular and face was not well studied.

[Error! Reference source not found.] then studies the cross-modal matching between periocular and face biometrics, in which that it is deemed non-trivial given the image size discrepancy that exists between both images. In particular, the HA-ViT, a variant of the Vision Transformer (ViT) [Error! Reference source not found.] was proposed with the concept of the multi-head self-attention [Error! Reference source not found.] such that cross-modal matching between both modalities was enabled. This was further backed up with the introduction of the Cross Face-Periocular Contrastive (CFPC) Loss, that leverages intra-modal and inter-modal negatives in a manner similar to label smoothing regularization [Error! Reference source not found.], which further boosts the cross-modal

matching between both modalities. However, the effect of cross-modal matching on intra-modal matching was not well studied.

In this work, we place an equal emphasis on intra-modal, multimodal, and cross-modal matching, thus fulfilling the CDB notion via the design of the CDB-Net. To be precise, the CDB-Net built upon a modified upgrade of the MobileFaceNet **Error! Reference source not found.**] with a CDB block accepting both periocular and face modalities as inputs, alongside the CDB loss were designed specifically to achieve this goal.

3. RESEARCH METHODOLOGY

3.1 Overview

Figure 2 illustrates the overview of the CDB-Net. In particular, the CDB-Net architecture is used as a feature extractor $f(\cdot)$ for images x^* , whereby $*$ = $\{f, p\}$, representing face and periocular images respectively. eliciting modality-specific feature embeddings $v^* = f(x^*; \phi) \in \mathbb{R}^d$ such that d is the dimension of the feature embedding. Here, we set $d = 1024$. These feature embeddings are appended with a Softmax predictor to elicit modality-specific losses L_* and the CDB loss L_{cdb} .

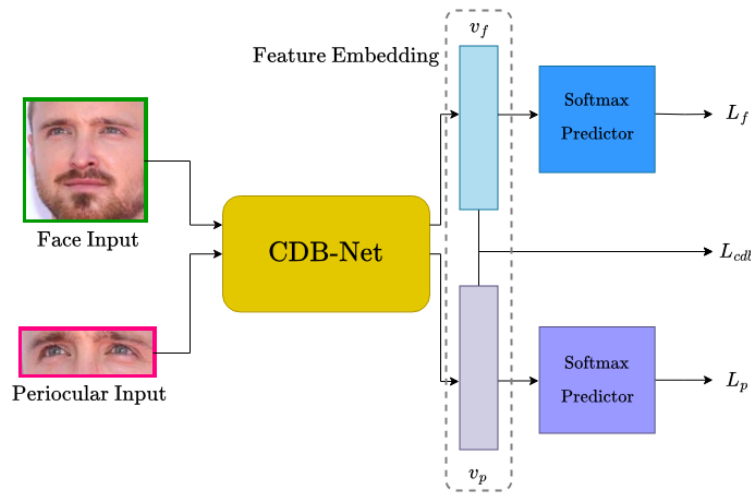


Figure 2. Overview Of CDB-Net

3.2 Network Architecture

Built upon the MobileFaceNet **Error! Reference source not found.**] architecture, we append the CDB block after the 6th layer, after the 4th bottleneck layer. In precise, the CDB block consists of two modules, whereby the first half is made up of a channel-wise multi head self-attention, while the second half is made up of the gated feed-forward network. Mathematically, given $Q = C_d^Q C_p^Q x$, $K = C_d^K C_p^K x$, $V = C_d^V C_p^V x$ whereby $C_d^{(\cdot)}$ and $C_p^{(\cdot)}$ represent the 1×1 depth-wise convolution and 3×3 point-wise convolution respectively, we first describe the channel-wise multi head self-attention as follows:

$$x = C_p \text{Attention}(Q, K, V) + x$$

$$\text{Attention}(Q, K, V) = V \cdot \text{Softmax}(K \cdot Q / \alpha)$$

Here, α is a learnable scaling parameter. Then, the gated feed-forward network is described as the following:

$$x = C_p^0 \text{Gating}(x) + x$$

$$\text{Gating}(x) = \Phi(C_d^1 C_p^1(x)) \odot (C_d^2 C_p^2(x))$$

where \odot represents element-wise multiplication while ϕ represents the GeLU gating function.

3.3 Loss Functions

Given a set of a set of N face (f) and periocular (p) images with shared identity labels $\{(x_i^*, y_i)\}$ where $i = \{1, \dots, N\}$, $x_i^* = \{* | f, p\}$ and C identities, the *Softmax predictors* can be computed as $\text{Softmax}(W^T v^*)$ and $v^* = f(x^*) \in \mathbb{R}^d$ where $W^T \in \mathbb{R}^{C \times d}$ is the prototype weight matrix for face or periocular images. We use **CosFace Error!** **Reference source not found.**] as the *Softmax predictors*.

3.3.1 Angular Margin Loss (CosFace)

Given a batch of B samples x_i^* with their corresponding identity labels y_i , the margin-based angular loss L_* is defined as follows:

$$L_* = \frac{1}{B} \sum_{i=1}^B -\log \frac{e^{s(\cos(\theta_{y_i,i})-m)}}{e^{s(\cos(\theta_{y_i,i})-m)} + \sum_i e^{s \cos \theta_i}}$$

3.3.2 CDB Loss

The CDB loss, L_{cdb} is computed as the following, whereby v_{*i} represents feature embedding of identity i , such that v_{*i}^+ and v_{*i}^- represent the positive and negative samples respectively.

$$L_{cdb} = -\frac{1}{B} \sum_{i=1}^B \log \frac{e^{\left(\frac{v_{*i}^+ v_{*i}^+}{\tau}\right)}}{e^{\left(\frac{v_{*i}^+ v_{*i}^+}{\tau}\right)} + \sum_{y_i^- \neq y_i} e^{\left(\frac{v_{*i}^+ v_{*i}^-}{\tau}\right)}}$$

Here, we set τ as the temperature hyperparameter, in which we set a value of 0.07.

3.3.3 Total Losses

Given the angular margin loss and the CDB loss functions, the total loss L is calculated as follows.

$$L = L_p + L_f + L_{cdb}$$

3.4 Inference

After training is completed, the two *Softmax predictors* are discarded. For deployment - comprising enrollment and query as shown in Figure 2—only the backbone network is used to extract the feature embedding $v^* = f(x^*) \in \mathbb{R}^d$ whereby $* = \{f, p, fp\}$ corresponds to the face (f), periocular (p) or both (fp) modalities, depending on the desired CDB configuration. This is applicable to images that are unseen to the model during training, thereby fulfilling the open-set setting during matching.

4. RESULTS AND DISCUSSIONS

4.1 Experimental Setup

The generalizability of the CDB-Net is evaluated using the rank-1 identification rate (IR) for identification tasks and the equal error rate (EER) for verification tasks, considering three different configurations: (1) intra-modal matching (periocular vs. periocular, and face vs. face), eliciting two rank-1 IR and EER values each, (2) multimodal matching (concatenated periocular and face vs. concatenated periocular and face), eliciting one rank-1 IR and EER value each, and (3) cross-modal matching (periocular vs. face, and face vs. periocular), eliciting two rank-1 IR values with a single EER value.

In the case of identification, each testing set is probed against the gallery, whereby we calculate the average rank-1 IR values obtained. In addition, we also include the Cumulative Matching Characteristic (CMC) Curve for rank-1 to rank-10 as a plot, as shown in Figures 3-5. As for verification, we sample 4 random images for each identity label from the gallery, such that a total of 12 ($= 4 \times (4 - 1)$) positive pairs and 16 ($= 4 \times 4$) negative pairs are elicited. Additionally, we include the Receiver Operating Characteristic (ROC) Curve as a plot, similarly as shown in Figure 3 to Figure 5. We proceed to describe the dataset distribution in the next section.

4.1.1 Dataset

In this section, we provide the dataset distribution for both our training and testing datasets. Our training set comprises face and periocular images sampled from the VGGFace [Error! Reference source not found.] and Ethnic [Error! Reference source not found.] datasets. In total, the training set includes 1,054 identities with 166,737 samples for each modality (periocular and face). To conduct more comprehensive experiments, we evaluate the CDB-Net on five testing datasets: Ethnic [Error! Reference source not found.], Pubfig [Error! Reference source not found.], FaceScrub [Error! Reference source not found.], IMDb Wiki [Error! Reference source not found.], and AR [Error! Reference source not found.]. It is important to note that these testing datasets are entirely separate from the training set, meaning there are no overlapping identities between the training and testing datasets.

The testing datasets are collected from the wild, which feature diverse lighting conditions, poses, and angles, except for the AR dataset, which is included because it contains additional challenging factors, such as occlusions. Ethnic covers various ethnicities, providing a robust benchmark for generalization. Pubfig is simpler compared to IMDb Wiki, while FaceScrub serves as a widely used intermediate dataset, more challenging than Pubfig but less so than IMDb Wiki. Although AR is not an in-the-wild dataset, it introduces challenges such as occlusion and blurring. Table 1 provides a summary of the data distribution across the training and testing sets.

Table 1. Distribution Summary Of Testing Data

	Ethnic	Pubfig	FaceScrub	IMDb Wiki	AR
# identities	328	200	530	2,129	100
# gallery	1,645	9,221	31,066	40,241	700
# probe 1	24,171	6,138	21,518	17,658	2,800
# probe 2	-	6,101	27,292	15,252	1,400
# probe 3	-	-	-	16,273	3,500
# probe 4	-	-	-	-	600

We apply aggressive data augmentations in our experiments for training CDB-Net. These include random in-plane rotation from -10 to 10 degrees, random scaling with a factor ranging from 1.0 to 1.2, and random horizontal flipping.

The datasets used for testing offer a broad range of conditions, including different demographics and environmental factors, which challenge the model's generalization capabilities. This diversity in testing data helps to thoroughly evaluate the robustness of CDB-Net across real-world scenarios.

4.1.2 Experiment and Hyperparameters

During the training, the CDB-Net built upon the MobileFaceNet was loaded with pre-trained MobileFaceNet [Error! Reference source not found.] weights that was trained on VGGFace2 [Error! Reference source not found.] images with a dimension of 1024. For the first 6 epochs, which we term as "pre-epochs", we freeze all the layers of MobileFaceNet with the exception of the CDB block. This allows the CDB block to attune itself to the architecture of the pre-trained weights, whereby the remaining 34 epochs (40 epochs in total) trains the whole network. Table 2 summarizes the hyperparameters used in the experiments.

Table 2. Hyperparameters Used For Training

Settings	Hyperparameters
Batch Size	24
Dropout	0.1
# Epochs	6 pre-epochs + 34 epochs
Learning Rate	0.001
Weight Decay	1e-5
Learning Rate Scheduler	[6, 18, 30]
Angular Margin (s, m)	(64.0, 0.35)
Temperature τ	0.07

4.2 Experimental Results

Following the experimental setup described in Section 4.1, we tabulate a summary of our experimental results in Table 3. In particular, we run these experiments on 3 different settings, namely the baselines, multitask learning (MTL) [Error! Reference source not found.], and the CDB-Net. In the case of the baselines, we run two modality-specific periocular and face networks, then perform the matching for these baseline networks. On the other hand, the MTL comprises a shared parameter network that accepts both periocular and face images, though we exclude the L_{cdb} loss function. Lastly, the CDB-Net configuration includes a shared-parameter network alongside L_{cdb} .

Table 3. Experimental Results Of CDB-Net With Baselines Averaged On Testing Datasets (Ethnic, Pubic, FaceScrub, IMDb Wiki, AR)

	Intra-Modal				Multimodal		Cross-Modal		
	Peri		Face		Peri + Face	Peri + Face	Peri G.	Face G.	Peri-Face
	Rank-1 IR (%)	EER (%)	Rank-1 IR (%)	EER (%)	Rank-1 IR (%)	EER (%)	Rank-1 IR (%)	Rank-1 IR (%)	EER (%)
Baselines	90.38	12.71	97.48	3.93	97.49	5.73	2.35	2.66	41.00
MTL	89.90	10.75	97.28	3.87	96.96	4.82	69.85	65.74	14.44
CDB-Net	90.50	10.01	97.35	3.86	97.13	4.56	84.84	82.03	10.75

In Table 3, notice that though the baselines have a relatively good performance in both intra-modal and multimodal matching settings, the performance for cross-modal matching shows a drastic decrease. This further justifies that both periocular and face are of different modalities, in which a considerable result would be elicited suppose that both are of the same modalities, sharing similar weights. On the other hand, the MTL exhibits a decent performance for periocular and face in terms of cross-modal matching, though it is outperformed by the baseline networks in intra-modal and multimodal matching. We credit this to the shared-parameter network that accepts both inputs, that embodies the concept of the conditional deployable biometrics (CDB), in which though the performance of cross-modal matching has increased significantly by roughly 60% in terms of rank-1 IR and 30% in terms of EER, there still lies room for improvement in the other matching settings. Therefore, the introduction of L_{cdb} as in the CDB-Net demonstrates the best improvement in terms of cross-modal matching, while minimizing the performance loss for face. In addition, the performance of periocular outperforms the baseline network. We attribute this to the CDB notion, whereby the face modality is able to guide the learning process of periocular, boosting its performance

significantly. We illustrate the CMC and ROC curves for intra-modal, multimodal, and cross-modal matching in Figures 3, 4, and 5 respectively.

In Figure 3, notice that the periocular performance for CDB-Net outperforms greatly the MTL and baseline networks. Though it can be said that the performance of face is negligibly lower, we treat this as a necessary trade-off for performance balance. This case is similar in Figure 4 that illustrates the multimodal matching setting, whereby it is shown that the CDB-Net has a slight performance degrade in comparison with the baseline network. It is believed that in order to significantly improve the performance of periocular, there needs to be a slight decrease in the performance of face. Considering that the multimodal setting considers both modalities, we attribute this slight performance degrade to the face modality. With reference to Table 3, though there is a performance degrade of 0.3% in terms of rank-1 IR, the performance improvement of 1.1% in terms of EER was able to make up for its degradation.

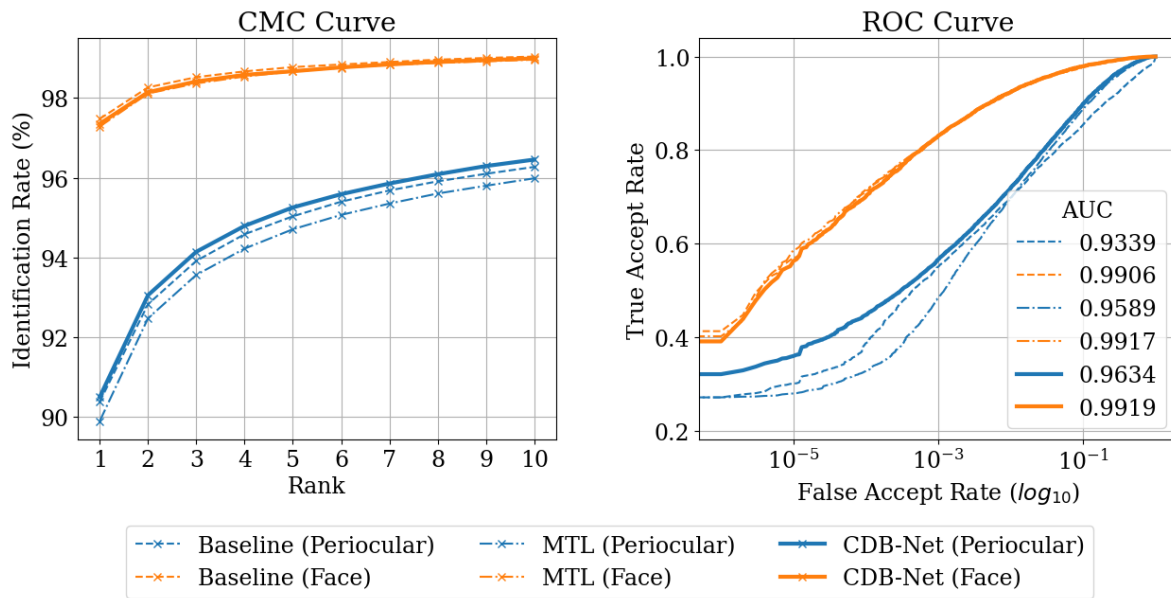


Figure 3. Cumulative Matching Characteristic (CMC) And Receiver Operating Characteristic (ROC) Curves For Intra-Modal Matching Averaged On Testing Datasets (Ethnic, Pubfic, FaceScrub, IMDb Wiki, AR)

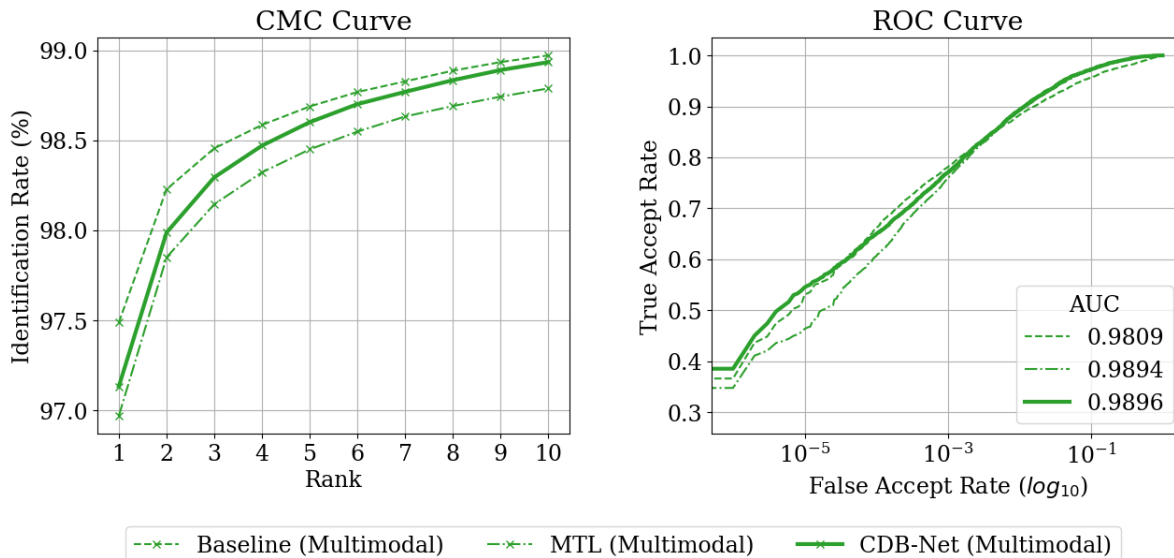


Figure 4. Cumulative Matching Characteristic (CMC) And Receiver Operating Characteristic (ROC) Curves For Multimodal Matching Averaged On Testing Datasets (Ethnic, Pubfic, FaceScrub, IMDb Wiki, AR)

Lastly, Figure 5 shows the cross-modal matching performances. As expected, it can be seen that the CDB-Net performs the best compared to the MTL and baseline settings. This exhibits a balanced performance of intra-modal, multimodal, and cross-modal matching of the CDB-Net compared to the other networks, thus fulfilling the CDB notion. In particular, enabling cross-modal matching allows a flexible face and periocular deployment in which either modalities can be enrolled and queried be it in a single modality or multiple modality case, similar to a combination of the work in **Error! Reference source not found.**] and **[Error! Reference source not found.]**.

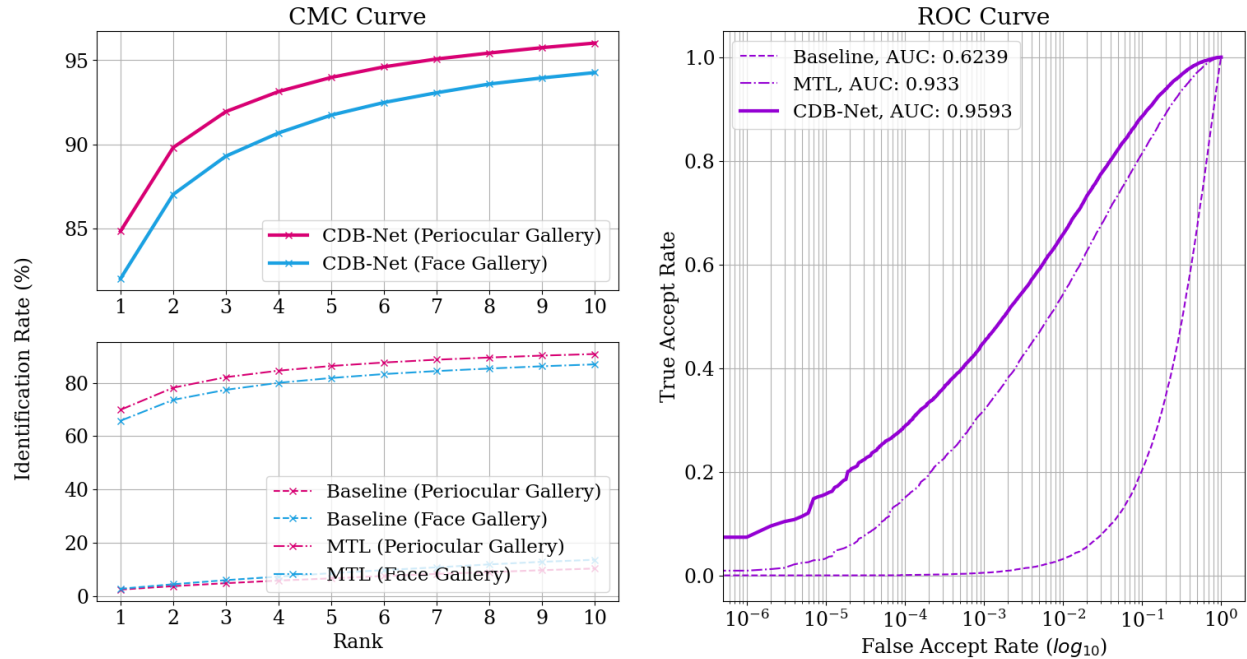


Figure 5. Cumulative Matching Characteristic (CMC) And Receiver Operating Characteristic (ROC) Curves For Cross-Modal Matching Averaged On Testing Datasets (Ethnic, Pubfic, FaceScrub, IMDb Wiki, AR)

5. CONCLUSION

In this paper, we presented the notion of the Conditional Deployable Biometrics (CDB), whereby we achieve a balanced performance for intra-modal, multimodal, and cross-modal matching settings. The notion of the CDB enables a deployable biometric authentication system that is flexible, in which matching authentication can be carried out regardless of the biometric modality. In particular, we realize the CDB regimen via the CDB-Net, a deep neural network that is designed for periocular and face modalities. Through our benchmark on five in-the-wild datasets that are deemed to be challenging, we demonstrate the effectiveness of the CDB-Net in realizing the CDB notion, whereby we observe a drastic performance improvement of the CDB-Net in comparison with the baseline networks. Some possible extensions to our work in the future may include consideration of other biometric modalities such as iris and fingerprint for a seamless biometric authentication system that is non-repudiable.

ACKNOWLEDGEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Jihyeon Kim: Conceptualization, Methodology and Experiments, Writing – Original Draft Preparation and Editing;
Tiong-Sik Ng: Experiments, Validation, Writing – Review & Editing;
Andrew Beng Jin Teoh: Project Administration, Supervision, Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS




Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. <https://publicationethics.org/>

REFERENCES

- [1] A. K. Jain, D. Deb, and J. J. Engelsma, "Biometrics: Trust, but verify," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 3, pp. 303-323, 2021, doi: 10.1109/TBIOM.2021.3115465.
- [2] P. Kumari and K. R. Seeja, "A novel periocular biometrics solution for authentication during COVID-19 pandemic situation," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 10321-10337, 2021, doi: 10.1007/s12652-020-02814-1.
- [3] J. Xu, M. Cha, J. L. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides, "Robust local binary pattern feature sets for periocular biometric identification," *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pp. 1-8, 2010, doi: 10.1109/BTAS.2010.5634504.
- [4] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a back-propagation network," *Advances in Neural Information Processing Systems*, vol. 2, 1989.
- [5] T. S. Ng, C. Y. Low, J. C. L. Chai, and A. B. J. Teoh, "Conditional multimodal biometrics embedding learning for periocular and face in the wild," *International Conference on Pattern Recognition (ICPR)*, pp. 812-818, 2022, doi: 10.1109/ICPR56361.2022.9956636.
- [6] S. Chen, Y. Liu, X. Gao, and Z. Han, "Mobilefacenets: Efficient CNNs for accurate real-time face verification on mobile devices," *Chinese Conference on Biometric Recognition*, pp. 428-438, 2018, doi: 10.1007/978-3-319-97909-0_46.
- [7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997, doi: 10.1109/34.598228.
- [8] H. Wang et al., "Cosface: Large margin cosine loss for deep face recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5265-5274, 2018, doi: 10.48550/arXiv.1801.09414.
- [9] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4690-4699, 2019. doi: 10.1109/CVPR.2019.00482.
- [10] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008. [Online]. Available: <https://inria.hal.science/inria-00321923/>
- [11] T. Zheng and W. Deng, "Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments," *Beijing University of Posts and Telecommunications Technical Report*, vol. 5, no. 7, 2018. [Online]. Available: <http://www.whdeng.cn/CPLFW/Cross-Pose-LFW.pdf>

- [12] R. Jillela and A. Ross, "Mitigating effects of plastic surgery: Fusing face and ocular biometrics," *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pp. 402-411, 2012, doi: 10.1109/BTAS.2012.6374607.
- [13] M. Karakaya, "Iris-ocular-periocular: Toward more accurate biometrics for off-angle images," *Journal of Electronic Imaging*, vol. 30, no. 3, pp. 033035-033035, 2021, doi: 10.1117/1.JEI.30.3.033035.
- [14] L. C. O. Tiong, S. T. Kim, and Y. M. Ro, "Implementation of multimodal biometric recognition via multi-feature deep learning networks and feature fusion," *Multimedia Tools and Applications*, vol. 78, pp. 22743-22772, 2019, doi: 10.1007/s11042-019-7618-0.
- [15] Y. G. Jung, C. Y. Low, J. Park, and A. B. J. Teoh, "Periocular recognition in the wild with generalized label smoothing regularization," *IEEE Signal Processing Letters*, vol. 27, pp. 1455-1459, 2020, doi: 10.1109/LSP.2020.3014472.
- [16] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215-244, 2021, doi: 10.1016/j.neucom.2020.10.081.
- [17] R. Sharma and A. Ross, "Periocular biometrics and its relevance to partially masked faces: A survey," *Computer Vision and Image Understanding*, vol. 226, p. 103583, 2023, doi: 10.1016/j.cviu.2022.103583.
- [18] L. C. O. Tiong, D. Sigmund, and A. B. J. Teoh, "Face-periocular cross-identification via contrastive hybrid attention vision transformer," *IEEE Signal Processing Letters*, vol. 30, pp. 254-258, 2023, doi: 10.1109/LSP.2023.3256320.
- [19] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020, doi: 10.48550/arXiv.2010.11929.
- [20] A. Vaswani et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017. [Online]. Available: <https://user.phil.hhu.de/~cwurm/wp-content/uploads/2020/01/7181-attention-is-all-you-need.pdf>
- [21] C. Szegedy et al., "Rethinking the inception architecture for computer vision," *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.
- [22] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," *British Machine Vision Conference*, 2015.
- [23] L. C. O. Tiong, A. B. J. Teoh, and Y. Lee, "Periocular recognition in the wild with orthogonal combination of local binary coded pattern in dual-stream convolutional neural network," *2019 Int. Conf. Biometrics (ICB)*, pp. 1-6, 2019, doi: 10.1109/ICB45273.2019.8987278.
- [24] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," *IEEE International Conference on Computer Vision*, pp. 365-372, 2009, doi: 10.1109/ICCV.2009.5459250.
- [25] H. W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," *IEEE International Conference on Image Processing*, pp. 343-347, 2014, doi: 10.1109/ICIP.2014.7025068.
- [26] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," *International Conference on Computer Vision Workshops*, 2015, pp. 252-257, doi: 10.1109/ICCVW.2015.41.
- [27] A. Martinez and R. Benavente, "The AR face database", *CVC technical report*, vol. 24, 1998.
- [28] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 67-74, 2018, doi: 10.1109/FG.2018.00020.
- [29] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 12, pp. 5586-5609, 2021, doi: 10.1109/TKDE.2021.3070203.

BIOGRAPHIES OF AUTHORS

	<p>Jihyeon Kim earned her Ph.D. in 2024 from the School of Electrical and Electronics Engineering at Yonsei University, South Korea. Her research focuses on Biometric Template Protection using Adversarial Neural Networks and Security Analysis. She can be reached at kim_jihyeon@yonsei.ac.kr.</p>
	<p>Tiong-Sik Ng received his BEng (Electronics Majoring in Computer) and MSc (Information Technology) from Multimedia University, Malaysia in 2016 and 2019 respectively. He is currently pursuing a Ph.D. degree in Electrical and Electronic Engineering in Yonsei University, South Korea. His research interests include biometric security and deep learning. He can be contacted at email: ngtionsik@yonsei.ac.kr.</p>
	<p>Andrew Beng Jin Teoh obtained his BEng (Electronic) in 1999 and a Ph.D. degree in 2003 from the National University of Malaysia. He is currently a full professor in the Electrical and Electronic Engineering Department, College Engineering of Yonsei University, South Korea. His research for which he has received funding focuses on biometric applications and biometric security. His current research interests are Machine Learning and Information Security. He has published more than 350 international refereed journal papers, conference articles, edited several book chapters, and edited book volumes. He served and is serving as a guest editor of the IEEE Signal Processing Magazine, associate editor of IEEE Transaction of Information Forensic and Security, IEEE Biometrics Compendium and Machine Learning with Applications, Elsevier. He can be contacted at email: bjteoh@yonsei.ac.kr.</p>