ZEROFACE: UNIFIED PRIVACY-PRESERVING FRAMEWORK FOR SECURE FACE VERIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Face recognition (FR) is widely used in authentication and surveillance applications due to its high accuracy for identity verification and recognition. However, there are potential privacy risks existing within FR systems at different stages, particularly when sensitive biometric data can be accessed without authorization. Existing countermeasures focus on protecting users' privacy by concealing identity information that underlies the facial embedding. However, recent generative privacy attacks are able to reconstruct high-quality and identity-preserving face images from facial embeddings. In this work, we propose a secure and efficient framework, ZeroFace, for embedding protection. Firstly, a compression network is proposed to reduce the embedding dimensionality by mapping it into low-dimensional representations, while maintaining nearly lossless recognition accuracy via learning the multi-scale identity information. Furthermore, identity-specific transformation is proposed to further obfuscate the compressed embedding. Lastly, a secure matching protocol is designed to enhance embedding security by secretly computing cosine distance without sharing the user's compressed embedding with the database holder. Extensive experiments show ZeroFace largely maintains recognition accuracy and enhances the embedding robustness compared to the state-of-the-art face template protection algorithms.

1 INTRODUCTION

Facial recognition (FR), characterized by its non-intrusive nature, convenience, and growing accuracy, has been extensively utilized in diverse domains including identity verification, security surveillance, mobile payments, and human-computer interaction, significantly altering various aspects of social life. FR systems generally do not store users' original facial images directly; rather, they utilize intricate algorithms (e.g., deep neural networks (DNNs)) to derive compact and distinctive digital feature vectors, referred to as face templates or face embeddings. These embeddings constitute the fundamental basis for subsequent identity comparison and verification. However, these highly concentrated digital identity indicators can present significant risks to users' personal privacy when compromised or maliciously utilized. For example, given a face embedding, it is possible to extract soft-biometric information such as age, gender, demographics and social traits as reported by Terhörst et al. (2021).

Moreover, researches have shown that attackers can reconstruct highly similar images to the original faces from leaked embeddings using specific algorithms such as template inversion (TI) attacks (Mai et al., 2018; Shahreza & Marcel, 2024; Otroshi Shahreza & Marcel, 2024; Zhang et al., 2024; Shahreza et al., 2025; Han et al., 2026). These reconstructed high-quality facial images can not only be used to impersonate legitimate users to gain unauthorized access to sensitive systems or services (e.g., unlocking a phone, accessing a bank account, or entering restricted areas), but they may also lead to serious privacy breaches. For example, reconstructed facial images can be used to infer an individual's identity, health status, or even certain sensitive attributes.

To protect biometric data, biometric template protection (BTP) has been widely studied. The core idea of BTP is to encrypt or transform the original template to generate a secure template representation that can be used for matching while reducing the risk of template leakage. Existing BTP methods can be broadly categorized into three groups: crypto-based methods, hash-based methods and transform-based methods. Transform-based methods offer a good trade-off between security

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

and utility. However, they rely on some pre-defined parameters or secret keys for transformation and verification. Raw embeddings might be revealed once such information is compromised. Moreover, transform-based methods are still vulnerable to the recent privacy attacks Zhang et al. (2024); Shahreza et al. (2025); Han et al. (2026) that based on generative models such as generative adversarial networks (GAN) and diffusion models (DM) to exploit protected embeddings or similarity scores to reconstruct face images with recognizable identity information.

To address the security and privacy issues in existing FR systems, we introduce a unified framework to effectively improve the robustness to various attacks while preserving recognition accuracy. Our key observation is that existing GAN-based and DM-based inversion attacks rely on recovering facial information from high-dimensional embedding distributions. By compressing face embeddings into compact low-dimensional representations, we can obfuscate the original distribution and reduce the information available for identity reconstruction. Inspired by this observation, we propose an adaptive compression network to transform face embeddings into low-dimensional representations for distribution obfuscation. However, compression alone is insufficient under a strong adversary with extensive query access to the target FR system. Therefore, we further introduce identity-specific transformation and secure distance computation protocol to obtain authentication results without disclosing protected embeddings. To comprehensively evaluate the effectiveness of the proposed method against privacy attacks, we consider multiple attack scenarios with different attacker locations and knowledge assumptions, as detailed in in Sec 3.1.

In summary, our contributions are as follows: **(1) Recognition Preservation:** ZeroFace maintains recognition accuracy close to that of the original FR system; **(2) Efficient Compression:** ZeroFace reduces embedding dimensionality, storage, and verification cost while retaining identity-discriminative information; **(3) Privacy Protection:** ZeroFace combines representation compression, identity-specific obfuscation, and secure distance computation to protect embeddings among differet parties; and **(4) Fast Adaptation:** the framework requires no modification or retraining of the target FR backbone and supports efficient adaptation to different recognition systems.

2 Related Work

2.1 Deep Face Embedding Inversion

TI attacks aim to reconstruct face image from leaked face embeddings that stored in database of a FR system. In general, CNN-based attacks Shahreza et al. (2022); Shahreza & Marcel (2023) rely on training reconstruction networks end to end on paired face images and embeddings by minimizing an identity loss. By leveraging GAN models, FaceTI Otroshi Shahreza & Marcel (2024) trains a mapping network to map face embeddings into latent spaces of StyleGAN3 Karras et al. (2021) through adversarial training. To consider limited query access to FR system, MAP2V proposes a training-free and query-efficient method based on StyleGAN2 Karras et al. (2020) to reconstruct face images based on only cosine distance by using initial latent minimizing and zeroth-order gradient estimation. Recently, diffusion-based methods Shahreza et al. (2025); Han et al. (2026) leverage the high-realistic face generation ability of face foundation models (Arc2Face Paraperas Papantoniou et al. (2024), IPA-FaceID Ye et al. (2023)) to reconstruct high-quality face image by training a lightweight adapter or mapping module.

2.2 Face Embedding Protection

BTP methods can be grouped into three main categories: crypto-based, hash-based, and transform-based methods. Homomorphic encryption (HE) Gentry (2009) is widely used in crypto-based methods Boddeti (2018), which encrypt embeddings with user-specific keys and perform matching directly on encrypted templates. However, their computational and storage overhead limits practical deployment. Recently, IDFace Kim et al. (2025) improves encrypted face identification through almost-isometric transformation and space-efficient encoding but still introduces ciphertext storage and communication overhead. Hash-based methods (such as locality-sensitive hashing (LSH) Dang et al. (2020), fuzzy commitment scheme (FCS) Teoh & Kim (2007), fuzzy vault Dong et al. (2021), etc.) use hash mapping functions to map face embeddings to irreversible hash values. As small variations on input face images can cause completely different hashes, hash-based methods often suffer noticeable degradation on recognition accuracy. IronMask Kim et al. (2021) is a modular face

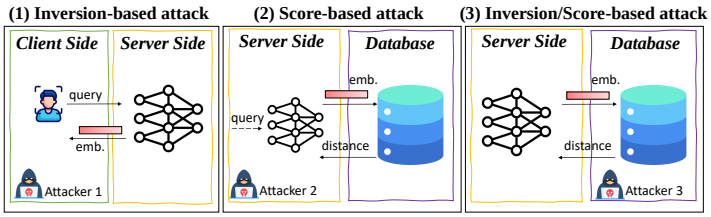


Figure 1: Attack scenarios. Attacker can be either client, server or database. Emb: embedding.

template protection architecture, which can be integrated with any FR system that uses angular distance metrics. Similarly, ASE Dusmanu et al. (2021) applies rotations and affine transformations to map embeddings into subspaces for protection. Transform-based methods protect embeddings using fixed or learnable transformations. PolyProtect Hahn & Marcel (2022) applies user-specific multivariate polynomial mappings, while MLP-Hash Shahreza et al. (2023) maps embeddings through an MLP followed by binarization. SlerpFace Zhong et al. (2025) transforms embeddings through learnable feature grouping and spherical interpolation to reduce inversion risk, but requires retraining the target FR model. Different from BTPs, privacy-preserving face recognition (PPFR) systems Mi et al. (2022); Ji et al. (2022); Mi et al. (2023; 2024) directly protect facial privacy of input face images during client-server communication rather than protecting face embeddings. We further evaluate the applicability of ZeroFace in PPFR systems in Sec. 5.

3 ZeroFace

This section defines the threat model and presents the technical design of ZeroFace.

3.1 Threat Model

We introduce threat models in different scenarios (see Figure 1) based on the location of attacker within the target FR system that involving three parties: client, server and database. **For simplicity, we consider there only one party can be the attacker at the same time.**

Attacker's goal. For a target unprotected FR system, the attacker aims to reconstruct a complete face image from a leaked face embedding or the system decision.

Attacker's knowledge and ability. We consider three attack cases according to the adversary's location and accessible information. In **case (1)**, a client-side attacker can query the target FR system to obtain the face embedding from the server side. Therefore, attacker can performs an *inversion-based attack* Shahreza et al. (2025); Han et al. (2026) to reconstruct the corresponding face image. In **case (2)**, a server-side attacker submits query embeddings to the database and observes their authentication scores to a target enrolled embedding. The attacker uses this feedback to iteratively optimize a candidate face image through a *score-based attack* Zhang et al. (2024). In **case (3)**, a database-side attacker can access an incoming query embedding and its matching scores against enrolled embeddings. The attacker can exploit the query embedding for an *inversion-based attack* or the matching scores for a *score-based attack* to reconstruct the face corresponding to the query.

3.2 Adaptive Matryoshka Compression

Curse of Dimension in Template Inversion. Recent TI attacks rely on learning a mapping between face image and embedding distribution from a target FR model using generative models. The encoded distinguished information of embedding is the critical prior for a successful reconstruction. Different dimensions in embedding capture different facial features. By only sharing partial of embedding, it can reduce the risk of potential privacy leakage. However, removing dimensions likely reduces discriminative power and hinders embedding utility, see in Appendix F.

Multi-scale Compression. As aforementioned, naive embedding dimension pruning can negatively affect its utility. Inspired by MRL Kusupati et al. (2022), we propose an *Adaptive Matryoshka Compression* (AMC) network to reduce dimension of face embeddings while maintain high usability for

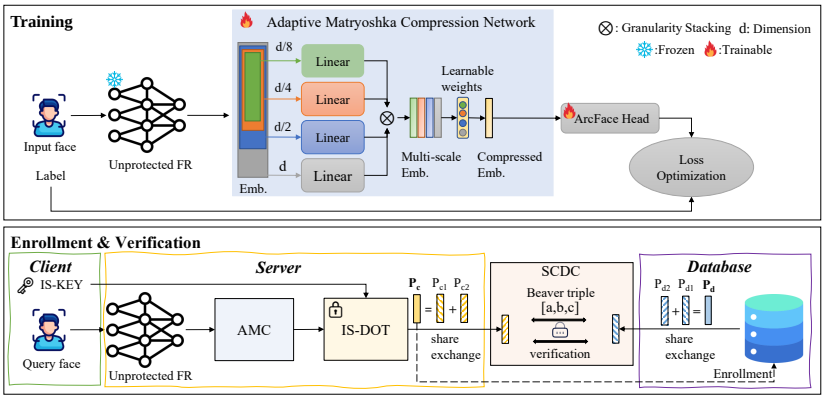


Figure 2: Overview of ZeroFace: (a) Multi-scale compression: During training, face embeddings are compressed from multi-granularity representations by jointly learning the AMC network with the classifier head. (b) Identity-specific protection: Each compressed embedding is transformed by IS-DOT with a unique key (IS-KEY). (c) Secure matching: Protected embeddings are further protected by sharing only the random split representations in verification phase via a designed SCDC protocol.

verification. MRL is originally proposed for learning a low-dimensional representation of rich information from deep representations to reduce computational costs for large-scale classification and retrieval tasks. Not all facial attributes encoded in an embedding contribute to identity discrimination Leroy et al. (2025). Moreover, locating identity-specific features is non-trivial because FR systems differ in their network architectures, training datasets, and loss functions Ren et al. (2024). To learn an approximation in low-dimensional representation from a face embedding e extracted from face image by a FR model $\Gamma(\cdot)$, AMC network leverages MRL for multi-scale information preservation (Figure 2 (a)). For a given face embedding vector e, it can be split into different slices with d dimensions $e[:d]$. Then each slice is transformed independently through a linear layer to capture multi-granularity feature $Linear(e[:d])$. Instead of aggregating losses on each feature granularity for training, we set the same output dimension for each linear layer and process it with granularity attention to fuse all outputs into single granularity dimension.

$$e' = \sum_{i=0} \alpha_i \cdot (Linear(e[:d]))_i \,, \tag{1}$$

where α_i is learnable fusion weight for i_{th} granularity.

To encourage compressed embeddings e' preserve distinguishable information of each identity, we utilize ArcFace loss Deng et al. (2019) for training. Hence, AMC is trained to effectively extract and compress identity information from original embedding into a low-dimensional representation through fusing granularity attention by learnable weights. For performance of different output dimensions, see in Sec. 5.

3.3 Identity-Specific DCPE-Orthogonal Transformation

Although AMC produces a sparse representation, it may still reveal the identity or soft-biometric information. This motivates an additional protection mechanism that obscures the retained representation while limiting distortion. We therefore propose the *Identity-Specific DCPE-Orthogonal Transformation* (IS-DOT), which combines bounded DCPE Fuchsbauer et al. (2022) perturbation with an identity-specific orthogonal transformation. The perturbation randomizes each compressed representation, while the orthogonal transformation maps embeddings into an identity-specific coordinate system. For each identity u, we generate an identity-specific protection key

$$\Theta_u = (K_u^{\mathrm{dcpe}}, s_u, \xi_u), \tag{2}$$

where K_u^{dcpe} is a secret pseudorandom function (PRF) Bellare et al. (1996) key, $s_u > 0$ is a secret scaling factor, and ξ_u is the seed used to generate the identity-specific orthogonal matrix R_u.

Given a compressed embedding $e'_i \in \mathbb{R}^d$ to be protected under identity u, a fresh public 128-bit nonce $\eta_i^{(u)}$ is sampled and combined with K_u^{dcpe} to generate a bounded perturbation $\lambda_i^{(u)}$ satisfying $\|\lambda_i^{(u)}\|_2 \leq s_u\beta/4$. β is the approximation parameter of DCPE that controls the perturbation magnitude. Following Fuchsbauer et al. (2022), the intermediate representation is

$$\bar{e}_i^{(u)} = s_u e'_i + \lambda_i^{(u)}. \tag{3}$$

The final protected embedding is obtained by

$$z_i^{(u)} = R_u \bar{e}_i^{(u)}, \tag{4}$$

where $R_u^\top R_u = I$. The same $(K_u^{\text{dcpe}}, s_u, R_u)$ is used for all embeddings protected under the same claimed identity u, while a fresh nonce produces an independent bounded perturbation for each protection operation. Thus, repeated protection of the same embedding is randomized rather than deterministic. The complete generation procedure is provided in Appendix C.

During 1:1 verification, both enrolled and query embeddings are protected by the key associated with the claimed identity u. This rule is applied to both *genuine* and *impostor* comparisons. IS-DOT preserves the angular geometry of the DCPE-transformed embeddings up to the bounded perturbation.

3.4 Secure Cosine Distance Computation

Although IS-DOT protects the compressed representation, exposure of the complete protected embedding may still enable inversion-based attack, especially when compromised database can access the query embedding (see **case 3** in Figure 1). To mitigate this residual risk, we propose a *Secure Cosine Distance Computation* (SCDC) protocol based on Secure Multiparty Computation (SMPC) Beaver (1991); Xia et al. (2021) and secret sharing (SS) Beimel (2011), which enables similarity computation without directly disclosing the complete protected embeddings. We consider a protected query embedding P_c provided by the client and an enrolled embedding P_d stored in the database, as illustrated in Figure 3. Before SS, the protected embeddings are ℓ_2-normalized such that $\|P_c\|_2 = \|P_d\|_2 = 1$. Therefore, their cosine similarity reduces to the dot product $\text{Cos}(P_c, P_d) = P_c^\top P_d$, and SCDC only needs to securely compute the dot product. During enrollment, the database samples a random vector P_{d1} and computes the complementary share $P_{d2} = P_d - P_{d1}$. The database retains P_{d2} and sends P_{d1} to the server. During each verification session, the server samples a fresh random vector P_{c2} and computes $P_{c1} = P_c - P_{c2}$. The server retains P_{c1} and sends P_{c2} to the database. To securely compute the dot product, the server and database execute the secure multiplication protocol using preprocessed Beaver triples. Specifically, the parties use a fresh Beaver triple for every comparison and exchange only the masked values (ϵ_1, δ_1) and (ϵ_2, δ_2). Then they both can reconstruct the public masked differences (ϵ, δ). For each comparison, the protocol produces additive shares $[Y_1]$ and $[Y_2]$ of the element-wise product satisfying $[Y_1] + [Y_2] = P_c \odot P_d$. The protected dot product can then be obtained by summing the corresponding elements:

$$P_c^\top P_d = \sum_j \left([Y_1]_j + [Y_2]_j\right). \tag{5}$$

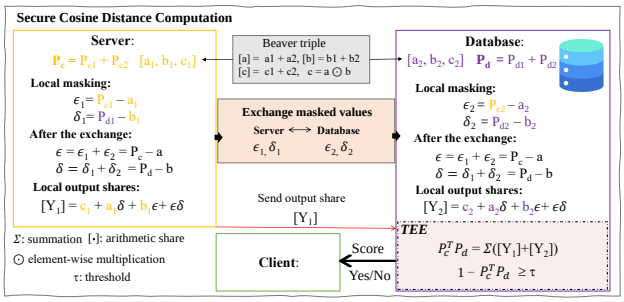


Figure 3: Details of the SCDC protocol. The TEE serves as an optional module for additional security.

The detailed Beaver-triple generation and secure multiplication procedure are provided in Appendix D. In the basic SCDC setting, the securely computed similarity score can be reconstructed for verification. SCDC further supports an optional trusted execution environment (TEE), where the similarity is reconstructed and compared with the verification threshold τ inside the TEE. In this setting, only the final matching decision is released, while the exact similarity score remains hidden from the participating parties outside the TEE, preventing its direct use as feedback for score-based attacks, including those launched by a compromised database.

Table 1: Comparison of recognition accuracy. ArcFace and Boddeti are reported as reference baselines. **Bold** indicates the best, while underlined denotes the second best. We report performance of our method using 128-dimensional embeddings.

Method	LFW	CFP-FP	AgeDB	CALFW	CPLFW	IJB-B	IJB-C
ArcFace	99.73	98.00	97.87	95.92	92.50	93.93	95.52
Boddeti	99.73	97.86	97.81	95.84	92.41	93.88	95.47
IronMask	84.42	52.70	53.22	50.00	50.00	N/A	N/A
ASE	98.77	86.80	88.48	84.12	83.42	0.00	0.00
MLP-Hash	98.82	91.40	93.67	93.07	87.12	N/A	N/A
PolyProtect	99.30	94.00	95.28	94.77	89.22	87.37	89.96
SlerpFace	99.42	92.79	95.70	94.82	88.90	89.96	92.25
Ours (w/o IS-DOT)	**99.73**	97.41	97.68	**95.95**	92.02	**92.43**	**94.47**
Ours	**99.73**	**97.51**	**97.72**	95.77	**92.08**	92.37	94.31

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Baselines. We compare proposed method with the unprotected ArcFace baseline and six BTP methods: the crypto-based Boddeti Boddeti (2018); hash-based IronMask Kim et al. (2021) and ASE Dusmanu et al. (2021); and transform-based MLP-Hash Shahreza et al. (2023), PolyProtect Hahn & Marcel (2022), and SlerpFace Zhong et al. (2025).

Utility Evaluation. Recognition performance is measured on 5 regular-size datasets: LFW Huang & Learned-Miller (2014), CFP-FP Sengupta et al. (2016), AgeDB Moschoglou et al. (2017), CPLFW Zheng & Deng (2018), and CALFW Zheng et al. (2017), and 2 large-scale benchmarks: IJB-B Whitelam et al. (2017) and IJB-C Maze et al. (2018). We conduct face verification experiment by comparing if two embeddings refer to the same person. We report 1:1 verification accuracy on the first five datasets and TAR@FAR=10^{-4} on IJB-B/C.

Robustness Evaluation. We evaluate privacy robustness against the inversion-based FEM Han et al. (2026) and score-based MAP2V Zhang et al. (2024) attacks. FEM-MLP is trained against each BTP method for 20 epochs on FFHQ Karras et al. (2019), following the official training settings. We employ MAP2V with its official default configuration. Attack success rate (ASR) is measured as the percentage of reconstructed faces accepted by independent FR models at FAR=10^{-2}. We evaluate the attacks on 1000 images (1000 unique identities) from CelebA-HQ Karras (2017).

Implementation Details. We employ an ArcFace IR-50 as target FR model to be protected, which is pre-trained on the MS1Mv2 Guo et al. (2016) and produces 512-dimensional face embeddings. Then we train AMC network with 4 levels of granularity setting (64, 128, 256, and 512) on VGGFace2 Cao et al. (2018) for 10 epochs while the whole target FR is kept frozen during training (takes around 32min per epoch), see AMC details in Appendix B. AMC training is conducted on 2 Tesla V100 GPU with 32G memory using PyTorch, with batch size 512. For optimizer, we use SGD with initial learning rate as 10^{-3}. For ArcFace loss, we set *scale* $s = 64$ and *margin* $m = 0.3$. During inference, β is set to 1 in IS-DOT. The implementation of SMPC protocol of SCDC is based on CrypTen Knott et al. (2021). Other implementation details are provided in Appendix.

4.2 RECOGNITION PERFORMANCE

As shown in Table 1, the FHE-based method of Boddeti achieves near-baseline recognition performance, while most hash-based and transform-based methods introduce larger accuracy degradation. IronMask and ASE exhibit inadequate performance owing to the low tolerance of their hash-based methodologies for intra-identity facial attribute variations. SlerpFace outperforms other transform-based methods across the majority of datasets. Nonetheless, it marginally lags behind PolyProtect on CFP-FP and CPLFW, which emphasize facial pose variations. Results for IronMask and MLP-Hash on IJB-B/C are "N/A" due to protocol incompatibility and prohibitive computational cost. Despite compressing the original 512-dimensional embedding to 128 dimensions, AMC preserves competitive recognition accuracy close to the ArcFace baseline, demonstrating that it preserves identity-discriminative information under substantial dimensionality reduction. After applying IS-DOT, it achieves recognition performance nearly identical to AMC, confirming that IS-DOT introduces protected-space obfuscation without greatly affecting verification utility.

4.3 Privacy Protection Against Privacy Attacks

Robustness to Facial Reconstruction Attacks. We evaluate model robustness against two TI attacks under the threat scenarios defined in Sec. 3.1. For all the BTP baselines, the attacker accesses the protected embeddings or similarity scores exposed by matching interface. As shown in Table 2, unprotected embeddings are highly vulnerable to reconstruction, reaching 92.6% and 99.3% inversion ASR on E.Face and M.Face, respectively. Existing BTP methods reduce reconstruction but remain vulnerable under at least one attack setting. M.Face is a lightweight FR model designed for mobile devices, whereas E.Face serves as a stronger, high-performance FR evaluator. Accordingly, M.Face evaluates the transferability of reconstructed identity cues to a compact mobile-oriented model, while E.Face provides a more demanding evaluation of whether the reconstructed faces retain sufficient identity information for successful recognition. Notably, AMC+IS-DOT substantially suppresses inversion-based attacks, but remains vulnerable to score-based reconstruction on M.Face with an ASR of 42.3%, highlighting the residual risk of exposing matching scores. Under a single-party compromise in which the adversary obtains only one additive share of the target protected embedding, SCDC further reduces ASR to at most 0.1% across all evaluated settings.

Table 2: Robustness to TI attacks. ASR (%) is evaluated on MobileFace Chen et al. (2018) (M.Face) and ElasticFace Boutros et al. (2022) (E.Face).

Method	Inversion-based ↓		Score-based ↓	
	E.Face	M.Face	E.Face	M.Face
Unprotected	92.6	99.3	87.3	98.1
MLP-Hash	54.8	80.4	11.6	78.6
PolyProtect	16.2	48.3	75.8	88.1
SlerpFace	7.5	78.6	17.0	74.0
Ours (w/o SCDC)	0.4	0.5	0.6	42.3
Ours	**0.0**	**0.1**	**0.0**	**0.0**

Robustness to Facial Attributes Attacks. In our work, we also consider the risk of soft-biometric information leakage from protected embeddings. We train MLP classifiers on protected embeddings for each attribute, e.g., gender (2 classes), age (9 classes), and race (7 classes) on the FairFace dataset Karkkainen & Joo (2021) (86744 images for training and 10954 images for evaluation). The detailed attribute categories are provided in Appendix E. Unprotected embeddings are extremely susceptible to attribute attacks, as seen in Table 3. AMC alone does not effectively reduce the attribute inference. In contrast, with IS-DOT applied, our method reduces the corresponding accuracies to 50.57%, 27.24%, and 15.70%. In specific, gender and race prediction approach the random-guessing baselines of 50.0% and 14.3%, respectively, while age prediction is also substantially reduced from the unprotected baseline. These results indicate that our method substantially suppresses soft-biometric information even without the additional protection provided by SCDC.

Table 3: Top-1 accuracy (%) of facial attribute attack on the FairFace dataset. A separate MLP classifier is trained for each attribute. Lower is better.

Method	Gender ↓	Age ↓	Race ↓
Unprotected	89.40	52.89	60.28
MLP-Hash	85.81	48.22	51.55
PolyProtect	88.21	<u>46.77</u>	55.85
SlerpFace	87.45	50.96	59.72
Ours (AMC only)	<u>82.40</u>	50.95	<u>42.35</u>
Ours (w/o SCDC)	**50.57**	**27.24**	**15.70**

4.4 Efficiency Analysis

We evaluate the computational and storage overhead of 1:1 face verification. All runtime experiments are repeated 10 times on CPU and averaged. For FHE-based Boddeti baseline, we follow the official implementation using the FV Fan & Vercauteren (2012) scheme with 512-dimensional face embeddings. As shown in Table 4, our method requires 2.96 ms per comparison. This includes AMC compression (0.12 ms), DCPE transformation (0.0335 ms), orthogonal transformation (0.0046 ms), and SCDC matching (2.80 ms). Unlike FHE-based matching, SCDC operates on additive shares and avoids ciphertext expansion, resulting in substantially lower computational overhead than Boddeti while remaining competitive with lightweight transform-based methods. For one million enrolled embeddings stored in *float64*, the protected database requires approximately 0.95 GiB, compared with 3.81 GiB for 512-dimensional plaintext embeddings and 5.84 GiB for SlerpFace. Our solution uses 128-dimensional AMC representation to minimize storage by 75% compared to plaintext, while retaining secure verification.

Table 4: Time analysis and storage cost comparison for 1 M enrolled identities in database.

Method	Time ↓	Size ↓
Plaintext	0.01 ms	3.81 GiB
Boddeti	14.13 ms	24 GiB
MLP-Hash	34.78 ms	**0.48** GiB
PolyProtect	2.93 ms	3.79 GiB
SlerpFace	**0.36** ms	5.84 GiB
Ours	2.96 ms	0.95 GiB

4.5 Security Analysis

We assess the security based on established ISO/IEC criteria for BTP: *irreversibility*, *revocability*, and *unlinkability*.

Irreversibility. We consider the worst-case scenario in which the attacker knows the compression algorithm and obtains compressed embeddings. Following Shahreza et al. (2023), we measure irreversibility using Success Attack Rate (SAR), i.e., whether inverted embeddings can access the unprotected FR system at FAR=10^{-3}. We consider embeddings protected only by AMC and use a numerical solver with 10 iterations to approximate the original embedding. On LFW and AgeDB, the attack results in **0%** SAR, indicating that the dimension reduction of AMC makes direct numerical inversion difficult.

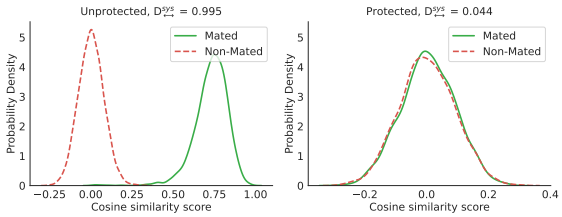


Figure 4: Unlinkability comparison for unprotected baseline and ours on LFW. $D^{sys}_{\leftrightarrow}$ indicates the overall system linkability.

Revocability. A compromised protected embedding can be replaced by generating a new identity-specific IS-DOT key to protect the enrolled embedding. SCDC further uses fresh randomness when generating secret shares and fresh Beaver triples for each comparison. Thus, compromised keys or embeddings can be revoked and replaced without retraining the underlying FR model.

Unlinkability. Following Gomez-Barrero et al. (2017), we evaluate unlinkability using *mated* and *non-mated* score distributions and the system-level linkability measure $D^{sys}_{\leftrightarrow} \in [0, 1]$, where lower values indicate stronger unlinkability. As shown in Figure 4, our method achieves $D^{sys}_{\leftrightarrow} = 0.044$ on LFW, compared with 0.995 for the unprotected baseline, indicating substantially reduced linkability.

5 Ablation Study and Analysis

Compression Baselines. To showcase the improvement of AMC, we compare it with other representative dimensionality reduction baselines, including PCA, a bottleneck MLP, and Naive MRL Kusupati et al. (2022). PCA performs a global linear projection, which is simple and stable but may discard discriminative directions when embeddings contain nonlinear variations caused by pose, age, and illumination. A bottleneck MLP introduces nonlinear compression, but it is optimized for a single compact representation and does not explicitly preserve multi-granularity information, leading to larger information loss after compression. Naive MRL improves over the bottleneck MLP by encouraging nested representations, but it lacks the adaptive compression mechanism of AMC. As shown in Table 5, AMC consistently achieves the highest accuracy across different benchmarks.

Table 5: Face verification accuracy (%) across different compression baselines. All methods compress 512-dimensional embeddings to 128 dimensions.

Method	LFW	CFP-FP	AgeDB
PCA	99.63	97.24	97.43
MLP	99.15	95.67	93.30
Naive MRL	99.55	96.05	96.25
AMC	**99.73**	**97.41**	**97.68**

Impact of Compressed Embedding Dimension. AMC reduces the embedding dimensionality to suppress dense identity information while preserving recognition utility. We evaluate this trade-off using different output dimensions. The utility is measured by recognition accuracy on AgeDB, while the privacy is assessed by inversion-based ASR on M.Face. As shown in Figure 5, reducing the embedding dimension consistently decreases the ASR, but aggressive compression also degrades recognition accuracy. In particular, 32-dimensional and 64-dimensional embeddings provide stronger resistance to inversion but lead to noticeable utility loss, whereas 256-dimensional representation still remains more susceptible to inver-

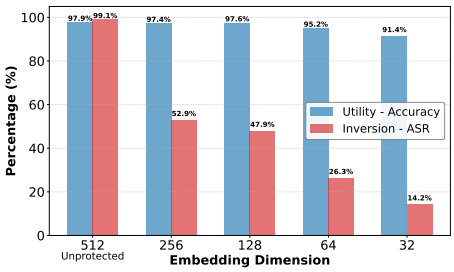


Figure 5: Privacy-utility trade-off of AMC.

sion. Considering both recognition performance and inversion robustness, we select 128 dimensions as the default setting, which maintains near-baseline accuracy while substantially reducing reconstruction.

Choice of Component in IS-DOT. We further analyze the contribution of each component in IS-DOT against TI attacks, as reported in Table 6. To isolate the effect of IS-DOT, all protection methods are directly applied to the original 512-dimensional face embeddings without AMC or SCDC. The results show that the two components of IS-DOT play different roles under different attack settings. On M.Face, Naive DCPE and Naive OT remain vulnerable to both inversion-based and score-based attacks. A similar trend is observed on E.Face for the score-based attack: Naive OT remains highly vulnerable with an ASR of 92.1%, while Naive DCPE substantially reduces the ASR to 7.6%. For inversion-based attacks on E.Face, all variants obtain very low ASRs (0.2-0.4%), indicating limited transferability. Overall, these results suggest that the randomized DCPE perturbation is particularly important for disrupting score-based optimization, whereas the identity-specific orthogonal transformation primarily provides additional obfuscation. Their combination yields the strongest overall robustness across the evaluated attacks and recognition models.

Table 6: The impact of IS-DOT against TI attacks. All variants use identity-specific settings and are without applying AMC or SCDC. ASR (%) is reported on E.Face and M.Face. OT: orthogonal transformation.

Method	Inversion-based ↓		Score-based ↓	
	E.Face	M.Face	E.Face	M.Face
Naive DCPE	0.4	76.8	7.6	74.3
Naive OT	**0.2**	79.0	92.1	97.7
IS-DOT	**0.2**	**52.7**	**0.7**	**47.5**

Generalization on Various Recognition Models. Unlike SlerpFace, which requires modifying and retraining the target FR model, our method can be applied to other recognition systems with minimal adaptation. We apply the 128-dimensional AMC pre-trained on IR-50 to different FR and PPFR systems, including MagFace, AdaFace, DuetFace, DCTDP, and MinusFace. All target models remain frozen, and IS-DOT is disabled to isolate the effect of AMC on recognition accuracy. As shown in Table 7, the transferred AMC maintains reasonable performance without fine-tuning. Fine-tuning AMC for only 1000 steps substantially improves performance. For example, the average drop on AdaFace decreases from **4.13%** to **0.84%**. This fine-tuning takes less than **5 min** on a single GPU. A similar trend is observed for PPFR systems, where the average drop remains below **1%** after fine-tuning. These results demonstrate that AMC generalizes across different FR backbones and PPFR frameworks, enabling fast adaptation without retraining the target model.

Table 7: Generalization to other recognition systems. All FRs and PPFRs are frozen. AMC is fine-tuned for only 1000 steps; IS-DOT is disabled to isolate AMC's effect on recognition accuracy (%). FT: fine-tuning. Avg Drop: average accuracy drop from the original model.

FR&PPFR	CFP-FP	AgeDB	CALFW	Avg Drop ↓
MagFace	96.89	97.75	95.87	-
MagFace+AMC (w/o FT)	94.96	96.30	94.85	1.47
MagFace+AMC (w/ FT)	96.60	97.68	95.80	**0.14**
AdaFace	97.24	97.67	95.95	-
AdaFace+AMC (w/o FT)	91.33	93.52	93.62	4.13
AdaFace+AMC (w/ FT)	95.81	97.20	95.33	**0.84**
DuetFace	94.37	97.37	95.83	-
DuetFace+AMC (w/o FT)	91.01	95.13	94.25	2.39
DuetFace+AMC (w/ FT)	93.96	96.88	95.40	**0.44**
DCTDP	97.83	94.78	94.12	-
DCTDP+AMC (w/o FT)	95.36	91.20	92.23	2.65
DCTDP+AMC (w/ FT)	97.45	93.55	93.63	**0.70**
MinusFace	97.74	92.43	93.17	-
MinusFace+AMC (w/o FT)	93.40	88.32	90.42	3.73
MinusFace+AMC (w/ FT)	95.91	92.16	92.72	**0.85**

6 Conclusion

We introduced ZeroFace, a training-efficient biometric template protection framework for secure face recognition. ZeroFace can be integrated into existing FR and PPFR systems without modifying or retraining the recognition backbone. Experiments demonstrate strong robustness against various attacks, while maintaining verification accuracy. Nevertheless, its protection depends on secure management of identity-specific keys, non-colluding computing parties, and does not completely eliminate the privacy leakage on disclosed matching scores. Future work will investigate stronger formal privacy guarantees and more scalable matching protocols.

AI USE STATEMENT

In this work, we used generative AI tools for checking grammar of manuscript.

ETHICS STATEMENT

This work investigates the face embedding protection in face recognition using publicly available face datasets and does not involve the collection of new human data or human subjects. The face reconstruction and inference attacks considered in this work are performed solely to evaluate potential privacy risks and the effectiveness of the proposed protection mechanisms. We do not use the reconstructed faces for identification or impersonation. We follow the respective usage conditions of the public datasets and report the security risks and limitations of our method to promote responsible development of privacy-preserving face recognition systems.

REPRODUCIBILITY STATEMENT

We provide comprehensive details to ensure the reproducibility of ZeroFace. The main paper describes the overall framework, threat model, training configuration, datasets, baselines, and evaluation protocols. Additional implementation details are provided in the appendix, including the training procedure of AMC, the generation and application of identity-specific IS-DOT transformations, the Beaver triple generation and secure computation used in SCDC, and the detailed ASR evaluation protocol.

REFERENCES

Donald Beaver. Efficient multiparty protocols using circuit randomization. In *Annual international cryptology conference*, pp. 420–432. Springer, 1991.

Amos Beimel. Secret-sharing schemes: A survey. In *International conference on coding and cryptology*, pp. 11–46. Springer, 2011.

Mihir Bellare, Ran Canetti, and Hugo Krawczyk. Keying hash functions for message authentication. In *Annual international cryptology conference*, pp. 1–15. Springer, 1996.

George Robert Blakley. Safeguarding cryptographic keys. In *Managing requirements knowledge, international workshop on*, pp. 313–313. IEEE Computer Society, 1979.

Vishnu Naresh Boddeti. Secure face matching using fully homomorphic encryption. In *2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS)*, pp. 1–10. IEEE, 2018.

Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1578–1587, 2022.

Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pp. 67–74. IEEE, 2018.

Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In *Chinese conference on biometric recognition*, pp. 428–438. Springer, 2018.

Ivan Damgård, Valerio Pastro, Nigel Smart, and Sarah Zakarias. Multiparty computation from somewhat homomorphic encryption. In *Annual cryptology conference*, pp. 643–662. Springer, 2012.

Thao M Dang, Lam Tran, Thuc D Nguyen, and Deokjai Choi. Fehash: Full entropy hash for face template protection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 810–811, 2020.

Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4690–4699, 2019.

Xingbo Dong, Soohyong Kim, Zhe Jin, Jung Yeon Hwang, Sangrae Cho, and Andrew Beng Jin Teoh. Secure chaff-less fuzzy vault for face identification systems. *ACM Transactions on Multimidia Computing Communications and Applications*, 17(3):1–22, 2021.

Mihai Dusmanu, Johannes L Schonberger, Sudipta N Sinha, and Marc Pollefeys. Privacy-preserving image features via adversarial affine subspace embeddings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14267–14277, 2021.

Junfeng Fan and Frederik Vercauteren. Somewhat practical fully homomorphic encryption. *Cryptology ePrint Archive*, 2012.

Georg Fuchsbauer, Riddhi Ghosal, Nathan Hauke, and Adam O'Neill. Approximate distance-comparison-preserving symmetric encryption. In *International Conference on Security and Cryptography for Networks*, pp. 117–144. Springer, 2022.

Craig Gentry. Fully homomorphic encryption using ideal lattices. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pp. 169–178, 2009.

Marta Gomez-Barrero, Javier Galbally, Christian Rathgeb, and Christoph Busch. General framework to evaluate unlinkability in biometric template protection systems. *IEEE Transactions on Information Forensics and Security*, 13(6):1406–1420, 2017.

Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European conference on computer vision*, pp. 87–102. Springer, 2016.

Vedrana Krivokuća Hahn and Sébastien Marcel. Towards protecting face embeddings in mobile face verification scenarios. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4 (1):117–134, 2022.

Dong Han, Yong Li, and Joachim Denzler. Realistic face reconstruction from facial embeddings via diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 4538–4546, 2026.

Gary B Huang and Erik Learned-Miller. Labeled faces in the wild: Updates and new reporting procedures. *Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep*, 14 (003), 2014.

Patrick Jauernig, Ahmad-Reza Sadeghi, and Emmanuel Stapf. Trusted execution environments: properties, applications, and challenges. *IEEE Security & Privacy*, 18(2):56–60, 2020.

Jiazhen Ji, Huan Wang, Yuge Huang, Jiaxiang Wu, Xingkun Xu, Shouhong Ding, ShengChuan Zhang, Liujuan Cao, and Rongrong Ji. Privacy-preserving face recognition with learnable privacy budgets in frequency domain. In *European Conference on Computer Vision*, pp. 475–491. Springer, 2022.

Kimmo Karkkainen and Jungseock Joo. Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1548–1558, 2021.

Tero Karras. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8110–8119, 2020.

Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. *Advances in neural information processing systems*, 34:852–863, 2021.

Marcel Keller, Emmanuela Orsini, and Peter Scholl. Mascot: faster malicious arithmetic secure computation with oblivious transfer. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 830–842, 2016.

Sunpill Kim, Yunseong Jeong, Jinsu Kim, Jungkon Kim, Hyung Tae Lee, and Jae Hong Seo. Ironmask: Modular architecture for protecting deep face template. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16125–16134, 2021.

Sunpill Kim, Seunghun Paik, Chanwoo Hwang, Dongsoo Kim, Junbum Shin, and Jae Hong Seo. Idface: Face template protection for efficient and secure identification. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13995–14005. IEEE, 2025.

Brian Knott, Shobha Venkataraman, Awni Hannun, Shubho Sengupta, Mark Ibrahim, and Laurens Van Der Maaten. Crypten: Secure multi-party computation meets machine learning. *Advances in Neural Information Processing Systems*, 34:4961–4973, 2021.

Aditya Kusupati, Gantavya Bhatt, Aniket Rege, Matthew Wallingford, Aditya Sinha, Vivek Ramanujan, William Howard-Snyder, Kaifeng Chen, Sham Kakade, Prateek Jain, et al. Matryoshka representation learning. *Advances in Neural Information Processing Systems*, 35:30233–30249, 2022.

Pierrick Leroy, Antonio Mastropietro, Marco Nurisso, and Francesco Vaccarino. Attributes shape the embedding space of face recognition models. *arXiv preprint arXiv:2507.11372*, 2025.

Guangcan Mai, Kai Cao, Pong C Yuen, and Anil K Jain. On the reconstruction of face images from deep face templates. *IEEE transactions on pattern analysis and machine intelligence*, 41 (5):1188–1202, 2018.

Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, Jordan Cheney, et al. Iarpa janus benchmark-c: Face dataset and protocol. In *2018 international conference on biometrics (ICB)*, pp. 158–165. IEEE, 2018.

Yuxi Mi, Yuge Huang, Jiazhen Ji, Hongquan Liu, Xingkun Xu, Shouhong Ding, and Shuigeng Zhou. Duetface: Collaborative privacy-preserving face recognition via channel splitting in the frequency domain. In *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 6755–6764, 2022.

Yuxi Mi, Yuge Huang, Jiazhen Ji, Minyi Zhao, Jiaxiang Wu, Xingkun Xu, Shouhong Ding, and Shuigeng Zhou. Privacy-preserving face recognition using random frequency components. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19673–19684, 2023.

Yuxi Mi, Zhizhou Zhong, Yuge Huang, Jiazhen Ji, Jianqing Xu, Jun Wang, Shaoming Wang, Shouhong Ding, and Shuigeng Zhou. Privacy-preserving face recognition using trainable feature subtraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 297–307, 2024.

Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: the first manually collected, in-the-wild age database. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 51–59, 2017.

Hatef Otroshi Shahreza and Sébastien Marcel. Face reconstruction from facial templates by learning latent space of a generator network. *Advances in Neural Information Processing Systems*, 36, 2024.

Foivos Paraperas Papantoniou, Alexandros Lattas, Stylianos Moschoglou, Jiankang Deng, Bernhard Kainz, and Stefanos Zafeiriou. Arc2face: A foundation model for id-consistent human faces. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024.

Min Ren, Yuhao Zhu, Yunlong Wang, Yongzhen Huang, and Zhenan Sun. Understanding deep face representation via attribute recovery. *IEEE Transactions on Information Forensics and Security*, 19:6949–6961, 2024.

Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs. Frontal to profile face verification in the wild. In *2016 IEEE winter conference on applications of computer vision (WACV)*, pp. 1–9. IEEE, 2016.

Hatef Otroshi Shahreza and Sébastien Marcel. Blackbox face reconstruction from deep facial embeddings using a different face recognition model. In *2023 IEEE International Conference on Image Processing (ICIP)*, pp. 2435–2439. IEEE, 2023.

Hatef Otroshi Shahreza and Sébastien Marcel. Breaking template protection: Reconstruction of face images from protected facial templates. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, pp. 1–7. IEEE, 2024.

Hatef Otroshi Shahreza, Vedrana Krivokuća Hahn, and Sébastien Marcel. Face reconstruction from deep facial embeddings using a convolutional neural network. In *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 1211–1215. IEEE, 2022.

Hatef Otroshi Shahreza, Vedrana Krivokuća Hahn, and Sébastien Marcel. Mlp-hash: Protecting face templates via hashing of randomized multi-layer perceptron. In *2023 31st European Signal Processing Conference (EUSIPCO)*, pp. 605–609. IEEE, 2023.

Hatef Otroshi Shahreza, Anjith George, and Sébastien Marcel. Face reconstruction from face embeddings using adapter to a face foundation model. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 5584–5593, 2025.

Andrew Beng Jin Teoh and Jaihie Kim. Secure biometric template protection in fuzzy commitment scheme. *IEICE Electronics Express*, 4(23):724–730, 2007.

Philipp Terhörst, Daniel Fährmann, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. On soft-biometric information stored in biometric face embeddings. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(4):519–534, 2021.

Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K Jain, James A Duncan, Kristen Allen, et al. Iarpa janus benchmark-b face dataset. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 90–98, 2017.

Zhihua Xia, Qi Gu, Wenhao Zhou, Lizhi Xiong, Jian Weng, and Naixue Xiong. Str: Secure computation on additive shares using the share-transform-reveal strategy. *IEEE Transactions on Computers*, 73(2):340–352, 2021.

Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*, 2023.

Hui Zhang, Xingbo Dong, YenLung Lai, Ying Zhou, Xiaoyan Zhang, Xingguo Lv, Zhe Jin, and Xuejun Li. Validating privacy-preserving face recognition under a minimum assumption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12205–12214, 2024.

Tianyue Zheng and Weihong Deng. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. *Beijing University of Posts and Telecommunications, Tech. Rep*, 5(7):5, 2018.

Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint arXiv:1708.08197*, 2017.

Zhizhou Zhong, Yuxi Mi, Yuge Huang, Jianqing Xu, Guodong Mu, Shouhong Ding, Jingyun Zhang, Rizen Guo, Yunsheng Wu, and Shuigeng Zhou. Slerpface: Face template protection via spherical linear interpolation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 10698–10706, 2025.

APPENDIX

A PRELIMINARIES

A.1 DISTANCE COMPARISON PRESERVING SYMMETRIC ENCRYPTION

Distance Comparison Preserving Symmetric Encryption (DCPE) Fuchsbauer et al. (2022) is a property-preserving encryption designed for encrypted nearest-neighbor search. When the plaintext distance gap strictly exceeds a tolerance value β, the β-approximate DCPE scheme preserves relative distance orderings instead than exact distances. The encryption function f ensures vectors x, y, z in a bounded Euclidean space:

$$\begin{aligned} & \|x-y\|_2 < \|x-z\|_2 - \beta \\ \Longrightarrow \quad & \|f(x)-f(y)\|_2 < \|f(x)-f(z)\|_2. \end{aligned} \tag{6}$$

The Scale-and-Perturb Construction. The core mechanism behind DCPE is the *Scale-and-Perturb* (SAP) scheme. This scheme encrypts a plaintext vector by multiplying it by a secret scaling factor $s > 0$ and adding a pseudorandom perturbation λ_η. Using a pseudorandom function (PRF) Bellare et al. (1996) key k and a newly generated public random number η, the encryption of the message vector m is defined as:

$$\mathsf{Enc}_{s,k}(m;\eta) = (c_m, \eta), \qquad c_m = sm + \lambda_\eta. \tag{7}$$

Here, distance comparisons are performed on the ciphertext vector component c_m, while η is public nonce used for decryption. The perturbation λ_η is sampled from a Euclidean ball of radius $s\beta/4$, ensuring its magnitude is strictly bounded:

$$\|\lambda_\eta\|_2 \leq \frac{s\beta}{4}. \tag{8}$$

As all randomness is deterministically derived from the PRF key and η, a legitimate decryptor can exactly regenerate the perturbation λ_η. Decryption simply reverses the operation by subtracting the noise and dividing by the scale factor:

$$\mathsf{Dec}_{s,k}(c_m, \eta) = \frac{c_m - \lambda_\eta}{s}. \tag{9}$$

The security and utility of this construction rely entirely on the strict bounding of the perturbation. And the approximate distance-comparison property follows from the bounded perturbation. Because the noise is constrained, the distance between any two ciphertexts $\mathsf{Enc}(x)$ and $\mathsf{Enc}(y)$ is mathematically bounded relative to their scaled plaintext distance:

$$s\|x-y\|_2 - \frac{s\beta}{2} \leq \|\mathsf{Enc}(x) - \mathsf{Enc}(y)\|_2 \leq s\|x-y\|_2 + \frac{s\beta}{2}. \tag{10}$$

As the maximum noise added to any pairwise ciphertext distance is strictly $\pm\frac{s\beta}{2}$, it cannot overcome a plaintext distance gap larger than β. Thus, the ciphertexts reliably preserve the ordinal relationship of Euclidean distances without leaking the exact underlying vectors. This is the key reason why SAP satisfies the β-DCP property and supports approximate nearest-neighbor search over encrypted vectors. In our work, we explore its' potential for face embedding protection.

A.2 SECURE MULTIPARTY COMPUTATION (SMPC)

Secure Multiparty Computation (SMPC) Beaver (1991) is a method where multiple parties collaboratively achieve a computational goal, ensuring that each participant cannot obtain any input information of others except for the computation result. Assuming there are *n* participants P_1, P_2, ...,

P_n, each with a private input data x_i. All participants jointly compute a function $f(x_1, x_2, ..., x_n)$, and it is required that at the end of the computation, each participant P_i can only obtain the output corresponding to its private input data x_i, and cannot access the input information or output results of other participants. Secret Sharing (SS) Beimel (2011) is one of popular schemes.

Secure Multiplication. Assuming party A and party B hold their own secret scalar values X and Y, the goal is to calculate multiplication XY without sharing their secrets. Before SS multiplication protocol, each party can split secret into two parts through additive sharing Blakley (1979), i.e., X = X_1 + X_2 and Y = Y_1 + Y_2. And they share a precomputed random Beaver triples Beaver (1991): $([a], [b], [c])$, where $c = ab = (c_1 + c_2) = (a_1 + a_2)(b_1 + b_2)$ and the shared value denoted as $[\cdot]$. Specifically, party A holds $([a_1], [b_1], [c_1])$ while party B holds $([a_2], [b_2], [c_2])$. **In SS multiplication protocol**, party A can compute $[\epsilon_1] = (X_1 - a_1)$, $[\delta_1] = (Y_1 - b_1)$ with Y_1 shared by party B and send $[\epsilon_1], [\delta_1]$ back to party B. Similarly, party B can compute $[\epsilon_2] = (X_2 - a_2)$, $([\delta_2] = Y_2 - b_2)$ with X_2 shared by party A and send $[\epsilon_2], [\delta_2]$ back to party A. Finally, the product XY can now be evaluated as: $c_1 + a_1([\delta_1] + [\delta_2]) + b_1([\epsilon_1] + [\epsilon_2])$ + $c_2 + a_2([\delta_1] + [\delta_2]) + b_2([\epsilon_1] + [\epsilon_2])$ + $([\delta_1] + [\delta_2])([\epsilon_1] + [\epsilon_2])$. At the end of protocol, both parties are able to obtain the product XY secretly, but neither party learns anything about the individual secrets of the other.

B Details of the Adaptive Matryoshka Compression Network

This section provides the detailed architecture and forward computation of the proposed Adaptive Matryoshka Compression (AMC) network. AMC is designed to compress the original high-dimensional face embedding while adaptively retaining identity-discriminative information from multiple nested granularities.

B.1 Facial Embedding and Nested Granularities

For 512-dimensional face embedding $\mathbf{x}$ extracted by the frozen face recognition backbone. AMC uses the following nested granularity dimensions:

$$\mathcal{N} = \{n_1, n_2, n_3, n_4\} = \{64, 128, 256, 512\}. \tag{11}$$

For each granularity n_k, AMC takes the first n_k dimensions of the original embedding:

$$\mathbf{x}^{(k)} = \mathbf{x}_{1:n_k} = [x_1, x_2, \ldots, x_{n_k}]^\mathsf{T} \in \mathbb{R}^{n_k}, \qquad k \in \{1, 2, 3, 4\}. \tag{12}$$

The nested inputs are therefore

$$\mathbf{x}^{(1)} = \mathbf{x}_{1:64}, \quad \mathbf{x}^{(2)} = \mathbf{x}_{1:128}, \quad \mathbf{x}^{(3)} = \mathbf{x}_{1:256}, \quad \mathbf{x}^{(4)} = \mathbf{x}_{1:512}. \tag{13}$$

Each larger-granularity input contains the information used by all preceding granularities. This nested design enables AMC to simultaneously model coarse, intermediate, and fine-grained information in the original face embedding, instead of selecting a single manually truncated representation.

B.2 Granularity-Specific Compression

For every nested input, AMC employs an independent linear transformation. All branches produce an output of dimension $d = 128$:

$$\mathbf{h}^{(k)} = \mathbf{W}_k \mathbf{x}^{(k)} + \mathbf{b}_k \in \mathbb{R}^{128}, \tag{14}$$

where

$$\mathbf{W}_k \in \mathbb{R}^{128 \times n_k}, \qquad \mathbf{b}_k \in \mathbb{R}^{128}. \tag{15}$$

In particular, the four AMC branches are

$$\mathbf{h}^{(1)} = \mathbf{W}_1 \mathbf{x}_{1:64} + \mathbf{b}_1, \qquad \mathbf{W}_1 \in \mathbb{R}^{128\times 64}, \tag{16}$$

$$\mathbf{h}^{(2)} = \mathbf{W}_2 \mathbf{x}_{1:128} + \mathbf{b}_2, \qquad \mathbf{W}_2 \in \mathbb{R}^{128\times 128}, \tag{17}$$

$$\mathbf{h}^{(3)} = \mathbf{W}_3 \mathbf{x}_{1:256} + \mathbf{b}_3, \qquad \mathbf{W}_3 \in \mathbb{R}^{128\times 256}, \tag{18}$$

$$\mathbf{h}^{(4)} = \mathbf{W}_4 \mathbf{x}_{1:512} + \mathbf{b}_4, \qquad \mathbf{W}_4 \in \mathbb{R}^{128\times 512}. \tag{19}$$

The branch-specific parameters are not shared across granularities. This allows each branch to learn a projection tailored to the amount of input information available at its corresponding granularity. Importantly, the branch outputs have the same dimension, which makes it possible to combine them without concatenating their features or increasing the final template dimension.

B.3 Adaptive Granularity Fusion

AMC adaptively combines the four branch outputs using learnable granularity weights. Let

$$\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \gamma_3, \gamma_4]^\mathsf{T} \in \mathbb{R}^4 \tag{20}$$

denote the learnable fusion parameters. They are initialized with the same value:

$$\gamma_k = 1, \qquad k \in \{1, 2, 3, 4\}. \tag{21}$$

The normalized contribution of the k-th granularity is computed using a softmax function:

$$\alpha_k = \frac{\exp(\gamma_k)}{\sum_{j=1}^{4} \exp(\gamma_j)}. \tag{22}$$

The initialization leads to uniform weights, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1/4$. During training, these weights are optimized jointly with the branch transformations, allowing AMC to automatically adjust the contribution of different embedding granularities.

The final compressed embedding is obtained by weighted summation:

$$\mathbf{e}_{\mathrm{AMC}} = \sum_{k=1}^{4} \alpha_k \mathbf{h}^{(k)}. \tag{23}$$

Substituting Equation 14 into Equation 23, the complete AMC mapping can be written as

$$\mathbf{e}_{\mathrm{AMC}} = \sum_{k=1}^{4} \alpha_k \left(\mathbf{W}_k \mathbf{x}_{1:n_k} + \mathbf{b}_k\right) \in \mathbb{R}^{128}. \tag{24}$$

Thus, AMC reduces the original 512-dimensional embedding to a compact 128-dimensional representation. The output dimension is independent of the number of granularity branches because all branch outputs are projected to the same 128-dimensional space before fusion.

B.4 Training Configuration

In all experiments, we use four independent AMC branches corresponding to the dimensions in Equation 11. Each branch contains a linear transformation from its respective nested input dimension to the common output dimension of 128. The four branch outputs are always fused using Equation 23. Therefore, the final AMC representation is produced solely through the following three operations:

1. extracting nested prefixes from the original face embedding;
2. independently projecting each prefix into a 128-dimensional representation;
3. adaptively fusing the projected representations using normalized learnable weights.

The resulting 128-dimensional AMC representation is then fed into the ArcFace classification head for identity supervision. During training, only AMC and the classification head are optimized, while the recognition backbone remains frozen. The $\mathbf{e}_{\mathrm{AMC}}$ is used as the compact face template for the subsequent protection and verification.

C Generation of Identity-Specific Orthogonal Matrices

This section provides the procedure for generating the identity-specific orthogonal matrix used in IS-DOT. For each enrolled identity u, the protection key is defined as

$$\Theta_u = \left(K_u^{\mathrm{dcpe}}, s_u, \xi_u\right), \tag{25}$$

where K_u^{dcpe} is the secret DCPE key, $s_u > 0$ is the secret scaling factor, and ξ_u is the seed used to generate the orthogonal matrix $R_u \in \mathbb{R}^{d\times d}$. The same Θ_u is used for the enrolled template and all queries claiming identity u, while a fresh nonce is generated for each protection operation.

Pseudorandom Matrix Generation. A pseudorandom generator initialized with ξ_u first produces a matrix $G_u \in \mathbb{R}^{d\times d}$ whose entries follow a standard Gaussian distribution:

$$G_u = \mathsf{PRG}_{\xi_u}(d, d), \qquad (G_u)_{ab} \sim \mathcal{N}(0, 1). \tag{26}$$

Using an identity-specific seed ensures that different identities obtain independent transformation matrices, while the same matrix can be deterministically regenerated whenever the corresponding seed is available.

Sign-corrected QR Decomposition. We apply QR decomposition to G_u:

$$G_u = Q_u T_u, \tag{27}$$

where $Q_u \in \mathbb{R}^{d\times d}$ is orthogonal and $T_u \in \mathbb{R}^{d\times d}$ is upper triangular. Since QR decomposition has sign ambiguity, we define

$$D_u = \mathrm{diag}\left(d_{u,1}, d_{u,2}, \ldots, d_{u,d}\right), \tag{28}$$

where

$$d_{u,r} = \begin{cases} 1, & (T_u)_{rr} \geq 0, \\ -1, & (T_u)_{rr} < 0. \end{cases} \tag{29}$$

The final identity-specific orthogonal matrix is

$$R_u = Q_u D_u. \tag{30}$$

The sign correction makes the generated matrix deterministic with respect to ξ_u and removes the sign ambiguity introduced by QR decomposition.

Orthogonality. Since both Q_u and D_u are orthogonal,

$$R_u^\top R_u = D_u^\top Q_u^\top Q_u D_u = I_d, \qquad R_u R_u^\top = I_d. \tag{31}$$

Therefore, for any column vectors $x, y \in \mathbb{R}^d$,

$$\|R_u x\|_2 = \|x\|_2, \tag{32}$$

$$\|R_u(x - y)\|_2 = \|x - y\|_2, \tag{33}$$

$$\langle R_u x, R_u y\rangle = \langle x, y\rangle. \tag{34}$$

Consequently, applying the same R_u to both embeddings preserves their cosine similarity:

$$\mathrm{Cos}(R_u x, R_u y) = \mathrm{Cos}(x, y). \tag{35}$$

Protected Embedding Generation. Given a compressed embedding $e'_i \in \mathbb{R}^d$ to be protected under the claimed identity u, IS-DOT first applies the scale-and-perturb (SAP) construction of DCPE Fuchsbauer et al. (2022):

$$\bar{e}_i^{(u)} = s_u e'_i + \lambda_i^{(u)}, \qquad \|\lambda_i^{(u)}\|_2 \leq \frac{s_u \beta}{4}, \tag{36}$$

where β is the DCPE approximation parameter controlling the perturbation magnitude. For each protection operation, a fresh public 128-bit nonce $\eta_i^{(u)}$ is sampled, and $\lambda_i^{(u)}$ is pseudorandomly derived from K_u^{dcpe} and $\eta_i^{(u)}$. The final protected embedding is

$$z_i^{(u)} = R_u \bar{e}_i^{(u)}. \tag{37}$$

D Beaver Triple Generation and Secure Multiplication in SCDC

We implement the secure multiplication in SCDC using CrypTen (version 0.4.1) Knott et al. (2021), a PyTorch-based secure multiparty computation framework. CrypTen represents private values using arithmetic secret sharing over a finite ring and employs fixed-point encoding for real-valued tensors. Following its default MPC construction, correlated randomness required for secure multiplication is generated in an offline preprocessing phase by a trusted third party (TTP).

Arithmetic Secret Sharing. For a private value x, two computation parties P_1 and P_2 hold additive shares $[x]_1$ and $[x]_2$ such that

$$x = [x]_1 + [x]_2 \pmod{Q}, \tag{38}$$

where Q denotes the underlying ring size. Neither party needs to reconstruct x during the secure computation.

Beaver Triple Generation. Before the online multiplication phase, the TTP samples two independent random values (or vectors)

$$a, b \leftarrow \mathbb{Z}_Q^d \tag{39}$$

and computes their element-wise product

$$c = a \odot b \pmod{Q}. \tag{40}$$

The resulting Beaver triple (a, b, c) therefore satisfies $c = a \odot b$. Each component is independently secret-shared between the two computation parties:

$$[a] = ([a]_1, [a]_2), \quad [b] = ([b]_1, [b]_2), \quad [c] = ([c]_1, [c]_2), \tag{41}$$

such that

$$a = [a]_1 + [a]_2, \quad b = [b]_1 + [b]_2, \quad c = [c]_1 + [c]_2 \pmod{Q}. \tag{42}$$

Only the corresponding shares are distributed to each party; the complete triple is not revealed to either computation party. A fresh Beaver triple is used for each secure multiplication.

Secure Multiplication. Let $[x]$ and $[y]$ denote the secret-shared vectors to be multiplied. Using a preprocessed Beaver triple $([a], [b], [c])$, the parties first compute

$$[\epsilon] = [x] - [a], \qquad [\delta] = [y] - [b]. \tag{43}$$

The masked values ϵ and δ are then reconstructed. Since a and b are independently random and unknown to the computation parties, opening these masked differences does not reveal the underlying values x and y under the semi-honest security model. Each party subsequently computes its share of the product as

$$[xy] = [c] + \epsilon[b] + \delta[a] + \epsilon\delta. \tag{44}$$

The final term $\epsilon\delta$ is public after opening ϵ and δ and can be assigned to one party's share (or equivalently handled according to the sharing convention used by the MPC implementation). Summing the resulting shares reconstructs

$$xy = c + \epsilon b + \delta a + \epsilon\delta = xy. \tag{45}$$

Application to SCDC. For SCDC, the above multiplication is applied element-wise to the protected query template $P_c \in \mathbb{R}^d$ and enrolled template $P_d \in \mathbb{R}^d$. The resulting multiplication shares are locally summed, yielding additive shares of their dot product:

$$[P_c^\top P_d] = \sum_{j=1}^{d} [P_{c,j} P_{d,j}]. \tag{46}$$

Thus, neither computation party needs to reconstruct the complete protected templates during dot-product evaluation. In our experiments, Beaver-triple preprocessing and arithmetic secret sharing are implemented based on CrypTen, while SCDC uses the resulting secure dot product for protected face matching.

E Details of Experimental Setups

Attack Strategy. In our work, we explore vulnerabilities of both clean FR system and the those protected by various BTPs. see in Figure 6. We mainly utilize two privacy attacks including inversion-based and score-based attacks. Inversion-based attack, e.g., FEM Han et al. (2026) focuses on executing an embedding-to-face attack to reconstruct realistic, high-resolution face images directly from leaked facial embeddings. FEM utilizes a Kolmogorov-Arnold Network (KAN) and MLP in conjunction with a pre-trained Identity-Preserving diffusion model Ye et al. (2023). Score-based attack model, e.g., MAP2V Zhang et al. (2024) framework simulates a highly realistic adversary operating under a threat model that the attacker acts as a normal user with only black-box access to observe the target system's output similarity scores. To reconstruct a face image without requiring network training by accessing limited queries, MAP2V formulates the attack as a zeroth-order optimization problem leveraging generic deep image priors. In our experiment, for the inversion-based attack mentioned in **case (1)/(3)**, FEM-MLP is trained directly on the protected embeddings on public dataset FFHQ. Then the trained FEM-MLP is used to evaluate ASR with leaked protected embeddings on CelebA-HQ dataset. Therefore, it is a transferable attack. For the score-based attack mentioned in **cases (2)/(3)**, we adapt MAP2V by replacing its original cosine-score oracle with surrogate similarities used in each corresponding BTP method while optimizing the StyleGAN2 latent space to reconstruct facial images from CelebA-HQ dataset. For detailed implementation, see original works Han et al. (2026); Zhang et al. (2024).

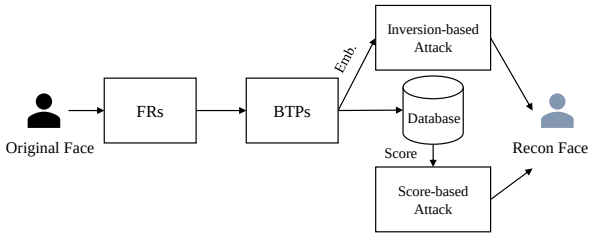


Figure 6: Threat models for BTP methods. Recon: reconstructed.

Attack Success Rate Evaluation. We evaluate the effectiveness of privacy attacks using the Attack Success Rate (ASR), following the evaluation protocol adopted in prior work Han et al. (2026); Zhang et al. (2024). Given a reconstructed face image, we extract its face embedding using an independent FR model and compare it with the target identity embedding. A reconstruction is considered successful if the resulting similarity score exceeds the decision threshold corresponding to a FAR of 10^{-2} on the corresponding FR model. The ASR is computed as:

$$\mathrm{ASR} = \frac{N_{\text{successful}}}{N_{\text{evaluated}}} \times 100\%. \tag{47}$$

For each privacy attack, we generate one reconstructed face for each target sample and evaluate the reconstructed images using two independent FR models, MobileFace Chen et al. (2018) and ElasticFace Boutros et al. (2022). The FR models used for evaluation are not involved in training the attack models. We report the ASR separately for each FR model, where a lower ASR indicates stronger privacy protection.

Facial Attributes Attacks on FairFace Dataset. We follow the attribute annotations provided by FairFace Karkkainen & Joo (2021). Gender contains two classes: `Male`, `Female`. Age contains nine classes: `0-2`, `3-9`, `10-19`, `20-29`, `30-39`, `40-49`, `50-59`, `60-69`, `more than 70`. Race contains seven classes: `East Asian`, `Indian`, `Black`, `White`, `Middle Eastern`, `Latino_Hispanic`, `Southeast Asian`. For each attribute, we train a dedicated MLP model, using the protected face embeddings as input. The output classes C are set to $C = 2$ for gender, $C = 7$ for race, and $C = 9$ for age.

The architecture consists of two fully connected (FC) layers:

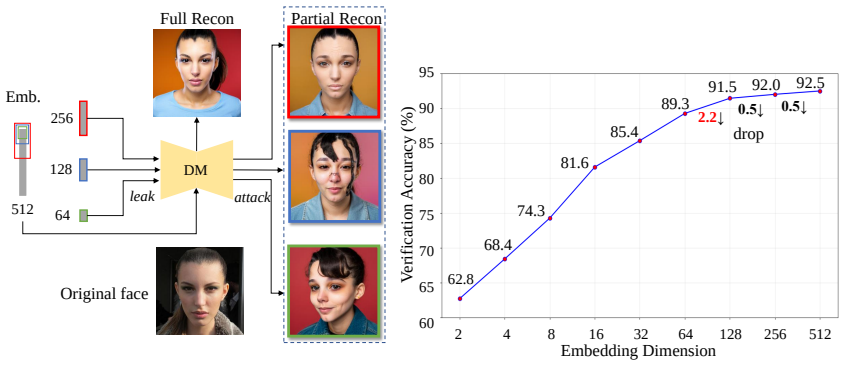
- **FC Layer 1:** Linear $(128 \to 512) \to$ ReLU $\to$ Dropout $(p = 0.3)$.
- **FC Layer 2:** Linear $(512 \to 256) \to$ ReLU $\to$ Dropout $(p = 0.3)$.
- **Output Layer:** Linear $(256 \to C)$ yielding prediction logits.

To assess multi-class classification performance across different attributes, we measure Top-1 accuracy:

$$\text{Accuracy} = \frac{\sum_{i=1}^{C} \text{TP}_i}{N} \tag{48}$$

where N is the total number of test samples, N_i is the number of true instances in class i, and TP_i denotes the true positives for class i.

F Additional Analysis



(a) Face reconstruction from partial embeddings

(b) Utility of partial embeddings

Figure 7: (a) Inversion-based attack on different size of embeddings. Recon: Reconstruction. (b) Face verification accuracy on CALFW Zheng et al. (2017) on various size of embeddings.

Dimension Truncation on Facial Embeddings. A straightforward approach for reducing the privacy risks associated with face embeddings is to truncate dimensions. Specifically, we can retain only the first d dimensions of the original embedding and use the resulting partial embedding for face verification. As shown in Figure 7(a), reducing the embedding dimension removes part of the identity information available to the inversion model and gradually degrades the quality and identity consistency of the reconstructed faces. For example, when facial reconstruction is performed using only 64-dimensional embeddings, the identity preservation capability is significantly lower than that based on full 512-dimensional embeddings. However, excessively truncating the embedding dimension also leads to loss of identity-discriminative information, which further causes a decline in face verification performance, as shown in Figure 7(b). The above phenomenon indicates that simple dimension truncation is difficult to simultaneously balance privacy protection and recognition utility. Based on this observation, we further proposes AMC, which learns to compress and aggregate key identity-discriminative information into a compact representation, rather than directly discarding partial embedding dimensions, thereby reducing the risk of privacy leakage while preserving face recognition performance as much as possible.

Visualization of Reconstructed Face Images. Figure 8 provides qualitative comparisons. Without protection, reconstructed images preserve strong identity consistency and visual attributes such as hairstyle, gender, and facial structure are often recovered. Images reconstructed from MLP-Hash, PolyProtect and SlerpFace still retain recognizable identity cues, especially under score-based attacks. After applying the proposed method, both inversion-based and score-based attacks fail to recover visually consistent target identities, demonstrating the stronger privacy protection.

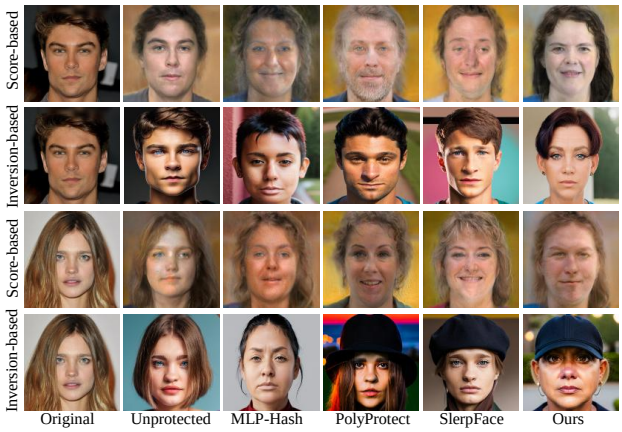


Figure 8: Reconstructed face images from different BTP methods.

Security of SCDC. In the SCDC protocol, the protected embeddings are ℓ_2-normalized before secret sharing, such that cosine similarity can be securely evaluated through the dot product without additionally exposing their norms. Under the standard "honest-but-curious" assumption, we assume that **all parties follow the prescribed protocol but may attempt to infer private information from their local observations**. The security of the secure multiplication protocol relies on additive secret sharing and fresh Beaver triples. Each party only observes its local shares and the masked values exchanged during secure multiplication, which do not reveal the complete protected embeddings. Distinct Beaver triples are generated for each comparison and are used only once. Importantly, the server and database are assumed to be non-colluding; they cannot be compromised simultaneously. Under this assumption, we further consider the compromised *(server, client)* scenario. The goal is to recover the enrolled embedding held by the database or reconstruct the corresponding face. In this case, the attacker has access to the query and the server-side observations. Obtaining the share of the enrolled protected embedding does not revealing the complete embedding. However, if exact similarity scores are released, it may still enable score-based attacks. For stronger malicious-adversary settings, authenticated secret sharing with message authentication codes (MACs) Damgård et al. (2012); Keller et al. (2016) can additionally be incorporated to detect manipulated shares or Beaver triples; we leave such active-security extensions to future work.

Comparison with TEE-only Design. A straightforward alternative to SCDC is to execute the entire matching pipeline inside a fully trusted TEE. However, such a configuration requires complete trust in hardware. Prior works have shown that TEEs are still vulnerable to side-channel and implementation-level attacks Jauernig et al. (2020). Once the TEE is compromised, embeddings from both the server and database sides may be exposed. In contrast, SCDC adopts a hybrid design in which MPC securely computes the dot product, while the TEE is optionally used only to reconstruct the similarity score and perform the final threshold comparison. This design reduces the amount of sensitive computation and data that must be entrusted to the TEE and prevents either the server or the database from directly accessing the exact similarity score. Such intermediate scores are sensitive: if the database stores enrolled protected embeddings and observes sufficiently many dot-product results for different enrolled embeddings, it may approximate the query embedding by solving an overdetermined linear system. A symmetric leakage risk arises if the server can directly observe similarity scores against database embeddings. Therefore, integrating MPC with a minimally used TEE provides a practical balance between efficiency and security, limiting information leakage while avoiding full reliance on trusted hardware.

G Additional Ablation Studies

Impact of IS-DOT on Recognition Accuracy. As shown in Table 8, we analyze the impact of IS-DOT on the recognition utility of compressed embeddings. AMC, Naive DCPE, and IS-DOT achieve average accuracies of 96.56%, 96.57%, and 96.56%, respectively, showing that the additional protection introduced by IS-DOT causes negligible degradation in recognition performance. Across all five benchmarks, the differences between AMC and IS-DOT remain within 0.18%. This behavior is expected because the same identity-specific orthogonal matrix is applied to both embeddings during verification, thereby preserving their cosine similarity. The bounded perturbation introduced by DCPE also has only a minor effect on the discriminative structure learned by AMC. Overall, these results demonstrate that IS-DOT provides additional embedding protection while largely preserving the recognition utility of AMC. We further investigate the impact of β on recognition utility in Figure 9. Small values of β preserve the discriminative structure of AMC, with $\beta = 1$ providing a favorable utility setting. As β increases, stronger perturbations gradually affect the protected representations, resulting in decreased recognition accuracy.

Table 8: Impact of IS-DOT on recognition accuracy. All methods use 128-dimensional embeddings. We report mean verification accuracy (%) and the standard deviation across evaluation folds.

Method	LFW	CFP-FP	AgeDB	CALFW	CPLFW
AMC	99.73 ($\pm$ 0.29)	97.41 ($\pm$ 0.64)	97.68 ($\pm$ 0.63)	95.95 ($\pm$ 1.19)	92.02 ($\pm$ 1.73)
Naive DCPE	99.73 ($\pm$ 0.29)	97.51 ($\pm$ 0.67)	97.63 ($\pm$ 0.61)	95.77 ($\pm$ 1.14)	92.20 ($\pm$ 1.67)
IS-DOT	99.73 ($\pm$ 0.29)	97.51 ($\pm$ 0.58)	97.72 ($\pm$ 0.65)	95.77 ($\pm$ 1.14)	92.08 ($\pm$ 1.58)

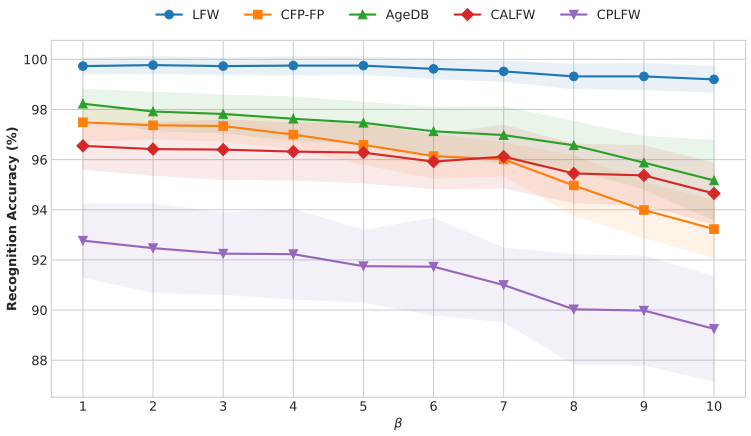


Figure 9: Impact of the IS-DOT parameter β on recognition accuracy.