frontiers | Frontiers in Imaging

TYPE Original Research
PUBLISHED 12 May 2025
DOI 10.3389/fimag.2025.1476377

OPEN ACCESS

EDITED BY
Christian Kraetzer,
Otto-von-Guericke University, Germany

REVIEWED BY
Surendra Singh,
Clarkson University, United States
Tobias Scheidat,
Anhalt University of Applied Sciences, Germany

*CORRESPONDENCE
Amina Bassit
✉ bassitam@msu.edu

RECEIVED 05 August 2024
ACCEPTED 01 April 2025
PUBLISHED 12 May 2025

CITATION
Bassit A, Hahn F, Rezgui Z, Shahreza HO, Veldhuis R and Peter A (2025) Template recovery attack on encrypted face recognition systems with unprotected decision using synthetic faces. *Front. Imaging* 4:1476377. doi: 10.3389/fimag.2025.1476377

Template recovery attack on encrypted face recognition systems with unprotected decision using synthetic faces

Amina Bassit[1,2]*, Florian Hahn[2], Zohra Rezgui[2], Hatef Otroshi Shahreza[3,4], Raymond Veldhuis[2,5] and Andreas Peter[2,6]

[1]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, United States, [2]Data Management & Biometrics and Semantics, Cybersecurity & Services, University of Twente, Enschede, Netherlands, [3]Biometrics Security & Privacy Group, Idiap Research Institute, Martigny, Switzerland, [4]School of Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [5]Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjøvik, Norway, [6]Safety-Security-Interaction Group, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

Introduction: Homomorphic encryption (HE) enables privacy-preserving face recognition by allowing encrypted facial embeddings to be compared without decryption. While efficient, these systems often reveal comparison scores in plaintext, introducing a security risk. Revealing these scores can potentially allow adversaries to reconstruct sensitive facial embeddings and infer demographic attributes, thus compromising user privacy.

Methods: This work proposes a training-less face template recovery attack leveraging the Lagrange multiplier optimization method. The attack requires only a small set of randomly generated synthetic facial images and their associated comparison scores with a target template. The method assumes attackers use spoofed synthetic faces and lack direct access to the face recognition system, aligning with real-world threat models.

Results: Experimental evaluation demonstrates the feasibility and effectiveness of the proposed attack. It shows that between 50 and 192 comparison scores and synthetic images are sufficient to recover the target face template with 100% success under strict system thresholds. The recovered templates closely resemble the original and retain identifiable soft biometric traits.

Discussion: The findings reveal a critical vulnerability in face recognition systems employing inner product similarity measures under homomorphic encryption. Even without system access or training data, attackers can exploit leaked comparison scores to compromise facial privacy. The study underscores the need to reassess how score leakage is handled in encrypted recognition systems and explore stronger protection mechanisms against template reconstruction.

KEYWORDS

homomorphic encryption, template recovery, biometric template protection, inner product-based score, biometric recognition, synthetic facial images

1 Introduction

Modern face recognition systems (Rathgeb et al., 2023) represent biometric samples by distinctive feature vectors that can be viewed as points spread out in an embedding space. They produce recognition decisions by evaluating whether a similarity score is above a predefined biometric threshold. Several studies (Acien et al., 2018; Zhang et al., 2020) report that facial feature vectors contain personally identifiable information, such as demographic information, and are susceptible to being inverted for retrieving raw facial images, threatening the privacy and security aspects of face recognition schemes.

To alleviate these privacy and security threats, biometric template protection schemes (BTPs; Sandhya and Prasad, 2017) were introduced as protection mechanisms withstanding these threats for a maintained recognition performance. Among existing BTPs, homomorphic encryption (HE)–based BTPs (Chitrapu and Kalluri, 2023) have gained traction in the biometrics community for their ability to process biometric data under encryption, aiming to prevent the biometric leakage. By comparing encrypted biometric templates, HE-based BTPs produce encrypted scores that are nearly identical to cleartext scores when decrypted. This makes HE-based BTPs maintain the unprotected system's recognition performance because HE allows arithmetic operations over encrypted data without decryption. However, they save links between pairs of unprotected templates and pairs of protected ones that are expressed as (dis)similarity scores. The heavy computational nature of homomorphic operations leaves no choice for most HE-based BTPs (Drozdowski et al., 2019; Kolberg et al., 2020; Engelsma et al., 2022; Boddeti, 2018; Bauspieß et al., 2022) but to break the protection and reveal the scores to carry out the comparison with the predefined biometric threshold in the unprotected cleartext domain. This approach is taken to avoid the significant additional runtime required for performing the comparison under encryption.

A similarity score is a strong indicator of how similar a facial feature vector is to another one, typically with one being freshly extracted (i.e., a probe) and the other stored (i.e., the reference). For a target facial feature vector, this score provides enough information to mount a template recovery attack in which a set of different facial feature vectors and their corresponding scores regarding the target vector form an optimization problem, which we tackle in this work. Figure 1 depicts an overview of our template recovery attack and its capabilities.

What are the potential dangers and implications of exposing cleartext scores?

Once a target facial embedding template is recovered, this cannot be effectively mitigated. Even if this recovered embedding template is replaced with a newly generated one, the original and its neighboring embeddings can still deceive the system due to their proximity in the embedding space. To mitigate this risk, minimizing biometric leakage by performing the comparison with the threshold within the encrypted domain is crucial to ensure that only the final decision is transmitted in encrypted form. Recovering raw templates from leaked comparison scores of homomorphically encrypted biometric templates has received limited attention in the biometrics community. It is often considered low-risk, as attackers are perceived to lack the necessary knowledge to successfully mount a recovery attack (Engelsma et al., 2022). However, our research demonstrates that semi-honest attackers can leverage HE-based BTPs' architecture to facilitate template recovery.

Approaches that exploit score leakage can be divided into two main categories: *adaptive* and *non-adaptive*. Adaptive approaches (Soutar et al., 2002; Adler, 2004, 2005; Galbally et al., 2009, 2010; Gomez-Barrero et al., 2011, 2012) rely on adaptive hill-climbing attacks. In these attacks, a target template is gradually recovered by adaptively adjusting a template and evaluating the resulting variations in comparison scores relative to the target template, eventually converging on the target template. Despite their effectiveness, adaptive hill-climbing attacks have notable drawbacks, such as being comparator-dependent or requiring a large number of iterations [e.g., 900 iterations as shown by Gomez-Barrero et al. (2012), for a 100-dim eigenface] for successful recovery. Additionally, changing the target necessitates repeating the process with new arbitrary templates.

The only known non-adaptive approach to exploiting score leakage is described in Mohanty et al. (2006). The authors proposed the first non-adaptive recovery attack, which involves inverting an affine transform that approximates an eigenface-based recognition model, from which specific scores are collected. However, similar to white-box adversarial attacks in machine learning, this approach requires knowledge of the underlying face recognition model's hyperparameters used during the training phase to approximate the model, making it model-dependent. Moreover, it relies on a set of real facial images from individuals different from the target identity to recover the target template.

In this article, we propose a non-adaptive facial template recovery attack that exploits the cleartext score disclosure vulnerability of many HE-based BTPs, further motivating research on HE-based BTPs that keep the scores encrypted (Bassit et al., 2021, 2022, 2023b; Ibarrondo et al., 2023). We focus on HE-based BTPs (Drozdowski et al., 2019; Kolberg et al., 2020; Engelsma et al., 2022; Boddeti, 2018; Bauspieß et al., 2022) that compare and reveal inner product–based scores. To simulate this attack, we choose a pretrained ArcFace-like feature extractor (Deng et al., 2019) as it is considered a state-of-the-art face recognition model to generate the feature vectors known as templates before they are encrypted using HE-based BTPs. These feature vectors are d-dimensional and normalized. They can be seen as points spread on the surface of the unit d-ball, for which the similarity measure is based on the inner product (i.e., the cosine similarity).

We begin by defining our attack model, which includes a one-to-one attack scenario as well as a one-to-many attack scenario. We consider a semi-honest attacker, such as a service provider, *who does not collude with the database (DB) server but legitimately receives cleartext scores.* We show that this attacker can recover raw templates of target users using auxiliary knowledge while remaining semi-honest. This scenario falls within the threat model of HE-based BTPs (Drozdowski et al., 2019; Kolberg et al., 2020; Engelsma et al., 2022; Boddeti, 2018; Bauspieß et al., 2022). Unlike previous work (Mohanty et al., 2006), we demonstrate that this recovery is feasible for inner product-based template comparisons without requiring real data or training.

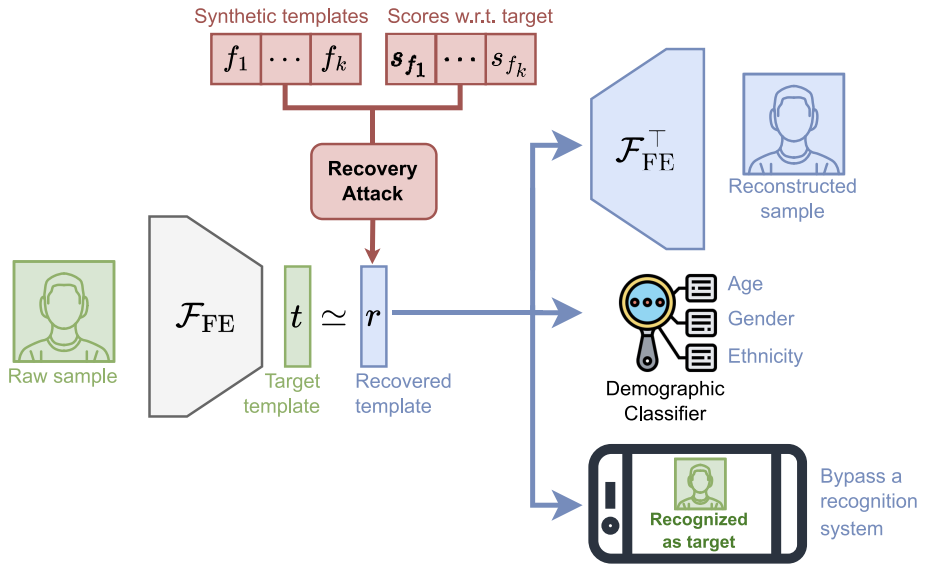


FIGURE 1
Overview of the capabilities of our template recovery attack using synthetic templates extracted from synthetic facial images. Further details of our attack can be found in Sections 2, 3.

Our attack follows the non-adaptive approach and consists of randomly selecting synthetic facial images (e.g., samples from the SFace dataset; Boutros et al., 2022) from which it derives their respective facial embeddings as normalized feature vectors of the same dimension as the target. Those synthetic facial embeddings[1] represent the attacker's auxiliary knowledge for which they will receive their respective scores. Using synthetic facial images allows us to demonstrate the real-world applicability of our attack, assuming that the synthetic facial images represent spoofed faces in reality. This assumption is relevant in scenarios in which attackers do not have direct access to face recognition systems or wish to avoid detection, thus resorting to external approaches that use spoofed faces to compromise these systems. In fact, this assumption is feasible as it was experimentally demonstrated in section IV-C by Shahreza and Marcel (2023a). The authors conducted a presentation attack using reconstructed facial images produced by a template-inversion model fed with facial templates. In this experiment, the reconstructed facial images were either printed photographs or displayed on the screen of an iPad and then placed in front of the capturing device (camera) to fool a face recognition system that tolerates 0.001% false match rate (FMR) with a success rate of approximately 85%.

Depending on the HE scheme, encrypted facial templates may require quantization, resulting in noisy scores from HE-based BTPs. To address this, we recover the target template by solving an optimization problem over synthetic templates and their noisy scores, employing the Lagrange multiplier method. This is done with the constraint that the target template is a normalized vector. Hence, this optimization's solution is a normalized vector approximate to the target.

Subsequently, we assess the effectiveness of our attack on three levels: (1) performance evaluation, (2) the amount of gender information recovered, and (3) the visual impact on raw sample reconstruction. Initially, we analyze our attack's performance by varying the number of synthetic facial templates across different quantization methods. Unlike (Mohanty et al., 2006) that tested a face recognition system with a threshold at 1% FMR, we choose stricter thresholds of 0.1%, 0.01%, and 0.001% FMR for evaluating our attack. Our results show that, in practice, an attacker has a 50% chance of bypassing such systems with a 100% success rate using only 50–195 revealed scores and synthetic facial templates to recover a 512-dim target vector, regardless of the quantization approach. Next, we analyze the risk of template recovery on gender information retrieval through the performance comparison of three gender classifiers logistic regression [LR], linear support vector machine [SVM], and radial basis function (RBF) kernel [SVM] tested on the original templates (our baseline) and recovered templates. Logistic regression and linear SVM to assess the linear separability of the templates according to gender, and the RBF kernel SVM to assess their non-linear separability. Our results show a difference between 0% and 13%; thus, more advanced gender classifiers would perform equally on the recovered and the original templates. Finally, we check the visual impact on reconstructed raw facial images from the recovered target templates, where we can visually see their resemblance, demographic information (gender expression and ethnicity), and facial artifacts (hair color), which is sufficient information for revealing the target's identity in real life.

In summary, we make the following contributions:

- We propose a non-adaptive template recovery attack that exploits score leakage using synthetic facial images as spoofed

1 In Bassit et al. (2023a), this attack was evaluated on random fake templates that are random normalized vectors.

faces, for which we define two attack scenarios for the one-to-one and one-to-many settings.
- We estimated that an attacker using only 50–192 revealed scores and synthetic facial templates can successfully bypass recognition systems with thresholds of various strictnesses.
- We analyze the recovered templates regarding facial image reconstruction and the amount of gender information they contain and demonstrate their accurate retrieval.

2 Attack model

Our focus is on HE-based BTPs that compute an inner product-based similarity measure under encryption to compare their encrypted templates. We base our attack model on the common architecture considered in most HE-based BTPs (Drozdowski et al., 2019; Kolberg et al., 2020; Engelsma et al., 2022; Boddeti, 2018; Bauspieß et al., 2022) that involves three semi-honest (a.k.a. honest-but-curious) parties, namely, the user, the DB server, and the service provider (SP).

Semi-honest security (Goldreich, 2009) is a widely used security model for multiparty cryptographic protocols. In this model, all the parties are assumed to strictly adhere to the protocol steps. Additionally, parties with private inputs aim to compute a joint function such that the protocol reveals nothing about their private inputs beyond the function's output. It is important to note that the semi-honest model makes no assumptions on (e.g., the correctness of) the parties' inputs; parties are allowed to freely choose or create their own inputs.

The user is a biometric data owner who wants to use a service (e.g., social service) offered by an SP (e.g., health care). The SP gives access to its services only if the user is biometrically recognized: biometric verification (one-to-one comparison) or biometric identification (one-to-many comparison). To comply with data privacy regulations (European Union, 2018) by not storing users' biometric data, in those HE-based BTPs (Drozdowski et al., 2019; Kolberg et al., 2020; Engelsma et al., 2022; Boddeti, 2018; Bauspieß et al., 2022), the SP delegates the biometric recognition task to the DB server as *a separate non-colluding entity* that stores encrypted reference templates for the SP and does not hold the decryption key. During the recognition, the DB server returns the encrypted scores to the SP. Subsequently, the SP uses its decryption key to reveal the scores on which it applies a threshold to make its decision.

Previous work (Drozdowski et al., 2019; Kolberg et al., 2020; Engelsma et al., 2022; Boddeti, 2018; Bauspieß et al., 2022) considers semi-honest security where they assume that the scores can be revealed to a semi-honest SP non-colluding with the DB server. They do not exclude any collusion assumptions between the user and the SP. In cryptography, a protocol secure against semi-honest attackers is a protocol that does not leak information about target parties' inputs other than what the protocol naturally outputs. In this case, a biometric recognition protocol (verification or identification) secure against semi-honest attackers (such as the SP) should not leak information about target parties' inputs (such as biometric reference or probe) other than what the protocol naturally outputs (such as a comparison scores revealed to the SP). Hence, an SP attacker that exactly follows the protocol, legitimately receives the comparison scores, and possesses auxiliary knowledge fits perfectly the definition of the semi-honest attacker, where they have no control over the protocol but has control over the inputs.

In this work, we study when revealing the scores to such an SP becomes a risk by defining two attack scenarios: (a) a non-adaptive attack in the one-to-one setting and (b) a non-adaptive template injection attack in the one-to-many setting, as illustrated in Figure 2. We define the attacker an SP that targets a specific user (or any user querying the system) in the one-to-one (or one-to-many) setting. The attacker aims to learn the target user's raw biometric template from the revealed scores they naturally receive and *some auxiliary knowledge they can acquire without colluding with the DB server and remaining semi-honest*.

Once this raw template is recovered, the attacker can use it to impersonate or retrieve the target user's demographic information or see its appearance. We define the auxiliary knowledge required by the attacker in both one-to-one and one-to-many scenarios as a set of synthetic facial templates obtained from a feature extractor similar to the system's extractor. These templates, which consist of feature vectors extracted from synthetic facial images assumed to mimic spoofed faces, are used by the SP attacker to obtain the respective scores in relation to the target user. Those scores represent the similarity between the target template and the synthetic templates. In other words, the attacker knows a set of synthetic templates and their similarity to the same target, which we demonstrate is sufficient information for recovering the target template. Note that such an attacker is captured by the threat model defined by Drozdowski et al. (2019), Kolberg et al. (2020), Engelsma et al. (2022), Boddeti (2018), and Bauspieß et al. (2022), where the DB server and the SP are non-colluding semi-honest parties.

The possible ways to acquire such auxiliary knowledge can be as follows: (a) For the one-to-one case, for instance, the SP can control synthetic users[2] to send synthetic probes, as one-to-one queries, to the DB server claiming the identity of the target user so that the SP attacker receives the scores relative to this reference target; and (b) for the one-to-many case, for example, the SP can control synthetic users to enroll only once in the DB server using synthetic references; thus, the SP is able to inject synthetic references into the DB so that it receives the scores relative to any target user's probe sent to the system as a one-to-many query.

In those examples, the SP attacker only acquires the auxiliary knowledge they need to run the attack without colluding with the DB server while sticking to the protocol and thus remaining semi-honest.

2.1 One-to-one attack scenario

In Figure 2a, the SP attacker's goal is to recover the raw reference template of a specific target user whose encrypted template is stored on the DB server. In this case, the SP attacker uses only the synthetic probes they have from their auxiliary knowledge and the respective scores they receives

2 If the architecture allows the SP to send queries to the DB server, then the SP can themselves send the synthetic probes to the DB server, claiming the target user's identity to receive the scores.

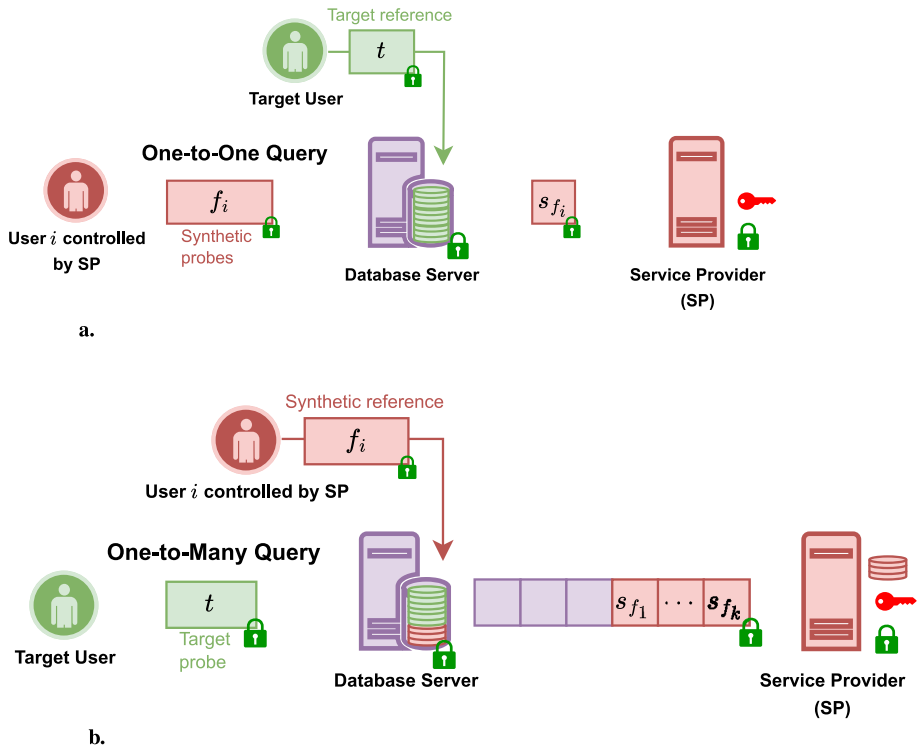


FIGURE 2
Illustration of our non-adaptive template recovery attack scenarios considering a semi-honest service provider attacker that has some auxiliary knowledge and legitimately receives cleartext scores. **(a)** One-to-one template recovery attack scenario in which the service provider attacker aims to recover the reference template of a specific target user using their auxiliary knowledge and the scores they legitimately receive. This auxiliary knowledge is a set of different synthetic probes that, for instance, the service provider can acquire by controlling synthetic users (spoofers) to send them to the database server claiming the target user's identity. **(b)** One-to-many template recovery attack scenario in which the attacker aims to recover the probe of any target user querying the system using their auxiliary knowledge and the scores they legitimately receive. This auxiliary knowledge is a set of different synthetic references that, for instance, can be acquired by controlling synthetic users (spoofers) to enroll in the database using synthetic references.

to recover the target reference template. This is a non-adaptive attack because the queries are sent independently from each other, unlike adaptive hill-climbing attacks, which adapt their actual query to its previous one. Note that some recognition systems limit the number of authentication attempts per user upon a limited number of rejections. To overcome this, the attacker sends their queries one batch after the other and waits between them for the number of authentication attempts to be reinitialized before sending the next batch.

2.2 One-to-many attack scenario

In Figure 2b, the SP attacker's goal is to recover the raw probe template of a target user who sends it encrypted to the DB server during the recognition phase. In this case, the SP attacker uses only the synthetic references they have from their auxiliary knowledge and the respective scores they receive and try to recover the target probe template. Thus, when a target user queries the DB server in a one-to-many manner, the SP attacker will receive the scores between the target template and the reference templates injected in the DB. Unlike the one-to-one attack scenario, this can be received in a single query, and the injected synthetic references are reused to recover any target probe.

Therefore, in both scenarios, we model the SP as a semi-honest attacker who does not collude with the database server. In Section 3, we describe our template recovery attack for this security model exploiting auxiliary knowledge and the cleartext comparison scores.

3 Our template recovery attack

Face recognition systems represent facial image samples as d-dim feature vectors resulting from well-trained deep neural networks (DNNs; i.e., feature extractors), which can be visualized as points spread in the embedding space. State-of-the-art DNNs, which demonstrate (near-)optimal recognition performance, use ArcFace-like loss (Deng et al., 2019) to train their models to represent samples as feature vectors on the surface of the unit d-ball, resulting in normalized feature vector representations. Thus, normalized vectors of dimension d can be seen as a point on the surface of the unit d-ball.

In this work, we consider the type of face recognition system that represents facial images as normalized d-dim feature vectors, such as DNNs trained with ArcFace-like loss. ArcFace loss enforces an angular margin between distinct identities on the surface of the

unit d-ball that ensures that feature vectors of the same identity are closely clustered, while different identities are further separated. This has achieved improved recognition performance compared to feature extractors trained with the conventional cross-entropy loss based on Softmax probabilities. Therefore, in our attack, we consider feature extractors that were trained using ArcFace-like losses. In those systems, the similarity between two normalized feature vectors is measured by their inner product to produce a similarity score. Our goal is to recover a target template T represented by a normalized d-dim feature vector resulting from those types of DNN for which the inner product is the similarity measure used to compare those vectors.

Let $F = [F_1, \cdots, F_k]$ be a column matrix of k synthetic templates F_i. Each synthetic template represents a normalized feature vector of dimension d. We denote $s_i =< T, F_i >= T^\intercal \cdot F_i$ the score corresponding to the inner product between the target T and a synthetic template F_i quantifying their similarity. We denote $S = (s_1, \cdots, s_k)$ a vector of k scores.

3.1 Optimization problem statement

Given k synthetic templates $\{F_i\}_{i\in[1,k]}$, which are normalized[3] d-dim vectors sampled at random, and their corresponding scores s_i with respect to the same target T, find the recovered template $\hat{T}$ such that $\hat{T}^\intercal \cdot \hat{T} = 1$. The recovered template $\hat{T}$ is found using Equation 1:

$$\min_{\hat{T}^\intercal \cdot \hat{T}=1} \quad \hat{T}^\intercal \cdot F - S. \tag{1}$$

By putting the constraint on $\hat{T}$ to be a normalized vector, we restrict the set of possible solutions to one solution that lies on the surface of the unit d-ball. This is because cosine similarity in most modern face recognition systems is measured over normalized vectors.

3.2 Our solution

This optimization problem translates to $\hat{T}^\intercal \cdot F_i = s_i$, which is basically minimizing $f(\hat{T}) = \sum_{i=1}^{k}(\hat{T}^\intercal \cdot F_i - s_i)^2$ subject to the constraint $g(\hat{T}) = \hat{T}^\intercal \hat{T} - 1$. We can solve this minimization problem in a simple way by using the Lagrange multiplier λ and forming the Lagrangian function $\mathcal{L}(\hat{T}, \lambda) = f(\hat{T}) + \lambda g(\hat{T})$. Given that

$$\mathcal{L}(\hat{T}, \lambda) = \sum_{i=1}^{k}(\hat{T}^\intercal \cdot F_i - s_i)^2 + \lambda(\hat{T}^\intercal \cdot \hat{T} - 1), \tag{2}$$

we partially derive Equation 2 with respect to the target vector's coordinates to obtain

$$\frac{\partial \mathcal{L}}{\partial \hat{t}_j} = \sum_{i=1}^{k} 2 \cdot F_{i,j} \cdot (\hat{T}^\intercal \cdot F_i - s_i) + 2\lambda \hat{t}_j. \tag{3}$$

3 We consider synthetic templates to be normalized to simulate their belonging to the same embedding space as the target.

We leave out the factor 2 from Equation 3 because we are interested in $\frac{\partial \mathcal{L}}{\partial \hat{T}} = 0$. Then, we calculate the partial derivative with respect to the target vector $\frac{\partial \mathcal{L}}{\partial \hat{T}}$ by assembling the partial derivatives with respect to its coordinates and rewriting them as $\frac{\partial \mathcal{L}}{\partial \hat{T}} = \sum_{i=1}^{k} F_i \cdot (\hat{T}^\intercal \cdot F_i - s_i) + \lambda \hat{T}$. Hence, the main equation to solve is

$$\left[FF^\intercal + \lambda I_d\right] \cdot \hat{T} - F \cdot S = 0. \tag{4}$$

As a result, the recovered template is $\hat{T}$, satisfying $\frac{\partial \mathcal{L}}{\partial \hat{T}} = 0$, that is, $\left[FF^\intercal + \lambda I_d\right] \cdot \hat{T} - F \cdot S = 0$ where I_d is the identity matrix of dimension d. We can solve Equation 4 for a large range of λ and choose the λ for which $\hat{T}^\intercal \cdot \hat{T} = 1$. For instance, this can be achieved using `scipy.optimize.fsolve` (The SciPy community, 2023). Hence, we can write $\hat{T} = \left[FF^\intercal + \lambda I_d\right]^{-1} \times F \cdot S$ for the recovered template. The pseudo-code of our attack is described in Algorithm 1.

Remark 3.1. (Solution [Non-] Unicity). Note that Equation 4 has a unique solution when $d \leq k$ and infinitely many solutions when $k < k$.

4 Experiments

We analyze revealed scores coming from HE-based BTPs, which, depending on the HE plaintext space, involve quantization techniques to adapt the feature values to the HE plaintext space. Hence, these quantization approaches can bring a certain level of noise to the actual score, making HE-based BTPs produce noisy scores. For instance, HE schemes (such as the BFV scheme) supporting only integers, their respective BTPs require a quantization either on feature-level, such as *precision-based quantization* (Boddeti, 2018), or on feature- and score-level *lookup table-based quantization* (Bassit et al., 2022, 2023b) to map their templates from floating points to integer values so that such encryption scheme can be applied. Other HE schemes (such as the CKKS scheme) supporting floating point encryption do not need quantization.

In the following, we vary the number of synthetic templates needed for the recovery to evaluate the performance of our attack in retrieving feature vectors that can either (1) bypass the

```
Input: k synthetic embeddings and their corresponding scores with respect to the target embedding
Output: T̂ the recovered target template
Combine the synthetic embeddings into a matrix F
Combine the scores into a vector S
Generate I_d an identity matrix of dimension d
Choose a λ for which T̂^⊺ · T̂ = 1          ▷ using scipy.optimize.fsolve
Solve [FF^⊺ + λI_d] · T̂ − F · S = 0        ▷ Equation 4
```

Algorithm 1. Pseudo-code of our attack.

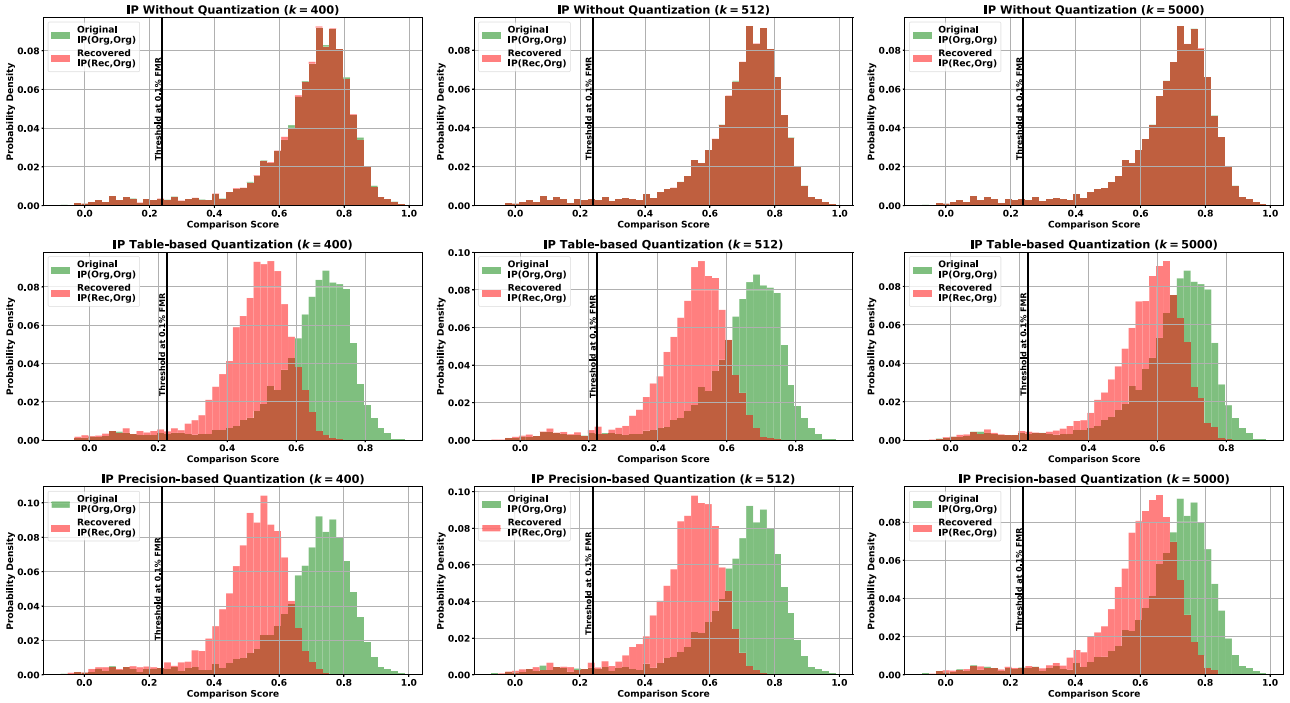


FIGURE 3
Comparison between the mated score distribution of original pairs (the green histogram) and mated score distribution of pairs where one of them is recovered (the red histogram) for the quantization cases: without **(first row)**, with table-based **(second row)**, and with precision-based **(third row)** using $k \in \{400, 512, 5,000\}$ synthetic templates. Original pairs from the LFW data set (Huang et al., 2008) produce similar distributions to those from the VGGFace2 data set plotted here. IP, inner product.

face recognition system (Section 4.1), (2) determine its gender (Section 4.2), or (3) reconstruct its raw facial image (Section 4.3).

For the quantization techniques over 512-dim normalized feature vectors, we use the same parameters reported in Bassit et al. (2022), Bassit et al. (2023b), and Boddeti (2018) with which they achieved good recognition performance, namely, the table-based quantization (the Multiplication-free Inner product (MFIP) table; Bassit et al., 2022, 2023b) with features quantized over 3 bits and a score quantization step of $\Delta = 0.001$, and for the precision-based quantization (Boddeti, 2018), precision $= 0.0025$. We have made the source code of our implementation publicly available.[4]

4.1 Performance evaluation

In our assessment, we use the VGGFace2 data set (Cao et al., 2018), a publicly available facial image dataset of real subjects captured in unconstrained settings, to simulate real-life template databases based on images of actual individuals. Additionally, we utilize the SFace data set (Boutros et al., 2022), a synthetic facial image data set, to model spoofing attack scenarios and the resources potentially available to attackers, enabling a comprehensive evaluation of both real and adversarial conditions. To evaluate our recovery attack, we use the ResNet-100 (He et al., 2016) pretrained with ArcFace (Deng et al., 2019) to extract normalized feature vectors of dimension 512 for (1) the target facial feature vectors from the VGGFace2 data set and (2) synthetic facial feature vectors as auxiliary information.

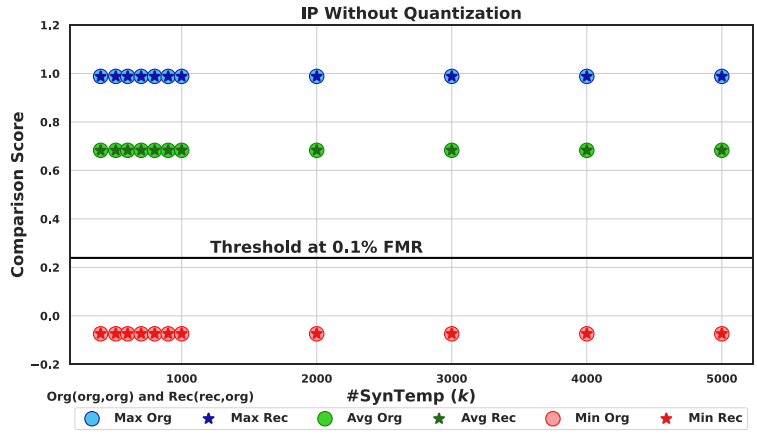
In this experiment, we aim to determine whether a recognition system will still recognize the recovered template as similar to the target template stored in the system. To replicate the case of a recovered template bypassing an inner product-based recognition system, in Figure 3, we compare the score distribution of mated pairs, where both pairs are original (the green histograms), with the score distribution of mated pairs, where one is original and the other is recovered using synthetic facial templates (the red histograms). We notice that starting from $k = 400$, both distributions completely overlap (first row of Figure 3) when no quantization interferes with the score computation. This means that the recovered templates will be treated (accepted or rejected) by a recognition system[5] as the original pairs, and thus, they can be used to bypass the system using a number of synthetic templates less than the dimensionality of the target feature vector, in this case, 400. Bassit et al.'s (2023a) study, using random normalized vectors as fake templates created a less noticeable overlap of both distributions at $k = 400$ but a full overlap starting from $k = 512$. Figure 4a confirms the complete overlap regarding the maximum (Max), average (Avg), and minimum (Min) scores, which is depicted by the superposition of the circles and the stars for all values of the number of synthetic templates k.

In the case of quantization, the second row (table-based) and the third row (precision-based) in Figure 3, the distributions are

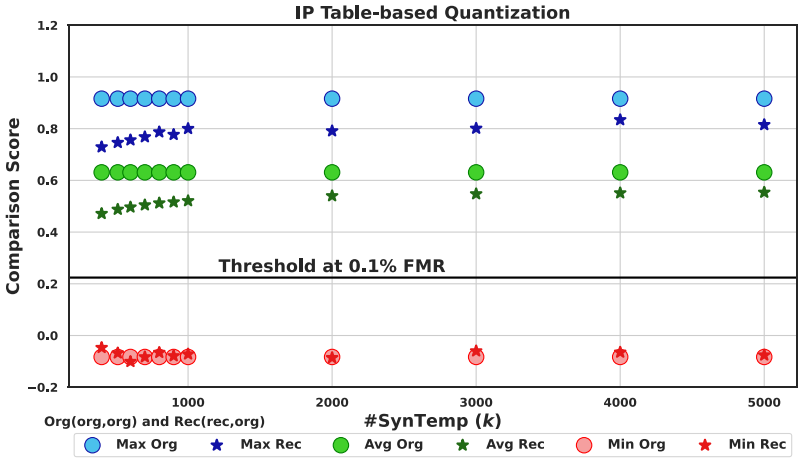
4 https://github.com/aminabassit/tra-he-btps

5 In this case, the system uses CKKS for encryption and reveals the scores.

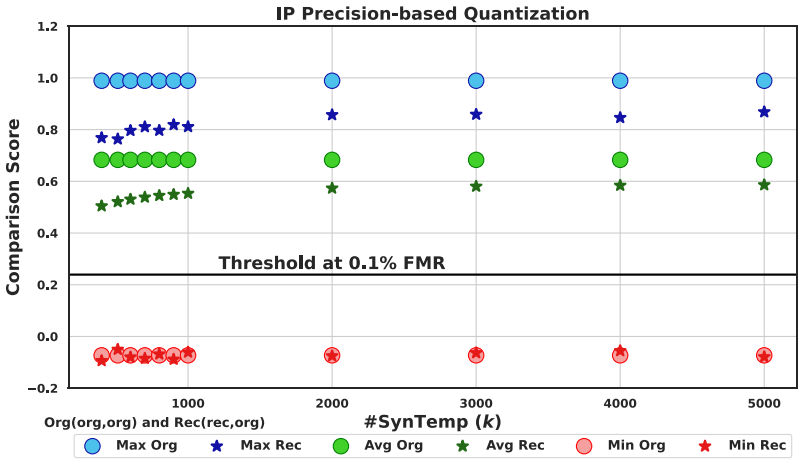
10.3389/fimag.2025.1476377



a. IP without quantization



b. IP with table-based quantization



c. IP with precision-based quantization

FIGURE 4
Mated score variation (maximum, average, and minimum) of original pairs (circles) and pairs with a recovered template (stars) using the inner product (IP) without quantization and with table-based and precision-based quantizations for $k \in [400, 5, 000]$. **(a)** IP without quantization. **(b)** IP with table-based quantization. **(c)** IP with precision-based quantization.

10.3389/fimag.2025.1476377

TABLE 1 Success rate in percentage (%) for various thresholds (θ).

	θ	400	512	600	900	1,000	3,000	4,000
Without	0.1% FMR	100	100	100	100	100	100	100
	0.01% FMR	100	100	100	100	100	100	100
	0.001% FMR	100	100	100	100	100	100	100
Table-based	0.1% FMR	98.39	98.73	99.00	99.21	99.23	99.46	99.46
	0.01% FMR	97.63	98.21	98.23	98.88	98.86	99.22	99.24
	0.001% FMR	68.08	74.59	78.43	83.92	85.10	90.77	91.44
Precision-based	0.1% FMR	98.62	98.87	98.96	99.06	99.23	99.31	99.41
	0.01% FMR	97.55	98.04	98.19	98.48	98.57	99.09	99.20
	0.001% FMR	69.80	76.77	79.90	84.68	85.65	90.70	91.75

clearly separable, and the distribution for the recovered case is gradually shifting to the right, increases, showing a slight increase of the overlap. This makes the produced scores more likely to be above the threshold (the black line) and thus are accepted, especially for $k = 5,000$. Unlike in Bassit et al.'s (2023a) study, where, for 5, 000 synthetic templates, table-based quantization approach achieves a full overlap and almost full overlap for precision-based one, which can be justified by the randomness effect and the increased number of fake templates. Figures 4b, c illustrate the gradual increase of the original-recovered mated score distributions for a number of synthetic templates varying between 400 and 5, 000. Although the overlap is not full, we notice that the average score for the recovered pairs is above a threshold chosen at 0.1% FMR, stricter than the threshold found in Mohanty et al.'s (2006) study.

4.1.1 Attack success rate

We denote x_{org} and y_{org} as two distinct mated templates and $\tilde{x}_{rec}$ as the recovered template of x_{org} using k synthetic templates. For a given threshold θ and a number of synthetic templates k, we define $N_{rec} = \left|\{\theta \leq \text{IP}(\tilde{x}_{rec}, y_{org})\}\right|$ as the number of $\tilde{x}_{rec}$ that would be accepted by the system and thus treated as mated templates and $N_{org} = \left|\{(\theta \leq \text{IP}(x_{org}, y_{org})\}\right|$ as the number of mated templates correctly accepted by the system. We define our attack success rate $\text{SR}(\theta, k)$ as

$$\text{SR}(\theta, k) = \frac{N_{rec}}{N_{org}} \tag{5}$$

the ratio of N_{rec} to N_{org}. The measured rate can be visually seen in the histograms of Figure 3 as the part of the red histogram on the right of the black line divided by the part of the green histogram on the right of the black line. Table 1 shows the success rate of our attack measured more htan 5,000 mated comparisons for different values of k. We notice that for the scores without quantization, starting at 400, the success rate reaches 100% for all three thresholds. In contrast, the success rate steadily increases in both quantization approaches. For a threshold at 0.1% FMR, it starts at approximately 98% at 400 to achieve approximately 99.4% at 4000. For a threshold at 0.01% FMR, it starts at approximately 97% at 400 to achieve approximately 99.2% at 4000. For a threshold at 0.001% FMR, it starts at approximately 68% at 400 to achieve approximately 91% at 4000. We observe that for stricter thresholds, at 0.01% FMR and 0.001% FMR, our attack requires more synthetic templates to reach confident success rates of 91% and 99%. However, our attack still achieves satisfying success rates ($\sim$ 74% $-$ 76%) using 512 synthetic templates, which corresponds to the dimensionality of the target vector.

In summary, our attack successfully recovers templates from leaked unquantized scores with a 100% success rate for a number of synthetic templates smaller than the target vector's dimensionality. However, for leaked quantized scores, our attack requires more synthetic templates to achieve a confident success rate. This suggests that quantization approaches act as a form of mitigation mechanism against our attack.

4.1.2 How much knowledge does an attacker need for a successful bypass?

In practice, when the number of synthetic templates k is inferior to d, then Equation 4 has an infinite number of solutions. However, our interest lies in solutions that satisfy the normality constraint, ensuring that the recovered template resides on the surface of the unit d-ball. This makes a recovered template using $k < d$ more likely to find a good enough approximation of the target that would be acceptable by the recognition system. To estimate this number in practice, we fix the success rate to 100% and select 200 distinct subjects. For each subject, we select two templates: a target template to be recovered and a mated template. We then evaluate whether the system recognizes the recovered template as the mated template, in other words, if the inner product of this mated template and the recovered one exceeds the system's threshold. For each subject, we iterate 10 times over a set of possible $k \in [50, 2, 000]$ with a step of 50. An iteration ends as soon as the first k satisfies the inner product of the mate template and a recovered template using k synthetic templates exceeds the threshold, reflecting the success rate that we fix to 100%.

Figure 5 shows bar plots of the median and mean numbers of synthetic templates measured over 10 iterations per subject for 200 different subjects. We observe that the mean is greater than the median for all bar plots, indicating that the mean overestimates the most frequent values of k. Thus, we rely on the median to estimate k. The median number of synthetic templates required to bypass an inner product-based recognition system with a threshold at 0.1% FMR for a 512-dim target vector from (1) non-quantized scores, (2) from table-based quantized scores, and (3) from precision-based quantized scores is 50 synthetic templates. For different quantization approaches, the estimated number of synthetic templates increases as the threshold gets stricter. For a threshold at 0.01% FMR, the median is between 55 and 60, while for a threshold at 0.001% FMR, it is between 130 and 192. This means that in practice, an attacker has a 50% chance of finding a sufficiently good approximation of the target template by using only between 50 and 195 synthetic templates to recover a 512-dim normalized target vector, comparable to Bassit et al.'s (2023a) study, where this attack required 60–165 fake templates.

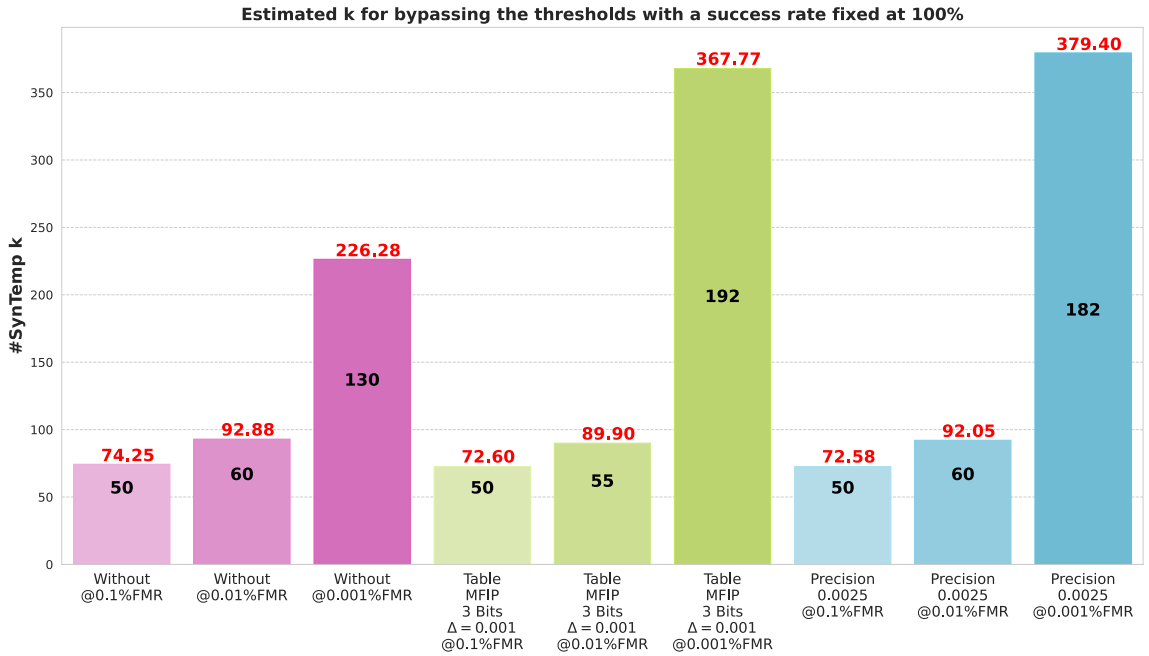


FIGURE 5
Estimated number of synthetic templates required for bypassing thresholds at 0.1%, 0.01%, and 0.001% false-match rate (FMR) for a 100% fixed success rate using different quantization approaches and measured for 200 subjects from the VGGFace2 data set with their median (numbers in black) and mean (numbers in red) over 10 iterations per subject to recover a 512-dim template.

4.2 Gender information analysis

In this section, we aim to analyze how much gender information can be retrieved from recovered templates using various numbers of synthetic templates and different quantization approaches. Our aim in this gender information analysis is to consider the worst-case scenario in which both genders are balanced. This approach allowed us to objectively evaluate the effectiveness of our attack in recovering gender information while avoiding potential biases that could arise from fixing the gender to one group. To remove risks of overfitting and bias on one class, in this analysis, we select a gender-balanced subset from the LFW data set (Huang et al., 2008), which is a publicly available data set with facial images captured in unconstrained settings that includes 5,934 images of females and males (2,967 for each gender). We use this subset to generate the original embeddings that are facial feature vectors of dimension 512 extracted using ResNet-100 (He et al., 2016) trained with ArcFace (Deng et al., 2019), which we then normalize. To simulate different levels of the attacker's resources, we generate the recovered templates for the three quantization cases using 512, 1,000, and 3,000 synthetic templates and their respective scores. Then, we test the linear SVM and RBF SVM and LR as three gender classifiers (two linear classifiers and one non-linear) trained with their default parameters to measure the performance difference between the original and recovered templates. This comparison serves as an indicator of how much gender information can be retrieved from the recovered templates.

Table 2 shows the gender classifiers' performance results expressed in terms of accuracy over a threefold cross-validation setting. The results in this table highlight the impact of different quantization approaches on the accuracy of gender classification over recovered templates from using synthetic templates extracted from the SFace data set.

In the absence of quantization, the recovered templates' accuracy is consistently high and close to the original accuracy for all classifiers, especially for higher values of $k = 3,000$.

TABLE 2 Gender classification performance measuring the difference between original and recovered templates in terms of accuracy (%) for different numbers of synthetic templates (k) and scores tested on a gender-balanced subset from the LFW data set.

Classifier	Original	k	Recovered		
			Quantization approach		
			Without	Precision-based	Table-based
LR	**70.86**	512	70.86	61.42	62.03
		1,000	70.84	63.41	62.67
		3,000	**70.87**	**65.09**	**65.21**
Linear SVM	**72.05**	512	71.97	62.18	63.04
		1000	72.02	64.39	63.97
		3,000	**72.05**	**65.52**	**67.07**
RBF SVM	**89.72**	512	89.63	76.55	76.76
		1,000	89.70	79.62	78.96
		3,000	**89.72**	**82.27**	**82.27**

The bold values indicate the best performances for a classifier among k per quantization approach. LR, linear regression; SVM, support vector machine; RBF, .

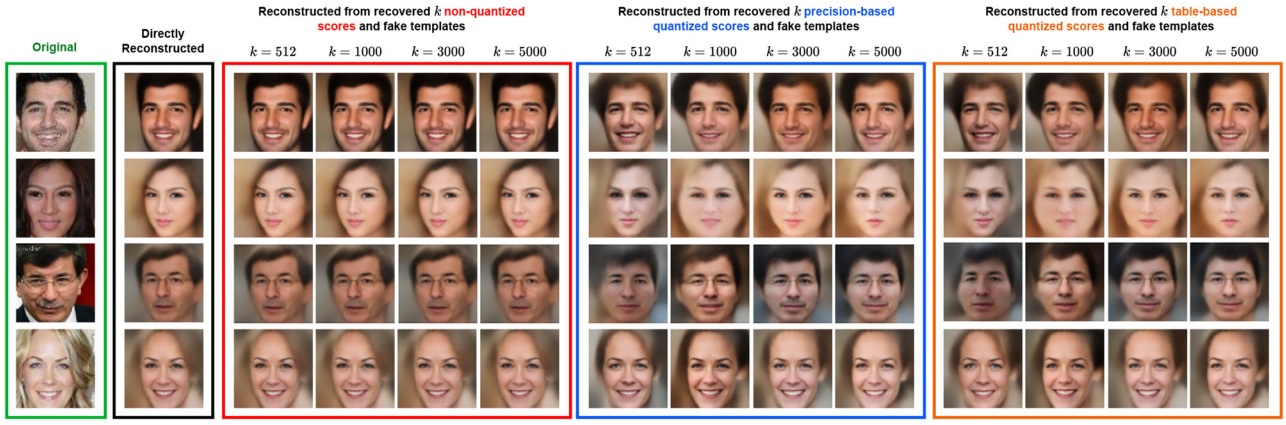


FIGURE 6
Image reconstruction of recovered target templates using $k \in \{512, 1,000, 3,000, 5,000\}$ synthetic templates and scores over three different quantization approaches. The target templates are 512-dim normalized feature vectors extracted using ResNet-100, and the images are from the VGGFace2 data set (https://www.kaggle.com/datasets/hearfool/vggface2) reconstructed using the inversion model for low-resolution images from Shahreza et al. (2024).

For precision-based and table-based quantization approaches, they generally reduce accuracy, although table-based quantization tends to perform slightly better than precision-based quantization across most of the values of k.

For classifier performance, LR has an original accuracy of 70.86%. The accuracy of recovered templates without quantization is nearly the same across all k values. However, both precision and table-based quantization approaches lead to a noticeable drop in accuracy, between 5% and 13%, with table-based quantization performing marginally better. The linear SVM has an original accuracy of 72.05%, and similar to the LR, the accuracy of recovered templates without quantization remains close to the original, while, the precision and table-based quantization approaches reduce the accuracy by 6%–9%, with table-based quantization showing a slight edge. The RBF SVM, with an original accuracy of 89.72%, the highest among the classifiers, shows that the accuracy in the absence of quantization remains very close to the original. Precision and table-based quantization approaches also decrease in accuracy by 7%–13%, but less so than for the LR and the linear SVM, with both approaches performing similarly well.

Overall, for each classifier, the best performance is achieved in the absence of quantization and remains very close to the original accuracy. Among the quantization approaches, table-based quantization often provides slightly better accuracy than precision-based quantization. The RBF SVM classifier is the most resilient to quantization approaches, maintaining higher accuracy compared to LR and linear SVM.

In summary, these results are comparable to Bassit et al.'s (2023a) and suggest that using quantization can make it more challenging for a linear classifier to correctly determine gender from recovered templates. However, for non-linear classifiers, gender classification from recovered templates remains relatively straightforward regardless of the quantization approach, highlighting the accurate retrieval of gender information using our recovery attack.

It is important to note that this analysis does not aim to enhance existing gender classifiers. Instead, it seeks to demonstrate the risk of template recovery on inferring gender information by comparing the accuracy difference between the original templates (our baseline) and the recovered templates, which ranges from 0% to 13%. Therefore, advanced gender classifiers would likely perform similarly on both recovered and original templates.

4.3 Image reconstruction analysis

In this section, we examine how much our recovery attack visually impacts the raw image reconstruction. We demonstrate this by reconstructing raw facial images from recovered target templates for various numbers of synthetic templates and scores for different quantization approaches. We emphasize the fact that our goal is not to improve on the existing work regarding image reconstruction attacks from feature vectors but rather to illustrate the visual quality of the recovered target template using our attack, regardless of the performance of the image reconstruction method used to invert feature vectors. For this experiment, we use the VGGFace2 data set for the target facial templates that we extract using ResNet-100.

It is important to note that our recovered templates can be inverted to raw images by any template-inversion model. To demonstrate the quality of our recovered templates, we tested them on two different state-of-the-art template-inversion models. We use the inversion model in (Shahreza and Marcel, 2023b) that is GAN-based and reconstructs high-resolution images and use the inversion model in (Shahreza et al., 2024) that is based on DSCasConv and reconstructs low-resolution images. We retrained both on normalized feature vectors over 10 epochs and 80 epochs, respectively.

Figure 6 shows the reconstructed facial images using the inversion model presented by Shahreza et al. (2024) and Figure 7 shows the reconstructed facial images using the inversion model presented by Shahreza and Marcel (2023b). In both figures, we selected the number of synthetic templates k starting from the dimensionality of the feature vector and progressively increasing

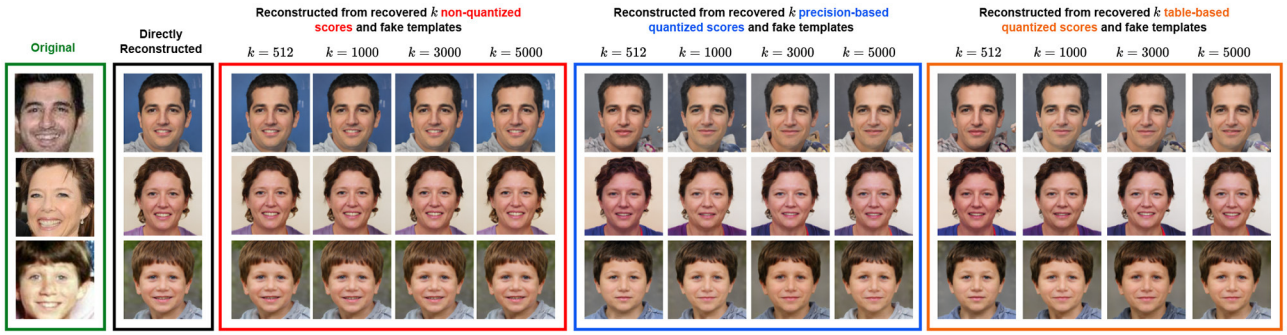


FIGURE 7
Image reconstruction of recovered target templates using $k \in \{512, 1,000, 3,000, 5,000\}$ synthetic templates and scores over three different quantization approaches. The target templates are 512-dim normalized feature vectors extracted using ResNet-100, and the images are from the VGGFace2 data set (https://www.kaggle.com/datasets/hearfool/vggface2) reconstructed using the inversion model for high-resolution images from Shahreza and Marcel (2023b).

it with large steps to evaluate whether the inversion improves as k increases. We use both pretrained models to reconstruct images and compare the directly extracted and non-recovered feature vectors against the recovered feature vectors. Note that our comparison references for this experiment are the reconstructed images from feature vectors directly extracted from ResNet-100 (the black box in both figures). In other words, both figures should be read by comparing the resemblance between the images in the black box and the images in the red, blue, and orange boxes. The comparison between images in the green box and the black box shows the performance of the inversion models, which is less relevant to the purpose of this experiment.

Both figures show identical gender expression and ethnicity in the reconstructed images from the directly extracted feature vectors and the recovered ones, regardless of the quantization approach. Also, the higher k is, the more similar the reconstructed images from recovered synthetic templates and scores become (the column $k = 5,000$ of the colored boxes). In both figures, we notice that the recovery from non-quantized scores for all k produces reconstructed images identical to the ones that are directly reconstructed from the original feature vector. For the recovery from precision and table-based quantized scores for all k, we observe that in Figure 6, the facial artifacts (i.e., hair) with a change in facial expression (i.e., smiling) are retrieved; however, the age is not fully retrieved as in the third row, the resulting facial images look younger than the original one. In Figure 7, the quantization approaches slightly the reconstructed images regarding the image background and hair color.

Overall, the recovered target templates from our attack can be inverted to reconstruct their corresponding raw facial images. These reconstructions retain enough demographic information and facial artifacts to enable an attacker to recognize or form a reasonable idea of the target subject's appearance in real life. Additionally, similarly to Shahreza and Marcel's (2023a) work, an attacker can also use the target reconstructed facial images to perform a presentation attack by either printing them as photographs or displaying them on the screen of an iPad and then presenting them in front of the capturing device (camera) to fool a face recognition system that tolerates 0.001% FMR.

5 Conclusion

In this article, we demonstrate that inner product-based homomorphically encrypted biometric recognition systems deliberately leaking cleartext scores are vulnerable to a template recovery attack. We present two non-adaptive template recovery attack scenarios suitable for different threat models, the one-to-one and the one-to-many settings. By using synthetic faces as spoofed faces, we show our attack's applicability in a real-world scenario in which, the adversary attacks the recognition system from outside. We assess the efficacy of our attack across three key levels: recovery performance in bypassing the system, retrieving gender information, and the visual impact on image reconstruction. Our results highlight the gravity of revealed scores as we showed that no matter which quantization approach HE-based BTPs use, an attacker needs at most 192 cleartext scores and synthetic templates for a successful recovery. Therefore, the scores must be hidden as a potential countermeasure to mitigate our attack. Throughout this work, our goal was to emphasize the significance of restricting the leakage level that HE-based biometric recognition systems can tolerate, as we demonstrated that even a minor leak, such as the score, can result in the recovery of a target template. Once a target template is retrieved, it cannot be mitigated even if it is replaced because the recovered template and its neighboring embeddings remain valid for bypassing the system.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in

this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s), nor the minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

AB: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. FH: Conceptualization, Supervision, Writing – review & editing. ZR: Software, Writing – review & editing. HS: Software, Writing – review & editing. RV: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Supervision, Validation, Writing – review & editing. AP: Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was partly supported by the PriMa project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 860315. This work was partly supported by the H2020 TReSPAsS-ETN Marie Skłodowska-Curie early training network (grant agreement 860813).

Acknowledgments

This paper is an extension of Bassit et al. (2023a) published at the IEEE International joint conference on biometrics (IJCB) 2023.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Acien, A., Morales, A., Vera-Rodriguez, R., Bartolome, I., and Fierrez, J. (2018). "Measuring the gender and ethnicity bias in deep models for face recognition," in *Iberoamerican Congress on Pattern Recognition* (Springer), 584–593. doi: 10.1007/978-3-030-13469-3_68

Adler, A. (2004). "Images can be regenerated from quantized biometric match score data," in *Canadian Conference on Electrical and Computer Engineering 2004 (IEEE Cat. No. 04CH37513)* (IEEE), 469–472. doi: 10.1109/CCECE.2004.1345057

Adler, A. (2005). "Vulnerabilities in biometric encryption systems," in *Audio-and Video-Based Biometric Person Authentication: 5th International Conference, AVBPA 2005, Hilton Rye Town, NY, USA, July 20-22, 2005* (Springer), 1100–1109.

Bassit, A., Hahn, F., Peeters, J., Kevenaar, T., Veldhuis, R., and Peter, A. (2021). Fast and accurate likelihood ratio-based biometric verification secure against malicious adversaries. *IEEE Trans. Inf. Forens. Secur.* 16:5045–5060. doi: 10.1109/TIFS.2021.3122823

Bassit, A., Hahn, F., Rezgui, Z., Kelly, U., Veldhuis, R., and Peter, A. (2023a). "Template recovery attack on homomorphically encrypted biometric recognition systems with unprotected threshold comparison," in *2023 IEEE International Joint Conference on Biometrics (IJCB)* IEEE. doi: 10.1109/IJCB57857.2023.10449211

Bassit, A., Hahn, F., Veldhuis, R., and Peter, A. (2022). "Multiplication-free biometric recognition for faster processing under encryption," in *2022 IEEE International Joint Conference on Biometrics (IJCB)*, 1–9. doi: 10.1109/IJCB54206.2022.10007958

Bassit, A., Hahn, F., Veldhuis, R., and Peter, A. (2023b). Improved multiplication-free biometric recognition under encryption. *IEEE Trans. Biometr. Behav. Ident. Sci.* 6, 314–325. doi: 10.1109/TBIOM.2023.3340306

Bauspieß, P., Olafsson, J., Kolberg, J., Drozdowski, P., Rathgeb, C., and Busch, C. (2022). "Improved homomorphically encrypted biometric identification using coefficient packing," in *2022 International Workshop on Biometrics and Forensics (IWBF)* (IEEE), 1–6. doi: 10.1109/IWBF55382.2022.9794523

Boddeti, V. N. (2018). "Secure face matching using fully homomorphic encryption," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)* (IEEE), 1–10. doi: 10.1109/BTAS.2018.8698601

Boutros, F., Huber, M., Siebke, P., Rieber, T., and Damer, N. (2022). "SFace: privacy-friendly and accurate face recognition using synthetic data," in *2022 IEEE International Joint Conference on Biometrics (IJCB)* (IEEE). doi: 10.1109/IJCB54206.2022.10007961

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., and Zisserman, A. (2018). "Vggface2: a dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)* (IEEE), 67–74. doi: 10.1109/FG.2018.00020

Chitrapu, P., and Kalluri, H. K. (2023). "A survey on homomorphic encryption for biometrics template security based on machine learning models," in *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)* (IEEE), 1–6. doi: 10.1109/SCEECS57921.2023.10062968

Deng, J., Guo, J., Xue, N., and Zafeiriou, S. (2019). "Arcface: additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4690–4699. doi: 10.1109/CVPR.2019.00482

Drozdowski, P., Buchmann, N., Rathgeb, C., Margraf, M., and Busch, C. (2019). "On the application of homomorphic encryption to face identification," in *2019 International Conference of the Biometrics Special Interest Group (Biosig)* (IEEE), 1–5.

Engelsma, J. J., Jain, A. K., and Boddeti, V. N. (2022). HERS: homomorphically encrypted representation search. *IEEE Trans. Biometr. Behav. Ident. Sci.* 4, 349–360. doi: 10.1109/TBIOM.2021.3139866

European Union (2018). *The General Data Protection Regulation (Regulation (EU) 2016/679)*. Available online at: https://gdpr-info.eu/ (accessed April 15, 2025).

Galbally, J., Fierrez, J., Ortega-Garcia, J., McCool, C., and Marcel, S. (2009). "Hill-climbing attack to an eigenface-based face verification system," in *2009 First IEEE International Conference on Biometrics, Identity and Security (BIdS)* (IEEE), 1–6. doi: 10.1109/BIDS.2009.5507530

Galbally, J., McCool, C., Fierrez, J., Marcel, S., and Ortega-Garcia, J. (2010). On the vulnerability of face verification systems to hill-climbing attacks. *Pattern Recogn.* 43, 1027–1038. doi: 10.1016/j.patcog.2009.08.022

Goldreich, O. (2009). *Foundations of Cryptography: volume 2, Basic Applications*. Cambridge: Cambridge University Press.

Gomez-Barrero, M., Galbally, J., Fierrez, J., and Ortega-Garcia, J. (2011). "Hill-climbing attack based on the uphill simplex algorithm and its application to signature verification," in *Biometrics and ID Management: COST 2101 European Workshop, BioID 2011, Brandenburg (Havel)* (Springer), 83–94. doi: 10.1007/978-3-642-19530-3_8

Gomez-Barrero, M., Galbally, J., Fierrez, J., and Ortega-Garcia, J. (2012). "Face verification put to test: a hill-climbing attack based on the uphill-simplex algorithm," in *2012 5th IAPR International Conference on Biometrics (ICB)* (IEEE), 40–45. doi: 10.1109/ICB.2012.6199756

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. doi: 10.1109/CVPR.2016.90

Huang, G. B., Mattar, M., Berg, T., and Learned-Miller, E. (2008). "Labeled faces in the wild: a database for studying face recognition in unconstrained environments," in *Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition*.

Ibarrondo, A., Chabanne, H., Despiegel, V., and Önen, M. (2023). "Grote: group testing for privacy-preserving face identification," in *Proceedings of the Thirteenth ACM Conference on Data and Application Security and Privacy*, 117–128. doi: 10.1145/3577923.3583656

Kolberg, J., Drozdowski, P., Gomez-Barrero, M., Rathgeb, C., and Busch, C. (2020). "Efficiency analysis of post-quantum-secure face template protection schemes based on homomorphic encryption," in *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)* (IEEE), 1–4.

Mohanty, P., Sarkar, S., and Kasturi, R. (2006). "Privacy security issues related to match scores," in *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)* (IEEE), 162. doi: 10.1109/CVPRW.2006.163

Rathgeb, C., Kolberg, J., Uhl, A., and Busch, C. (2023). Deep learning in the field of biometric template protection: an overview. *arXiv preprint arXiv:2303.02715*.

Sandhya, M., and Prasad, M. V. (2017). "Biometric template protection: a systematic literature review of approaches and modalities," in *Biometric Security and Privacy. Signal Processing for Security Technologies*, eds. R. Jiang, S. Al-maadeed, A. Bouridane, P. D. Crookes, A. Beghdadi (Cham: Springer).

Shahreza, H. O., Hahn, V. K., and Marcel, S. (2024). Vulnerability of state-of-the-art face recognition models to template inversion attack. *IEEE Trans. Inf. For. Secur.* 19, 4585–4600. doi: 10.1109/TIFS.2024.3381820

Shahreza, H. O., and Marcel, S. (2023a). Comprehensive vulnerability evaluation of face recognition systems to template inversion attacks via 3D face reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 14248–14265. doi: 10.1109/TPAMI.2023.3312123

Shahreza, H. O., and Marcel, S. (2023b). "Face reconstruction from facial templates by learning latent space of a generator network," in *Thirty-seventh Conference on Neural Information Processing Systems*.

Soutar, C. (2002). *Biometric system security*. White Paper, Bioscrypt. Available online at: http://www.bioscrypt.com

The SciPy community (2023). *Scipy: open-source software for mathematics, science, and engineering*. Available online at: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.fsolve.html (accessed February 17, 2025).

Zhang, Y., Jia, R., Pei, H., Wang, W., Li, B., and Song, D. (2020). "The secret revealer: generative model-inversion attacks against deep neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 253–261. doi: 10.1109/CVPR42600.2020.00033