PRO-Face C: Privacy-Preserving Recognition of Obfuscated Face via Feature Compensation

Lin Yuan, Wu Chen, Xiao Pu, Yan Zhang, Hongbo Li, Yushu Zhang, *Senior Member, IEEE*, Xinbo Gao, *Fellow, IEEE*, and Touradj Ebrahimi, *Fellow, IEEE*

***Abstract*— The advancement of face recognition technology has delivered substantial societal advantages. However, it has also raised global privacy concerns due to the ubiquitous collection and potential misuse of individuals' facial data. This presents a notable paradox: while there is a societal demand for a robust face recognition ecosystem to ensure public security and convenience, an increasing number of individuals are hesitant to release their facial data. Numerous studies have endeavored to find such a utility-privacy trade-off, yet many struggle with the dilemma of prioritizing one at the expense of the other. In response to this challenge, this paper proposes PRO-Face C, a novel paradigm for privacy-preserving recognition of obfuscated faces via a dedicated feature compensation mechanism, aimed at optimizing the equilibrium between privacy preservation and utility maximization. The proposed approach is characterized by a specialized client-server architecture: the client transmits only obfuscated images to the server, which then performs identity recognition using a pre-trained model in conjunction with a suite of privacy-free complementary features. This framework facilitates accurate face identification while safeguarding the original facial appearance from explicit disclosure. Furthermore, the obfuscated image retains its visualization capability, crucial for image preview functionalities. To ensure the desired properties, we have developed an identity-guided feature compensation mechanism, complemented by several privacy-enhancing techniques. Extensive experiments conducted across multiple face datasets underscore the effectiveness of the proposed approach in diverse scenarios.**

***Index Terms*— Face recognition, image obfuscation, privacy protection, utility.**

Manuscript received 20 October 2023; revised 13 February 2024 and 25 March 2024; accepted 5 April 2024. Date of publication 16 April 2024; date of current version 7 May 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62201107, Grant U22A2096, Grant 62036007, Grant 62221005, Grant 62441601, and Grant 62376047; in part by the Natural Science Foundation of Chongqing under Grant CSTB2022NSCQ-MSX1265, Grant CSTB2022NSCQ-MSX1342, and Grant CSTB2023NSCQ-LZX0061; and in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJQN202300606 and Grant KJQN202300619. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fernando Alonso-Fernandez. *(Lin Yuan and Wu Chen contributed equally to this work.) (Corresponding authors: Xinbo Gao; Xiao Pu.)*

Lin Yuan, Xiao Pu, Yan Zhang, Hongbo Li, and Xinbo Gao are with Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: yuanlin@cqupt.edu.cn; puxiao@cqupt.edu.cn; yanzhang1991@cqupt.edu.cn; lihongbo@cqupt.edu.cn; gaoxb@cqupt.edu.cn).

Wu Chen is with the School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: chenwu.c@foxmail.com).

Yushu Zhang is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 210016, China (e-mail: yushu@nuaa.edu.cn).

Touradj Ebrahimi is with the Multimedia Signal Processing Group, École Polytechnique Fédérale de Lausanne (EPFL), 1005 Lausanne, Switzerland (e-mail: touradj.ebrahimi@epfl.ch).

Digital Object Identifier 10.1109/TIFS.2024.3388976

I. Introduction

THE technology of face recognition has brought about significant benefits across various applications including authentication, access control and criminal pursuit. Negatively, it has also sparked global concerns regarding privacy due to the excessive collection and potential abuse of individuals' facial data. Debates surrounding the privacy and legality issues of facial capture and recognition systems persist [1], [2], highlighting a profound paradox between public utility and personal privacy. On one hand, it is clearly in the public interests to maintain a functioning face recognition ecosystem to address security needs and enhance social convenience. On the other hand, we have seen a decreasing number of individuals who wish to relinquish their biometric data freely, especially when it comes to automatic recognition services [3]. In essence, while facial recognition offers undeniable benefits, its implementation must be guided by robust privacy safeguards and ethical considerations to mitigate potential privacy risks. However, achieving a balanced *Utility-Privacy* trade-off is inherently complex due to the perpetual tension between data analytics and information preservation.

A considerable amount of research efforts have been devoted to deal with the privacy issues in the context of face capturing and recognition. Proposed approaches involve image sensors with built-in privacy enhancement [4], [5], adversarial methods to confound recognition systems [6], [7], [9], [10], privacy-preserving photo sharing [11], [12], face template protection [13], [14], cancelable biometrics [15], and generative face anonymizations [16], [17], [18], [20], [21], [22], [23]. Yet, many of them attempt to mitigate privacy concerns by reducing the biometric recognition capability of captured images, resulting in a severe compromise on either privacy or utility. With the aim of achieving a balance between facial privacy and recognition utility, two groups of methodologies have been developed, namely *Privacy-preserving Face Recognition (PPFR)* and *Identifiable Face Anonymization (IDFA)*. As Fig. 1 illustrates, the former primarily carries out face recognition in a transformed domain, such as encrypted codes or frequency components, the performance of which is usually

1556-6021 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

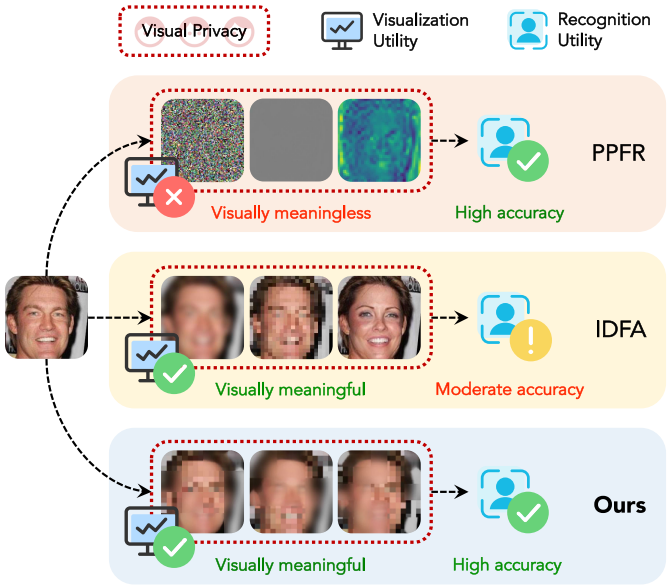

Fig. 1. Comparison of existing paradigms for visual privacy protection in the context of face recognition reveals distinct strategies: Privacy-Preserving Face Recognition (PPFR) prioritizes recognition performance while sacrificing the visual interpretability of protected images. Identifiable Face Anonymization (IDFA) concentrates on obfuscating or anonymizing the facial appearance against humans while retaining specific identifiable features in the image, but at the cost of machinery recognition accuracy. In contrast, the proposed approach seeks to bridge the gap between the two to strike a balance between the visual privacy, image utility, and recognition performance.

high thanks to the lossless transformations applied. However, this type of methods retain a low level of visualization utility as the protected data mostly presents highly randomized visual patterns. The latter attempts to morph the visual appearance of a given face in a more natural way, with face recognition achieved by retaining certain identifiable features inside the protected content. While, its recognition performance has to be compromised, due to the intrinsic contradiction between visual pattern analytics and anonymization.

To bridge the gap between the two, we propose a novel paradigm to solve the privacy-utility paradox in a more balanced manner. Our proposal is **PRO-Face C**, an acronym for **P**rivacy-preserving **R**ecognition of **O**bfuscated **Face** image via feature **C**ompensation, which essentially features a client-server framework that accomplishes face recognition in a collaborative manner. In PRO-Face C, a face image captured by the client device (smartphone, webcam, etc.) is first obfuscated by conventional and intuitive privacy filters, resulting in privacy-protected content that can be safely uploaded to an untrusted server for preview or recognition. To enable the recognition functionality, a feature compensation mechanism is utilized to integrate a set of privacy-free feature maps as the complementary information for the server-side recognition system to infer the correct identity from the obfuscated face with high confidence. In such a way, the framework allows for face recognition on the untrusted server without exposing the original facial appearance in any clear form. Meanwhile, the obfuscated image still presents meaningful visual content as a face, for the purpose of image preview. Furthermore, the functioning of the framework only relies on a pre-trained recognition backbone on the server, featuring a good level of generality in terms of system deployment. The contributions of this paper are summarized as follows:

- A novel paradigm, founded on a specialized client-server architecture, has been introduced to address the privacy-utility trade-off within the realm of face capture and recognition. This paradigm ensures comprehensive visual privacy, high recognition accuracy, effective image visualization, and broad applicability simultaneously.
- An identity-guided feature compensation mechanism accompanied by a suite of distinct privacy-enhancing techniques have been developed. These innovations are designed to uphold high recognition performance while ensuring visual privacy throughout the internal pipeline of the framework.
- Comprehensive evaluations conducted across multiple image datasets have been presented to showcase the effectiveness of the proposed approach in simultaneously enhancing privacy protection and maximizing utility across diverse scenarios, surpassing state-of-the-art methods.

II. Related Work

In this section, we review two categories of research highly relevant to ours and introduce the motivations of our design.

A. Privacy-Preserving Face Recognition (PPFR)

To tackle the privacy risk associated with face recognition systems, several efforts have been devoted to *privacy-preserving face recognition (PPFR)*, a category of methodologies that perform recognition in a protected domain. The PPFR methods can be roughly categorized into two groups:

1) Cryptography-Based Methods: These methods primarily conduct face recognition in the encrypted domain, where core operations such as feature extraction and similarity calculation are either performed directly on encrypted images or executed under specific secure protocols. Cryptographic primitives applied include garbled circuits (GC) [24], one-time-pad [25], matrix encryption [26], functional encryption [27], stable hash [28], homomorphic encryption (HE) [29], and secure multiparty computation (MPC) [30], [31].

Cryptographic methods cause little degradations to the recognition accuracy since most operations are strictly lossless. Their security is also strongly guaranteed by provable cryptographic primitives. However, the image visualization utility is lost significantly due to the encryption applied. The practical usage of such methods is also limited due to intolerable computation and communication overheads. Moreover, most methods have to be closely coupled with specific face recognition systems.

2) Transformation-Based Methods: Another mainstream of research attempts to perform face recognition based on transformed image formats. Frequency-domain learning is an actively investigated topic, where visually indistinguishable frequency-domain components are utilized to obtain meaningful representations for recognition. Wang et al. [32] proposed PPFR-FD to identify those frequency channels in DCT domain

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

that are important for visualization but less crucial for recognition, based on which, a masking mechanism is applied to retain only parts of the DCT channels for recognition. Ji et al. [33] learned privacy budget allocation via differential privacy, and added noise to the high-frequency DCT components to achieve a good balance between privacy and recognition. Mi et al. [34] proposed DuetFace, a collaborative client-server framework where high-frequency components are sent to server along with compensatory information to assist the server in achieving high recognition accuracy. Later on, PartialFace [35] has been proposed as an extension to DuetFace [34], which further conceals the visual information by pruning human-perceivable low-frequency components. Generative methods are also utilized to learn obscure facial features that are recognizable. Li et al. [36] proposed LightFace, a lightweight face recognition algorithm based on synthetic facial data generated by Bayesian GAN and depth-wise separable convolutions. Wang et al. [37] proposed FaceMAE, where face recognition is performed on images reconstructed from randomly masked face relying on masked auto-encoders (MAE). Recently, Mi et al. [38] proposed MinusFace, which achieves a strong privacy protection effect with high recognition accuracy based on trainable feature subtraction in image compression domain. In addition to [33], differential privacy is widely employed in various methods [25], [39], [40], [41] where raw images or representations are perturbed by slight noises to make them less differentiable visually.

Most transformation-based methods have to make a compromise between privacy and recognition accuracy since the transformations applied would inevitably induce information loss to different degrees. Moreover, the transformed data usually loses visualization utility similar to the cryptography-based methods, making it inapplicable for image preview.

B. Identifiable Face Anonymization (IDFA)

Face anonymization has long been studied with the aim of morphing the facial appearance to make it anonymous in terms of facial attributes such as identity. A good number of studies exist, including image obfuscations [4], [5], [42] and face synthesis [16], [17], [18], [19], [21], [22], [23]. A distinct subset of these methods aims to preserve specific identifiable characteristics within anonymized visual content or to develop models capable of identifying privacy-protected faces. This is essentially possible based on the face recognition studies towards challenging or wild cases [43], [44], [45], [46], demonstrating the potential of employing deep models to recognize faces in extremely low-quality or highly obscured images. We collectively call these methods *identifiable face anonymization (IDFA)*. Li et al. [47] first proposed identity-preserved face camouflage, which disentangles facial features into the appearance and the identification codes and replaces the former towards another target. The same authors later proposed two more advanced approaches [48], [49], where an identity-aware region discovery module is devised to determine the facial attributes sensitive to humans but less crucial to machines, based on which, facial anonymization is performed via conditional GAN. Yuan et al. [50] proposed PRO-Face, a generic framework for recognizable face obfuscation inspired by image steganography. Zhang et al. [51] proposed RAPP, a reversible privacy protection scheme for facial attributes based on stream cipher and GAN. Yuan et al. [52] proposed IVFG, an identifiable virtual face generator to achieve accurate recognition on virtual face domain. Similarly, Wang et al. [53] devised a novel scheme via virtual identity transformation based on disentangled facial representation, which guarantees strong privacy and high identifiability.

In contrast to PPFR methods, IDFA approaches prioritize the goal of "face anonymization," considering face recognition as a secondary outcome of the anonymization process. As anonymization inherently opposes recognition, the recognition performance of IDFA approaches typically suffers in comparison to PPFR methods. However, IDFA offers an advantage over PPFR in that the anonymized face retains its visual integrity, resembling a face rather than random patterns. This feature fulfills basic requirements for image preview in specific applications, such as video surveillance.

Building upon the foundation laid by pioneering researchers, our approach endeavors to bridge the gap between two distinct families of methods, striving to attain a harmonious balance of robust visual privacy, optimal recognition utility, image preview functionality, and broad applicability simultaneously.

III. The PRO-Face C Framework

A. Overview and Formulation

The PRO-Face C framework provides a novel paradigm for privacy-preserving recognition of obfuscated faces with the aim of achieving the highest trade-off between visual privacy and recognition utility. An overview of PRO-Face C is shown in Fig. 2, highlighting a specialized client-server architecture that performs identity recognition in a collaborative way based on obfuscated image data only. The threat model and the core principles of the framework are given below.

1) Threat Model and Assumptions: First, we outline the following the security models and assumptions:

- We postulate the existence of a client-server architecture behaving as follows: The client is lightweight and thus tasked with image capture and simplified processings, which demand minimal computational resources. The backend server (or edge) is more powerful and is ready for undertaking more resource-consuming computations, such as image analytics and identity inference.
- Any end device (e.g., surveillance camera or smartphone) on the client side is trusted, indicating that the client can manage the image data in any clear format.
- The server is *honest-but-curious*, which means the online service provider should follow the defined recognition protocol honestly but may attempt to inspect the visual content of a face or invert the image for curiosity.

2) Working Principles: Based on the above assumptions, the client in PRO-Face C is responsible for performing privacy protection on an input image, while the server is for recognizing the facial identity based on the protected data only. Specifically, the client device transforms an input image into two parts: 1) a visually obfuscated image meant to support

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

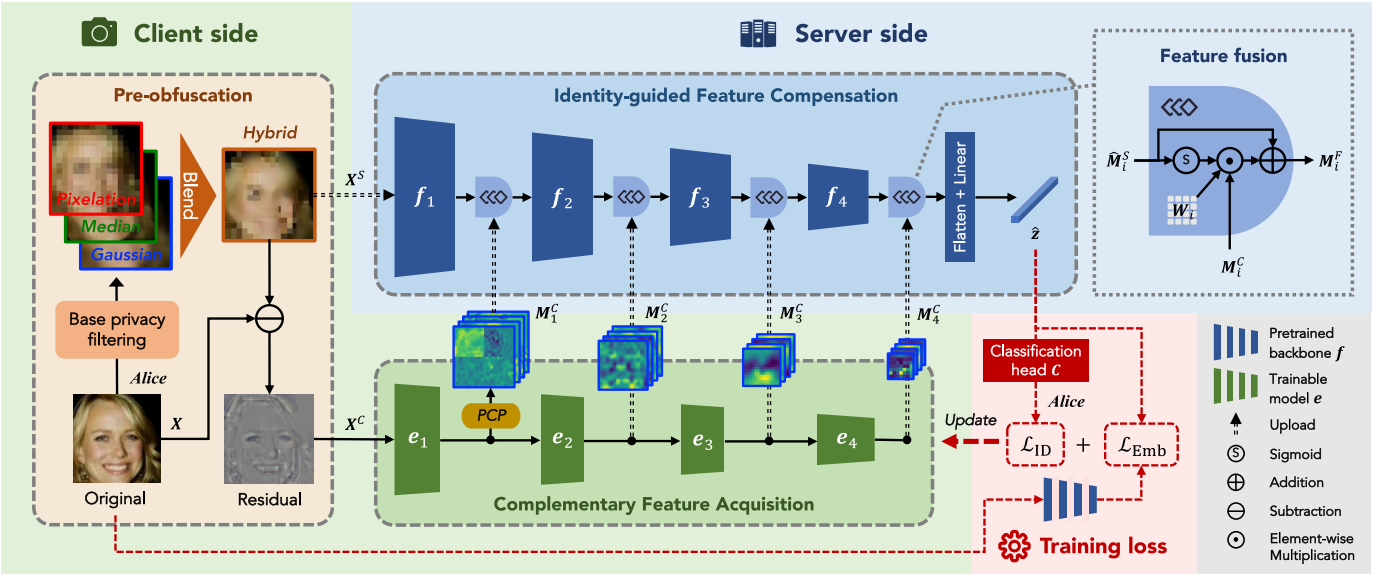

Fig. 2. The comprehensive framework of PRO-Face C comprises a client-server architecture composed of three fundamental components: 1) pre-obfuscation generating privacy-protected image $\boldsymbol{X}^S$, 2) complementary feature acquisition $\boldsymbol{e}$ obtaining a set of complementary feature maps $\boldsymbol{M}_i^C$, and 3) identity-guided feature compensation integrating the complementary features inside the internal inference pipeline of the pre-obfuscated image within an online recognition backbone $\boldsymbol{f}$, which finally produces the recognizable identity representation $\hat{\boldsymbol{z}}$. Accurate face recognition is achieved on the pre-obfuscated image with the assistance of privacy-preserving complementary features.

image preview functionality without leaking the original face clearly, and 2) a suite of privacy-free feature maps served as complementary information to the obfuscated image for identification. On the server, the complementary features are integrated into the inference pipeline of an existing recognition system, which outputs the facial embedding that signals the original identity. The entire workflow therefore achieves accurate face identification without exposing the original face in any clear form to the server. For broader applicability, PRO-Face C operates solely on a pre-trained recognition backbone on the server, eliminating the need to retrain the online recognition model from scratch. In this manner, the balance between utility and privacy is maximized. The two competing aspects of PRO-Face C, namely privacy protection and identity recognition, are formulated as follows.

a) Privacy protection: Given a captured face image $\boldsymbol{X}$, the client first obfuscates it using an operator denoted as $\boldsymbol{o}$, which can be drawn from a set of privacy filters $\mathbb{O}$:

$$\boldsymbol{X}^S = \boldsymbol{o}(\boldsymbol{X}), \quad \boldsymbol{o} \in \mathbb{O}, \tag{1}$$

where $\boldsymbol{X}^S$ is defined as the pre-obfuscated image. To ensure the visual privacy protection, our aim is for the identity of the obfuscated image to remain unrecognizable by both machines and humans with high confidence:

$$\delta_f(\boldsymbol{X}^S, \boldsymbol{X}) < T_f, \tag{2}$$
$$\delta_v(\boldsymbol{X}^S, \boldsymbol{X}) < T_v, \tag{3}$$

where δ_f and δ_v denote specific similarity metrics on facial embedding and visual perception respectively, and T_f and T_v denote two thresholds.

b) Identity recognition: As Eq. (2) implies, it is obviously difficult for an online recognition model (denoted as $\boldsymbol{f}$) to infer the face identity directly from the obfuscated image due to the missing facial details. To enable identity recognition as a utility service without sacrificing visual privacy, a specialized mechanism is proposed to compensate the lost information in the internal inference pipeline of $\boldsymbol{f}$. To this end, a group of feature maps $\boldsymbol{M}^C$ are extracted from the residual between the obfuscated image and its original by a lightweight model (denoted as $\boldsymbol{e}$) on the client device:

$$\boldsymbol{M}^C = \boldsymbol{e}(\boldsymbol{X}^C), \tag{4}$$

where $\boldsymbol{X}^C = \boldsymbol{X} - \boldsymbol{X}^S$. The model $\boldsymbol{e}$ is designed in such a way that the extracted features $\boldsymbol{M}^C$ can assist $\boldsymbol{f}$ in inferring the correct identity of $\boldsymbol{X}$:

$$\boldsymbol{f}^*\left(\boldsymbol{X}^S, \boldsymbol{M}^C\right) = \boldsymbol{f}\left(\boldsymbol{X}\right), \tag{5}$$

where $\boldsymbol{f}^*$ represents a specially crafted recognition interface that combines the obfuscated image $\boldsymbol{X}^S$ and the features $\boldsymbol{M}^C$ utilizing the identical pre-trained parameters as model $\boldsymbol{f}$. The feature maps $\boldsymbol{M}^C$ serve as complementary information to the obfuscated image, supposed to be privacy-free as well:

$$\delta_f(\boldsymbol{M}^C, \boldsymbol{X}) < T_f, \tag{6}$$
$$\delta_v(\boldsymbol{M}^C, \boldsymbol{X}) < T_v. \tag{7}$$

3) Privacy-Utility Requirements: In conclusion, we assert the following privacy and utility requirements within the proposed framework:

- ***Privacy against visual inspection:*** As the fundamental requirement, operators (humans) on the server side should not be able to extract the clear visual appearance of the original face from either the obfuscated image or the complementary information.
- ***Privacy against reconstruction attacks:*** Considering the possibility of server operators attempting to reconstruct the original image from the received inputs, we stipulate

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

that the obfuscated image and complementary features cannot be utilized to generate high-quality reconstructions effectively.

- ***Utility of identity recognition:*** Utilizing solely the privacy-preserved inputs (the obfuscated image and the complementary feature maps), the server should be capable of accurately conducting identity recognition through a pre-trained recognition model deployed on the server.
- ***Utility of photo preview:*** To fulfill the requirement for visual inspection or image visualization, the obfuscated image should be presented in a visually meaningful form, such as a thumbnail, rather than being completely randomized, e.g., encrypted patterns.

Several distinctive components have been developed to fulfill the proposed properties, described in detail as follows.

B. Hybrid Image Obfuscation

In PRO-Face C, the visual privacy is primarily ensured by a pre-obfuscation module that carries out empirical privacy filtering to obfuscate the captured face. The privacy filters applied can be in various forms, including but not limited to Gaussian blur, median filter and pixelation. To achieve stronger visual confusability and to reduce the risk of image reconstruction, we propose ***hybrid obfuscation***, a simple yet effective mechanism that blends multiple conventional privacy filters into one, which presents different obfuscation effects in different spatial locations. Specifically, given the input image initially obfuscated by Gaussian blur, median filter, and pixelation, denoted as $\boldsymbol{X}^G$, $\boldsymbol{X}^M$, and $\boldsymbol{X}^P$ respectively, the hybrid obfuscated image $\boldsymbol{X}_S$ can be computed as:

$$\boldsymbol{X}^S = \boldsymbol{m}_1 \odot \boldsymbol{X}^G + \boldsymbol{m}_2 \odot \boldsymbol{X}^M + \boldsymbol{m}_3 \odot \boldsymbol{X}^P, \tag{8}$$

where $\odot$ denotes element-wise multiplication. $\boldsymbol{m}_i$ denotes a boolean mask matrix composed of $M \times M$ patches each with either all ones or all zeros drawn uniformly, and the sum of the three matrices equals an all-ones matrix $\boldsymbol{J}_{W,H}$:

$$\begin{aligned} &\boldsymbol{m}_i \in \{\boldsymbol{J}_{K,L}, \boldsymbol{O}_{K,L}\}^{M\times M}, 1 \leqslant i \leqslant 3, \\ &\text{s.t.} \quad K = \left\lfloor \frac{W}{M} \right\rfloor, \ L = \left\lfloor \frac{H}{M} \right\rfloor, \ \sum_{i=1}^{3} \boldsymbol{m}_i = \boldsymbol{J}_{W,H}, \end{aligned} \tag{9}$$

where W and H denote the width and height of the input image, and $\boldsymbol{J}_{K,L}$ and $\boldsymbol{O}_{K,L}$ indicate all-ones and all-zeros matrices respectively, both in the dimension of $K \times L$. If W and H are not evenly dividable by M, the division results are lower bounded and the last patches on the boundary occupy a slightly large size. A varying integer of M greater than 1 can be used to tune the obfuscation strength. An illustrative example of 5×5 hybrid obfuscation is shown in Fig. 3.

During training, we use hybrid obfuscation images only, and randomly vary the strength of each base privacy filter, the value of M, and the blending masks $\boldsymbol{m}_i$. During inference, any type of obfuscation including the base privacy filters and the hybrid one can be used. The advantages of hybrid obfuscation are twofold, which have been verified in our experiments:

1) It offers stronger visual privacy protection compared to a single obfuscation especially against the image reconstruction attacks.
2) It provides a means of data argumentation in the training phase to improve the model performance and generality, compared to using single obfuscations separately.

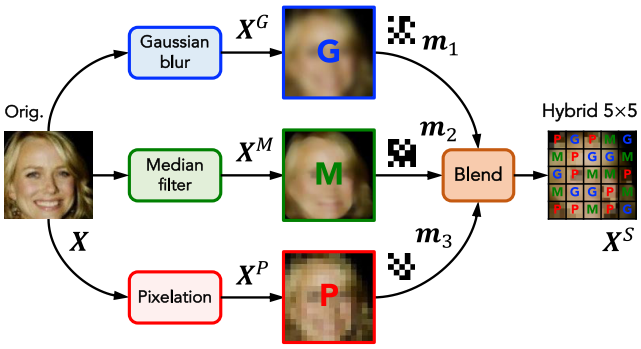

Fig. 3. Illustration of 5×5 hybrid obfuscation, with the obfuscated image constructed from three base privacy filters, namely Gaussian blur, median filter, and pixelation.

C. Complementary Feature Acquisition

Most face recognition backbones employ a stage-by-stage architecture to learn identifiable features at different scales. Inspired by [34], we also structure the complementary features in a multi-scale format. We assume the recognition backbone $\boldsymbol{f}$ deployed on the server has N sequential stages of feature extraction, and each generates a set of feature maps $\boldsymbol{M}_i^S$ at a different resolution:

$$\boldsymbol{M}_i^S = \boldsymbol{f}_i(\boldsymbol{M}_{i-1}^S) \in \mathbb{R}^{C_i \times W_i \times H_i}, \quad i \in [1..N], \tag{10}$$

where (C_i, W_i, H_i) denotes the channel volume, width, and height of the feature maps at the i-th output stage of $\boldsymbol{f}$, and $\boldsymbol{M}_0^S$ is equivalent to $\boldsymbol{X}^S$, aka the obfuscated image uploaded from the client. Accordingly, we employ a lightweight model (denoted as $\boldsymbol{e}$) on the client side that comprises N stages of feature extraction corresponding to the server-side model $\boldsymbol{f}$. The client model $\boldsymbol{e}$ takes as input the residual between the obfuscated image and its original, and sequentially produces N suites of feature maps $\boldsymbol{M}_i^C$ from its intermediate layers:

$$\boldsymbol{M}_i^C = \boldsymbol{e}_i(\boldsymbol{M}_{i-1}^C) \in \mathbb{R}^{C_i \times W_i \times H_i}, \quad i \in [1..N], \tag{11}$$

where $\boldsymbol{M}_i^C$ has the same dimension as the online feature $\boldsymbol{M}_i^S$, and $\boldsymbol{M}_0^C$ is equivalent to $\boldsymbol{X}^C$, which is the initial input to $\boldsymbol{e}$, defined as the element-wise residual between the obfuscated and original image. $\boldsymbol{M}^C = \{\boldsymbol{M}_i^C, i \in [1..N]\}$ are formally known as *complementary feature maps*, which are utilized to assist the online recognition model $\boldsymbol{f}$ to accurately infer the correct identity from the obfuscated face. To minimize the computational burden of the client, we employ MobileFaceNets [54] as the backbone for $\boldsymbol{e}$, which is a typical lightweight model designed for resource-friendly devices.

D. Patch-Wise Channel Permutation (PCP)

The complementary feature maps directly generated from the client-side model $\boldsymbol{e}$ may still disclose the original facial appearance to a certain degree, especially for those from the shallow layers such as $\boldsymbol{M}_1^C$. In response to this issue, we propose *Patch-wise Channel Permutation (PCP)*, an obfuscation

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

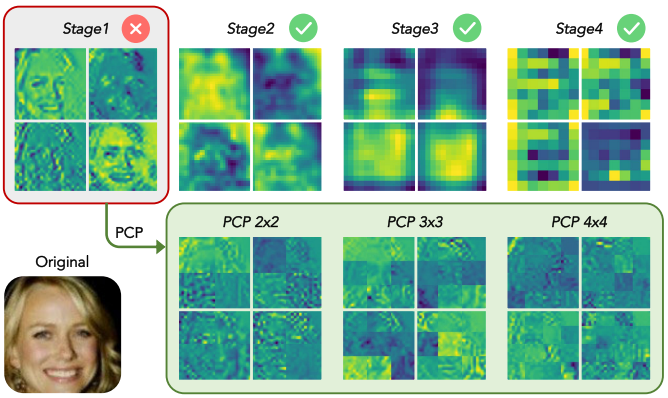

Fig. 4. Illustration of Patch-wise Channel Permutation (PCP) with different values for P applied to the first-stage complementary feature maps.

mechanism applied to the complementary feature maps aimed at mitigating the associated privacy risk. Specifically, we split each channel of $\boldsymbol{M}_i^C$ into $P \times P$ patches and permute those patches corresponding to each position across channels. The order of permutation is independent and distinct for each patch position, leading to semi-randomized visual effects. The permutation remains fixed by a random seed during training, ensuring that the same permutation order can be replicated during inference. Correct recognition results during inference are only achievable when the permutation order applied is identical to that during training. Illustrative examples of PCP for $P \in \{2, 3, 4\}$ are shown in Fig. 4. In our work, only the complementary feature maps at the first layer (i.e., $\boldsymbol{M}_1^C$) are permuted since the rest can hardly disclose any original facial detail. A mild level of PCP with 2×2 patches is utilized for a balance between visual privacy and recognition accuracy, according to the results in Section IV-A.3. The proposed PCP mechanism offers at least three proven advantages:

1) It effectively diminishes the visibility of the complementary feature maps, therefore enhancing visual privacy.
2) It reduces the likelihood of malicious image reconstruction from the complementary feature maps.
3) It functions as a "secret-key" mechanism, ensuring that the inference process operates accurately only under the condition of utilizing the correct seed (or secret) employed during training.

E. Identity-Guided Feature Compensation

To achieve accurate face identification on the server, we have developed an identity-guided compensation mechanism, which integrates the complementary feature maps into the internal recognition pipeline of the obfuscated face, enabling the output of the recognition model to indicate the real identity of the original face. Inspired by DuetFace [34], a similar feature fusion mechanism is employed to combine the complementary features $\boldsymbol{M}^C$ with the intermediate features of obfuscated image $\boldsymbol{X}^S$ extracted by $\boldsymbol{f}$ corresponding to each feature extraction stage i:

$$\boldsymbol{M}_i^F = \boldsymbol{W}_i \odot \left(\mathrm{S}\left(\hat{\boldsymbol{M}}_i^S\right) \odot \boldsymbol{M}_i^C + \hat{\boldsymbol{M}}_i^S\right), \; i \in [1..N], \quad (12)$$

where $\boldsymbol{W}_i \in \mathbb{R}^{C_i \times W_i \times H_i}$ denotes a set of 2D matrices in the same dimension as the feature maps at each stage. $\odot$ stands for element-wise multiplication and S() for the Sigmoid function. $\hat{\boldsymbol{M}}_1^S$ is equivalent to $\boldsymbol{M}_1^S$. The fusion result $\boldsymbol{M}_i^F$ at each stage is fed into the next stage $\boldsymbol{f}_{i+1}$, producing $\hat{\boldsymbol{M}}_{i+1}^S$ to replace the original feature $\hat{\boldsymbol{M}}_1^S$:

$$\boldsymbol{M}_{i+1}^S \leftarrow \hat{\boldsymbol{M}}_{i+1}^S = \boldsymbol{f}_{i+1}\left(\boldsymbol{M}_i^F\right). \quad (13)$$

The final fusion result at the last stage is flattened and linearly projected into an embedding $\hat{\boldsymbol{z}}$ used for identity recognition:

$$\hat{\boldsymbol{z}} = \mathrm{Linear}\left(\mathrm{Flatten}\left(\boldsymbol{M}_N^F\right)\right) \in \mathbb{R}^{512}. \quad (14)$$

To uphold the recognition performance, we train the client-side feature extractor $\boldsymbol{e}$ concurrently with the weight matrices $\boldsymbol{W}_i$, ensuring that the facial embedding $\hat{\boldsymbol{z}}$ can be classified into the correct identity label. To this end, we constrain the embedding $\hat{\boldsymbol{z}}$ with the softmax loss proposed by ArcFace [55]:

$$\mathcal{L}_{\mathrm{ID}} = \mathcal{L}_{\mathrm{ArcFace}}\left(\hat{\boldsymbol{z}}, \boldsymbol{C}\right), \quad (15)$$

where $\mathcal{L}_{\mathrm{ArcFace}}$ denotes the standard ArcFace loss defined in [55] and $\boldsymbol{C}$ denotes a classification head. To further refine the facial embedding, we leverage the pre-trained recognition model and introduce an embedding loss facilitated by knowledge distillation. Essentially, we encourage the learned embedding $\hat{\boldsymbol{z}}$ to approximate that of the original face extracted from the pre-trained model $\boldsymbol{f}$:

$$\mathcal{L}_{\mathrm{Emb}} = \left\|\hat{\boldsymbol{z}} - \boldsymbol{f}\left(\boldsymbol{X}\right)\right\|_2, \quad (16)$$

where $\|\cdot\|_2$ denotes the Euclidean distance. This loss function can encourage our model to effectively exploit the knowledge of the original identity acquired from a substantially larger pre-trained model. Furthermore, the utilization of this loss function is particularly beneficial to cross-domain face verification scenarios [50], wherein there is a practical need to match anonymized faces with unprotected plain templates. The final loss function is computed as the weighted sum of the aforementioned two functions:

$$\mathcal{L}_{\mathrm{Total}} = \mathcal{L}_{\mathrm{ID}} + \lambda \cdot \mathcal{L}_{\mathrm{Emb}}, \quad (17)$$

where λ represents a hyper-parameter balancing the two loss components.

F. Generality by Design

To maintain the generality of the proposed approach, we adopt two distinct designs aimed at optimizing the model deployment complexity and reducing overhead in data communications within practical scenarios.

1) Seamless Integration With Pre-Trained Models: First, we presume the proposed method should seamlessly integrate with any existing face recognition system with minimal deployment efforts. This requires the client-side feature extractor to function effectively with the majority of pre-trained recognition models hosted on the server. Hence, in the architecture of PRO-Face C, we only train the "tiny" client-side model $\boldsymbol{e}$ concurrently with the weight matrices $\boldsymbol{W}_i$, without updating the much "heavier" server-side backbone $\boldsymbol{f}$. This

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

not only alleviates training burdens but also maximizes the utilization of knowledge acquired from the pre-trained face recognizer. Furthermore, it enables straightforward deployment of the proposed framework as a "plugin" module into an existing recognition system, requiring minimal effort to overhaul the entire system. This represents one of the most distinctive property of the proposed approach, setting it apart from other PPFR methods, such as DuetFace [34] and PartialFace [35], which necessitate retraining the entire recognition model.

2) Managing Communication Overhead: The mechanism of feature compensation inevitably introduces an overhead to the server-client communications. To mitigate this impact, we propose several measures to reduce the data volume of the complementary features required for transmission, without significantly sacrificing the recognition performance. The first measure involves quantizing feature maps into integer values, which can be compressed at a high rate by a lossless compression algorithm before transmission to the server. This quantization can be performed solely during the inference stage. The second measure entails pruning certain layers of the feature maps, also known as *dropout*. This can be implemented randomly or by removing feature maps from the last three stages, thus preserving only $\boldsymbol{M}_1^C$. The dropout mechanism must be applied during both the training and inference processes. Combining these two measures allows for joint control of communication overhead.

IV. Experiments

In this section, we report the performance of the proposed approach with respect to identity recognition and privacy protection. For recognition performance, 1:1 face verification was primarily conducted. For privacy protection, both quantitative and qualitative results are presented.

A. Experimental Settings

1) Datasets and Benchmarks: For training and validation, we primarily use the CelebA [56] dataset, which contains 202'599 face images of 10'177 celebrity identities. For testing, several more datasets have been employed to be inline with existing state-of-the-art solutions. To compare with the PPFR methods, we utilize the most commonly used benchmarks in face recognition, including LFW [57], CFP-FP [58], AgeDB [59], CPLFW [60], CALFW [61], IJB-B [62], and IJB-C [63]. Following the prior studies, the average verification accuracy is reported for the former five datasets while the TAR@FAR = 0.01% is used for the latter two (IJB-B and IJB-C). For fair comparisons with major IDFA methods, we made testing experiments on LFW [57], the test split of CelebA [56], and the test split of VGGFace2 [64], using TAR@FAR = 0.01 as the evaluation metric. For the verification benchmark implementation, we relied on face.evoLVe[1] [65], a high-quality open-source library that incorporates a wide range of popular methods, datasets, and pre-trained models for face recognition.

TABLE I

Hyper-Parameter Selection: Average Verification Accuracy on CelebA for Different Values of λ and the PCP Parameter P. Bold Values Indicate Those Finally Chosen. *N.A. Indicates the Case Without the Utilization of PCP

Param.	Weight of embedding loss λ					PCP parameter P		
	0.6	0.8	**1.0**	1.2	1.4	N.A.*	**2×2**	4×4
Acc. (%)	96.85	96.98	**97.08**	97.02	96.92	97.06	**97.00**	94.87

2) Pre-Obfuscations: We experimented with four types of obfuscations, including three base privacy filters (i.e., Gaussian blur, median filter, and pixelation), and the proposed hybrid obfuscation, constructed from the former three. During training, only the **hybrid obfuscation** was used, with the parameter M randomly sampled from $\{2, 3, 4, 5\}$, and image patches randomly selected from the three filtering results:

- **Gaussian blur:** the kernel size was fixed to 31 and the sigma is chosen uniformly at random in range of 2 to 8.
- **Median filter:** the median kernel was chosen uniformly at random between odd number 7 to 19.
- **Pixelation:** the pixelation block size was selected uniformly at random between integer 4 to 10.

During testing, the hybrid obfuscation with $M = 4$ formed by the three filters with a median strength level was primarily evaluated, i.e., Gaussian sigma of 5, median kernel of 13, and pixelation block size of 7. The performance with respect to each single base filter was also evaluated for comparison.

3) Implementation Details: As preprocessing, we aligned and cropped all images into a uniform resolution of 112×112, in line with other face recognition works. For the server-side recognition backbone $\boldsymbol{f}$, we primarily experimented with a ResNet100 model pre-trained by a recent recognition method AdaFace[2] [46]. For the client-side feature extractor $\boldsymbol{e}$, we employed the MobileFaceNets [54], a typical lightweight backbone designed for low-resource devices. Following [34], the stochastic gradient descent (SGD) optimizer with an initial learning rate of 0.1 is used for training, with the momentum of 0.9 and the weight decay of 0.0005. We experimented with multiple values for the hyper-parameter λ defined in Eq. 17 and finally chosen $\lambda = 1.0$ due to its optimal validation performance on the CelebA dataset. Furthermore, we opted for a PCP setting of $P = 2$ as it induced a noticeable obfuscation effect with minimal impact on the verification accuracy. This choice was made after comparing scenarios where PCP was not utilized and where a finer-grained PCP of $P = 4$ was employed. The validation performance of face verification for different hyper-parameter settings is shown in Table I. All experiments were conducted using two NVIDIA RTX A5000 GPUs within the PyTorch framework.

B. Recognition Performance

We first report the identity recognition performance of the proposed approach in comparison with literature methods for privacy-preserving face recognition (PPFR) and identifiable face anonymization (IDFA).

[1] https://github.com/ZhaoJ9014/face.evoLVe

[2] https://github.com/mk-minchul/AdaFace

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

TABLE II

FACE VERIFICATION PERFORMANCE COMPARED WITH UNPROTECTED FACE RECOGNITION (FR) ANDPRIVACY-PRESERVING FACE RECOGNITION (PPFR) METHODOLOGIES. THE AVERAGE VERIFICATION ACCURACY (%) IS REPORTED.BOLD HIGHLIGHTS THE OPTIMAL SCORE WITHIN EACH COLUMN

Methodology		Dataset							
		LFW [57]	CFP-FP [58]	AgeDB [59]	CPLFW [60]	CALFW [61]	*Avg.*	IJB-B [62]	IJB-C [63]
FR	ArcFace [55]	99.83	99.19	97.95	94.35	96.00	97.46	95.75	97.16
	AdaFace [46]	99.80	99.17	97.90	94.64	96.05	97.51	96.03	97.39
PPFR	PEEP [40]	98.41	74.47	87.47	79.58	90.06	86.00	5.82	6.02
	Cloak [66]	98.91	87.97	92.60	83.43	92.18	91.02	33.58	33.82
	InstaHide [29]	96.53	83.20	79.58	81.03	86.24	85.32	61.88	69.02
	CPGAN [67]	98.87	94.61	96.98	90.43	94.79	95.14	92.67	94.31
	PPFR-FD [32]	99.68	95.04	97.37	90.78	95.72	95.72	/	94.10
	DCTDP [33]	99.77	96.97	97.72	91.37	96.05	96.38	93.29	94.43
	MinusFace [38]	99.78	96.92	97.57	91.90	95.59	96.35	93.37	94.70
	PartialFace [35]	99.80	97.63	97.79	92.03	96.07	96.67	93.64	94.93
	DuetFace [34]	99.82	97.79	**97.93**	92.35	**96.10**	96.80	93.66	95.30
PRO-Face C	*Hybrid*	**99.82**	98.50	97.30	93.65	95.93	97.04	94.93	96.49
	Hybrid (int)	99.80	**98.63**	97.40	**93.75**	95.87	**97.09**	**95.15**	**96.49**
	Gaussian	99.75	98.77	97.32	93.72	95.85	97.08	95.06	96.60
	Median	99.75	98.69	97.57	93.77	96.02	97.16	95.11	96.58
	Pixelation	99.77	98.50	96.95	93.28	95.87	96.87	94.59	96.49

TABLE III

FACE VERIFICATION PERFORMANCE COMPARED WITH IDENTIFIABLE FACE ANONYMIZATION (IDFA) METHODS. TAR@FAR(0.01) IS REPORTED TO BE CONSISTENT WITH COMPETING METHODS

Methodology		Dataset		
		LFW	CelebA	VGGFace2
IDFA	Li [48]	/	78.39	61.28
	Li [49]	/	82.00	89.00
	Wang [53]	/	86.00	80.47
	PRO-Face (*Gaussian*) [50]	94.52	86.64	73.38
	PRO-Face (*Pixelation*) [50]	89.04	70.12	62.03
PRO-Face C	*Hybrid*	99.83	94.90	**92.17**
	Hybrid (int)	**99.83**	**94.93**	92.03
	Gaussian	99.83	95.00	92.20
	Median	99.87	94.90	92.07
	Pixelation	99.80	94.93	92.03

1) Comparison With PPFR Methods: We initially compare our approach with state-of-the-art PPFR methods, which are expected to provide high recognition accuracy. Following the standard benchmark protocol, the average accuracy is reported for the datasets LFW [57], CFP-FP [58], AgeDB [59], CPLFW [60], and CALFW [61], and the TAR@FAR(0.01%) is presented for IJB-B [62] and IJB-C [63]. The performance of the two recent face recognition methods (i.e., ArcFace [55] and AdaFace [46]) on unprotected plain images is also provided for reference. Table II shows the comparison results. It is clear that PRO-Face C with hybrid obfuscation outperforms major PPFR methods on most datasets, except for AgeDB [59] and CALFW [61], which show slight degradations compared to DuetFace [34]. Yet, the average accuracy of our method over the fist five datasets and the TAR@FAR scores on the IJB-B/C datasets remain the optimal among all the listed methods. Furthermore, quantizing complementary feature maps into integer does not impact the overall recognition performance and, in certain cases, even demonstrates improvements on specific datasets. In comparison to the conventional face recognition models without privacy protection, our verification performance experiences a slight degradation. Nonetheless, the difference in average verification accuracy between our method and AdaFace [46] is less than 0.5%, which is reasonable and acceptable considering multiple privacy protection measures integrated into the proposed approach to safeguard visual information. In addition, we evaluated the performance of the proposed approach when employing three base privacy filters individually, observing consistently high overall accuracies. Remarkably, the model maintains its performance even though it has not been trained using those privacy filters exclusively.

2) Comparison With IDFA Methods: We further compare the proposed approach with pioneering IDFA methods, which retain identifiable information in image while performing face anonymization. To align with common experimental setups utilized in prominent IDFA methods, we report TAR@FAR(0.01) on three datasets (LFW, CelebA and VGGFace2) in Table III. The results demonstrate a significant improvement of our approach over the listed IDFA methods, despite each method employing distinct measures to obscure facial appearance. Particularly, when compared with PRO-Face [50], a previous work that also employs the blur-based obfuscation, our recognition performance exhibits significant improvement, attributed to the complementary features introduced in auxiliary channels. On the contrary, the recognition mechanism underlying PRO-Face [50] involves retraining identifiable visual features within the obfuscated face, thereby sacrificing a certain level of visual privacy for recognition utility. A comparison of the obfuscation effect between PRO-Face [50] and the proposed PRO-Face C is shown in Fig. 5. One can clearly observe certain visible details on facial landmarks (such as the nose and eyes) in the obfuscations produced by PRO-Face [50], which are not present in PRO-Face C.

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.



(a) (b) (c) (d) (e) (f)

Fig. 5. Comparison of the obfuscation effect between PRO-Face [50] and PRO-Face C. Columns from left to right indicate (a) original images, (b) PRO-Face [50] Gaussian blur, (c) PRO-Face [50] pixelation, (d) PRO–Face C Gaussian blur, (e) PRO-Face C pixelation, and (f) PRO-Face C hybrid obfuscation, respectively. It is clear that the proposed PRO-Face C keeps the faces completely obfuscated, but PRO-Face [50] has to preserve certain amount of facial details (i.e., eyes, nose) for identity recognition.

TABLE IV

PERFORMANCE COMPARISON OF CROSS-DOMAIN FACE VERIFICATION WITH/WITHOUT UTILIZING $\mathcal{L}_{\text{Emb}}$.TAR@FAR(0.01) IS REPORTED

Method		Dataset		
		LFW	CelebA	VGGFace2
PRO-Face [50]	*Gaussian*	95.92	93.13	91.50
	Pixelation	94.03	85.69	73.23
PRO-Face C (w/o $\mathcal{L}_{\text{Emb}}$)	*Hybrid*	99.70	94.70	92.07
	Gaussian	99.77	94.73	91.87
	Median	99.77	94.63	92.07
	Pixelation	99.77	94.50	91.87
PRO-Face C (with $\mathcal{L}_{\text{Emb}}$)	*Hybrid*	**99.87**	**95.03**	**92.07**
	Gaussian	99.83	94.93	92.27
	Median	99.83	94.83	92.03
	Pixelation	99.80	94.83	92.00

3) Cross-Domain Face Verification Performance: In practice, protecting the template image in a recognition system already in operation may not always be feasible due to legal restrictions or complex legacy infrastructures. Therefore, face verification between privacy-protected images and unprotected face templates, also known as cross-domain verification [50], is a valuable yet non-trivial property to explore. Unfortunately, most existing methods have not considered this aspect, with the exception of PRO-Face [50]. In our work, we consider cross-domain verification and introduce the embedding loss $\mathcal{L}_{\text{Emb}}$ to facilitate this objective. Table IV showcases the results of cross-domain verification performance, highlighting the superiority of the proposed approach over PRO-Face [50] with substantial margins. Furthermore, a slight improvement is observed when using $\mathcal{L}_{\text{Emb}}$ compared to the case without it, indicating the effectiveness of the introduced embedding loss in cross-domain verification.

C. Privacy Analysis

From another perspective, the privacy protection capability of the proposed approach is analyzed both quantitatively and qualitatively.

TABLE V

AVERAGE VERIFICATION ACCURACY FOR HYBRID OBFUSCATION OVER THE FIVE TYPICAL DATASETS IN CASES WHERE THE PCP SEED USED IN INFERENCE IS EQUAL (s_0) OR UNEQUAL ($s_i, i > 0$) TO TRAINING (s_t)

PCP seed	$s_0 = s_t$	$s_1 \neq s_t$	$s_2 \neq s_t$	$s_3 \neq s_t$	$s_4 \neq s_t$
Acc. (%)	**97.04**	75.36	79.37	81.61	78.26

1) Server-Side Data Exposure Analysis: As a first glance, we visualize in Fig. 6 various types of image data exposed on server, including the obfuscated image $\boldsymbol{X}^S$, the complementary feature maps $\boldsymbol{M}^C$, and the fusion feature maps $\boldsymbol{M}^F$. Subjective inspection indicates that all obfuscations effectively eliminate facial details perceptible to the human eye. With PCP 2×2 applied, the first layer of the complementary feature maps, namely $\boldsymbol{M}^C_1$, exhibits minimal meaningful visual information about the original face. Similar visual effects can be observed in the fusion feature $\boldsymbol{M}^F$. Moreover, applying the pre-trained recognition model directly on the obfuscated image or the complementary feature maps results in a significant deterioration in accuracy, as illustrated in Table VIII and discussed further in Section IV-E. Therefore, we affirm the effectiveness of the proposed approach in concealing visual privacy from the untrusted server.

2) Effectiveness of PCP: As mentioned earlier, the proposed patch-wise channel permutation (PCP) mechanism offers the advantage of enhancing the privacy of the complementary feature maps. In this section, we exclusively analyze the effectiveness of PCP. Initially, samples of the first-stage complementary feature maps (i.e., $\boldsymbol{M}^C_1$) generated with and without PCP are visualized in Fig. 7. Additionally, the facial mask of DuetFace [34] is included as a reference for comparison, since it also serves as complementary information for recognition. It is evident that the facial details presented by our feature maps are less visible than the facial mask of DuetFace [34]. When PCP 2×2 is applied, the facial clarity of the feature maps further decreases. This demonstrates the effectiveness of proposed PCP in enhancing visual privacy of complementary feature maps. To verify the "secret-key" functionality of the PCP mechanism claimed in Section III-D, we compute the face verification accuracy in the scenarios where the same or different PCP seeds are used across training and testing. As the results in Table V indicate, when the PCP seed used in inference differs from that in training (multiple seeds being tested), face recognition consistently fails with significantly reduced accuracy. Another benefit of employing PCP is its ability to prevent malicious image reconstruction from the complementary feature maps, which is discussed collectively in the following section.

3) Privacy Analysis Under Reconstruction Attacks: Conventional privacy filters are known to be vulnerable to image reconstruction attacks [68]. To verify the effectiveness of the proposed hybrid obfuscation and PCP mechanism in defending against reconstruction attacks, we conducted experiments involving image reconstructions from various intermediate data of PRO-Face C. We utilized the state-of-the-art image restoration method SwinIR [69] and trained

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

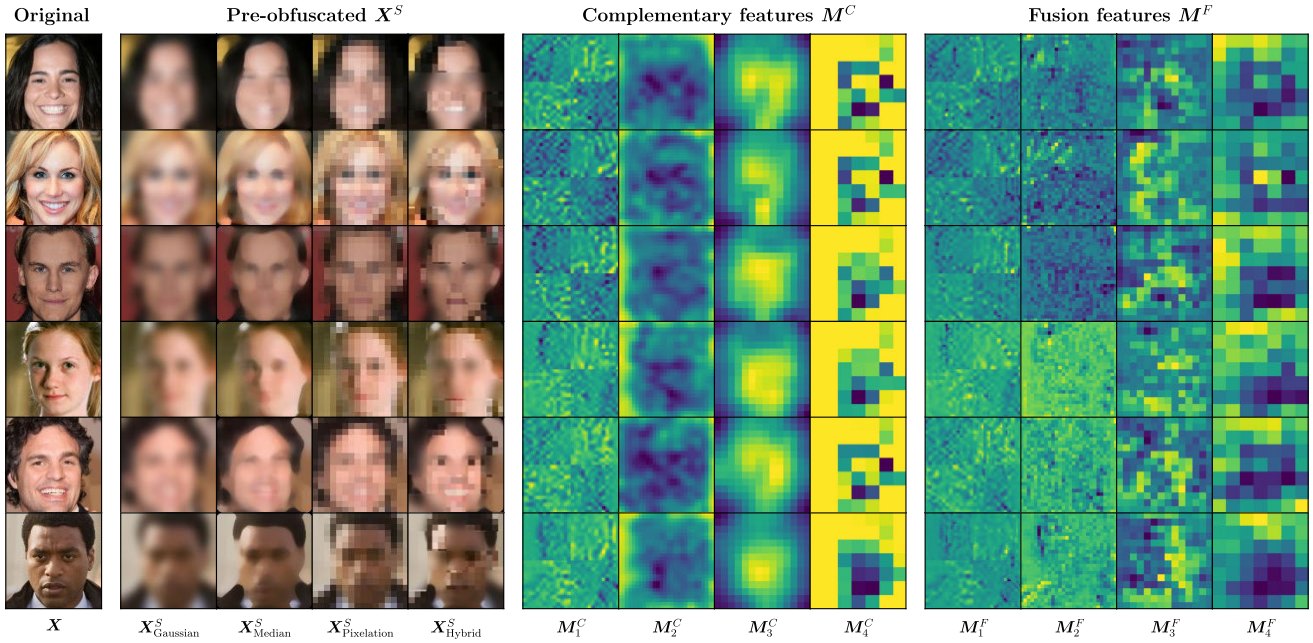

Fig. 6. Visualization of data exposed to the server: the four image groups from left to right indicate those original ones, four different pre-obfuscated forms, the complementary feature maps for hybrid obfuscation, and the corresponding fusion feature maps, respectively.

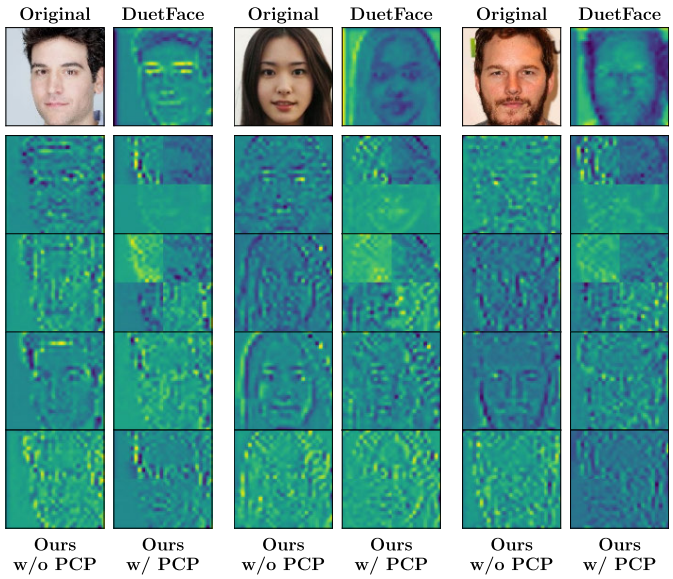

Fig. 7. Visualization of the first-stage complementary feature maps $\boldsymbol{M}^C_1$ with or without PCP applied, in comparison with the original face and the complementary feature map of DuetFace [34].

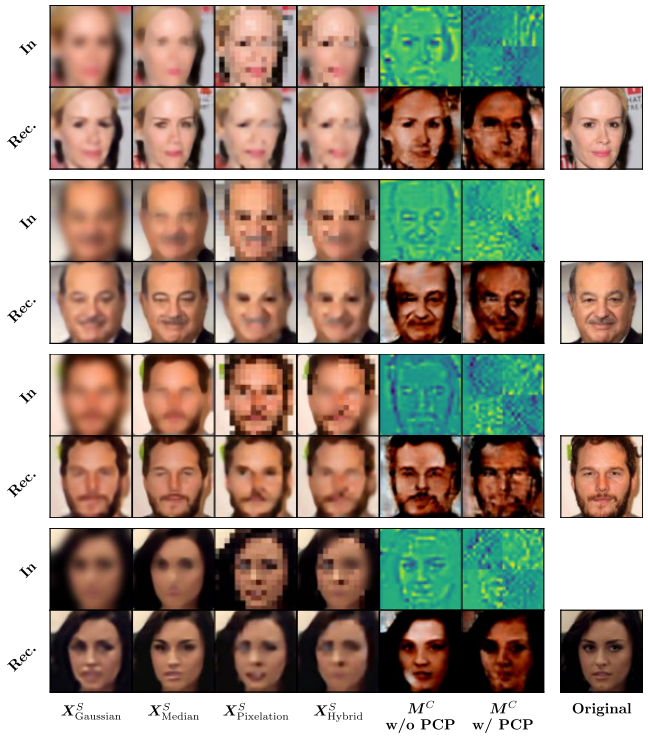

Fig. 8. Examples of reconstruction attacks using different input data.

multiple reconstruction models corresponding to different types of image obfuscations and complementary feature maps. The quality of reconstructed images measured by PSNR, SSIM [70], and LPIPS [71], along with the results of face verification based on the reconstructed images are reported in Table VI. In parallel, several reconstructed image samples are visualized in Fig. 8. From both objective and subjective results, images presenting a substantial amount of facial details can be recovered from Gaussian blurred and median filtered faces, attributed to the strong restorability of SwinIR [69]. Pixelation and the proposed hybrid obfuscation result in worse reconstruction results that still preserve visual privacy to a good extent, with the hybrid obfuscation providing slightly stronger protection effects in terms of both the quality metrics and the verification accuracies. Regarding reconstruction attacks from the complementary features, applying PCP can effectively reduce the recovery quality, again demonstrating the advantage of PCP for enhancing visual privacy.

4) Utility-Privacy Trade-off: Furthermore, the trade-off effect between utility and privacy is inspected. For each type

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

TABLE VI

PERFORMANCE OF RECONSTRUCTION ATTACKS ON OBFUSCATED IMAGES X^S AND COMPLEMENTARY FEATURE MAPS M^C, IN TERMS OF THE IMAGE QUALITY AND THE VERIFICATION ACCURACY OF THE RECONSTRUCTED FACES. R-VS-R STANDS FOR FACE VERIFICATION BETWEEN TWO RECONSTRUCTED FACES, WHILE R-VS-O FOR THAT BETWEEN THE RECONSTRUCTED AND THE ORIGINAL ONE

Reconstruction from		Image quality			Face recognition acc.	
		PSNR ↓	SSIM ↓	LPIPS ↑	R-vs-R ↓	R-vs-O ↓
X^S	*Gaussian*	22.95	0.704	0.322	82.08	80.38
	Median	22.97	0.767	0.247	81.85	80.25
	Pixelation	**20.15**	0.618	0.403	77.57	78.48
	Hybrid	20.84	**0.617**	**0.440**	**76.87**	**77.17**
M^C	*w/o PCP*	10.21	0.307	0.511	62.20	52.20
	w/ PCP	**9.07**	**0.248**	**0.578**	**60.63**	**51.93**

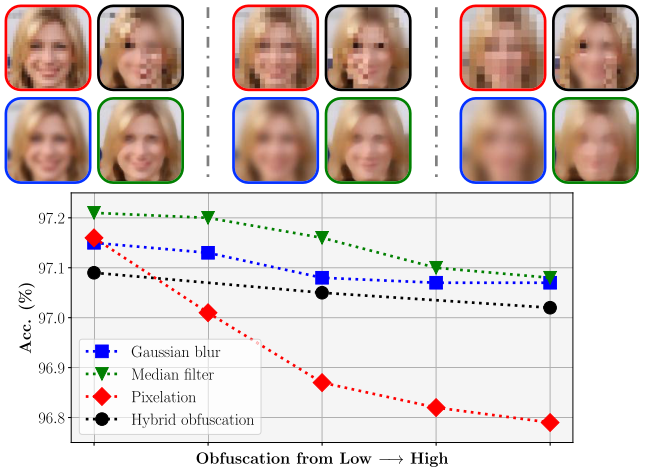

Fig. 9. Illustration of the trade-off between recognition accuracy and obfuscation strength. For hybrid obfuscation, three strength levels corresponding to $M \in \{2, 4, 5\}$ are demonstrated.

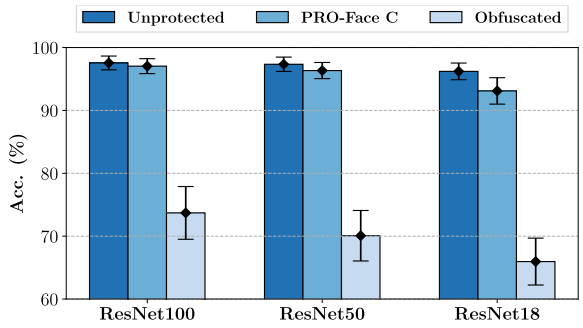

Fig. 10. Average verification accuracy (over the five typical datasets) w.r.t. different pre-trained recognition backbones on the server, evaluated on unprotected (original), PRO-Face C, and obfuscated images, respectively.

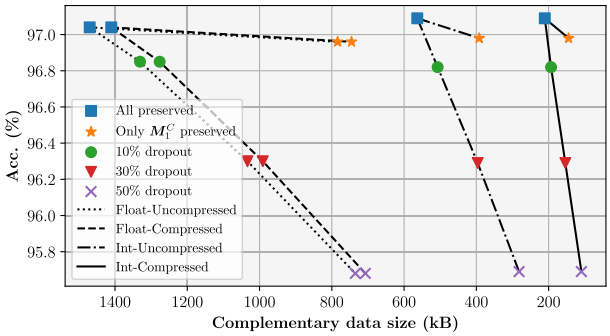

Fig. 11. Average verification accuracy (over five datasets) versus complementary data size, for various overhead control methods. Quantizing feature maps to integers facilitates data compression without substantially impacting the recognition performance. Dropping out features in the last three stages additionally contributes to data compression while causing slight degradation in recognition accuracy.

of image obfuscation, the average verification accuracies for obfuscations with an increasing strength are visualized in Fig. 9, in parallel with corresponding sample images shown on the top. As expected, with stronger obfuscation applied, the overall verification accuracy decreases, which is particularly prominent for image pixelation. Nonetheless, even with the strongest obfuscation applied, the overall recognition accuracy remains above 96.7%, for all types of obfuscations. Considering that all the results presented in the paper are based on only the mid-level strength for each obfuscation, it's conceivable that stronger visual privacy can be achieved at the expense of a slight compromise in recognition utility.

D. Generality and Complexity Analysis

Moreover, we analyze and demonstrate the generality of the proposed approach in terms of model compatibility, communication overhead, and the inference cost.

1) Compatibility With Pre-Trained Recognition Backbones: To verify the compatibility of the proposed approach with more recognition systems, we experimented with two more backbones (ResNet50 and ResNet18) as the recognition model hosted on the server. We obtained their pre-trained checkpoints by AdaFace [46] as well. The comparison between the three backbones in different recognition scenarios is shown in Fig. 10. For all three backbones, we achieve substantially high recognition performance, with very slight accuracy degradation compared to that of unprotected images. Yet, the PRO-Face C alway bridges the huge gap between the obfuscated and original images, demonstrating the effectiveness of our core mechanism of feature compensation. Considering all the backbones are pre-trained without being updated when building the PRO-Face C framework, the proposed approach offers a good level of generality in terms of seamless deployment with existing face recognition models.

2) Communication Overhead: We further analyze the communication overhead of the PRO-Face C system, and validate the proposed measures for constraining such overhead. Herein, we computed the face verification accuracy corresponding to different methods for overhead reduction proposed in Section III-F.2. The average accuracy over the five image datasets for hybrid obfuscation versus the average data size of the complementary features for a single image is plotted in Fig. 11. Clearly, when applying the lossless compression on the complementary features (bzip2[3] is our case), the overall data volume can be reduced without impacting the recognition performance. The compression performance greatly depends on the data type. For example, feature maps in float have limited space for compression. By quantizing the feature maps

[3]https://sourceware.org/bzip2/

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

TABLE VII
COMPLEXITY AND COST OF PRO-FACE C REQUIRED FOR INFERENCE

Component		# Param.	Time (ms)
Server-side	Face recognizer $\boldsymbol{f}$	**65.15M**	**41.3**
	Feature compensation - all $\boldsymbol{M}_i^C$	0.38M	0.01
	Feature compensation - only $\boldsymbol{M}_1^C$	0.2M	0.006
Client-side	Hybrid obfuscation	/	9.59
	Gaussian blur	/	1.4
	Median filter	/	7.2
	Pixelation	/	1.0
	Feature extractor $\boldsymbol{e}$	1.94M	15.1
	Patch-wise channel permutation	/	2.2

TABLE VIII
RESULTS OF ABLATION STUDY BY FACE VERIFICATION ACCURACY

Scenario	Dataset					
	LFW	CFP-FP	AgeDB	CPLFW	CALFW	*Avg.*
w/o $\boldsymbol{M}^C$	87.82	70.81	65.40	71.72	72.77	73.70
w/o $\boldsymbol{X}^S$	97.90	86.76	86.40	82.55	89.52	88.83
w/o $\boldsymbol{W}$	99.73	98.26	96.93	93.42	95.80	96.83
w/o $\mathcal{L}_{\text{Emb}}$	99.77	98.46	97.15	93.25	95.83	96.89
w/o Hybrid	99.77	97.01	96.85	92.25	95.57	96.30
w/ Facial mask	99.70	97.84	96.55	92.15	95.35	96.32
w/ Mask+Norm	91.53	74.57	67.50	75.00	74.63	76.65
Final setup	**99.82**	**98.50**	**97.30**	**93.65**	**95.93**	**97.04**

into integers, the data can be significantly compressed from over 1400 kB to just 200 kB, without impacting the recognition accuracy. Dropping out feature maps can indeed reduce the data size, but it often comes at the expense of significant loss in recognition performance when being done randomly. However, an acceptable trade-off can be achieved by dropping the features of the last three stages, thereby preserving only $\boldsymbol{M}_1^C$. In this scenario, the data size can be reduced to approximately 144 kB with only a slight degradation in recognition accuracy. These results demonstrate the effectiveness of the proposed measures for managing the communication overhead.

3) Model Size and Inference Cost: Last but not least, we demonstrate the resource-friendliness of the proposed method by analyzing the size of models and the time required for inference, with results shown in Table VII. It is evident that the most time-consuming component remains the server-side face recognition model f, with other components incurring only slight and tolerable inference costs. Specifically, the feature compensation process is negligible as it essentially involves a weighted sum. The hybrid obfuscation requires more time than any single base privacy filter because the hybrid blending relies on the presence of all three base obfuscations. The overall time consumption of hybrid obfuscation remains within a reasonable scale compared to the recognition model. The client-side feature extractor $\boldsymbol{e}$ is lightweight enough, resulting in much less time consumption compared to the server-side model $\boldsymbol{f}$. PCP involves the simple permutation of feature patches and therefore requires minimal time compared to feature extraction. In summary, the proposed approach incurs reasonable costs in terms of model complexity and inference time consumption.

E. Ablation Study

Finally, the ablation study was conducted to evaluate the effectiveness of the distinct components of the proposed framework. The cases under examination, each removing a specific component, are defined below:

- *w/o* $\boldsymbol{M}^C$ indicates the face recognition scenario using only the obfuscated image (hybrid obfuscation) without the complementary feature maps. This corresponds to the privacy requirement prescribed in Eq. (2).
- *w/o* $\boldsymbol{X}^S$ represents the case opposite to the above, where face recognition is carried out using only the complementary feature maps, corresponding to the privacy requirement prescribed Eq. (6).
- *w/o* $\boldsymbol{W}$ represents the model trained without using the weight matrices $\boldsymbol{W}$ in Eq. (12), equivalent to setting all weights to 1.
- *w/o* $\mathcal{L}_{\text{Embed}}$ denotes the case without using the embedding loss $\mathcal{L}_{\text{Emb}}$ in training.
- *w/o Hybrid* denotes the scenario of training the model using the three base privacy filters scheduled at random, without using hybrid obfuscation.

The function of PCP has been discussed in Section IV-C.2 so is not repeated here. In addition, we experimented with two more designs proposed by DuetFace [34] in our approach:

- *w/ Facial mask*: adding facial masks to remove the face background in the complementary feature maps.
- *w/ Mask+Avg.*: the scenario where multiple complementary feature maps for each stage are averaged over channels into one, with facial mask jointly applied. This is identical to the processing pipeline of DuetFace [34].

The ablation study results are shown in Table VIII. It is evident that the verification performance is significantly affected when either the obfuscated image or the complementary features are removed. This demonstrates the effectiveness of the proposed feature compensation mechanism and verifies the objective privacy requirements prescribed in Eq. (2) and (6). By removing each of the other training-phase components, such as the weight matrices $\boldsymbol{W}$ in feature compensation, the embedding loss $\mathcal{L}_{\text{Embed}}$, and the hybrid obfuscation, degradations in recognition performance to different degrees are observed. Moreover, employing the same processing pipeline for the complementary feature maps as DuetFace [34] does not yield improved results. All of these pieces of evidence validate our final framework setups.

V. CONCLUSION

In this paper, we introduce PRO-Face C, a novel paradigm designed to address the privacy-utility trade-off inherent in face recognition and facial capturing systems. Within this framework, a face image captured by a client device (such as a smartphone or webcam) initially undergoes a pre-obfuscation using conventional and intuitive privacy filters, resulting in a visually-protected form. A feature compensation mechanism is then employed to integrate a set of privacy-preserving feature maps as complementary information into the online recognition system. This integration enables the system to infer the real identity of the obfuscated face with high confidence.

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

Consequently, the framework facilitates accurate face recognition on an untrusted server, ***ensuring recognition utility***, without exposing the original face in clear form, thus ***safeguarding visual privacy***. Furthermore, the obfuscated image retains meaningful visual content for low-level functions like photo preview, instead of being completely randomized, thereby ***preserving visualization utility***. Additionally, the high recognition performance achieved relies solely on a pre-trained recognition backbone on the server, eliminating the need for re-training or fine-tuning the entire system, and thereby ***supporting high generality***. As a result, the proposed approach offers a highly general, usable, and well-balanced solution for visual privacy protection in most client-server structured face recognition systems. Extensive experiments conducted on multiple face image datasets under distinct recognition and privacy breach scenarios demonstrate the superiority of the proposed approach over competitive privacy-preserving face recognition (FFPR) and identifiable face anonymization (IDFA) methods.

Nevertheless, there still exist several limitations: 1) While the proposed hybrid obfuscation provides stronger protection than conventional privacy filters, it remains vulnerable to reconstruction attacks to some extent. Exploring more advanced obfuscation methods could further enhance privacy protection against malicious reconstruction. 2) Despite implementing overhead control mechanisms, the communication and storage overhead induced by the complementary feature maps are still non-negligible. This highlights the need for the development of more compressive approaches. 3) This work mainly presents a privacy-preserving face recognition approach in general cases without specifically addressing extreme and challenging samples. Future research will explore dedicated methods to address the same task in real-world scenarios. 4) Lastly, this paper primarily focuses on protecting visual privacy in the inference stage of face recognition, without addressing the privacy risks associated with the facial embeddings or the training data, which are common challenges for broader recognition systems. Therefore, solutions to these limitations will be pursued in our future work.

REFERENCES

[1] A. Satariano. (2019). *Police Use of Facial Recognition is Accepted By British Court*. New York Times. [Online]. Available: https://www.nytimes.com/2019/09/04/business/facial-recognition-uk-court.html

[2] K. Hill. (2020). *The Secretive Company That Might End Privacy as We Know It*. [Online]. Available: https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

[3] W. Lei. (2021). *China's First Lawsuit on Facial Recognition Made Verdict*. [Online]. Available: https://news.cgtn.com/news/2021-04-10/China-s-first-lawsuit-on-facial-recognition-made-verdict-ZlvYRrL2la/index.html

[4] T. Winkler and B. Rinner, "TrustCAM: Security and privacy-protection for an embedded smart camera based on trusted computing," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug. 2010, pp. 593–600.

[5] H. Wu et al., "PECAM: Privacy-enhanced video streaming and analytics via securely-reversible transformation," in *Proc. 27th Annu. Int. Conf. Mobile Comput. Netw.*, 2021, pp. 229–241.

[6] S. Shan, E. Wenger, J. Zhang, H. Li, H. Zheng, and B. Y. Zhao, "Fawkes: Protecting privacy against unauthorized deep learning models," in *Proc. 29th USENIX Secur. Symp.*, 2020, pp. 1589–1604.

[7] V. Mirjalili, S. Raschka, and A. Ross, "PrivacyNet: Semi-adversarial networks for multi-attribute face privacy," *IEEE Trans. Image Process.*, vol. 29, pp. 9400–9412, 2020.

[8] X. Yang et al., "Towards face encryption by generating adversarial identity masks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3877–3887.

[9] S. Hu et al., "Protecting facial privacy: Generating adversarial identity masks via style-robust makeup transfer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 14994–15003.

[10] Y. Zhong and W. Deng, "OPOM: Customized invisible cloak towards face privacy protection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3590–3603, Mar. 2023.

[11] L. Yuan, P. Korshunov, and T. Ebrahimi, "Privacy-preserving photo sharing based on a secure JPEG," in *Proc. IEEE Conf. Comput. Commun. Workshops*, Apr. 2015, pp. 185–190.

[12] L. Yuan, P. Korshunov, and T. Ebrahimi, "Secure JPEG scrambling enabling privacy in photo sharing," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, vol. 4, May 2015, pp. 1–6.

[13] J. Zhou, D. Shang, H. Lang, G. Ye, and Z. Xia, "Face template protection through residual learning based error-correcting codes," in *Proc. 4th Int. Conf. Control Comput. Vis.*, Aug. 2021, pp. 112–118.

[14] G. Mai, K. Cao, X. Lan, and P. C. Yuen, "SecureFace: Face template protection," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 262–277, 2021.

[15] V. M. Patel, N. K. Ratha, and R. Chellappa, "Cancelable biometrics: A review," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 54–65, Sep. 2015.

[16] H. Hukkelås, R. Mester, and F. Lindseth, "DeepPrivacy: A generative adversarial network for face anonymization," in *Proc. Int. Symp. Vis. Comput.*, 2019, pp. 565–578.

[17] M. Maximov, I. Elezi, and L. Leal-Taixé, "CIAGAN: Conditional identity anonymization generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5446–5455.

[18] X. Gu, W. Luo, M. S. Ryoo, and Y. J. Lee, "Password-conditioned anonymization and deanonymization with face identity transformers," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 727–743.

[19] J. Cao, B. Liu, Y. Wen, R. Xie, and L. Song, "Personalized and invertible face de-identification by disentangled identity information manipulation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3314–3322.

[20] Z. Kuang et al., "Effective de-identification generative adversarial network for face anonymization," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 3182–3191.

[21] H. Proença, "The UU-Net: Reversible face de-identification for visual surveillance video footage," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 2, pp. 496–509, Feb. 2022.

[22] D. Li, W. Wang, K. Zhao, J. Dong, and T. Tan, "RiDDLE: Reversible and diversified de-identification with latent encryptor," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 8093–8102.

[23] Y. Wen, B. Liu, J. Cao, R. Xie, L. Song, and Z. Li, "IdentityMask: Deep motion flow guided reversible face video de-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8353–8367, Dec. 2022.

[24] A.-R. Sadeghi, T. Schneider, and I. Wehrenberg, "Efficient privacy-preserving face recognition," in *Information, Security and Cryptology—ICISC 2009*. Berlin, Germany: Springer, 2009, pp. 229–244.

[25] O. O. Ergun, "Privacy preserving face recognition in encrypted domain," in *Proc. IEEE Asia–Pacific Conf. Circuits Syst. (APCCAS)*, Nov. 2014, pp. 643–646.

[26] X. Kou, Z. Zhang, Y. Zhang, and L. Li, "Efficient and privacy-preserving distributed face recognition scheme via FaceNet," in *Proc. ACM Turing Award Celebration Conf.—China (ACM TURC)*. New York, NY, USA: ACM, 2021, pp. 110–115.

[27] M. Abdalla, F. Bourse, A. De Caro, and D. Pointcheval, "Simple functional encryption schemes for inner products," in *Public-Key Cryptography—PKC 2015*. Berlin, Germany: Springer, 2015, pp. 733–751.

[28] D. Osorio-Roig, C. Rathgeb, P. Drozdowski, and C. Busch, "Stable hash generation for efficient privacy-preserving face identification," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 4, no. 3, pp. 333–348, Jul. 2022.

[29] Y. Huang, Z. Song, K. Li, and S. Arora, "InstaHide: Instance-hiding schemes for private distributed learning," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 4507–4518.

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.

[30] Z. Ma, Y. Liu, X. Liu, J. Ma, and K. Ren, "Lightweight privacy-preserving ensemble classification for face recognition," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5778–5790, Jun. 2019.

[31] X. Yang, H. Zhu, R. Lu, X. Liu, and H. Li, "Efficient and privacy-preserving online face recognition over encrypted outsourced data," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jul. 2018, pp. 366–373.

[32] Y. Wang, J. Liu, M. Luo, L. Yang, and L. Wang, "Privacy-preserving face recognition in the frequency domain," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2022, vol. 36, no. 3, pp. 2558–2566.

[33] J. Ji et al., "Privacy-preserving face recognition with learnable privacy budgets in frequency domain," in *Computer Vision—ECCV 2022*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds. Cham, Switzerland: Springer, 2022, pp. 475–491.

[34] Y. Mi et al., "DuetFace: Collaborative privacy-preserving face recognition via channel splitting in the frequency domain," in *Proc. 30th ACM Int. Conf. Multimedia*. New York, NY, USA: Association for Computing Machinery, 2022, pp. 6755–6764.

[35] Y. Mi et al., "Privacy-preserving face recognition using random frequency components," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 19673–19684.

[36] Y. Li, Y. Wang, and D. Li, "Privacy-preserving lightweight face recognition," *Neurocomputing*, vol. 363, pp. 212–222, Oct. 2019.

[37] K. Wang et al., "FaceMAE: Privacy-preserving face recognition via masked autoencoders," 2022, *arXiv:2205.11090*.

[38] Y. Mi et al., "Privacy-preserving face recognition using trainable feature subtraction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, to be published.

[39] Y. Mao, S. Yi, Q. Li, J. Feng, F. Xu, and S. Zhong, "A privacy-preserving deep learning approach for face recognition with edge computing," in *Proc. USENIX Workshop Hot Topics Edge Comput. (HotEdge)*, 2018, pp. 1–6.

[40] M. A. P. Chamikara, P. Bertok, I. Khalil, D. Liu, and S. Camtepe, "Privacy preserving face recognition utilizing differential privacy," *Comput. Secur.*, vol. 97, Oct. 2020, Art. no. 101951.

[41] C. Zhang, X. Hu, Y. Xie, M. Gong, and B. Yu, "A privacy-preserving multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *Frontiers Neurorobotics*, vol. 13, p. 112, Jan. 2020.

[42] P. Korshunov and T. Ebrahimi, "Using face morphing to protect privacy," in *Proc. 10th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug. 2013, pp. 208–213.

[43] J. Zhao et al., "Dual-agent GANs for photorealistic and identity preserving profile face synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Red Hook, NY, USA, 2017, pp. 65–75.

[44] J. Zhao et al., "Towards pose invariant face recognition in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2207–2216.

[45] J. Zhao, L. Xiong, J. Li, J. Xing, S. Yan, and J. Feng, "3D-aided dual-agent GANs for unconstrained face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2380–2394, Oct. 2019.

[46] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality adaptive margin for face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 18729–18738.

[47] J. Li, L. Han, H. Zhang, X. Han, J. Ge, and X. Cao, "Learning disentangled representations for identity preserving surveillance face camouflage," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 9748–9755.

[48] J. Li et al., "Identity-preserving face anonymization via adaptively facial attributes obfuscation," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 3891–3899.

[49] J. Li, H. Zhang, S. Liang, P. Dai, and X. Cao, "Privacy-enhancing face obfuscation guided by semantic-aware attribution maps," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 3632–3646, 2023.

[50] L. Yuan, L. Liu, X. Pu, Z. Li, H. Li, and X. Gao, "PRO-face: A generic framework for privacy-preserving recognizable obfuscation of face images," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 1661–1669.

[51] Y. Zhang, T. Wang, R. Zhao, W. Wen, and Y. Zhu, "RAPP: Reversible privacy preservation for various face attributes," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 3074–3087, 2023.

[52] Z. Yuan, Z. You, S. Li, Z. Qian, X. Zhang, and A. Kot, "On generating identifiable virtual faces," in *Proc. 30th ACM Int. Conf. Multimedia (MM)*. New York, NY, USA: ACM, 2022, pp. 1465–1473.

[53] T. Wang, Y. Zhang, R. Zhao, W. Wen, and R. Lan, "Identifiable face privacy protection via virtual identity transformation," *IEEE Signal Process. Lett.*, vol. 30, pp. 773–777, 2023.

[54] S. Chen, Y. Liu, X. Gao, and Z. Han, "MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices," in *Proc. Chin. Conf. Biometric Recognit.*, 2018, pp. 428–438.

[55] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4685–4694.

[56] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3730–3738.

[57] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Proc. Workshop Faces 'Real-Life' Images, Detection, Alignment, Recognit.*, 2008.

[58] S. Sengupta, J. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, "Frontal to profile face verification in the wild," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–9.

[59] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, "AgeDB: The first manually collected, in-the-wild age database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1997–2005.

[60] T. Zheng and W. Deng, "Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments," Pattern Recognit. Intell. Syst. Lab., Beijing Univ. Posts Telecommun., Beijing, China, Tech. Rep. 18-01, 2018.

[61] T. Zheng, W. Deng, and J. Hu, "Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments," 2017, *arXiv:1708.08197*.

[62] C. Whitelam et al., "IARPA Janus benchmark-B face dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 592–600.

[63] B. Maze et al., "IARPA Janus benchmark-C: Face dataset and protocol," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 158–165.

[64] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 67–74.

[65] Q. Wang, P. Zhang, H. Xiong, and J. Zhao, "Face.evoLVe: A cross-platform library for high-performance face analytics," *Neurocomputing*, vol. 494, pp. 443–445, Jul. 2022.

[66] F. Mireshghallah, M. Taram, A. Jalali, A. T. T. Elthakeb, D. Tullsen, and H. Esmaeilzadeh, "Not all features are equal: Discovering essential features for preserving prediction privacy," in *Proc. Web Conf.*, New York, NY, USA: Association for Computing Machinery, 2021, pp. 669–680.

[67] B.-W. Tseng and P.-Y. Wu, "Compressive privacy generative adversarial network," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2499–2513, 2020.

[68] B. Meden et al., "Privacy–enhancing face biometrics: A comprehensive survey," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 4147–4183, 2021.

[69] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1833–1844.

[70] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, pp. 600–612, 2004.

[71] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.



Lin Yuan received the B.Eng. degree in electronic science and technology from the University of Electronic Science and Technology of China (UESTC) in 2011 and the Ph.D. degree in electrical engineering from École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2017. He is currently an Associate Professor with the School of Cyber Security and Information Law, Chongqing University of Posts and Telecommunications. His research interests include image and video analysis, multimedia privacy protection, and media forensics.

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.



Wu Chen received the B.S. degree from the School of Intelligence Technology and Engineering, Chongqing University of Science and Technology, Chongqing, China, in 2021. He is currently pursuing the M.S. degree in software engineering with the School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing. His current research interests include multimedia security and privacy protection.



Xiao Pu received the Ph.D. degree in electrical engineering from École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2018. She is currently with the School of Cyber Security and Information Law, Chongqing University of Posts and Telecommunications. Her research interests include cross-modal semantic understanding and natural language processing.



Yan Zhang received the B.E. degree from the National University of Defense Technology, Changsha, China, in 2014, the M.S. degree from the University of Birmingham, U.K., in 2017, and the Ph.D. degree from Chongqing University, Chongqing, China, in 2020. He is currently a Lecturer with Chongqing University of Posts and Telecommunications. His research interests include image processing, machine learning, and pattern recognition.



Hongbo Li is currently a Senior Engineer and the Graduate Supervisor of the School of Innovation and Entrepreneurship Education, Chongqing University of Posts and Telecommunications. He has published two books, around 50 technical articles in refereed journals and proceedings, and has been granted 15 patents. His research interests include digital media analysis, computer vision, and information engineering.



Yushu Zhang (Senior Member, IEEE) received the Ph.D. degree from the College of Computer Science, Chongqing University, Chongqing, China, in December 2014. He held various research positions with the City University of Hong Kong, Hong Kong; Southwest University, Chongqing; University of Macau, Macau, China; and Deakin University, Geelong, VIC, Australia. He is currently a Professor with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China. His research interests include multimedia security, artificial intelligence security, and blockchain. He has published over 150 refereed journal articles and conference papers in these areas.



Xinbo Gao (Fellow, IEEE) received the B.Eng., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively. From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Post-Doctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. From 2001 to 2020, he was with the School of Electronic Engineering, Xidian University. He is currently a Cheung Kong Professor with the Ministry of Education and the President of Chongqing University of Posts and Telecommunications. He has published more than 200 articles in refereed journals and proceedings. His current research interests include multimedia analysis, computer vision, pattern recognition, machine learning, and wireless communications.



Touradj Ebrahimi (Fellow, IEEE) received the M.Sc. and Ph.D. degrees from École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 1989 and 1992, respectively. He is currently a Professor with EPFL, where heads the Multimedia Signal Processing Group (MMSPG). He is the Co-Founder of a startup company RayShaper, based in Crans-Montana, Switzerland, which is active in the computational vision for machine- and human-centric applications. He has published over 500 technical articles in refereed journals and proceedings. His research interests include image and video compression, quality of experience in multimedia, and media security. He is a fellow of SPIE and EURASIP. He was a recipient of the Society of Motion Picture and Television Engineers (SMPTE) Progress Medal-the most coveted award in the field of image processing. He has been the Convener of the JPEG Standardization Committee since 2014.

Authorized licensed use limited to: University Duisburg Essen. Downloaded on May 26,2025 at 11:39:22 UTC from IEEE Xplore. Restrictions apply.