Balancing Privacy and Performance: A Many-in-One Approach for Image Anonymization

Xuemei Jia[1,2*], Jiawei Du[3], Hui Wei[1,2], Ruinian Xue[1,2], Zheng Wang[1,2†], Hongyuan Zhu[4], Jun Chen[1,2†]

[1] National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University, China
[2] Hubei Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University, China
[3] Centre for Frontier AI Research (CFAR) & Institute of High Performance Computing (IHPC), A*STAR, Singapore
[4] Centre for Frontier AI Research (CFAR) & Institute for Infocomm Research (I^{2}R), A*STAR, Singapore

Abstract

The effective utilization of data through Deep Neural Networks (DNNs) has profoundly influenced various aspects of society. The growing demand for high-quality, particularly personalized, data has spurred research efforts to prevent data leakage and protect privacy in recent years. Early privacy-preserving methods primarily relied on instance-wise modifications, such as erasing or obfuscating essential features for de-identification. However, this approach highlights an inherent trade-off: minimal modification offers insufficient privacy protection, while excessive modification significantly degrades task performance. In this paper, we propose a novel Recombining for Obfuscation (FRO) approach to address this trade-off. Unlike existing methods that generate one anonymized instance by perturbing the original data on a *one-to-one* basis, our FRO approach generates an anonymized instance by reassembling mixed id-related features from multiple original data sources on a *many-in-one* basis. Instead of introducing additional noise for de-identification, our approach leverages the existing non-polluted features from other instances to anonymize data. Extensive experiments on identity identification tasks demonstrate that FRO outperforms previous state-of-the-art methods, not only in utility performance but also in visual anonymization.

Introduction

With rapid advances in deep learning, Deep Neural Networks (DNNs) have gained the exceptional ability to learn from vast amounts of data (Jia et al. 2022; Wan et al. 2024; Liu et al. 2025), driving the development of more customized and efficient services. However, the increasing demand for data, particularly personalized data, raises significant concerns about data leakage and privacy breaches during collection. As a result, securing data privacy has become a crucial challenge in the ongoing deployment and advancement of these technologies.

Early research (Zhou and Pun 2021; Li and Choi 2021; Alshaibani and Quinn 2021; Seneviratne et al. 2022) on privacy preservation typically focused on erasing or obfuscating identity-related information to defend against privacy breaches. For instance, Zhou and Pun (2021) introduced perturbations to anonymize the original data, while Alshaibani and Quinn (2021) employed facial blurring to protect privacy. However, the enhanced privacy often comes at the cost of degraded performance in downstream tasks. Although subsequent research (Ahmad, Morerio, and Bue 2023; Kansal, Wong, and Kankanhalli 2024) has proposed bi-level optimization methods to generate adaptive perturbations aimed at improving performance, some degree of performance degradation remains unavoidable.

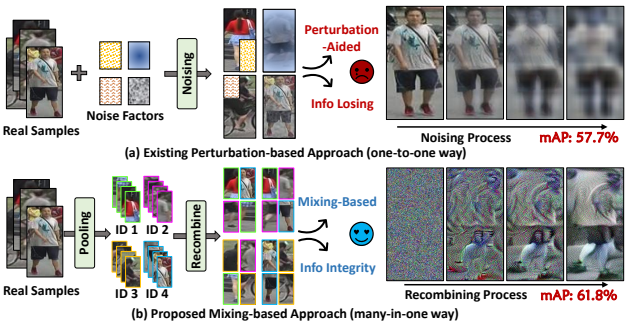


Figure 1: Pipeline comparison. (a) In a *one-to-one* way, existing methods typically append perturbations caused by different noise factors to transform real data into anonymized versions, often resulting in identity-related information loss. The extent of this information loss causes a trade-off between privacy protection and data utility. (b) Our proposed *many-in-one* approach preserves the integrity of identity information without external noise. Recombining features from mixed fragments of multiple identities at the feature level effectively mitigates the risk of personal privacy breaches.

The primary factor contributing to performance degradation is the loss of information caused by these added perturbations. Ideally, these perturbations are intended to erase only visual identity-related features. However, they often inadvertently remove or distort other crucial features necessary for downstream tasks, leading to significant information loss. This creates an inherent trade-off between erasing visual identity-related features and preserving the remaining essential features in privacy preservation efforts.

To address this trade-off, we aim to introduce less external noise but leverage existing features from other identities to

*This work was done during her visiting to A*STAR.
†Corresponding authors.

obfuscate the visual features. The existing methods follow the *one-to-one* basis for generating protected data, i.e., each data instance is processed individually by adding perturbation (Li and Lin 2019; Ahmad, Morerio, and Bue 2023). By focusing on each instance in isolation, this *one-to-one* basis fails to utilize the inter-instance relationships that could be used to enhance privacy preservation without compromising performance. Figure 1 (a) illustrates how the *one-to-one* image constraint introduced perturbations to the original data to create a protected version, inevitably causing information loss/degradation.

Recognizing the limitations of the *one-to-one* approach, we explore a novel *many-in-one* strategy to mitigate information loss. Recent advances in deep learning interpretability (Ribeiro, Singh, and Guestrin 2018; Goyal et al. 2019) demonstrate that local critical regions of images can still yield accurate predictions when combined with perturbed parts of other instances. Building on this insight, we consider using unaltered features from multiple other identities as perturbations to obscure the visual features. By reassembling id-specific features from various identities to generate a new instance, the *many-in-one* approach minimizes the need for external noise while preserving essential features.

In this paper, we introduce a novel approach called Feature Recombining for Obfuscation (FRO), which embodies the *many-in-one* strategy to balance identity privacy with task performance. FRO extracts features from original data to build a knowledge pool of fragmented features, which are used to optimize a randomly initialized instance into a new protected sample. To ensure anonymization, we apply label softening to regularize feature acquisition and use Kullback-Leibler divergence with varying temperatures to minimize information loss during reassembly. Finally, we assess the effectiveness of the protected samples through a process we term knowledge replay. Figure 1 illustrates how our FRO method compares to the previous *one-to-one* strategy, demonstrating its superior ability to preserve essential features while enhancing privacy. The proposed FRO method demonstrates enhanced utility performance in identity-related tasks, person re-identification and face recognition, surpassing state-of-the-art methods. Concurrently, computational results and human perception evaluations support its visual anonymization effectiveness.

To conclude, our contribution can be summarized:

- We address the limitations of existing privacy-preserving methods that rely on introducing external noise. Our work highlights the inherent trade-off in traditional *one-to-one* strategies between erasing visual identity-related features and preserving essential features.
- We propose a novel *many-in-one* strategy for privacy preservation that avoids the need for external noise. This strategy is realized through our Feature Recombining for Obfuscation (FRO) method, which combines identity-specific features from a knowledge pool to generate protected data instances.
- Extensive experiments on two identity-related tasks have demonstrated the superior performance of FRO in maintaining the utility of anonymized data. Moreover, quantitative and qualitative results highlight the substantial capability of the method in safeguarding visual privacy.

Preliminaries and Related Works

Throughout this paper, we use f_θ to denote a neural network with the corresponding weight parameters θ. The network f_θ is normally trained on a real dataset $\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{|\mathcal{X}|}$. Traditionally, model weight parameters θ is trained by,

$$\theta_{\mathcal{X}} = \arg\min_{\theta} \mathbb{E}_{x,y\in\mathcal{X}} \mathcal{L}_{\text{ID}(x,y)}, \tag{1}$$

where $\mathcal{L}_{\text{ID}}$ is specified identity-related loss, *e.g.*, classification loss and metric loss (Schroff, Kalenichenko, and Philbin 2015), or a combination of multiple identity losses.

Early privacy-protecting techniques often degraded the quality of original face images by simple operations, such as masking (Liu, Kong, and Wang 2018; Meden et al. 2021; Seneviratne et al. 2022), based on the assumption that only facial regions in images contain sensitive personal information. Later, it was widely recognized among researchers that entire images of individuals contain personal privacy information. Zhou and Pun (2021); Li and Choi (2021) opted to blur specific details to prevent privacy breaches, yet obscured images still preserve primary body structures and clothing style. Wang, Kelly, and Veldhuis (2021) treated only certain aspects of identity information, such as gender, as privacy, altering images to reflect only gender changes.

Recently, focusing on face images, Li and Lin (2019); Barattin et al. (2023) implemented identity anonymization while preserving facial attributes using GAN-optimized latent codes. Pan et al. (2024) isolated depression-related signals from facial images, protecting identities with a simulated privacy lens. Targeting person images, Ahmad, Morerio, and Bue (2023) developed an anonymization network with similarity suppression loss to prevent identity recognition. Likewise, Kansal, Wong, and Kankanhalli (2024) produced protected images using adversarial identity constraints and similarity guidance. Furthermore, (Hanisch et al. 2024) proposed to evaluate the performance of anonymization methods in the worst-case scenario.

The privacy protection techniques outlined above can be categorized as perturbation-based techniques. They ensure privacy by injecting external noise into the original data and subsequently optimizing it to obtain the *one-to-one* privacy-preserving version. As shown Figure 1 (a), such optimization can be formulated as:

$$\theta_{\mathcal{X}} = \arg\min_{\theta}\max_{\delta} \mathbb{E}_{x,y\in\mathcal{X}}[\mathcal{L}_{\text{ID}(x,y)} + \mathcal{L}_{\text{sim}}(x,\delta)], \tag{2}$$

where $\mathcal{L}_{\text{sim}}(x,\delta)$ constrains the similarity between the original data x and the perturbation-based $x+\delta$. The objective is to obtain a model that is well-trained on original data and a protected/anonymized set that is visually dissimilar to the original ones. The protected image $x+\delta$ generated on a *one-to-one* basis is highly dependent on the guidance of $\mathcal{L}_{\text{sim}}$, inevitably complying with the biometric structures of original data x. Besides, introducing δ to modify the real data will unconditionally bring about direct information loss.

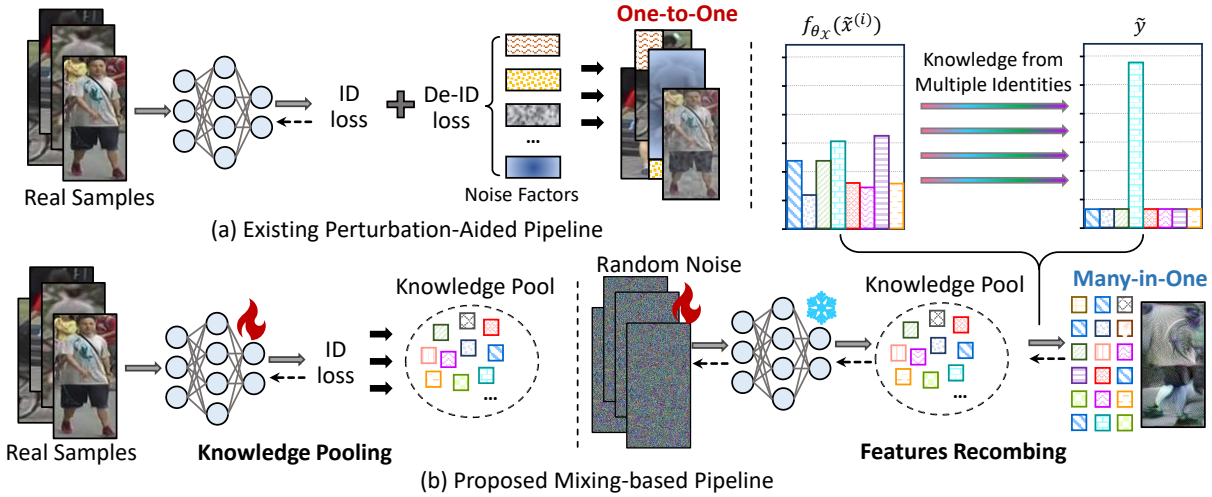


Figure 2: Overview of existing perturbation-based approach (a) and the proposed mixing-based Features Recombining for Obfuscation (FRO) (b). (a) Perturbation-based approaches often rely upon the noise factor to add perturbation to the original data. They generate *one-to-one* anonymized version with information loss. (b) FRO fundamentally resolves the issue of incomplete anonymization by recombining features to generate new instances in a *many-in-one* manner. Furthermore, it preserves the latent information inherent in the original distribution.

Methodology

Considering the limitations of the one-to-one approach, we explore a novel *many-in-one* strategy to mitigate information loss. In this section, we begin generating protected instances by blending patches from multiple identities to induce visual obfuscation, a simple pixel-level technique we term Patches Recombining for Obfuscation (PRO). However, the protected samples remain vulnerable to recovery through a finite exhaustive search, indicating a limitation in privacy preservation. To enhance privacy security, we reassemble new instances at the feature level, still avoiding introducing external noises. Subsequently, we detail the derivation and introduction of Features Recombining for Obfuscation (FRO), to achieve comprehensive anonymization while maintaining competitive utility performance. Figure 2 (b) shows the overall pipeline of the proposed FRO.

Patches Recombining for Obfuscation

To balance reduced recognizability with preserved data integrity, we introduce Patches Recombining for Obfuscation (PRO), a straightforward pixel-level method for preliminary exploration. PRO segments person images into grids and rearranges them into puzzled samples (Figure 3), reducing identity loss and enhancing visual ambiguity without external noise.

Through training real data and puzzled data generated by PRO, we demonstrate the feature extraction capabilities of models, visualized by t-SNE (Van der Maaten and Hinton 2008) in Figure 3. The results indicate that puzzled samples do not train the model to distinguish obfuscated identities. However, the model still achieves classification performance on the target validation set, closely approximating that of a model trained on real data. This confirms the approach's ability to maintain utility performance.

However, the patch fragments used in pixel-level obfuscation may still lead to potential data breaches and are vulnerable to recovery through finite patch-level searches, failing to ensure complete visual anonymization. To address this, we further explore feature-level combinations as a more robust privacy-preserving approach, effectively mitigating information leakage.

Figure 3: Visualization of identity differentiation capability on real data training and that of Patches Recombining for Obfuscation.

Features Recombining for Obfuscation

To overcome the limitations of PRO in fully safeguarding personal privacy information, we introduce Features Recombining for Obfuscation (FRO) at the feature level. FRO leverages unaltered features from multiple other identities as perturbations to acquire new obscured instances. Specifically, identity-related feature fragments from all identities are stored within a knowledge pool. These fragments are then reassembled into visually obfuscated instances without

Algorithm 1: Feature Resembling Algorithm

Input: $f_{\theta_{\mathcal{X}}}$: Pre-trained knowledge pool;
Parameter: θ: pre-trained parameter weight; t: iteration to update the obfuscated samples; C: the number of original data classes; temperature scale T; label softening parameter ϵ; learning rate η;
Output: protected person images

for $k \leftarrow 1$ to C **do**
  $\tilde{x}_k^0 \sim \mathcal{N}(0,1)$ {Initialize the input}
  $\tilde{y}_k^0 \sim \mathcal{Y} = \{0, 1, ..., C-1\}$ {Specify main class}
  $\tilde{y}_k = (1-\epsilon)\tilde{y}_k^0 + \epsilon/C$ {Soften the labels}
  for $i \leftarrow 1$ to t **do**
    $\tilde{x}_k^{(i+1)} = \tilde{x}_k^{(i)} - \eta \nabla_{\tilde{x}} (\mathbb{E}[\mathcal{L}_{\tilde{\mathcal{X}}}] + \alpha \mathcal{R}_{\text{BN}}(\tilde{x}))$
      {Update the obfuscated samples}
  end for
end for
return protected person images $\tilde{\mathcal{X}}$.

introducing external noise. Finally, the obfuscated instances undergo a knowledge replay process to ensure that utility performance is maintained.

Knowledge Pooling. The knowledge pooling step aims to extract identity-related information from the original dataset and encapsulate it within a deep neural network. The trained model serves as a knowledge pool, embedding the rich information the original data distribution provides. The model is trained on the original dataset $\mathcal{X}$ using the formulation in Eq. 1, employing cross-entropy loss with one-hot labels, $\mathcal{L}_{\text{ID}} = -y \log(p(x))$, $p(x)$ is the predicted probability. Following the knowledge pooling training, the data distribution knowledge inherent in the original dataset is distilled and encapsulated into feature fragments, parameterized by $f_{\theta_{\mathcal{X}}}$.

Feature Assembling. For the data utility and recovering the knowledge of real data distribution, we propose assembling new instances by leveraging the dispersed features stored within the knowledge pool. We integrate features from multiple identities to meet privacy-preserving requirements, generating new instances with visually obfuscated characteristics. The recombined instances encapsulate extensive details about the data distribution, encompassing abundant inter-class and intra-class features. The core mechanism of Feature Assembling is detailed in

To meet privacy-preserving requirements, we initiate the optimization of protected data $\tilde{\mathcal{X}}$ from chaos, ensuring no visual cues from the original data $\mathcal{X}$ are retained. We initial the protected images of the specified class k as $(\tilde{x}_k^0, \tilde{y}_k^0)$, where $\tilde{x}_k^0 \sim \mathcal{N}(0,1)$, $\tilde{y}_k^0 \sim \mathcal{Y} = \{0, 1, ..., C-1\}$.

Rather than relying on the cross-entropy loss over a hard target label, we employ a softer approach using Kullback-Leibler divergence (Kullback and Leibler 1951) to distill rich information from all identities within the knowledge pool. Blending information from multiple identities effectively induces visual obfuscation and prevents information loss. Here, the knowledge pool $f_{\theta_{\mathcal{X}}}$ computes the soft logits z_i of the updating sample. Normally, a "softmax" output layer is utilized to obtain probability distribution, $p_i = e^{z_i} / \sum_j e^{z_j}$. We employ the distribution as a softer target to guide sample updates, encompassing probability information across all classes. To further enrich the supervision signals provided by the target, we implement a temperature scaling strategy, following the principle outlined by Hinton, Vinyals, and Dean (2015). The temperature scaling makes the feature reassembling process pay more attention to matching logits that are much more negative than the average. It enables the feature recombination to incorporate more extensive information from other classes with lower logits as the temperature increases. The probability is adjusted by,

$$\tilde{p}_i = \frac{e^{z_i/T}}{\sum_j e^{z_j/T}}, \tag{3}$$

where T is a temperature that is normally set to 1.

To further enhance multi-class feature mixing and mitigate distribution information loss, we utilize a softened label rather than the one-hot label Szegedy et al. (2016), $q = \tilde{y}_k, \tilde{y}_k = (1-\epsilon)\tilde{y}_k^0 + \epsilon/C$, ϵ is the softening parameter. Hence, the objective to transfer fragmented knowledge to new instances can be formulated as:

$$\mathcal{L}_{\tilde{\mathcal{X}}} = \Sigma_{0 \le i < C} \mathcal{L}_{\text{KL}}(\tilde{p}_i \parallel q_i). \tag{4}$$

To improve the quality of the recovered images, we employ the Batch Normalization (BN) feature distribution regularization term (Yin et al. 2020), formulated as,

$$\begin{aligned} \mathcal{R}_{\text{BN}}(\tilde{x}) &= \sum_l \|\mu_l(\tilde{x}) - \mathbb{E}(\mu|\mathcal{X})\|_2 + \sum_l \|\sigma_l^2(\tilde{x}) - \mathbb{E}(\sigma_l^2|\mathcal{X})\|_2 \\ &= \sum_l \|\mu_l(\tilde{x}) - \text{BN}_l^{\text{RM}} \|_2 + \sum_l \|\sigma_l^2(\tilde{x}) - \text{BN}_l^{\text{RV}} \|_2, \end{aligned} \tag{5}$$

where l is the index of BN layer, $\mu_l(\tilde{x})$ and $\sigma_l^2(\tilde{x})$ are mean and variance. BN_l^{RM} and BN_l^{RV} are running mean and variance in the pre-trained knowledge pool $f_{\theta_{\mathcal{X}}}$ at l-th layer, which is globally counted.

The overall learning objective can be formulated as:

$$\tilde{\mathcal{X}} = \arg\min_{\tilde{x}} \mathbb{E}[\mathcal{L}_{\tilde{\mathcal{X}}}] + \alpha \mathcal{R}_{\text{BN}}(\tilde{x}). \tag{6}$$

Knowledge Replay. Finally, we replay the knowledge by assessing the new model's performances on the protected datasets $\tilde{\mathcal{X}} = \{(x_i)\}_{i=1}^{|\tilde{\mathcal{X}}|}$. It illustrates how recombined instances convey data distribution information to a neural model. Since these new instances are synthesized from mixed information across multiple identities, we avoid assigning them fixed labels based on the predominant component. Instead, we obtain the soft targets for samples $\tilde{x} \in \tilde{\mathcal{X}}$ by $\tilde{y} = f_{\theta_{\mathcal{X}}}(\tilde{x})$. We utilize a teacher-student training framework, employing Kullback-Leibler divergence to replay the knowledge embedded in the recombined instances effectively. The output of the newly trained model f_ϕ, *i.e.*, referred to as the student, is denoted as $f_{\phi_{\tilde{\mathcal{X}}}}(\tilde{x})$. Finally, we define the following objective as the identity loss for training the model $f_{\phi_{\tilde{\mathcal{X}}}}$ on the protected dataset $\tilde{\mathcal{X}}$:

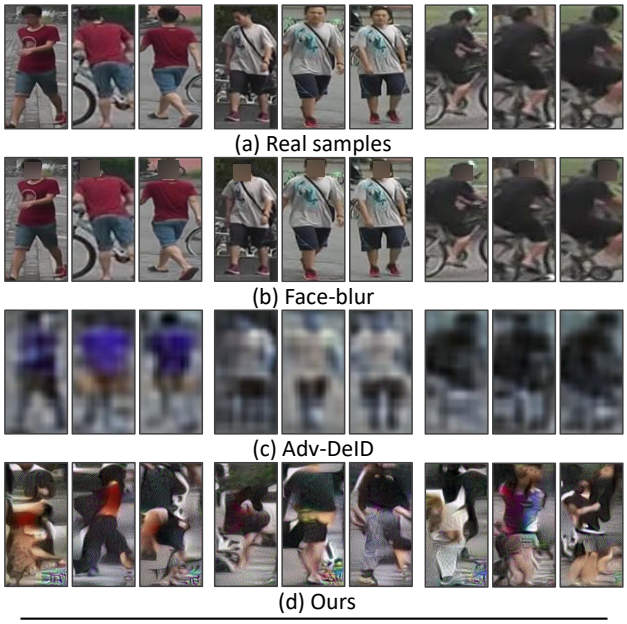
$$\phi_{\tilde{\mathcal{X}}} = \arg\min_{\phi} \mathbb{E}_{\tilde{x} \in \tilde{\mathcal{X}}} \mathcal{L}_{\text{KL}}(f_{\theta_{\mathcal{X}}}(\tilde{x}) \parallel f_{\phi_{\tilde{\mathcal{X}}}}(\tilde{x})). \tag{7}$$

Methods	Models	Backbone	mAP	Rank-1
Base Backbone	SE-Net	R-50	59.1	80.1
	ResNet-50	R-50	77.5	90.1
	ResNet-101	R-101	81.0	92.4
Perturbation -based	DP-SGD	R-50	4.5	17.6
	Federated-Camera	R-50	36.6	61.1
	Face-blur MuDeep	ConvNet	44.8	69.6
	PIS	R-50	51.9	74.9
	Face-blur HACNN	Inception	71.3	87.5
	Face-blur PCB	R-50	72.9	88.2
	De-ID	R-50	57.7	77.6
Mixing -based	FRO (ours)	R-50	61.8	80.3
	FRO (ours)	R-101	65.7	84.6

Table 1: Comparisons with different baselines and privacy-based SOTA methods on the Market-1501 dataset. mAP, Rank-1 (%) accuracy (%) are reported.

Methods	Models	Backbone	mAP	Rank-1
Base Backbone	SE-Net	R-50	52.3	70.3
	ResNet-50	R-50	70.0	84.1
	ResNet-101	R-101	69.8	84.3
Perturbation -based	Face-blur MuDeep	ConvNet	34.8	54.7
	Face-blur HACNN	Inception	56.7	73.1
	Face-blur PCB	R-50	65.4	80.2
	De-ID	R-50	49.3	67.0
Mixing -based	FRO (ours)	R-50	51.4	70.0
	FRO (ours)	R-101	50.9	68.7

Table 2: Comparisons with different baselines and privacy-based SOTA methods on the DUKEMTMC-REID dataset. mAP, Rank-1 (%) accuracy (%) are reported.



Methods	Face-blur	Adv-DeID	FRO (ours)
SSIM	0.84	0.35	0.07
Human Perception	100%	65%	2%

Figure 4: Examples visualization of original and synthetic data samples from MARKET-1501. The table below illustrates computational and human-perceived similarity evaluations between original and synthetic data.

Experiments

Datasets and Experiment Setup

Datasets. MARKET1501 (Zheng et al. 2015) is a large-scale re-id benchmark comprising 32,668 images of 1,501 pedestrians, 751 for training and 750 for testing, captured by six cameras. DUKEMTMC-REID (Zheng, Zheng, and Yang 2017) dataset contains 36,441 images of 1,812 persons captured by eight cameras, 702 identities are used as the training set, and 702 persons are used as the query and gallery, respectively. A subset of CASIA (Yi et al. 2014) containing 30,726 face images of 952 identities are selected in our experiments. And LFW (Huang et al. 2008), CFP-FF and CFP-FP (Sengupta et al. 2016), and AGEDB (Moschoglou et al. 2017) are used for downstream face verification.

Training Details. The models are trained with one NVIDIA GeForce RTX 4090 GPU using Pytorch. We use ResNet-50 (He et al. 2016) as the backbone with Adam Optimizer. The input images are resized to 256×128. The mini-batch size is set to 64, containing 32 persons with 4 images each. The initial learning rate is $3e{-}4$ and is reduced by following an exponentially decaying training schedule. α is set as 0.05.

Evaluation Metrics. We use Rank-1 accuracy and mean average precision (mAP) to evaluate the Re-ID utility performance. SSIM (Wang et al. 2004) and Human Perception are adopted to evaluate the visual anonymization capability of protected data. Additionally, we employ the Silhouette Score (SS) and Adjusted Rand Index (ARI) to evaluate the quality of clustering extracted features.

Privacy Assumption. Our proposed method, FRO, targets real-world cloud-sharing systems, such as surveillance or facial recognition in online banking. In this context, attackers can intercept anonymized data streams with labels but cannot access the server, model, or auxiliary data. The goal is to prevent mapping intercepted data to real individuals, reducing fraud, financial risks, and information misuse.

Results on Person Re-Identification

We compare the ReID utility performance of three types of methods: base backbone (SE-Net (Hu, Shen, and Sun 2018), ResNet-50 and ResNet-101 (He et al. 2016)), *one-to-one* perturbation-based privacy-preserving methods (DP-SGD (Abadi et al. 2016), federated learning (Zhuang et al. 2020, 2023), Face-blur MuDeep (Dietlmeier et al. 2020), PIS (Dou et al. 2022), Adv-DeID (Kansal, Wong, and Kankanhalli 2024)), and the proposed *many-in-one* mixing-based privacy-preserving methods on MARKET-1501 and DUKEMTMC-REID datasets in Table 2. We retrain the base backbone with cross-entropy loss for classification training. As demonstrated, both perturbation-based and mixing-based methods exhibit a performance drop compared to the baseline models without privacy preservation. Among these

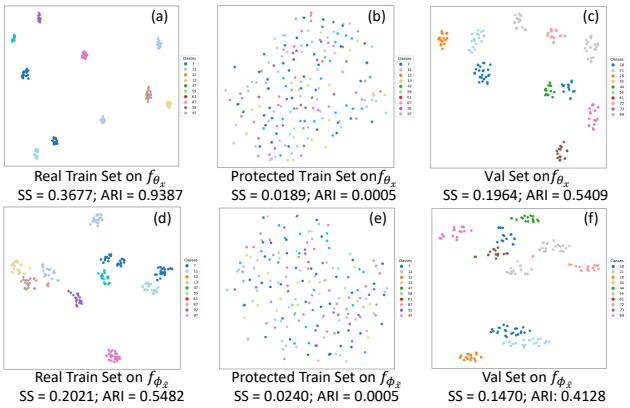
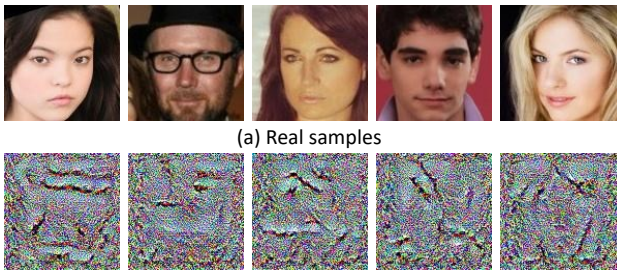


Figure 5: Visualization of model's identity differentiation capability on real data training and that of Features Recombining for Obfuscation (FRO).



Methods	LFW	CFP-FF	CFP-FP	AgeDB
ResNet	78.3	79.3	63.9	64.6
FRO (ours)	69.0	69.8	57.6	55.0

Figure 6: Examples visualization of original and protected faces from CASIA dataset. The table below gives face verification results (%) of the baseline model ResNet-50 and the proposed FRO on four datasets.

methods, face-blur approaches achieve relatively higher mAP and Rank-1 scores than others, and the proposed method achieved second-tier accuracy performance.

To comprehensively evaluate privacy protection capabilities, we present some protected instances of MARKET-1501 dataset in Figure 4, including (a) real samples, (b) face-blurred samples (Dietlmeier et al. 2020), (c) samples generated by Adv-DeID (Kansal, Wong, and Kankanhalli 2024), and (d) our proposed FRO. Face-blurred samples (a) conceal facial features but fail to protect information from other body parts, despite higher identification accuracy. Adv-DeID (c) applies adversarial constraints to alter images but often produces samples resembling the originals due to color distribution limitations. In contrast, the proposed FRO method, benefiting from recombining feature fragments from multiple identities, exhibits superior performance in privacy preservation and identification utility.

To further assess anonymization effectiveness, we present qualitative comparisons using the SSIM metric and visual perception results in Figure 4. Specifically, we measure visual similarity across the dataset by calculating the overall average SSIM between image pairs in the original and protected data within the same category. Our approach achieves an SSIM score of 0.07, significantly lower than that of Adv-DeID. Additionally, we conducted a visual evaluation with 10 volunteers to determine whether the protected images could still be recognized as their true identities. As expected, the recognition rates for Face-blur, Adv-DeID, and our method were 100%, 65%, and 2%, respectively. These findings confirm that our generated data surpasses the *one-to-one* perturbation-based method, providing stronger visual privacy protection than real data.

Additionally, in Figure 5, we illustrate the feature extraction capability of the knowledge pool $f_{\theta_{\mathcal{X}}}$, and the knowledge replay model $f_{\phi_{\tilde{\mathcal{X}}}}$. The model $f_{\theta_{\mathcal{X}}}$ exhibits strong class feature discrimination ability on both train and val sets in (a) and (c). ARI achieves approximately 0.94 and 0.54 on train and val sets, respectively. However, as depicted in (b), it loses clear discernment on the privacy-protected set, with SS and ARI metrics indicating a distribution resembling random noise. In contrast, the model $f_{\phi_{\tilde{\mathcal{X}}}}$ in the second row cannot directly identify the anonymized training set (e) but can distinguish between the original training set (d) and the target validation set (f). The smaller reductions in SS and ARI, compared to those seen in the model trained on real data, suggests that discriminative information at the feature level is effectively preserved and transmitted through FRO.

Results on Face Recognition

We also evaluate our privacy-preserving approach on another identity-related task, face recognition, as shown in Figure 6. In this section, we select a subset of CASIA containing 30,726 face images of 952 identities to train the knowledge pool. The visual examples confirm that the proposed FRO effectively protects facial features while retaining some implicit identity-related textures. The table below demonstrates that anonymized samples from the CASIA dataset maintain the differentiation ability of the pre-trained knowledge pool, achieving notable verification results. Although there is an accuracy loss of less than 10% across four datasets, this trade-off is considered acceptable given the significant improvement in visual privacy protection.

Results on Other Combination Strategies

Besides the proposed Patches Recombining for Obfuscation (PRO), we also evaluate another pixel-level obfuscation method, *i.e.*, mix-up (Zhang et al. 2018). However, mix-up faces similar limitations as PRO in achieving complete privacy protection. Beyond concerns about visual privacy, we illustrate the data processed by the mix-up could be easily recognized by the model $f_{\theta_{\mathcal{X}}}$ trained on the original dataset. Figure 7 illustrates the clustering performance of extracted features using different mixing ratios in the mix-up operation, $r \in \{0.9, 0.8, 0.7, 0.6, 0.5\}$. The results indicate that even at a ratio of $r = 0.7$, the model $f_{\theta_{\mathcal{X}}}$ can still effectively distinguish features from mix-up samples. This finding suggests that pixel-level merging of information from other identity samples does not substantially impair the model's

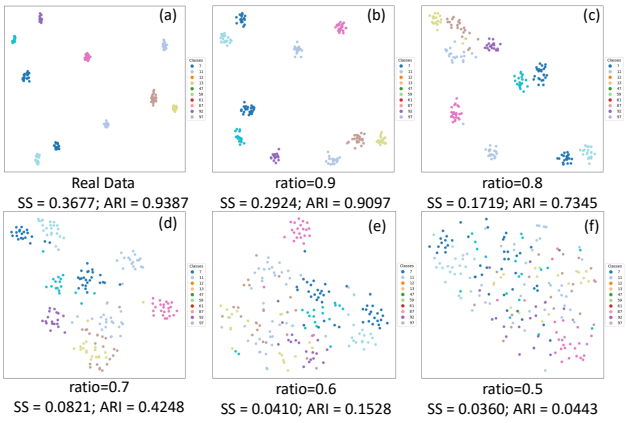


Figure 7: Visualization of real data trained model's identity differentiation capability on mixup processed data with varying ratio.

recognition capabilities, thereby limiting the effectiveness of this method in preserving identity privacy.

Parameter Studies

Effect of Temperature Scale and label softening parameter. In Features Recombining for Obfuscation (FRO), the temperature scaling parameter T and label softening parameter ϵ are pivotal in modulating the proportion of identity information integrated into the obfuscated images. We examine their combined effects on performance in Figure 8, evaluating various configurations where $T \in \{0.5, 1.0, 5, 10, 20, 30\}$ and $\epsilon \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ on MARKET-1501 dataset. In this figure, we use the knowledge pool $f_{\theta_{\mathcal{X}}}$ trained with ResNet-50 backbone on the real training set using cross-entropy as identity loss, achieving 77.5%/90.1% mAP/Rank-1 accuracy on MARKET-1501 as shown in Table 1. Overall, when the label softening parameter ϵ is set to values of 0.1, 0.2, 0.3, or 0.4, variations in temperature exert a relatively minimal influence on the retention of knowledge within the synthesized samples. The overall trend reveals that the label softening parameter considerably influences the final accuracy, particularly at extreme temperature values. Specifically, when $T = 0.5$ and $\epsilon = 0.5$, the low temperature sharpens the distribution, resulting in overly confident predictions for the predominant class. Meanwhile, the large label softening parameter seeks to soften the labels. This combination creates a scenario where the mixed images are less effective in transmitting comprehensive identity distribution information.

Effect of Data Scale. The proposed method has the potential to generate an unlimited amount of protected data. However, given practical resource constraints, we investigate the impact of varying the scale β of the protected dataset, exploring different values of $\beta \in 0.5, 0.75, 1.0, 1.5, 2.0$ to assess how the quantity of generated data influences the effectiveness and performance of the model. We illustrated the results of two combinations of temperature scale and label smooth parameter, $[T = 0.5, \epsilon = 0.2]$ and $[T = 5, \epsilon = 0.0]$, in Figure 9. They showcase the same trend in Re-ID utility performance. When the scale of generated data exceeds

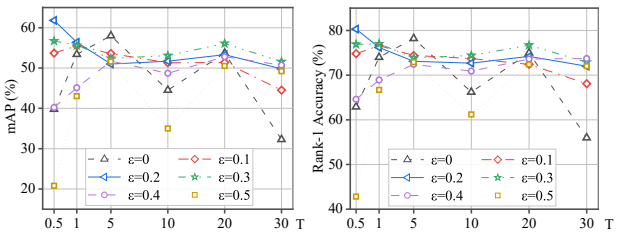


Figure 8: mAP and Rank-1 accuracy on different combinations of temperature scale T and label softening parameter ϵ on MARKET-1501 dataset.

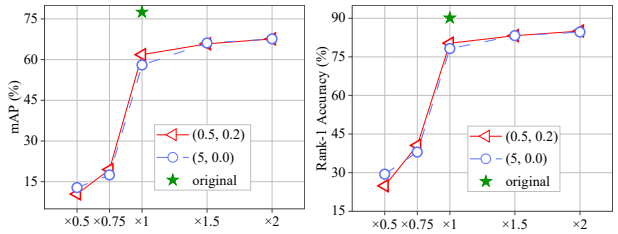


Figure 9: Comparison on different protected data scales on MARKET-1501 dataset. Green stars denote the performances obtained by original data. $\times N$ indicates the approximate ratio to the original training data.

that of the original training set, both mAP and Rank-1 accuracy progressively approach the performance achieved by the original data. This observation indicates that the proposed method can effectively and incrementally capture real data distribution, closely approximating the original accuracy in identity-related tasks. For fair comparisons and to conserve resources, we maintain a data scale similar to the original set in all other experiments.

Conclusion

This paper highlighted the inherent trade-off in traditional *one-to-one* strategies between erasing visual identity-related features and preserving essential features. In response, a novel *many-in-one* method, Feature Recombining for Obfuscation (FRO), was designed, leveraging identity-specific feature fragments from a knowledge pool to reassemble protected instances. Extensive experiments on two identity-related tasks demonstrated the superior performance of FRO in maintaining the utility of anonymized data. Concurrently, quantitative and qualitative results showcased the substantial capability of the method in safeguarding visual privacy.

Ethics Statement. This research focuses on privacy-preserving techniques within deep learning, specifically aimed at protecting personal identity information during image-based tasks. This work contributes to the broader objective of mitigating risks associated with the misuse of sensitive data. We acknowledge the potential ethical implications of manipulating identity-related information and are committed to the responsible use of such techniques.

Acknowledgements

We thank our anonymous reviewers for their valuable feedback. This research was supported by China Scholarship Council Program (202306270235), National Natural Science Foundation of China (62171325, 62071338), A*STAR Career Development Fund (CDF) C233312004, and the EDB Space Technology Development Programme under Project S22-19016-STDP.

References

Abadi, M.; Chu, A.; Goodfellow, I. J.; McMahan, H. B.; Mironov, I.; Talwar, K.; and Zhang, L. 2016. Deep Learning with Differential Privacy. In *ACM SIGSAC Conference on Computer and Communications Security*, 308–318.

Ahmad, S.; Morerio, P.; and Bue, A. D. 2023. Person Re-Identification without Identification via Event Anonymization. In *IEEE/CVF International Conference on Computer Vision*, 11098–11107.

Alshaibani, A.; and Quinn, A. J. 2021. Pterodactyl: Two-Step Redaction of Images for Robust Face Deidentification. In *AAAI Conference on Human Computation and Crowdsourcing*, 27–34.

Barattin, S.; Tzelepis, C.; Patras, I.; and Sebe, N. 2023. Attribute-Preserving Face Dataset Anonymization via Latent Code Optimization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8001–8010.

Dietlmeier, J.; Antony, J.; McGuinness, K.; and O'Connor, N. E. 2020. How important are faces for person re-identification? In *International Conference on Pattern Recognition*, 6912–6919.

Dou, S.; Jiang, X.; Zhao, Q.; Li, D.; and Zhao, C. 2022. Towards Privacy-Preserving Person Re-identification via Person Identify Shift. arXiv:2207.07311.

Goyal, Y.; Wu, Z.; Ernst, J.; Batra, D.; Parikh, D.; and Lee, S. 2019. Counterfactual Visual Explanations. In *International Conference on Machine Learning*, volume 97, 2376–2384.

Hanisch, S.; Todt, J.; Patino, J.; Evans, N. W. D.; and Strufe, T. 2024. A False Sense of Privacy: Towards a Reliable Evaluation Methodology for the Anonymization of Biometric Data. *Proc. Priv. Enhancing Technol.*, 2024(1): 116–132.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Hinton, G. E.; Vinyals, O.; and Dean, J. 2015. Distilling the Knowledge in a Neural Network. arXiv:1503.02531.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-Excitation Networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 7132–7141.

Huang, G. B.; Mattar, M.; Berg, T.; and Learned-Miller, E. 2008. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*.

Jia, X.; Zhong, X.; Ye, M.; Liu, W.; and Huang, W. 2022. Complementary Data Augmentation for Cloth-Changing Person Re-Identification. *IEEE Trans. Image Process.*, 31: 4227–4239.

Kansal, K.; Wong, Y.; and Kankanhalli, M. S. 2024. Privacy-Enhancing Person Re-identification Framework - A Dual-Stage Approach. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, 8528–8537.

Kullback, S.; and Leibler, R. A. 1951. On information and sufficiency. *The annals of mathematical statistics*, 22(1): 79–86.

Li, T.; and Choi, M. S. 2021. DeepBlur: A Simple and Effective Method for Natural Image Obfuscation. arXiv:2104.02655.

Li, T.; and Lin, L. 2019. AnonymousNet: Natural Face De-Identification With Measurable Privacy. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, 56–65.

Liu, S.; Kong, L.; and Wang, H. 2018. Face Detection and Encryption for Privacy Preserving in Surveillance Video. In *Pattern Recognition and Computer Vision, Part III*, volume 11258 of *Lecture Notes in Computer Science*, 162–172.

Liu, W.; Jia, X.; Zhong, X.; Jiang, K.; Yu, X.; and Ye, M. 2025. Dynamic and static mutual fitting for action recognition. *Pattern Recognit.*, 157: 110948.

Meden, B.; Rot, P.; Terhörst, P.; Damer, N.; Kuijper, A.; Scheirer, W. J.; Ross, A.; Peer, P.; and Struc, V. 2021. Privacy-Enhancing Face Biometrics: A Comprehensive Survey. *IEEE Trans. Inf. Forensics Secur.*, 16: 4147–4183.

Moschoglou, S.; Papaioannou, A.; Sagonas, C.; Deng, J.; Kotsia, I.; and Zafeiriou, S. 2017. AgeDB: The First Manually Collected, In-the-Wild Age Database. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 1997–2005.

Pan, Y.; Jiang, J.; Jiang, K.; Wu, Z.; Yu, K.; and Liu, X. 2024. OpticalDR: A Deep Optical Imaging Model for Privacy-Protective Depression Recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Ribeiro, M. T.; Singh, S.; and Guestrin, C. 2018. Anchors: High-Precision Model-Agnostic Explanations. In *AAAI Conference on Artificial Intelligence*, 1527–1535.

Schroff, F.; Kalenichenko, D.; and Philbin, J. 2015. FaceNet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, 815–823.

Seneviratne, S.; Kasthuriarachchi, N.; Rasnayaka, S.; Hettiachchi, D.; and Shariffdeen, R. 2022. Does a Face Mask Protect My Privacy?: Deep Learning to Predict Protected Attributes from Masked Face Images. In Long, G.; Yu, X.; and Wang, S., eds., *Advances in Artificial Intelligence*, volume 13151 of *Lecture Notes in Computer Science*, 91–102.

Sengupta, S.; Chen, J.; Castillo, C. D.; Patel, V. M.; Chellappa, R.; and Jacobs, D. W. 2016. Frontal to profile face verification in the wild. In *IEEE Winter Conference on Applications of Computer Vision*, 1–9.

Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; and Wojna, Z. 2016. Rethinking the Inception Architecture for Computer Vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2818–2826.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Wan, Z.; Wang, Z.; Wang, Y.; Wang, Z.; Zhu, H.; and Satoh, S. 2024. Contributing Dimension Structure of Deep Feature for Coreset Selection. In Wooldridge, M. J.; Dy, J. G.; and Natarajan, S., eds., *AAAI Conference on Artificial Intelligence*, 9080–9088.

Wang, S.; Kelly, U. M.; and Veldhuis, R. N. J. 2021. Gender Obfuscation through Face Morphing. In *International Workshop on Biometrics and Forensics*, 1–6.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4): 600–612.

Yi, D.; Lei, Z.; Liao, S.; and Li, S. Z. 2014. Learning Face Representation from Scratch. arXiv:1411.7923.

Yin, H.; Molchanov, P.; Álvarez, J. M.; Li, Z.; Mallya, A.; Hoiem, D.; Jha, N. K.; and Kautz, J. 2020. Dreaming to Distill: Data-Free Knowledge Transfer via DeepInversion. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8712–8721.

Zhang, H.; Cissé, M.; Dauphin, Y. N.; and Lopez-Paz, D. 2018. mixup: Beyond Empirical Risk Minimization. In *International Conference on Learning Representations*.

Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; and Tian, Q. 2015. Scalable Person Re-identification: A Benchmark. In *IEEE International Conference on Computer Vision*, 1116–1124.

Zheng, Z.; Zheng, L.; and Yang, Y. 2017. Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in Vitro. In *IEEE International Conference on Computer Vision*, 3774–3782.

Zhou, J.; and Pun, C. 2021. Personal Privacy Protection via Irrelevant Faces Tracking and Pixelation in Video Live Streaming. *IEEE Trans. Inf. Forensics Secur.*, 16: 1088–1103.

Zhuang, W.; Gan, X.; Wen, Y.; and Zhang, S. 2023. Optimizing Performance of Federated Person Re-identification: Benchmarking and Analysis. *ACM Trans. Multim. Comput. Commun. Appl.*, 19(1s): 38:1–38:18.

Zhuang, W.; Wen, Y.; Zhang, X.; Gan, X.; Yin, D.; Zhou, D.; Zhang, S.; and Yi, S. 2020. Performance Optimization of Federated Person Re-identification via Benchmark Analysis. In *ACM International Conference on Multimedia*, 955–963.