SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

January 2026, Vol. 69, Iss. 1, 112103:1–112103:14
https://doi.org/10.1007/s11432-023-4583-x

Person identity shift for privacy-preserving person re-identification

Shuguang DOU[1], Xinyang JIANG[2], Qingsong ZHAO[1], Yansen WANG[2],
Dongsheng LI[2] & Cairong ZHAO[1]*

[1]*School of Computer Science and Technology, Tongji University, Shanghai 201804, China*
[2]*Microsoft Research Asia, Shanghai 200232, China*

Received 20 September 2023/Revised 5 February 2024/Accepted 13 October 2024/Published online 4 January 2026

Abstract Recently, privacy concerns of person re-identification (ReID) have raised more and more attention, and protecting personal information in the privacy-sensitive images used by ReID methods has become essential. In order to utilize data from video surveillance without leaking pedestrians' private information, person de-identification (DeID) is a simple and effective method of alleviating privacy issues by removing identity-related information from the data. Most of the existing DeID methods focus on identity-irrelevant tasks such as pose and action recognition and tend to remove all identity-related information. However, this compromises the usability of de-identified data in the ReID task. In this paper, we aim to develop a technique to achieve a good trade-off between privacy protection and data usability for person ReID. To achieve this, we propose a novel de-identification method designed explicitly for person ReID, named person identity shift (PIS). PIS removes the absolute identity in a pedestrian image while preserving the identity relationship between image pairs by exploiting the interpolation property of the variational auto-encoder. Experimental results show that our method has a better trade-off between privacy-preserving and model performance compared to existing de-identification methods and can defend against human and model attacks for data privacy. The codebase of PIS is available at https://github.com/Vill-Lab/2025-SCIS-PIS.

Keywords person re-identification, person de-identification, privacy protection

Citation Dou S G, Jiang X Y, Zhao Q S, et al. Person identity shift for privacy-preserving person re-identification. Sci China Inf Sci, 2026, 69(1): 112103, https://doi.org/10.1007/s11432-023-4583-x

1 Introduction

Due to the advances in cameras and web technology, it is easy to capture and share large amounts of video surveillance data, which facilitates the research and application of person re-identification (ReID) [1, 2] in recent years. However, ReID has introduced severe concerns about personal privacy leakage. A person ReID development process contains four stages [3]: (1) data collection, (2) data annotation, (3) model training, and (4) model deployment and inference, each of which carries the risk of exposing the privacy to attackers. Specifically, in the data collection and annotation phase, the attackers may obtain the personal images of a pedestrian directly, which could easily reveal personal information such as daily individual whereabouts or personal activities. In the model training and inference stage, there is still a risk of the attacker obtaining the ReID model or its outputs and recovering the original training or inference data with techniques like model inversion [4, 5], membership inference attacks [6–8] or backdoor attack [9].

Data protection and machine learning seem to be in natural conflict, so one of the challenges for both users and service providers is to achieve privacy protection within a machine learning framework that balances privacy and benefits, which also applies to person ReID. Most of the existing privacy protection methods mainly focus on the privacy protection of the training and inference stage, such as Homomorphic encryption (HE) [10] and differential privacy (DP) [11]. Furthermore, HE and DP also face difficulties applying to large deep models. As a result, in this paper, we focus on a type of simple yet effective approach that can be applied to the privacy protection of all four stages, namely person de-identification (DeID) [12].

DeID aims at removing identity-related features about pedestrians from an image or video while retaining as much information as possible for various downstream tasks, such as pose or action recognition. Since DeID removes the identity features directly from raw data, the de-identified data can be applied to any stage of the ReID development

[*] Corresponding author (email: zhaocairong@tongji.edu.cn)

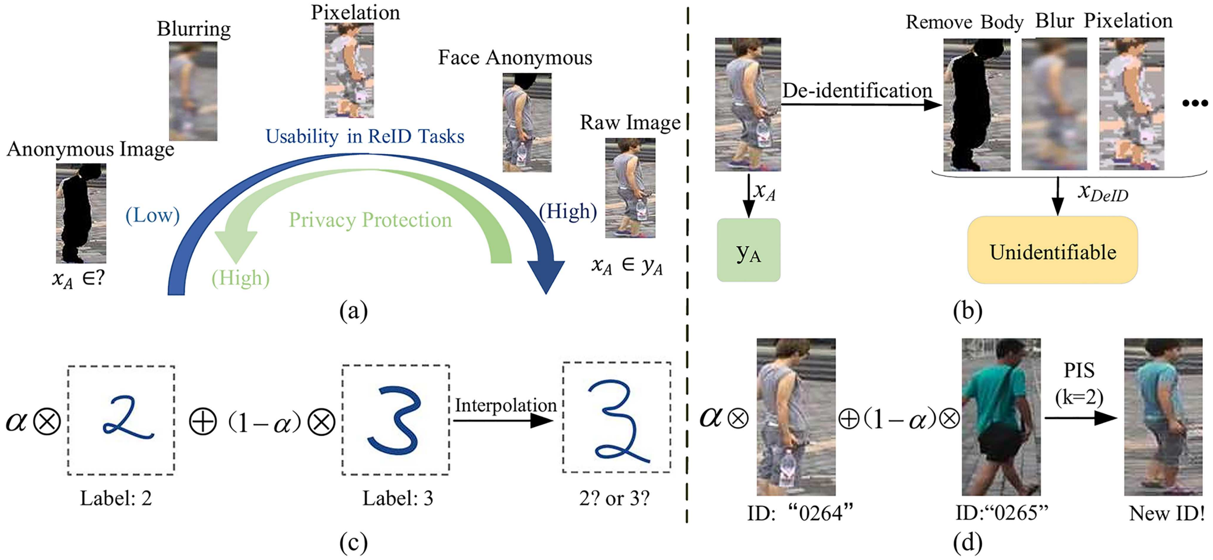

Figure 1 (Color online) The challenge of privacy protection and our solution. (a) The trade-off between privacy protection and usability in ReID tasks; (b) person de-identification (DeID); (c) our motivation; (d) person identity shift (PIS).

process, including data collection and annotation. However, existing DeID methods mainly focus on identity-irrelevant classification tasks, and ignore the usability of de-identified data on person re-identification. As shown in Figures 1(a) and (b), it is extremely challenging for the DeID method to achieve a good trade-off between privacy protection and ReID data usability, because existing DeID methods aim at removing all personal identity-related information in each pedestrian image, while the ReID model relies on this removed identity information to conduct training and inference. As a result, we seek to explore: is it possible to protect privacy without compromising the usability of the data for the ReID task?

An important observation to resolve this challenging task is the gap between what models rely on to perform re-identification and what we want to protect. Person ReID models are learned to make predictions based on the similarity or the relationship between pedestrian image pairs (i.e., whether two images belong to the same identity or not). On the other hand, we would like to protect the absolute identity of an image from possible attackers (i.e., who is the person in the image), which is not required by the ReID models. This leaves us room to solve the problem by removing the absolute privacy information while preserving the relationship between image pairs.

Inspired by the interpolation ability of variational auto-encoder [13] (as shown in Figure 1(c)), we aim to mix the original image with several reference images in the latent representation space, leading to a generated image with a mixed identity. Variational auto-encoder (VAE) has been extensively utilized in various studies on person ReID [14–16]. In contrast to the aforementioned studies that primarily concentrate on feature decoupling, our focus is on utilizing the interpolation properties of VAE to generate a new identification.

As a result, we propose a novel DeID-based method called person identity shift (PIS) that only removes the absolute or real identity of the images, while still preserving the relationship between image pairs. PIS ensures images originally belonging to the same identity still have the same new identity after de-identification. Specifically, as shown in Figure 1(d), PIS transforms a set of images with the same identity into a new set of images with a different identity (i.e., identity shift). In this way, the adversary will not obtain the absolute or real identity of the images. However, since the generated images originally belonging to the same identity still have the same new identity (i.e., preserve the relative identity), it can still be used in person re-identification.

We introduce the similarity preservation branch to constrain all images of the same ID to have the same distribution as much as possible, to indirectly make the distribution of all images of the same ID after shifting as much as possible. On the other hand, we design a cycle training strategy to establish a consistency constraint similar to that of the original image and the reconstructed image. In this way, the image is shifted from the current identity to another human-readable pedestrian image with a new identity. Notably, the images shifted by PIS can be directly understood by humans and used directly for model training without any additional processing.

In conclusion, we try to address the challenge of privacy-preserving person re-identification via the following contributions.

- We propose a novel DeID method that achieves a good trade-off between privacy protection and data usability for privacy-preserving ReID, called PIS, which shifts pedestrians' original identity to a new identity by sampling from the interpolation of Gaussian distribution, so that only trained with the generated new data to achieve similar ReID performance.

- We conduct detailed experiments to quantify the trade-off between the ReID performance and privacy protection ability of existing data de-identification methods and the proposed PIS. The results show that our method can outperform existing methods substantially on the two commonly used ReID datasets.
- We demonstrate that the proposed approach can defend against human and model attacks to some extent through user study and model inversion experiments. Compared with DP-SGD [11], PIS is more efficient and does not require changes to the model.

The rest of the paper is organized as follows. We introduce privacy protection methods for surveillance video, related person ReID methods general privacy protection methods, and privacy attack methods in Section 2. Section 3 presents the problem definition of privacy-preserving person re-identification and the details of PIS. Experiments in Section 4 demonstrate the effectiveness and safety of our approach. Finally, we conclude this paper in Section 5.

2 Related work

2.1 Privacy protection methods for surveillance video

De-identification. Many of the existing DeID studies remove privacy information from the face region by image distortion methods or generation models [17–19]. Recently, many studies have focused on identity-preserving face camouflage and anonymization [20, 21]. However, the face is only one of the recognizable features of the human body; other biological features, such as body structure, silhouette, gait, gender, and race, could also reveal sensitive personal information. To protect an individual's privacy in video, person de-identification [12, 22] hides the person's private information in video surveillance systems. Although the person DeID methods [12, 22] can remove personal information of individuals other than face region features, these methods only focus on privacy protection and do not consider the usability of the de-identified data on the ReID task, leading to unsatisfactory ReID performance.

Feature encryption. Different from person DeID, which focuses on raw data, some privacy-preserving approaches focus on protecting the ReID phase by introducing cryptography. Following the privacy-preserving image retrieval solutions, similarity metrics between encrypted feature vectors are commonly computed using Euclidean distance [23, 24] or Hamming distance [25]. These similarity metrics are then used to rank the images using a kNN algorithm. However, the above solutions cannot support kRNN ranking over encrypted feature vectors. Although FREED [26] supports kRNN sorting, it focuses only on the ReID phase and cannot protect the raw data and training process. Zhao et al. [27] put forward FREED enables the cloud server to perform state-of-the-art operations of person Re-ID on encrypted feature vectors directly and outputs person Re-ID results.

Federated person ReID. Federated person ReID methods (FedReID) [28–30] represent approaches employing federated learning, a decentralized training method, for person re-identification tasks. FedReID aims to protect data privacy by aggregating model updates instead of sharing raw data between clients and a central server. However, FedReID also focuses on only one stage of the ReID development process.

Different from the above methods, there are two types of approaches similar to ours that try to start with raw data at the beginning of the ReID development process. Event-driven ReID [31] focuses on integrating privacy into person ReID and explores the potential of using event-camera networks for ReID tasks for the first time. However, the performance of event-driven ReID drops largely compared to RGB-based ReID. Zhang et al. [32] proposed a learnable privacy-preserving anonymization method to explore the privacy-utility trade-off for pedestrian images. Privacy is protected by anonymizing the image, and then the restored image is obtained by the recovery decoder. However, this method cannot defend against model inversion attacks and is difficult to use for real-time video surveillance because anonymized images make it hard to observe whether their behavior is suspicious of a crime. In addition, the recovery decoder contains private information that could also be a target for attackers. In contrast, the anonymized images of our method can be used directly by humans and models (such as pedestrian detection) without recovery, as shown in Table 1 [12, 22–26, 28–32], and have less potential for privacy leakage.

2.2 Person re-identification

Existing person ReID methods mostly focus on performance improvements on public ReID datasets while neglecting privacy protection. DeID focuses only on privacy protection and ignores data availability for the ReID task, i.e., they fail to maintain the ReID performance on the de-identified data. Unlike the above tasks, the privacy-preserving person re-identification task proposed in this paper lies in balancing data availability and privacy protection.

Table 1 Comparing PIS with existing privacy-preserving person ReID methods. The "Readable" and "Trainable" denote that the data can be read directly by humans and be used to train for specific tasks after processing by these privacy-preserving methods, respectively. The "Human" and "Model" denote that the method can defend against human and model attacks, respectively.

Method	Protected			Practicality			Defensibility	
	Raw data	Training	ReID	Readable	Detectable	Trainable	Model	Human
Person DeID [12, 22]	✓						✓	✓
Feature encryption [23–26]			✓					
FedReID [28–30]		✓				✓		
Event-driven ReID [31]	✓	✓	✓			✓	✓	✓
Reversibly anonymization [32]	✓	✓	✓			✓	✓	✓
PIS (ours)	✓	✓	✓	✓	✓	✓	✓	✓

Some of the existing disentangled representation-based person ReID methods [33–35] disentangle the appearance and posture feature so that it can also generate a new image using interpolation. For example, DG-Net [33] can switch the appearance encoder or structure code to generate high-quality composed images. Compared with the disentangled representation-based person ReID methods, PIS is fundamentally different from them in two ways. (1) These methods mostly focus on changing colors and clothes and neglect other identity-related high-level attributes, such as gender. (2) Based on disentangled representation, our proposed PIS uses VAE and the two proposed losses to further exploit its interpolation capability, and the interpolation capability of the PIS is proven to be superior to these methods through quantitative and qualitative results.

2.3 Differential privacy

Training a model with (ϵ, δ)-DP [36] can protect the privacy of all training examples, where (ϵ, δ)-DP means that the probabilities of outputting a model M trained on two datasets D and D' that differ in a single example are close. However, it is challenging to apply differential privacy to deep learning. Abadi et al. [11] proposed DP-SGD to train deep learning models with DP. DP-SGD proved that privacy preservation for deep neural networks can be achieved with moderate cost in terms of training efficiency and model quality. However, DP-SGD only achieves good results on MNIST, dropping performance by about 20% on Cifar10.

Federated learning and DP primarily focus on privacy protection during the training phase. Unlike these methods, our proposed PIS directly operates on the raw pedestrian images, encrypting the data at the beginning of the ReID process. This ensures that neither during the training phase nor the inference phase can attackers extract private information from pedestrian images. Additionally, while DP requires modifications to the model, PIS operates without any need for model alterations.

2.4 Privacy attack methods

Model inversion attack. Since deep learning models always overfit the training data to some extent, the trained models can cause leakage of private training data [37]. One common approach to attack model privacy is model inversion. Model inversion infers training data from trained model or training process [5,38]. Hitaj et al. [39] trained a GAN to generate prototypical samples of the training sets during the learning process. In federated learning or collaborative learning systems, people believe that sharing gradients will not expose private training data. However, Zhu et al. [4] introduced the deep leakage from gradient (DLG) to access private data in the training set from the publicly shared gradient. Zhao et al. [40] found that ground-truth labels can be leaked from shared gradients and proposed the improved DLG (iDLG) to extract accurate data from shared gradients.

Adversarial attacks. Recently, considerable research has been conducted on adversarial attacks in the context of reid [41–44], leading to the emergence of several defense approaches [45]. For example, Gao et al. [7] proposed a joint adversarial defense model based on feature-invariant against adversarial metric attacks. Nevertheless, our specific focus is solely on privacy-related attacks, and assessing the model's robustness falls outside the scope of this paper.

3 Method

3.1 Problem definition of privacy preserving person ReID

We define a ReID dataset as $D_{\mathrm{reid}} = \{x_i, y_i\}_{i \in N}$ containing N pedestrian images, where x_i denotes the i-th pedestrian image and y_i denotes its corresponding identity. Privacy-preserving reid (PPReID) aims at conducting person

ReID with minimum privacy information leakage.

Following previous privacy-preserving methods [46], to evaluate the trade-off between privacy protection and the ReID performance, PPReID involves two tasks, i.e., a target utility task T_u that aims at achieving good ReID performance and a privacy budget T_p task that quantifies the privacy leakage by applying some privacy attack algorithms. A good de-identification method should balance these two tasks by finding a data anonymization function F_{anony} to transform a raw data image x into de-identified data $\hat{x} = F_{\text{anony}}(x)$ and perform well on both tasks.

Target utility task. For DeID methods, the target utility task T_u reflects how well the de-identified data can be used for the ReID task. Intuitively, the performance of T_u can be evaluated by the ReID model's performance trained on de-identified data, and a higher ReID performance indicates better usability of the de-identified data. Specifically, in our experiments, we chose several state-of-the-art ReID models to train on the de-identified data and evaluate their Rank-1 and mean average precision (mAP).

Privacy budget task. The privacy budget task T_p evaluates the amount of privacy information remained after de-identification. A lower score means a lower privacy budget cost, indicating a better ability of the anonymization function F_{anony} to prevent data from leaking identity information. Following the previous privacy budget setting [46], we propose to adopt two types of privacy attack methods: the pedestrian identity attack task and the pedestrian attribute attack task and evaluate how much privacy information the attack model can infer from the de-identified data.

The pedestrian identity attack aims to reveal the identity information from de-identified data. Previous identity attacks mainly focus on classification or detection tasks and cannot be directly applied in ReID. Therefore, we reformulate the identity attack as a retrieval task. Assuming that an attacker holds a query image of a target ID, the attack model will try to query the de-identified ReID dataset and retrieve images with the same ID as the query image. Under this setting, the privacy budget cost can be evaluated as the retrieval performance with metrics including Rank-1 and mean average precision (mAP).

The pedestrian attribute attack aims at recognizing personal attributes from the de-identified pedestrian images, such as gender, age, and hairstyle. Specifically, we apply binary classifiers that consider each attribute as an independent binary classification problem following [47] and treat the average F1 score as the privacy budget cost.

The target utility task T_u and the privacy budget task T_p can be contradictory to each other. Hence, we propose a new general metric that considers the performances of both tasks with adjustable weights to balance the importance of privacy and ReID, called PU-score (privacy utility score):

$$\text{PU-Score} = \frac{2}{\frac{I_u}{S_{T_u}} + \frac{I_p}{(1-S_{T_p})}}, \tag{1}$$

where S_{T_u} is the average ReID performance, S_{T_p} is the privacy budget cost, I_u and I_p denote the importance of the corresponding task in the PU-score and $I_u + I_p = 2$.

3.2 Overall framework of person identity shift

Overall pipeline. Inspired by the interpolation ability of variational auto-encode (VAE) [48], F_{anony} is designed as a novel VAE that shifts the identity of the input images. Specifically, given a pedestrian image x, PIS de-identifies it by mixing its latent representation with the images from other k pedestrians and decoding the mixing embedding to a person with a new identity.

The overall pipeline of the PIS is shown in Figure 2. The PIS network takes the pedestrian image x and its estimated pose s as the input. For pose s, a trained multi-person key point estimation [49] is exploited to obtain the pedestrian posture. Following VUNet [50], We model $p(z|s)$ and $q(z|x,s)$ as Gaussian distributions and $p(x|s,z)$ as a Laplace. Three neural networks, E_ψ, F_ϕ, and D_θ with trainable weights are designed to estimate the parameters of $p(z|s)$, $q(z|x,s)$, and $p(x|s,z)$ respectively. The body structure s extracted from an image x is fed into a prior-encoder E_ψ, which models the structure prior distribution $p(z|s)$, where z is the latent variable. Following [50], conditioned on the image x and body structure s, F_ϕ models the posterior distribution $q(z|x,s)$. The decoder models the likelihood $p(x|s,z)$, and reconstructs a new image x' given the concatenation of latent variable z and pedestrian pose s.

Notably, we force PIS not to change the pose of the person in the original image by treating it as the condition. This is because the DeID methods need to keep as much information as possible other than identity-related information, and human behavior is important for video continuity and downstream tasks including ReID.

Backbone network. Similar to conditional U-Net [51], a skip connection is used between the structure decoder with the encoder. The encoder and decoder both use the same residual structure in ResNet [52] with m residual

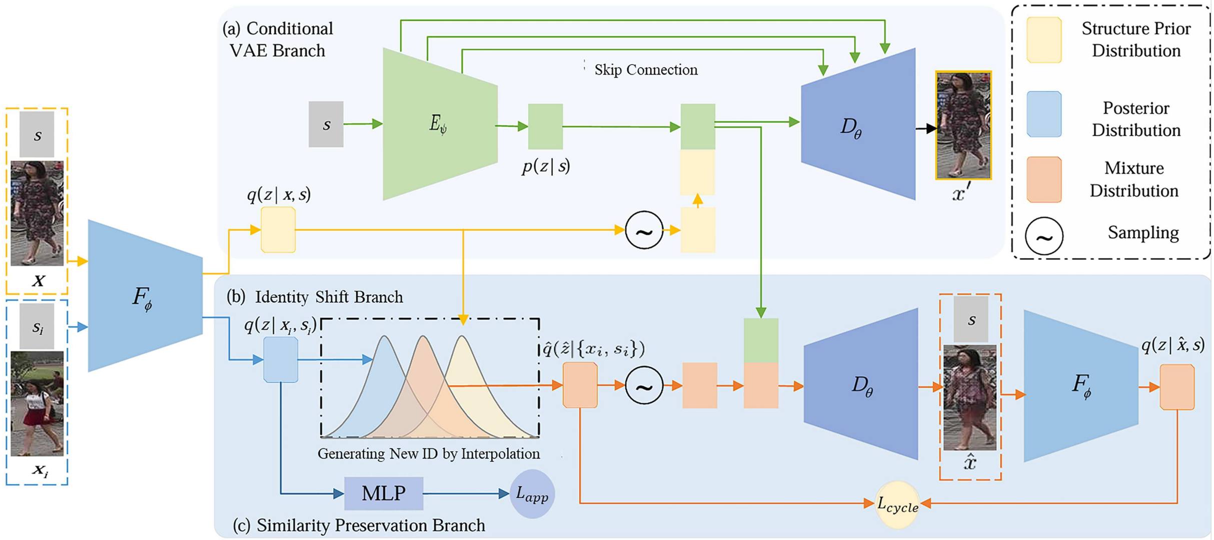

Figure 2 (Color online) Framework of the proposed person identity shift ($k = 2$). The training pipeline of PIS mainly contains two branches. The conditional VAE branch (a) conducts self-reconstruction. A latent representation z is obtained based on the original image x and pose s by sampling from the posterior distribution estimated by encoders E_ψ and F_ϕ, which is then reconstructed back to an image with decoder D_θ. The identity shift branch (b) mixes the latent representation of k images encoded separately by F_ϕ, and a mixed representation $\hat{z}$ is fed into D_θ to reconstruct an image with a new identity. The conditional VAE branch (c) is first trained.

blocks. Each residual block contains an ELU [53] activation function and a 3×3 convolutional layer with a residual connection way. The residual blocks are connected with a downsampling layer, where the downsampling layer is a convolutional layer with a stride size of 2×2. After the downsampling layer, the feature map is reduced to 1/4 of the original. The decoder also uses m residual blocks, while the sub-pixel convolution [54] between residual blocks is used as an up-sampling layer to change the feature map to four times the original.

Generating new ID by interpolation. Given a target image $x(=x_1)$ to be de-identified, its identity is shifted to a new one by mixing with $k-1$ images $\{x_i\}_{i\in(2,\ldots,k)}$ with different identities. Formally, the latent representation $\hat{z}$ is sampled from a Gaussian distribution $\hat{q}$, which is a mixture of k posteriors from the images as

$$\hat{z} \sim \hat{q}\left(\hat{z}|\{x_i, s_i\}_{i\in(1,\ldots,k)}\right) = \mathcal{N}\left(\sum_{i=1}^{k}\lambda_i\mu_i, 1\right), \tag{2}$$

where the mean of $\hat{q}$ is a Gaussian distribution is the weighted average of the means $\{\mu_i\}_{i\in(1,\ldots,k)}$ of the k distributions. λ_i is the interpolation weight of the i-th image. Then, the mixed latent representation $\hat{z}$ is fed into D_θ to generate an image $\hat{x}$ with a new identity, and hence F_{anony} de-identifies x by encrypting x with $k-1$ other images as follows:

$$\hat{x} = F_{\text{anony}}(x) = D_\theta(\hat{z}, E_\psi(s)). \tag{3}$$

In this way, the $k-1$ images used for creating a new ID can be seen as a random "one-time private key". That weakly encrypts the target image x. Two constraints are made for the coefficients λ_i. (1) The sum of the coefficients λ_i is 1, i.e., $\sum_i \lambda_i = 1$. (2) To prevent excessive coefficients from revealing private information of individual IDs, all coefficients are restricted to be less than d. We assign a new label to this new ID, distinct from the labels within the entire training set, i.e., not the label of the original image x nor x_i.

Similarity preservation branch. To indirectly constrain that the images with the shifting identity have the same appearance, we use the image's corresponding identity label for supervised learning. To make the distribution $q(z|x,s)$ of the images with the same ID as identical as possible, the mean of the distribution $q(z|x,s)$ is fed into an identity classifier containing a fully-connected layer.

3.3 Optimization

Conditional VAE loss. The perceptual loss, calculating the differences of activations on each layer of the backbone model, can enhance the perceptual quality of generated images [55]. The reconstruction term usually uses an L1-norm that often leads to blurry and less satisfying results in practice. Therefore, we substitute the reconstruction term with the perceptual loss and formulate our conditional VAE loss as

$$\mathcal{L}_{\text{cvae}} = -\text{KL}(q(z|x,s)||p(z|s)) + \alpha_{\text{rec}} \sum_{i}^{m} ||\mathcal{V}_i(x) - \mathcal{V}_i(x')||_1, \tag{4}$$

where $\mathcal{V}_i$ is the i-th layer of the backbone network, α_{rec} denotes a hyper-parameter that controls the weight of the perceptual loss. We set α_{rec} to 5 in this paper. To enable back-propagation to optimize the parameters in E_ψ, F_ϕ and D_θ, the re-parameterization trick [48] is used to sample z from the distributions. We use the softmax cross-entropy loss to supervise the similarity preservation branch:

$$\mathcal{L}_{\text{app}} = \text{SCE}(f_c(\mu), y), \tag{5}$$

where μ denotes the mean of distribution $q(z|x,s)$, SCE is the softmax cross-entropy loss, y is the ID of the image, and f_c denotes the identity classifier.

Cycle training for new ID generation. Since there is no corresponding ground truth image for the generated image $\hat{x}$, we propose to adopt a cyclical training loss to enforce better-mixed generation results. Specifically, $\hat{x}$ is fed back into F_ϕ to get its embedding distribution $q(z|\hat{x},s)$. Then, we minimize the Kullback-Leibler divergence between the embedding distribution and the original mixed distribution $\hat{q}$:

$$\mathcal{L}_{\text{cycle}} = D_{\text{KL}}(\hat{q}(\hat{z}|\{x_i, s_i\}_{i\in(1,\ldots,k)})||q(z|\hat{x},s)). \tag{6}$$

Finally, the overall loss of the PIS network is computed as the sum of the losses:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cvae}} + \mathcal{L}_{\text{cycle}} + \mathcal{L}_{\text{app}}. \tag{7}$$

By jointly optimizing all the losses, both reconstruction quality and the relative identity of the encrypted images are ensured.

The details of training. The training process is divided into two stages. In the first stage, we exclusively train F_ϕ and conditional VAE branches, with the network's optimization objective focused solely on minimizing $\mathcal{L}_{\text{cvae}}$. Once the conditional VAE branch achieves satisfactory performance in image reconstruction, we proceed to the second stage. In this phase, the remaining two branches are integrated, and the overall network is trained with the optimization objective updated to minimize $\mathcal{L}_{\text{total}}$.

3.4 Hardness of attacking PIS

Now we consider the difficulty of recovering information about x given a single encryption $\hat{x}$. Assume that the attackers all know the way to protect privacy, i.e., they can get the encoder and decoder used for our encryption.

We consider the naive attack where the attacker tries to figure out the set of all k images. To find the k correct images for shift, the attacker has to try all $\binom{m}{k}$ combinations with different coefficients, where m is the size of the dataset. Therefore, it is difficult to attack PIS even though the attacker knows as much information as possible, while for learning reversible anonymization [32], once the attacker steals the authentication, it will directly leak the privacy.

4 Experiments

First, we compare PIS with de-identification methods to verify that the proposed PIS is effective for PPReID. Second, we implement the user study to show the performance of our approach under human recognition attacks. Even though human recognition is far more expensive than model attacks, we also consider this possible attack. Third, if the attacker only has access to trained models or participates in distributed training such as federation learning, we mimic model reversal attacks to demonstrate the security of PIS. Finally, we compare with the differential privacy and provide visualization results.

4.1 Experimental setting

PPReID contains two tasks, namely the target utility task and the privacy budget task. For the target utility task T_u, four different state-of-the-art deep models, DenseNet121 [56], PCB [57], ISP [58] and TransReID [59] are trained on the de-identified data. The performance of T_u is the average MAP and Rank-1 of these four models. For the privacy budget task T_p, the pedestrian identity attack trains ISP on a dataset separated from the training

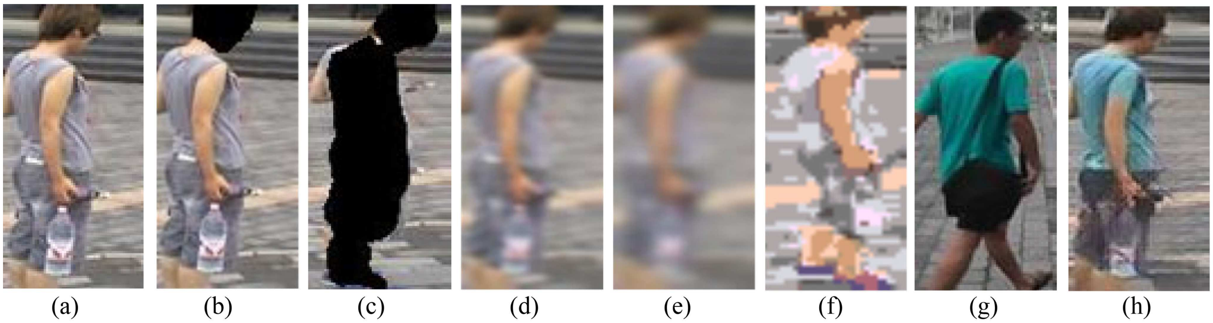

Figure 3 (Color online) Visualization of different de-identification methods and PIS. (a) Origin image; (b), (c) removing pixels from the head and body, respectively; (d), (e) Gaussian blur with different size kernels; (f) pixelation; (g) randomly selected image; (h) PIS ($k = 2$).

Table 2 Performance comparison between PIS and other methods in terms of target utility (in %) T_u, privacy budget (in %) T_p (identity attack), and PU-Score on Market-1501. Gaussian blur (S) and (L) denote the use of 11×5 and 21×11 Gaussian kernels. Given the width limitation, we do not show the results of DenseNet.

Method	PCB [57]		ISP [58]		TransReID [59]		$T_u \uparrow$		$T_p \downarrow$		PU-Score ↑	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Origin	71.8	89.5	85.5	93.3	89.0	95.1	**79.1**	**91.2**	90.0	95.1	17.8	9.3
Removing head	68.6	87.1	79.4	90.7	85.6	93.6	75.8	89.5	87.8	94.0	21.0	11.2
Removing body	25.8	49.3	27.4	54.2	27.6	48.0	21.1	39.8	**8.6**	**21.2**	34.2	52.8
Blur (S)	34.0	63.3	27.9	55.1	83.2	92.5	50.5	72.7	83.5	92.3	24.9	13.9
Blur (L)	11.5	28.7	30.2	58.1	72.3	87.6	34.7	56.0	34.7	59.3	45.3	47.1
Pixelation	48.7	73.3	65.4	83.7	77.2	89.5	58.8	78.7	56.2	76.9	50.2	35.7
DG-Net [33]	37.9	62.1	48.2	68.9	59.4	77.0	45.6	66.8	71.7	73.1	34.9	38.3
VUNet [50]	45.6	70.9	59.2	79.6	72.8	87.6	55.8	76.6	36.3	58.5	59.5	53.8
PIS (ours)	47.9	73.1	61.3	81.9	74.8	88.2	57.6	78.3	35.1	48.5	**61.0**	**62.1**

set of T_u and uses it to query a set of target images from the de-identified training data. The performance of T_p is evaluated based on whether the target images are successfully retrieved, and hence mAP and Rank-1 are used as evaluation metrics. Since PIS de-identifies images by mixing identities, an attack is considered successful if either one of the mixed images is retrieved. For the pedestrian attribute attack, we train a Densenet [56] to infer the privacy attributes and use the F1 score to evaluate the attack performance. The results of the trade-off between target utility and privacy budget are shown in Tables 2–4 where $I_u = I_p = 1$. Bold fonts indicate the best results. Up and down arrows indicate that higher values are better and lower values are better, respectively.

Dataset. We conduct experiments on the two commonly used ReID datasets. The details of the two ReID datasets are as follows. (1) Market-1501 [60] contains 12936 training images of 750 IDs, 3368 query images of 750 IDs, and 15913 gallery images of 751 IDs. Following [50], only 9939 training images of 730 IDs are used to train PIS. (2) DukeMTMC-reID [61] contains 16522 training images, 2228 query images, and 17661 gallery images.

Implementation details. The PIS framework is implemented by Tensorflow [62]. All images of the training set are resized to 128×64 and padded to 128×128. The optimizer of PIS is the Adam optimizer with a learning rate of 1.0e−3. The learning rate decays with training iteration. The batch size is 16, and the balance weight for the reconstruction term in the conditional VAE loss $\alpha_{\rm rec}$ is 5.

As for the interpolation of identities, we set $k = 2$ and $\lambda_1 = 0.5$, where the mixed identities are farthest from the original ones. When choosing the image used to de-identify the target, images of the same identity are mixed with the same image from another random identity. In this way, we expect the relationship among identities can be preserved. Moreover, in the pedestrian attribute attack experiment, we select images of another ID with the opposite attribute to achieve the attribute change. In order to ensure fairness, we use consistent identity selection strategies for methods and variants that are directly compared.

4.2 Results and analyses

We compare our PIS method with existing person DeID approaches, as shown in Figure 3 and person generation methods, DG-Net [33] and VUNet [50], to demonstrate the advantages of PIS in balancing target utility and privacy budget. The DeID-based methods include images with heads or bodies removed, Gaussian blurs with different sizes, and pixelation. For a fair comparison, DG-Net and VUNet use the same way to interpolate the appearance encoding

Table 3 Performance comparison between PIS and other methods in terms of target utility (in %) T_u, privacy budget (in %) T_p (identity attack), and PU-Score on DukeMTMC-reID. Gaussian blur (S) and (L) denote the use of 11×5 and 21×11 Gaussian kernels. Given the width limitation, we do not show the results of DenseNet.

Method	PCB [57]		ISP [58]		TransReID [59]		$T_u \uparrow$		$T_p \downarrow$		PU-Score ↑	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Origin	67.1	81.1	76.9	83.8	78.4	88.8	**71.4**	**83.2**	80.5	85.4	30.6	24.8
Removing head	63.8	78.0	67.5	81.2	76.1	88.1	66.7	81.4	77.4	81.0	33.8	30.8
Removing body	22.6	41.9	22.4	45.5	23.1	41.4	17.7	33.8	**7.9**	**18.2**	29.7	47.8
Blur (S)	28.9	55.1	23.6	46.8	74.2	82.0	43.7	63.0	75.7	83.4	31.2	26.3
Blur (L)	8.3	24.3	26.0	47.7	57.8	70.8	27.6	45.1	31.3	53.0	39.4	46.0
Pixelation	38.7	57.8	50.2	65.4	60.3	75.8	46.4	63.3	51.6	65.6	47.4	44.6
DG-Net [33]	29.0	50.4	39.0	56.6	47.3	64.7	36.6	55.3	64.0	62.2	36.3	44.9
VUNet [50]	38.1	61.0	45.0	60.8	55.7	70.2	44.3	60.9	32.9	52.6	53.4	53.3
PIS (ours)	38.1	63.5	49.2	66.2	61.5	74.7	46.6	64.1	32.0	43.7	**55.3**	**59.9**

Table 4 Attribute attack performance comparison between PIS and other methods in terms of target utility (in %) T_u, privacy budget (in %) T_p, and PU-Score on Market-1501.

Method	Teenager	Adult	Hair	Gender	$T_p \downarrow$	$T_u \uparrow$	PU-Score ↑
Origin	88.4	41.7	77.2	87.7	73.8	**89.9**	40.6
Removing head	87.9	39.3	69.1	77.8	68.5	88.2	46.4
Blur (S)	87.2	42.6	73.9	82.1	71.5	66.1	39.9
Blur (L)	85.1	34.3	64.7	68.7	63.2	45.5	40.7
Pixelation	**80.5**	39.6	63.5	70.6	63.6	75.1	49.1
DG-Net [33]	86.4	29.6	64.7	73.9	63.7	63.3	46.2
PIS (ours)	86.0	**20.9**	**55.3**	**58.4**	**55.2**	74.9	**56.1**

that generates similar images.

Person identity attack. As shown in Tables 2 and 3, removing all the pixels in the head region is not enough to defend against ReID model attacks, and intuitively the same conclusions go to other face de-identification methods. Removing pixels from the human body from the video is the most effective way to protect individual privacy, i.e., mAP of 8.6% under the ReID model attack. However, this method destroys the vast majority of information about the pedestrian in the image, resulting in an unacceptably low T_u performance. Minimal blurring does not protect privacy, while severe blurring brings a loss of utility, and hence likewise fails to balance T_u with T_p. Pixelation also does not perform well in the balance metric PU-Score. On the other hand, generation-based methods are better at balancing the two tasks. Among all the results, the mAP and Rank-1 of PIS in PU-Score are much better than existing methods while keeping a reasonable utility score and privacy budget cost.

Moreover, since the importance of T_u and T_p varies based on different applications, we also compare PIS and other methods under other importance weights, showing PIS's superiority over other methods under all privacy-performance importance weights in Figure 4. Note that we do not choose face recognition as an identity attack task because faces are not always visible in pedestrian images and the visible faces are most likely to be too blurred to recognize.

Person attribute attack. Table 4 compares PIS with other DeID methods with another privacy budget task called person attribute attack, which evaluates how many privacy attributes the attack model can infer from the de-identified data. Here, we select several seldom-changing attributes that highly relate to identity to infer, including gender, age, and hair length. As shown in Table 4, PIS significantly reduces the inference performance on adult, hair, and gender attributes in terms of F1-score by around 20 percentage points, outperforming other DeID methods. This result also shows that PIS not only changes clothes but also changes other high-level semantic information related to personal identity.

Ablation study. This paper proposes two novel losses to generate images with mixed identities while maintaining appearance consistency within the same identity mixture, namely the cycle training loss (CT) and similarity preservation branch (SP). We conduct an ablation study to analyze how different losses affect the ReID performance and privacy protection. Our baseline is a conventional conditional VAE [50] with only VAE loss. We also compare with ACAI [13] that adds the proposed adversarial regular term to the baseline. The results are shown in Table 5. First, the adversarial regular term can enhance the interpolation ability of the AE model, but it is not applicable for PPReID because it greatly increases the privacy budget. Second, both cycle training and similarity preservation branch perform well on T_u, but using both alone makes T_p lower than the baseline method. Finally, cycle training

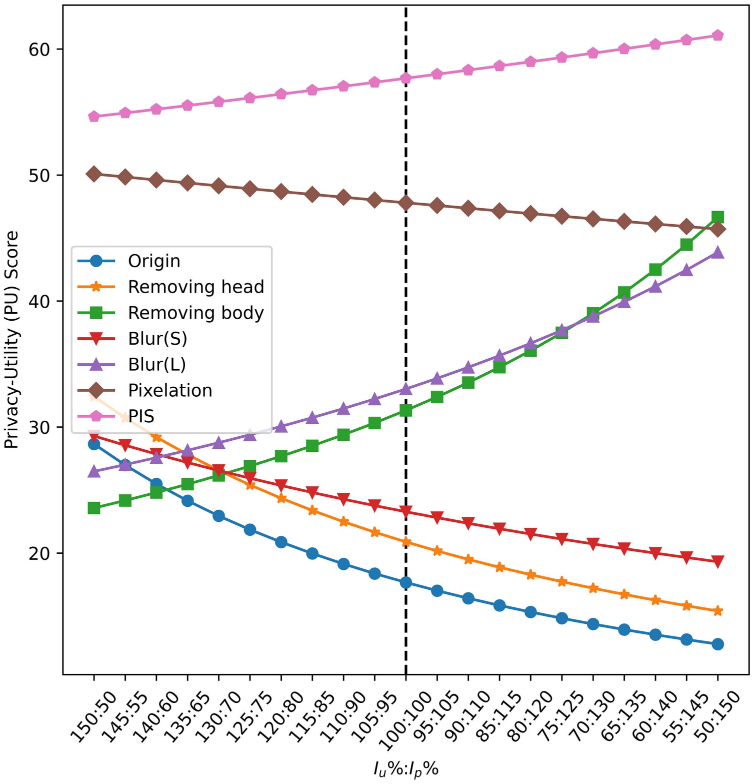

Figure 4 (Color online) The trade-off between ReID performance and protecting data privacy under different $I_p : I_u$ on Market-1501.

Table 5 Ablation study. AR denotes the adversarial regular term of ACAI [13], CT denotes cycle training, and SP denotes the similarity preservation branch. T_u uses the performance of PCB, and T_p uses the results of a person identity attack.

Baseline	AR	CT	SP	$T_u \uparrow$		$T_p \downarrow$		PU-Score ↑	
				mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
✓				45.6	70.9	36.3	50.8	53.2	58.1
✓	✓			45.5	73.0	44.4	65.4	50.0	46.9
✓		✓		**51.0**	75.1	42.6	58.7	54.0	53.3
✓			✓	50.0	**75.7**	39.9	53.3	54.6	57.8
✓		✓	✓	47.9	73.1	**35.1**	**48.5**	**55.1**	**60.4**

Table 6 Inside-dataset PIS vs. cross-dataset PIS on Market-1501. T_u uses the performance from PCB and T_p uses the results of a person identity attack.

Datasets	$T_u \uparrow$		$T_p \downarrow$		PU-Score ↑	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Inside	47.9	73.1	35.1	48.5	55.1	60.4
Cross	45.5	71.3	39.4	55.5	52.0	54.8

and similarity preservation branch are both effective in the PU-Score and are used together to reduce the privacy budget while enhancing the target utility.

Cross-dataset vs. inside-dataset. The previous experiments used identities from the same dataset to shift each other's identities, i.e., inside-dataset PIS. To verify the generalization ability of PIS on data outside the training set domain, we select images from another dataset, DukeMTMC-reID [61], to encrypt the original images in the Market-1501, i.e., cross-dataset PIS. As shown in Table 6, there is a decrease in target utility and an increase in privacy burden for cross-dataset PIS compared to inside-dataset PIS, but overall the impact of introducing a new dataset is limited. Therefore, in practice, PIS can encrypt the identities either within the private ReID dataset or with large-scale public ReID datasets to increase security.

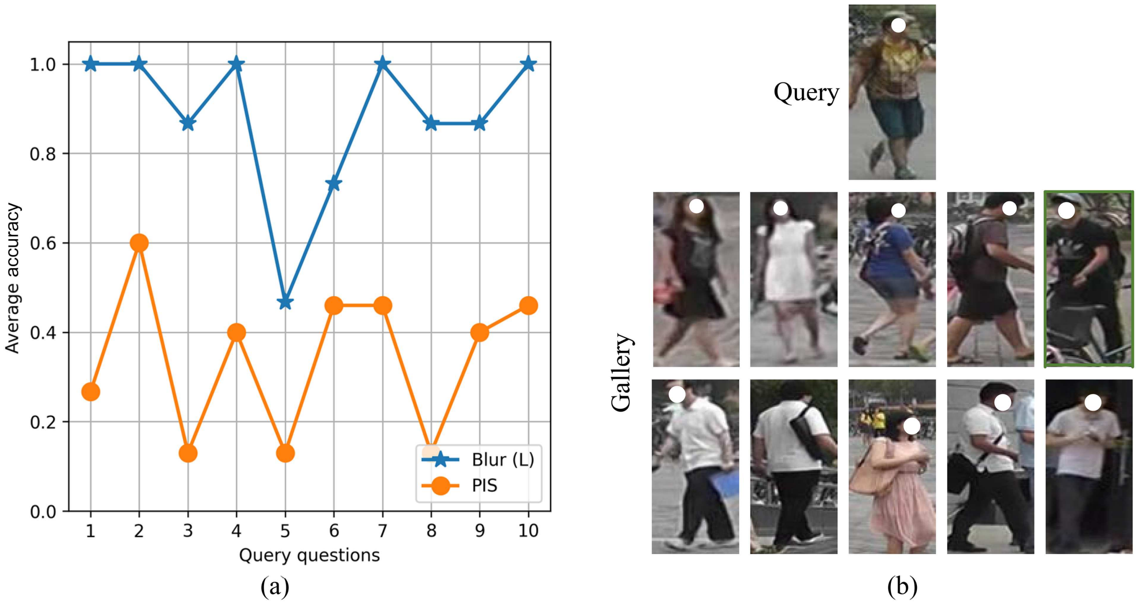

Figure 5 (Color online) User study. (a) The average accuracy of 15 human observers on 10 query questions for PIS and Blur (L); (b) the successful case of person attack for PIS. The pedestrian in the green box is the original identity corresponding to the query.

4.3 User study

To further evaluate the privacy protection ability of PIS, we also conducted a user study to verify if humans can manually identify the image generated by PIS. Directly using human recognition ability to attack the de-identification method is one of the common methods, but human recognition ability makes it difficult to quantitatively state the efficacy of the method. Here, we focus on the differences between human and model attacks.

Specifically, 15 human observers are asked to identify one image in the gallery with the same identity as a query image. The average accuracy of the 15 participants for each query is shown in Figure 5(a), PIS achieves a way better privacy protection ability compared to blur DeID in the user study. We also visualize the testing case in which the participants achieve the highest accuracy in Figure 5(b). In this test case, there is only one person in the gallery wearing a hat as the query image, while in this case, PIS may fail to shift personal items such as backpacks and hats.

4.4 Defend against model inversion attack

To simulate the possible existence of model inversion attacks, we use iDLG [40], an improved version of DLG [4], to recover the private train set by the shared gradient. According to the experimental setup of [4], we made two changes to the ReID model. Since iDLG requires the model to be twice differentiable, we replace all the ReLU activation functions with Sigmoid and remove the strides as well. The L-BFGS [63] is used as an optimizer with a learning rate of 1.0 and 200 iterations.

The attacking process on Market-1501 [60] is shown in Figure 6. The stolen pedestrian images start as Gaussian noise, and the gap between the dummy data and the input to the model gets smaller and smaller as the training iterates. For the vanilla training method, iDLG directly recovers the corresponding private data, while PIS uses the identity-shifted images to train the model, and iDLG is only able to recover the transformed image instead of the original private data set, making it impossible for the attacker to directly steal the private data.

4.5 PIS vs. differential privacy

PIS and deferential privacy are not directly comparable in terms of privacy protection, because one directly encrypts training data while the other focuses on preventing model leaking training data information. Following the setting of [10], we only report the target utility performance comparison between PIS and DP-SGD.

Comparison with DP-SGD. DP-SGD [11] implements differential privacy in deep learning by clipping the gradient first before adding noise during model training. The batchnorm [64] layer computes the mean and variance of the batch to create a dependency between samples in a batch that violates privacy. To implement DP-SGD in person ReID, we use opacus [65] to modify the ReID model by replacing all the batchnorm (BN) layers in ReID

Table 7 Performance comparison between PIS and DP-SGD in terms of target utility on Market-1501. The GN indicates whether to replace the batchnorm layer with the groupnorm layer.

Method	GN	ResNet-50 [52]		PCB [57]	
		mAP	Rank-1	mAP	Rank-1
DP-SGD [11]	✓	3.7	10.6	4.5	17.6
PIS (ours)	x	46.5	67.0	47.9	73.1

Table 8 Performance comparison between appearance noise and Laplace noise in terms of target utility (in %) T_u, privacy budget (in %) T_p, and PU-Score on Market-1501. T_u only uses the performance of PCB.

Noise	$T_u \uparrow$		$T_p \downarrow$		PU-Score↑	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Appearance noise	47.9	73.1	35.1	48.5	55.1	60.4
Laplace noise	52.4	77.0	60.0	78.7	45.4	33.4

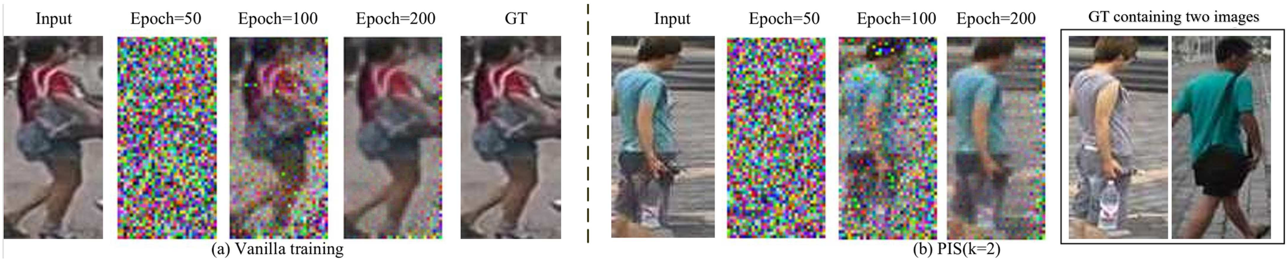

Figure 6 (Color online) Stealing training ReID images from the shared gradient. (a) Vanilla training; (b) training on images encrypted by PIS. iDLG can steal images, but what it steals are generated images, so the original images are still safe.

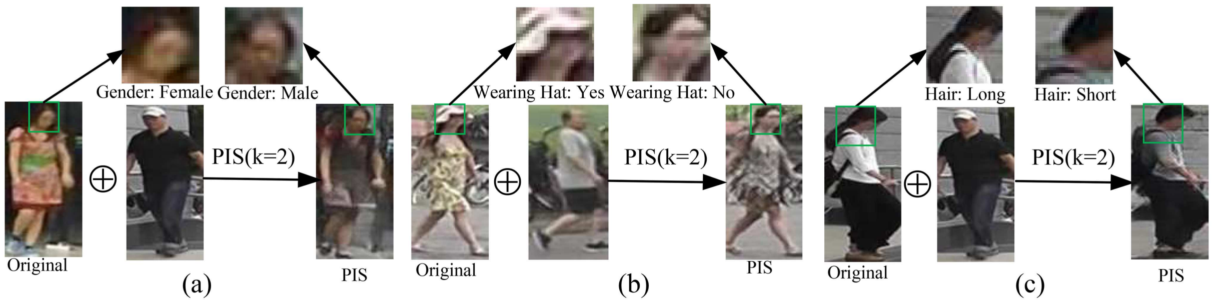

Figure 7 (Color online) Examples of PIS. (a) Change the gender; (b) change the appearance of whether to wear a hat; (c) change the hair's length.

models with groupnorm (GN) [66]. As shown in Table 7, our method shows good performance on ResNet50 and PCB without GN. In contrast, DP-SGD essentially fails to converge on the ReID dataset even with ϵ set to 50 (a very small noise) or changing the learning rate. Moreover, another advantage of our method over the DP-based method is that no modifications to the regularization layer or the activation function [67] of the model are required.

Comparison with adding random noise to appearance codes. Another typical approach to achieve differential privacy is to add Laplace noise directly to the image. In this section, we add the Laplace noise to the appearance coding of pedestrian images and compare the structured appearance noise $n_A = \sum_{i=2}^{k} \lambda_i E_a(x_i)$ with the random Laplace noise n_L where the mean of n_A is equal to n_L. As shown in Table 8, adding random Laplace noise does not make the encrypted identity well away from the original one.

4.6 Visualization results

We provide qualitative results showing what aspects of a pedestrian our approach has changed. As shown in Figure 7, it is most obvious that our approach changes the color, texture, and type of clothing. In addition to clothing, PIS also changes other accessories. For example, PIS removes the hat from a person in the shifted image in Figure 7(b). More importantly, PIS is also able to change other seldom-changing privacy attributes. For example, as shown in Figures 7(a) and (c), PIS changes the gender and hair length of a person. The qualitative results show that PIS can shift a person's identity by changing various attributes and appearance features, including constantly changing clothing or accessories, and more permanent attributes like age, hairstyle, and gender.

5 Conclusion

In this paper, we propose a novel privacy-preserving method called PIS towards privacy-preserving person re-identification. PIS generates weakly encrypted and human-readable pedestrian images by shifting each image to a new image with a different identity. Since generated images with the same identity still have high appearance similarity, PIS can preserve the relative identity and hence is suitable for the ReID task. The experiments show PIS achieves a better trade-off between privacy protection and ReID performance. With some problems left open, we hope that this study will raise more attention to the privacy risk of person re-identification.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. U25A20527, 62473286).

References

1 Gong S G, Cristani M, Yan S C, et al. Person Re-Identification. Berlin: Springer, 2014

2 Shu X, Wang X, Zang X, et al. Large-scale spatio-temporal person re-identification: algorithms and benchmark. IEEE Trans Circ Syst Video Technol, 2022, 32: 4390–4403

3 Ye M, Shen J, Lin G, et al. Deep learning for person re-identification: a survey and outlook. IEEE Trans Pattern Anal Mach Intell, 2022, 44: 2872–2893

4 Zhu L, Liu Z, Han S. Deep leakage from gradients. In: Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS), 2019

5 Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS), 2015. 1322–1333

6 Backes M, Berrang P, Humbert M, et al. Membership privacy in microRNA-based studies. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS), 2016. 319–330

7 Gao J, Jiang X, Zhang H, et al. Similarity distribution based membership inference attack on person re-identification. ArXiv:2211.15918

8 Gao J, Jiang X, Dou S, et al. Re-ID-leak: membership inference attacks against person re-identification. Int J Comput Vision, 2024, 132: 4673–4687

9 Sun W, Jiang X, Dou S, et al. Invisible backdoor attack with dynamic triggers against person re-identification. IEEE Trans Inform Forensic Secur, 2024, 19: 307–319

10 Huang Y, Song Z, Li K, et al. Instahide: instance-hiding schemes for private distributed learning. In: Proceedings of the International Conference on Machine Learning (ICML), 2020

11 Abadi M, Chu A, Goodfellow I J, et al. Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016. 308–318

12 Agrawal P, Narayanan P J. Person de-identification in videos. IEEE Trans Circ Syst Video Technol, 2011, 21: 299–310

13 Berthelot D, Raffel C, Roy A, et al. Understanding and improving interpolation in autoencoders via an adversarial regularizer. In: Proceedings of the 7th International Conference on Learning Representations, 2019

14 Pu N, Chen W, Liu Y, et al. Dual Gaussian-based variational subspace disentanglement for visible-infrared person re-identification. In: Proceedings of the 28th ACM International Conference on Multimedia, Seattle, 2020. 2149–2158

15 Wu C, Ge W, Wu A, et al. Camera-conditioned stable feature generation for isolated camera supervised person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022. 20206–20216

16 Wang Z, Wang Z, Zheng Y, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019. 618–626

17 Gross R, Sweeney L, la Torre F D, et al. Semi-supervised learning of multi-factor models for face de-identification. In: Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition, 2008

18 Wu Y, Yang F, Xu Y, et al. Privacy-protective-GAN for privacy preserving face de-identification. J Comput Sci Technol, 2019, 34: 47–60

19 Gafni O, Wolf L, Taigman Y, et al. Live face de-identification in video. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019. 9377–9386

20 Li J, Han L, Zhang H, et al. Learning disentangled representations for identity preserving surveillance face camouflage. In: Proceedings of the 25th International Conference on Pattern Recognition, 2020. 9748–9755

21 Li J, Han L, Chen R, et al. Identity-preserving face anonymization via adaptively facial attributes obfuscation. In: Proceedings of ACM Multimedia Conference, 2021. 3891–3899

22 Zhang W, Cheung S S, Chen M. Hiding privacy information in video surveillance system. In: Proceedings of International Conference on Image Processing, 2005. 868–871

23 Wang X, Ma J, Liu X, et al. Search in my way: practical outsourced image retrieval framework supporting unshared key. In: Proceedings of the 2019 IEEE Conference on Computer Communications, 2019. 2485–2493

24 Zhang L, Jung T, Liu C, et al. POP: privacy-preserving outsourced photo sharing and searching for mobile devices. In: Proceedings of the 35th IEEE International Conference on Distributed Computing Systems, 2015. 308–317

25 Cheng H, Wang H, Liu X, et al. Person re-identification over encrypted outsourced surveillance videos. IEEE Trans Dependable Secure Comput, 2021, 18: 1456–1473

26 Zhao B, Li Y, Liu X, et al. FREED: an efficient privacy-preserving solution for person re-identification. In: Proceedings of the IEEE Conference on Dependable and Secure Computing, 2022. 1–8

27 Zhao B, Li Y, Liu X, et al. Identifiable, but not visible: a privacy-preserving person reidentification scheme. IEEE Trans Rel, 2023, 72: 1295–1307

28 Zhuang W, Wen Y, Zhang X, et al. Performance optimization of federated person re-identification via benchmark analysis. In: Proceedings of the 28th ACM International Conference on Multimedia, 2020. 955–963

29 Zhuang W, Gan X, Wen Y, et al. Optimizing performance of federated person re-identification: benchmarking and analysis. ACM Trans Multimedia Comput Commun Appl, 2023, 19: 1–18

30 Zhuang W, Wen Y, Zhang S. Joint optimization in edge-cloud continuum for federated unsupervised person re-identification. In: Proceedings of the ACM Multimedia Conference, 2021. 433–441

31 Ahmad S, Scarpellini G, Morerio P, et al. Event-driven re-ID: a new benchmark and method towards privacy-preserving person re-identification. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, 2022. 459–468

32 Zhang J, Ye M, Yang Y. Learnable privacy-preserving anonymization for pedestrian images. In: Proceedings of the 30th ACM International Conference on Multimedia, 2022. 7300–7308

33 Zheng Z, Yang X, Yu Z, et al. Joint discriminative and generative learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019. 2138–2147

34 Eom C, Ham B. Learning disentangled representation for robust person re-identification. In: Proceedings of Annual Conference on Neural Information Processing Systems, 2019. 5298–5309

35 Eom C, Lee W, Lee G, et al. Disentangled representations for short-term and long-term person re-identification. IEEE Trans Pattern Anal Mach Intell, 2022, 44: 8975–8991

36 Dwork C, Kenthapadi K, McSherry F, et al. Our data, ourselves: privacy via distributed noise generation. In: Proceedings of Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2006

37 Papernot N, Abadi M, Erlingsson Ú, et al. Semi-supervised knowledge transfer for deep learning from private training data. In: Proceedings of the 5th International Conference on Learning Representations, 2017

38 Melis L, Song C, Cristofaro E D, et al. Exploiting unintended feature leakage in collaborative learning. In: Proceedings of the 2019 IEEE Symposium on Security and Privacy, 2019. 691–706

39 Hitaj B, Ateniese G, Pérez-Cruz F. Deep models under the GAN: information leakage from collaborative deep learning. In: Proceedings of ACM SIGSAC Conference on Computer and Communications Security (CCS), 2017. 603–618

40 Zhao B, Mopuri K R, Bilen H. IDLG: improved deep leakage from gradients. ArXiv:2001.02610

41 Bai S, Li Y, Zhou Y, et al. Adversarial metric attack and defense for person re-identification. IEEE Trans Pattern Anal Mach Intell, 2021, 43: 2119–2126

42 Ding W, Wei X, Ji R, et al. Beyond universal person re-identification attack. IEEE Trans Inform Forensic Secur, 2021, 16: 3442–3455

43 Bouniot Q, Audigier R, Loesch A. Vulnerability of person re-identification models to metric adversarial attacks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020. 3450–3459

44 Wang L, Zhang W, Wu D, et al. Attack is the best defense: towards preemptive-protection person re-identification. In: Proceedings of the 30th ACM International Conference on Multimedia, New York, 2022. 550–559

45 Liu D, Wu L Y, Hong R, et al. Generative metric learning for adversarially robust open-world person re-identification. ACM Trans Multimedia Comput Commun Appl, 2023, 19: 1–19

46 Wu Z, Wang H, Wang Z, et al. Privacy-preserving deep action recognition: an adversarial learning framework and a new dataset. IEEE Trans Pattern Anal Mach Intell, 2022, 44: 2126–2139

47 Lin Y, Zheng L, Zheng Z, et al. Improving person re-identification by attribute and identity learning. Pattern Recogn, 2019, 95: 151–161

48 Kingma D P, Welling M. Auto-encoding variational Bayes. In: Proceedings of the 2nd International Conference on Learning Representations, 2014

49 Cao Z, Simon T, Wei S, et al. Realtime multi-person 2D pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. 1302–1310

50 Esser P, Sutter E, Ommer B, et al. A variational U-net for conditional appearance and shape generation. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018. 8857–8866

51 Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the Medical Image Computing and Computer-Assisted Intervention, 2015. 234–241

52 He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, 2016. 770–778

53 Clevert D, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus). In: Proceedings of the International Conference on Learning Representations, 2016

54 Shi W, Caballero J, Huszar F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016. 1874–1883

55 Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of the European conference on computer vision, 2016. 694–711

56 Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. 2261–2269

57 Yifan S, Liang Z, Yi Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline). In: Proceedings of European Conference of Computer Vision, 2018. 501–18

58 Kuan Z, Haiyun G, Zhiwei L, et al. Identity-guided human semantic parsing for person re-identification. In: Proceedings of European Conference of Computer Vision, 2020. 346–63

59 He S, Luo H, Wang P, et al. Transreid: transformer-based object re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021. 15013–15022

60 Zheng L, Shen L, Tian L, et al. Scalable person re-identification: a benchmark. In: Proceedings of the IEEE International Conference on Computer Vision, 2015. 1116–1124

61 Zheng Z, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In: Proceedings of the IEEE International Conference on Computer Vision, 2017. 3774–3782

62 Abadi M, Agarwal A, Barham P, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. ArXiv:1603.04467

63 Liu D C, Nocedal J. On the limited memory BFGS method for large scale optimization. Math Programm, 1989, 45: 503–528

64 Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of International Conference on Machine Learning, 2015. 448–456

65 Yousefpour A, Shilov I, Sablayrolles A, et al. Opacus: user-friendly differential privacy library in PyTorch. ArXiv:2109.12298

66 Wu Y, He K. Group normalization. Int J Comput Vis, 2020, 128: 742–755

67 Papernot N, Thakurta A, Song S, et al. Tempered sigmoid activations for deep learning with differential privacy. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence, 2021. 9312–9321