Learnable Privacy-Preserving Anonymization for Pedestrian Images

Junwu Zhang
junwuzhang@whu.edu.cn
School of Computer Science, Wuhan University
Wuhan, China

Mang Ye*
yemang@whu.edu.cn
School of Computer Science, Wuhan University
Wuhan, China

Yao Yang
yangyao@zhejianglab.com
Zhejiang Lab
Hangzhou, China

ABSTRACT

This paper studies a novel privacy-preserving anonymization problem for pedestrian images, which preserves personal identity information (PII) for authorized models and prevents PII from being recognized by third parties. Conventional anonymization methods unavoidably cause semantic information loss, leading to limited data utility. Besides, existing learned anonymization techniques, while retaining various identity-irrelevant utilities, will change the pedestrian identity, and thus are unsuitable for training robust re-identification models. To explore the privacy-utility trade-off for pedestrian images, we propose a joint learning reversible anonymization framework, which can reversibly generate full-body anonymous images with little performance drop on person re-identification tasks. The core idea is that we adopt desensitized images generated by conventional methods as the initial privacy-preserving supervision and jointly train an anonymization encoder with a recovery decoder and an identity-invariant model. We further propose a progressive training strategy to improve the performance, which iteratively upgrades the initial anonymization supervision. Experiments further demonstrate the effectiveness of our anonymized pedestrian images for privacy protection, which boosts the re-identification performance while preserving privacy. Code is available at https://github.com/whuzjw/privacy-reid.

CCS CONCEPTS

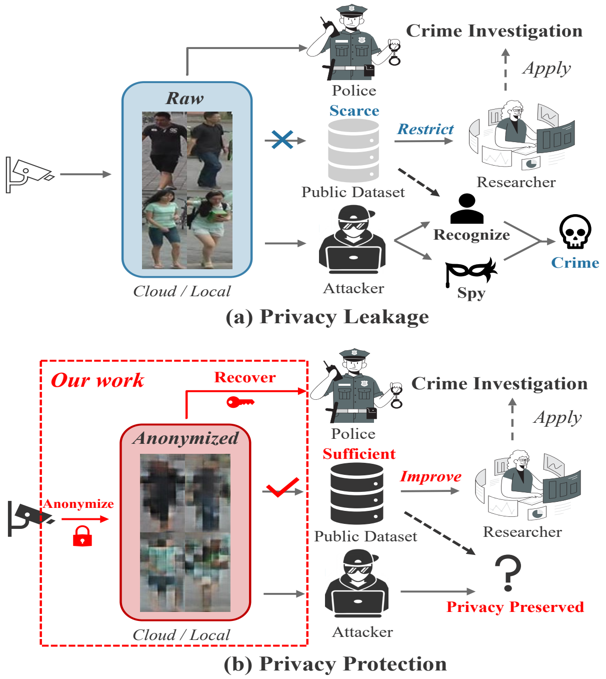
• **Security and privacy → Privacy protections**.

KEYWORDS

pedestrian image, privacy protection, person re-identification

1 INTRODUCTION

With the development of machine learning and the expansion of personal private data, various intelligent applications emerge and bring lots of utility value to individuals and the society. However, sensitive personal information raises serious privacy issues that are becoming increasingly prominent. Due to the privacy concerns, a lot of organizations have to make a compromise, e.g., Meta (Facebook) is shutting down its facial recognition software recently [7]. Moreover, public research datasets are also influenced by the privacy concerns, e.g., DukeMTMC dataset [23] and Tiny Image dataset [26] were taken down and all the facial areas in ImageNet [33] are blurred. As a result, privacy protection provides security while restricting technology improvement. Therefore, finding a privacy-utility trade-off point is of great importance.



(a) Privacy Leakage

(b) Privacy Protection

Figure 1: Motivation of our privacy-preserving system. (a) represents the risk of privacy leakage in current surveillance systems. The potential abuse of sensitive identity information leads to the deficiency of public datasets, which restricts related research. (b) represents our privacy protection method in surveillance systems. Our anonymized pedestrian images not only protect sensitive information from abuse, but also are suitable for person re-identification research that can be applied to crime investigation. Besides, original raw images cannot be recovered from our anonymized images by authorized users.

In surveillance scenarios, privacy issues are particularly evident, as shown in Fig. 1(a). Ubiquitous surveillance systems take a huge number of raw pedestrian images and videos, which are stored in local storage or uploaded to the cloud servers. On one hand, it is useful for legal users in many scenarios, e.g., in crime investigation. On the other hand, this raises serious privacy concerns for individuals and public safety, since original images or videos contain sensitive information about the pedestrians, e.g., realistic identity information of a specific person or special community. Without

*Corresponding author.

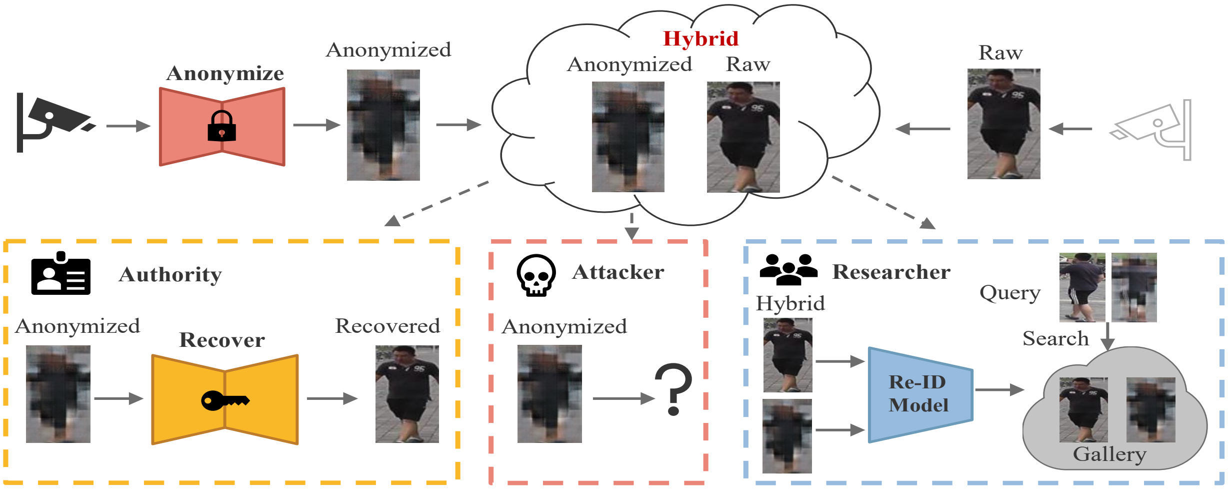


Figure 2: Various utilities of our surveillance systems for different targets. Our anonymized images can not only protect privacy from abuse by malicious attackers, but also retain utilities for authority and the researchers.

careful protection, the highly sensitive information might be leaked and abused by malicious parties for nefarious purposes. For example, malicious attackers may recognize individual realistic identities and spy on the individuals for further crime or even forge them via Deepfake [2]. Moreover, due to potential privacy concerns, public surveillance datasets are scarce and sometimes taken down like DukeMTMC-reid [23]. The deficiency of public datasets restricts the improvement of related research like person re-identification, and thus limits the development of intelligent video surveillance. Therefore, there is an urgent need to address the privacy issues for pedestrian images while retaining the utility value.

To tackle the above issues, a feasible anonymization solution is illustrated in Fig. 1(b). To prevent abuse after leakage, such an anonymization method is expected to have a satisfying visual obfuscation effect to ensure malicious parties cannot draw identity information from the anonymized images by human eyes. With the privacy preserved, the anonymized images can be securely used as public datasets. To retain the data utility for various users, the original raw images should be able to be recovered from the protected images for authorized utility, e.g., for police officers to investigate crime. Moreover, considering person re-identification (Re-ID) is imperative in intelligent surveillance systems with significant research impact and practical importance [36], the anonymized images are supposed to retain necessary information for researchers to perform Re-ID tasks. In summary, *an ideal anonymization method for pedestrian images should be reliable for privacy security, reversible for authorized utility and suitable for person re-identification research.*

Lots of research work has been done on image anonymization. *Conventional anonymization methods*, e.g., blurring, pixelation and Gaussian noise adding, face the problem of semantic information loss, causing significant declines in utility value. To explore the privacy-utility trade-off, new techniques and mechanisms are proposed to de-identify images or videos to fool identification models while achieving various identity-irrelevant utility goals [1, 17, 19, 21, 22, 38], such as reversibility [21], privacy preserving action detection [22], smile recognition [38] and so on. However, these anonymization techniques change the original individual identity to get a low identification rate by recognition model. In terms of crime investigation scenario, the identity variance is unsuitable for person re-identification task which is identity-relevant and requires that original and anonymized images of a specific pedestrian share the same virtual identity. Recently, Dietlmeier et al. [4] propose an anonymization dataset for Re-ID, which detect and blur the facial regions. However, non-face regions may also cause privacy leakage and the original images cannot be reconstructed from their anonymized images.

To achieve the goal illustrated in Fig. 1(b), we propose a new reversible anonymization framework for pedestrian images, which can reversibly generate full-body anonymous images with little performance drop on Re-ID task. As shown in Fig. 2, the identity information of our anonymized pedestrian images can be invisible to attackers, but recoverable for authorized users and computable for researchers to perform Re-ID task on hybrid domains (raw and anonymized). The core idea of our work is that we first desensitize raw images by conventional methods (i.e., blurring, pixelation, or noise adding). Then these desensitized images are adopted as initial supervision images for an anonymization encoder which can translate raw images to privacy-preserving images in a learnable manner. To preserve necessary features for recovery and person re-identification, we jointly optimize the anonymization encoder with a recovery decoder and a Re-ID model. Through supervised and joint learning, our anonymized images can achieve good performance on privacy protection, recovery, and person re-identification. Besides, to further improve Re-ID performance, we propose a progressive training strategy referred to as *supervision upgradation*. The supervision is upgraded by replacing the original desensitized images with the learned anonymized images, which are constrained by both privacy protection and Re-ID performance. Our main contributions are summarized as follows:

- To the best of our knowledge, we are the first to explore the privacy-utility trade-off for pedestrian images from a Re-ID perspective, in which anonymized images cannot be recognized by third parties, but are recoverable for authorized users and suitable for person re-identification research.
- We propose a reversible anonymization framework for Re-ID, which jointly optimizes an anonymization encoder with a recovery decoder and achieves the goal of obfuscating the image while keeping the identity for the authorized model.
- We design a progressive training strategy called *supervision upgradation*, which improves Re-ID performance by progressively upgrading the supervision of anonymization target.
- We experimentally show that our anonymized images achieve good performance for privacy protection, recovery, and person re-identification tasks.

2 RELATED WORK

Person Re-IDentification. Re-ID aims at retrieving a person of interest across multiple non-overlapping cameras [36]. With the development of deep neural networks, many works adopt deep convolutional neural networks (CNNs) as the backbone to extract the features of person images [34, 35, 37], and incorporate domain generalization [40, 43] to generalize better to unseen domains [41, 42]. The CNN-based baselines [18, 27, 36], such as AGW [36] and so on, achieve great success and play a key role in Re-ID community. However, public Re-ID datasets face the challenge of privacy concerns, e.g., DukeMTMC-reID [23] dataset was taken down due to privacy issues. To tackle this problem, we propose an anonymization method for Re-ID research, which can protect privacy while retaining necessary features for Re-ID tasks.

Image-to-Image Translation. Image-to-image translation is the task of transforming original images into the target images with a different style. Zhu *et al.* proposed Pix2pix network [12] and its unsupervised variant CycleGAN [44] which achieved impressive performance on paired and unpaired cross-domain image translation. In this work, we use two Pix2pix networks for anonymization and recovery. Other advanced models can also be applied.

Face Anonymization. Conventional face anonymization methods include pixelation, blurring, and noise adding. However, these methods cause semantic information loss, leading to performance degradation in detection and recognition. Therefore, researchers proposed many learnable anonymization methods based on face swapping [1, 8, 11, 15–17, 19, 21, 22, 24, 25, 32] to preserve important features for various utilities. However, in some special cases, generated faces might overlap with realistic faces. Therefore, You *et al.* [38] proposed a reversible face privacy-preserving framework based on a learned mosaic for smile recognition. Although the sole smile recognition task is easy and insufficient, the impressive results show that the semantic information for recovering and recognition can be invisibly embedded in protected images. This idea inspires us to anonymize pedestrian images for person re-identification.

Privacy-Preserving Methods. To tackle the privacy concerns, a lot of approaches are proposed, including differential privacy [5, 6], federated learning [13, 20], and so on. By contrast, our method protects privacy from the source, and thus the protected data can be stored securely and centrally for easy access.

3 PROPOSED METHOD

In this section, we detail the methodology to reversibly anonymize images for Re-ID. As illustrated in Fig. 3, our framework contains the following four components. 1) *Anonymized Image Generation* § *3.1.* We exploit the power of image-to-image translation with conditional adversarial networks to generate anonymized images in a learnable manner. With learnable anonymization, further objectives can be achieved by joint learning. 2) *Raw Image Recovery* § *3.2.* To achieve reversibility, a recovery model is added to embed necessary recovery information into the anonymized images. 3) *Joint Learning with a Re-ID Model* § *3.3.* To preserve features for re-identification, we jointly optimize the anonymization generator with a Re-ID model. 4) *Progressive Supervision Upgradation* § *3.4.* In order to further improve Re-ID performance, the supervision images are progressively upgraded according to the performance of anonymized images on privacy protection and Re-ID.

3.1 Anonymized Image Generation

To generate anonymized images, our inspiration comes from the method of image-to-image translation, which can transform an input image into an output image of a specified form in a learnable manner. Our goal is to convert the original input image into a protected image form, and blurred, pixelated, or noise-added images can be adopted as initial supervision to guide the generation of privacy-preserving images. Specifically, the pix2pix [12] framework for image translation based on GAN [9] is introduced.

The training samples contain $\{x_i\}_{i=1}^{n} \in X$ with labels $\{label_i\}_{i=1}^{m}$, where X are original images. We denote $\{y_i\}_{i=1}^{n} \in Y$, where Y are supervision images which are initialized by conventionally desensitized images and can be further updated. The goal of the anonymization generator G_X is to learn a mapping function $G : X \longrightarrow Y$. To achieve this goal, the generator G_X is trained in an adversarial manner with a discriminator D_Y, where G_X tries to generate images $G_X(x)$ similar to y while D_Y aims to distinguish between $G_X(x)$ and y. The adversarial objective of G_X can be expressed as

$$\begin{aligned}\mathcal{L}_{adv_1} = \frac{1}{n}\sum_{i=1}^{n} \log(D_Y(x_i, y_i)) +\\ \frac{1}{n}\sum_{i=1}^{n} \log(1 - D_Y(x_i, G_X(x_i))),\end{aligned} \tag{1}$$

where n represents the number of training samples within each batch and G_X tries to minimize the objective against an adversarial D_Y that tries to maximize it.

Besides, L_1 loss between $G_X(X)$ and Y is adopted to guarantee that the learned function can map an individual input x_i to a desired output y_i [12].

$$\mathcal{L}_{1_{ano}} = \frac{1}{n}\sum_{i=1}^{n} \|y_i - G_X(x_i)\|_1. \tag{2}$$

The total loss of the anonymization generator is

$$\mathcal{L}_{ano} = \mathcal{L}_{adv_1} + \lambda_{L_1}\mathcal{L}_{1_{ano}}, \tag{3}$$

where λ_{L_1} is a hyperparameter to reduce the artifacts [12].

3.2 Raw Image Recovery

To make the process of generating anonymized images reversible, we design the raw image recovery to obtain generative raw images

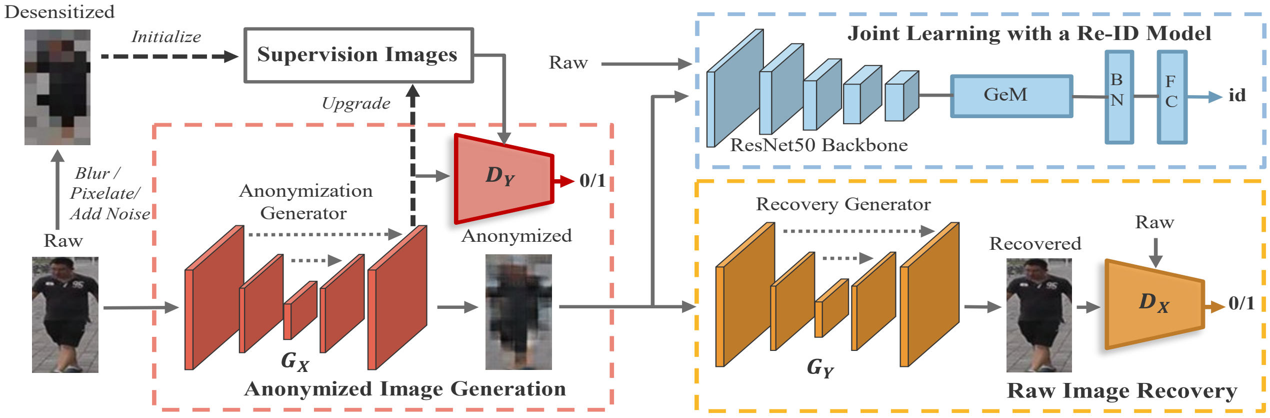


Figure 3: Framework of the proposed method. The framework consists of four components. The anonymization model aims to produce privacy-preserving images in a learnable manner. The recovery model and Re-ID model are added to jointly optimize the anonymization model. The supervision (desensitized images) are progressively upgraded by learned anonymized images to further improve Re-ID performance. Different colors are corresponding to the different utilities illustrated in Fig. 2.

by inputting anonymized images. The pix2pix [12] framework can be used to jointly optimize the anonymization generator.

The raw image recovery is similar to anonymization process. In contrast to G_X, the recovery generator G_Y is trained to learn a mapping function $F : Y \longrightarrow X$ and produce recovered images $G_Y(G_X(x))$, which cannot be distinguished from original raw images x. The adversarial objective function is similar to Eq. 1:

$$\begin{aligned}\mathcal{L}_{adv_2} = & \frac{1}{n}\sum\nolimits_{i=1}^{n}\log(D_X(G_X(x_i),x_i)) + \\ & \frac{1}{n}\sum\nolimits_{i=1}^{n}\log(1-D_X(G_X(x_i),G_Y(G_X(x_i)))),\end{aligned} \tag{4}$$

where G_Y and D_X oppose each other like G_X and D_Y.

To make G and F forward cycle-consistent, i.e., $x \longrightarrow G(x) \longrightarrow F(G(x)) \approx x$, a cycle consistency loss [44] is adopted:

$$\mathcal{L}_{1_{rec}} = \frac{1}{n}\sum\nolimits_{i=1}^{n}\|x_i - G_Y(G_X(x_i))\|_1. \tag{5}$$

The total loss of the recovery generator is:

$$\mathcal{L}_{rec} = \mathcal{L}_{adv_2} + \lambda_{L_1}\mathcal{L}_{1_{rec}}. \tag{6}$$

3.3 Joint Learning with a Re-ID Model

We design to embed the Re-ID model into our architecture for joint learning, which follows the powerful baseline AGW [36] in the existing Re-ID research. In the field of Re-ID, anonymization is a solution to the privacy issues. However, directly adopting desensitized images with conventional obfuscation methods will greatly affect the performance of Re-ID. Therefore, we propose to apply hybrid images (original and anonymized) to jointly train the Re-ID model and the anonymization generator.

The Re-ID model takes paired inputs: raw images and their corresponding anonymized images with the same labels. By training on paired data, the Re-ID model learns to map original raw images and anonymized images of a specific person to the same virtual identity. In detail, the Re-ID model contains three main components. *a) backbone.* ResNet50 [10] pre-trained on ImageNet [3] is adopted as the backbone with the stride of the last spatial down-sampling operation changed from 2 to 1. *b) Generalized-mean (GeM) pooling.* The Global Average Pooling in the original ResNet50 is replaced with GeM [36] whose output is adopted for computing center loss and triplet loss during training process. *c) BNNeck.* BNNeck [18] is added as a BN layer between features and FC layers.

The loss function is denoted by $\mathcal{L}_{AGW}(x)$, which combines three commonly used losses in Re-ID tasks including identity classification loss ($\mathcal{L}_{id}$), center loss ($\mathcal{L}_{ct}$) [31] and weighted regularization triplet loss ($\mathcal{L}_{wrt}$) [36] for optimization. To make our Re-ID model adaptive to both raw and privacy-preserving scenarios, both raw and anonymized images are added as the input. Therefore, the total loss of the Re-ID model is

$$\mathcal{L}_{reid} = \mathcal{L}_{AGW}(x) + \mathcal{L}_{AGW}(G_X(x)). \tag{7}$$

In summary, the final objective of our anonymization model for Re-ID on hybrid images is:

$$\mathcal{L} = \mathcal{L}_{ano} + \mathcal{L}_{rec} + \mathcal{L}_{reid}. \tag{8}$$

3.4 Progressive Supervision Upgradation

In § 3.3, a Re-ID model is added for joint learning to make anonymized images $G_X(x)$ suitable for identity preserving. However, initial supervision images (i.e., blurred, pixelated or noise-added images) are not optimal as final supervision images because the semantic information loss leads to identity variance, and thus restricts the improvement of Re-ID performance. Therefore, as shown in Fig.3 briefly and Fig. 4 in detail, we progressively upgrade the supervision images during the training process to satisfy both privacy protection (unrecognizability) and Re-ID constraints. Specifically, the privacy constraint is met when PSNR and SSIM calculated between anonymized images and raw images (i.e., $PSNR_{ano}$ and $SSIM_{ano}$) are lower than those calculated between desensitized images and raw images (i.e., $PSNR_{des}$ and $SSIM_{des}$) plus a small positive value (i.e., ϵ_{psnr} and ϵ_{ssim}). The Re-ID constraint is met when the rank-1 accuracy (i.e., $R1_{raw-ano}$) with raw images as

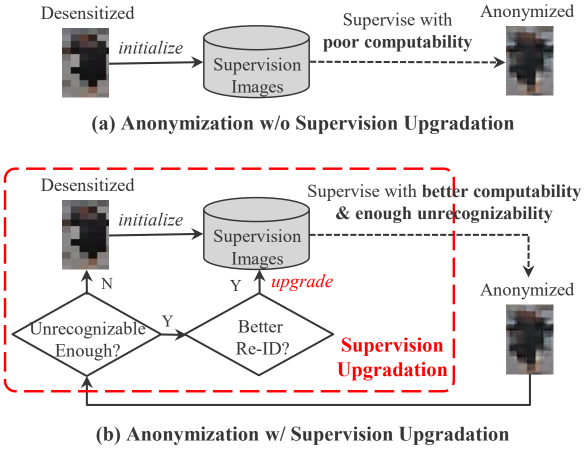

Desensitized
initialize
Supervision Images
Supervise with poor computability
Anonymized
(a) Anonymization w/o Supervision Upgradation
Desensitized
initialize
Supervision Images
Supervise with better computability & enough unrecognizability
Anonymized
N
Y
upgrade
Unrecognizable Enough?
Y
Better Re-ID?
Supervision Upgradation
(b) Anonymization w/ Supervision Upgradation

Figure 4: Illustration of supervision upgradation. The upgradation is based on the performance of protection and Re-ID.

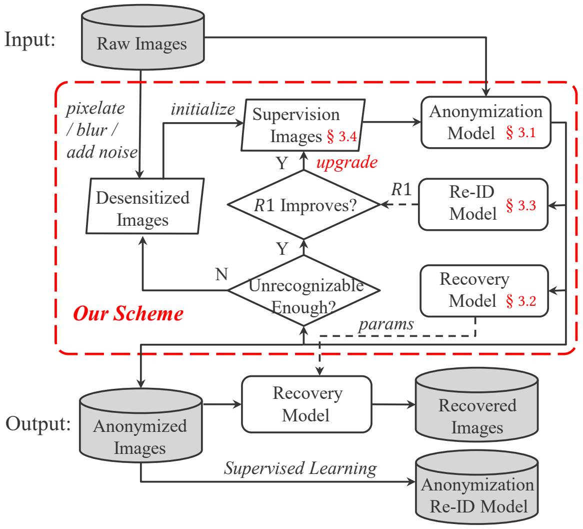


Figure 5: Flowchart of our training process. "R1" represents rank-1 accuracy with anonymized images as *query* and desensitized images as *gallary*.

query and anonymized images as *gallery* is higher than the previous maximum rank-1 value. To begin with, we train a Re-ID model with raw and desensitized images as input and get the rank-1 value (i.e., $R1_{raw-des}$) with raw images as *query* and desensitized images as *gallery*. Then, the desensitized images are adopted as the initial supervision images and the maximum rank-1 value is initialized to a $R1_{raw-des}$ plus small negative value ($-\epsilon_{r1}$). While training, the supervision images will keep as desensitized images to guarantee the unrecognizability if the privacy need is not met and will be upgraded to $G_X(x)$ only when both constraints are satisfied. Through the supervision upgradation, our supervision images are becoming more adequate for Re-ID research while preserving privacy.

3.5 Scheme of Training Process

In Fig. 5, we show the flowchart of the training process. The input raw images are first desensitized by conventional methods, which initialize the supervision images. Under the supervision, the anonymization model learns to translate raw images to anonymized images. The anonymized images are then fed into the Re-ID model and recovery model to jointly learn to preserve necessary features for recovery and retrieval. Then our training process is continued with the supervision images upgraded according to the performance of our anonymized images on privacy protection and person re-identification. After training, the output anonymized images can be used to recover original raw images with the parameters of the trained recovery model. Meanwhile, proper Re-ID performance can be achieved after supervised learning on these anonymized images.

4 EXPERIMENTAL RESULTS

Preliminary. In the following parts, we denote "OI" as the original raw images, "PI" as the protected images, "w/ U" and "w/o U" as anonymization with/without using supervision upgradation.

4.1 Datasets and Evaluation Metrics

We conduct experiments on three widely used datasets: Market-1501 [39], MSMT17 [30] and CUHK03 [30]. The Market-1501 dataset comprises 32,668 annotated bounding boxes under six cameras. The MSMT17 dataset consists of 4,101 identities and 126,441 bounding boxes taken by a 15-camera network. The CUHK03 dataset contains 1,467 identities and 14,097 detected bounding boxes.

We evaluate our model under image quality and re-identification metrics. For privacy protection and recovery, we adopt two widely used metrics: PSNR and SSIM [29]. For Re-ID performance, Cumulative Matching Characteristics (*a.k.a.*, Rank-k matching accuracy) [28], mean Average Precision (mAP) [39], and a new metric mean inverse negative penalty (mINP) [36] are used in our experiments.

4.2 Implementation Details

Training setup. We first split the original training set into a new training set and a validation set in a ratio of 4 : 1. Then we further split the validation set into a gallery set and a query set in the same ratio. All the performance while training is obtained by testing on the validation set. In all experiments, we jointly trained our three models for 120 epochs with batch size 64. All input images are resized to 256 × 128 and then desensitized by *blurring* 12 × 12, or *pixelation* 24 × 24, or adding *Gaussian noise* $N(0, 0.5)$. We use Adam optimizer [14] with $\beta_1 = 0.5, \beta_2 = 0.999$ for two generators and with default values $\beta_1 = 0.9, \beta_2 = 0.999$ for the Re-ID model. Learning rate is linearly increasing from 3.5×10^{-5} to 3.5×10^{-4} in the first 10 epochs, and then is decayed to 3.5×10^{-5} and 3.5×10^{-6} at 40th epoch and 80th epoch respectively.

Models of anonymization, recovery and Re-ID. Our implementation for the anonymization and recovery models follow Pix2pix network [12] and λ_{L_1} is set to 100 as suggested by [12], while the Re-ID model follows the practice in [36] and uses the same hyper-parameters. These might be replaced by other advanced methods.

Supervision Upgradation. To get a competitive effect of privacy protection, we set ϵ_{psnr} and ϵ_{ssim} (see in Fig. 4) to small values of 1.0 and 0.05. And ϵ_{r1} is set to a small value of 0.05.

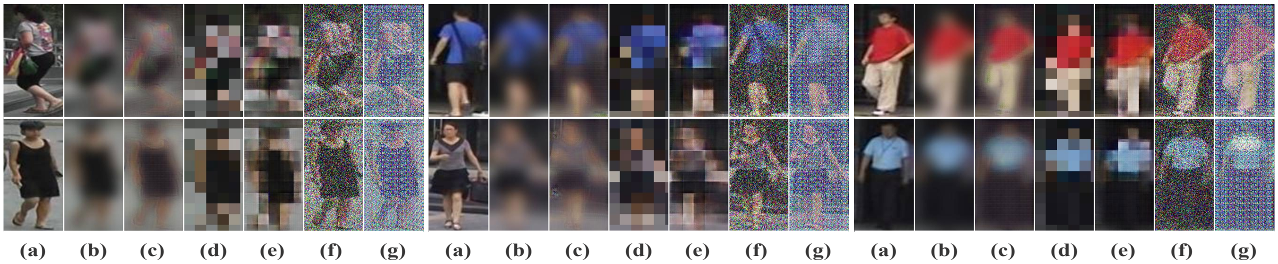


Figure 6: Qualitative comparison on privacy protection. (a) raw images; (b)/(d)/(f) blurred/pixelated/noise-added images; (c)/(e)/(g) anonymized images guided by blurring/pixelation/noise adding. Best view in color. Zoom in for details.

Table 1: Re-ID performance of common AGW on protected images. "Base" means traditionally desensitized images.

Dataset	Market-1501		MSMT17		CUHK03	
Images	rank-1	mAP	rank-1	mAP	rank-1	mAP
(a) Evaluation of blurring.						
Base	20.6	8.7	3.9	1.4	1.9	2.5
Ours	18.4	7.6	8.3	2.6	3.9	3.9
(b) Evaluation of pixelation.						
Base	20.2	9.1	1.8	0.7	2.1	2.3
Ours	17.5	7.7	6.7	2.1	1.3	1.6
(c) Evaluation of Gaussian noise.						
Base	0.6	0.4	0.2	0.1	0.1	0.3
Ours	1.4	0.6	0.4	0.1	0.2	0.4
Raw	95.7	88.6	68.6	49.8	67.3	65.8

Table 2: Human evaluation results. "Base" represents conventional anonymization methods. Privacy value(%) denotes verification accuracy by human eyes. Lower privacy value indicates better privacy protection, while higher Re-ID rank-1 accuracy means better Re-ID performance.

Image	A		B		Privacy value ↓	Re-ID rank-1 ↑
Method	OI	PI	OI	PI		
(a) Evaluation of blurring.						
Base	✓			✓	79	40.1
		✓		✓	82	67.3
Ours	✓			✓	83	88.2
		✓		✓	82	89.2
(b) Evaluation of pixelation.						
Base	✓			✓	71	75.3
		✓		✓	75	64.3
Ours	✓			✓	71	88.5
		✓		✓	64	87.0
(c) Evaluation of Gaussian noise.						
Base	✓			✓	84	50.8
		✓		✓	83	68.7
Ours	✓			✓	88	83.5
		✓		✓	84	91.2
Upper	✓		✓		92	95.7

4.3 Results of Privacy Protection

Qualitative Results. In Fig. 6, a qualitative comparison of privacy protection performance is conducted. Compared to raw images, our anonymized images achieve good visual privacy protection performance. The individual's body contour line and details of face and clothes are all concealed, and thus one cannot obtain the identity from the anonymized images by human eyes. Compared with the desensitized images, our corresponding anonymized images attain a competitive visual obfuscation effect in a different style.

Quantitative Results. Table 1 shows the Re-ID performance of unprotected AGW model on protected images. Compared with raw images (i.e., Raw), our anonymized images obtain extremely low rank-1 values and mAP values, showcasing that the common Re-ID model is not able to correctly identify our anonymized images. Compared with baselines, our anonymized images achieve similar Re-ID performance, indicating our anonymization method can obtain close privacy protection performance.

Human Evaluation. Table 2 shows the human evaluation of privacy protection effects of our method and baselines (blurring, pixelation and noise adding). We randomly sampled a pair of images from the raw or privacy-preserving Market-1501 testing set and ask participants whether the pair corresponds to the same person. The image pairs were divided into 13 groups (i.e., the 13 rows in Table 2) according to the protection method. Each group sampled 100 images that are distributed equally to 10 participants. The optimal privacy effect is when the privacy value equals 50%, which indicates random guessing. Compared with raw image pairs (i.e., Upper), the pairs with our anonymized image achieve a substantially lower privacy value and a slight decrease in Re-ID accuracy. Compared with baselines, our method obtains comparable verification accuracy (i.e., privacy value) by human eyes and significantly better Re-ID rank-1 accuracy by trained Re-ID models.

4.4 Results of Recovery

Qualitative Results. In Fig. 7, we qualitatively compare recovered images with raw images. Our recovered images achieve similar visual quality to original images. It is difficult to distinguish the recovered images from the original raw images by human eyes.

Quantitative Results. In Table 3, we show the image quality of recovered images. Approximately, in all three datasets, the PSNR and SSIM values of recovered images are higher than 25 and 0.9, indicating that our recovered images have good image quality. Moreover, as shown in Table 4, both the common AGW model and our

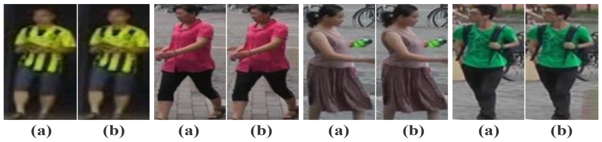


Figure 7: Qualitative results of the recovered images. (a) represents raw images; (b) represents recovered images.

Table 3: Image quality of the recovered image. PSNR and SSIM are reported. Higher value indicates better quality.

Dataset	Market-1501		MSMT17		CUHK03	
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
(a) Evaluation of blurring.						
Ours	26.78	0.92	30.02	0.93	23.67	0.89
(b) Evaluation of pixelation.						
Ours	29.74	0.93	25.94	0.89	27.00	0.93
(c) Evaluation of Gaussian noise.						
Ours	26.80	0.92	27.80	0.92	23.00	0.91
Upper	+∞	1	+∞	1	+∞	1

Table 4: Re-ID performance of the recovered images. Rank at r accuracy(%) and mAP(%) are reported.

	images		Market-1501		MSMT17		CUHK03	
Model	OI	RI	r=1	mAP	r=1	mAP	r=1	mAP
AGW	✓		95.7	88.6	68.6	49.8	67.3	65.8
		✓	93.8	84.0	63.2	43.1	64.6	62.2
Ours	✓		91.7	78.2	48.6	29.8	38.8	42.4
		✓	90.4	75.5	49.8	29.5	33.3	33.1

protected Re-ID model can obtain Re-ID performance on our recovered images comparable to that of original raw images. Besides, compared to AGW model, our model suffers a slight degradation of performance on raw and recovered images since it is also trained to improve performance on the anonymized images whose style is obviously different from raw and recovered images.

4.5 Results of Person Re-identification

Experiments under Four Test Settings. As shown in Table 5, we test the Re-ID performance under four settings with different queries and galleries. These four settings represent different scenarios: 1) **Original Setting** ($query = OI$, $gallery = OI$): This setting represents that we use original raw images for both query and gallery sets. The result shows that our proposed model achieves comparable performance to the existing widely used setting (Rank-1:91.7% *v.s.* 95.7% on the Market1501 dataset) with only a minor performance drop. This demonstrates that the model trained on our anonymized dataset can still be applied to practical scenarios when the testing pedestrian images are not anonymized. It brings in another interesting research topic, i.e., designing algorithms on the anonymized dataset without the invasion of privacy, and testing in practical non-anonymized scenarios, which alleviates the major ethical concern of recent research on human subjects. 2) **Protected Setting** ($query = PI$, $gallery = PI$): This setting indicates that we use privacy-preserving images for both query and gallery sets. Compared to baselines of blurring, pixelation and noise adding, our model achieves an average improvement of 26.8%, 26.3% and 28.0% in Rank-1 on three datasets. Compared to the original setting, our model under protected setting achieves comparable results. The results indicate that our anonymized images are suitable for Re-ID research and can be applied to practical scenarios when the testing pedestrian images are anonymized. 3) **Crossed Settings** ($query = OI$, $gallery = PI$ and $query = PI$, $gallery = OI$): These settings represent that we use different types of images for query and gallery sets. Compared to the baselines, our model significantly improves the performance on three datasets averagely by 34.2% and 41.1% for blurring, 22.4% and 21.7% for pixelation and 29.2% and 33.0% for noise adding. This indicates that our anonymization model is robust against privacy protection on the query and gallery sets. Besides, the performance is also comparable to the original setting, indicating that our model can be applied to Re-ID on hybrid images. It can be inferred that the feature distribution of our anonymized images has a close distance to that of raw images. Additionally, our model performs averagely better in the protected setting than in the crossed settings, probably because there still exists a minor domain gap between raw images and anonymized images. In summary, our anonymized images are suitable for Re-ID and the method can be applied to scenarios when testing images containing both raw and privacy-preserving images.

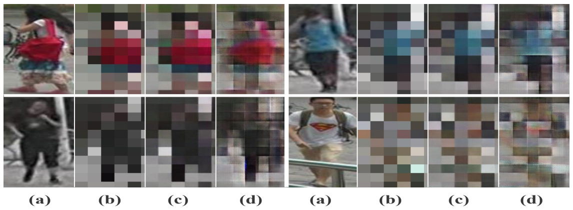


Figure 8: Qualitative comparison on supervision upgradation. (a) raw images; (b)pixelated images; (c) anonymized images w/o U; (d) anonymized images w/ U.

Effect of Supervision Upgradation. As shown in Table 5, compared to our model without upgradation, the model with upgradation generally performs better under all metrics and settings, e.g., the average Rank-1 increases 19.4% under the evaluation of blurring on Market1501 dataset. This shows that utilizing supervision upgradation indeed helps in improving Re-ID performance on our anonymized images. Fig. 8 shows that the anonymized images with upgradation retain a strong visual obfuscation effect. Instead of those without upgradation which follow the style of pixelation, raw images are obfuscated in a different learned style.

Discussion. In our privacy-preserving system, given a frame of raw video, pedestrians can be anonymized based on the detected bounding boxes. These anonymized bounding boxes can further be used to recover original raw images by police officers and adopted as public datasets for researchers. However, given an anonymized frame, anonymized pedestrians are unable to be detected by standard person detectors without training on them. It needs further joint learning with pedestrian detection tasks.

Table 5: Evaluation of Re-ID performance on three Re-ID datasets. "Base" means AGW model trained on desensitized images. "Upper" indicates original AGW model. Rank at r accuracy(%), mAP (%) and mINP (%) are reported.

	query		gallery		Market-1501				MSMT17				CUHK03			
	OI	PI	OI	PI	$r = 1$	$r = 5$	mAP	mINP	$r = 1$	$r = 5$	mAP	mINP	$r = 1$	$r = 5$	mAP	mINP
(a) Evaluation of blurring.																
Base	✓		✓		84.8	94.1	67.4	32.2	30.5	44.0	17.1	2.8	30.4	50.4	31.5	22.3
	✓			✓	40.1	59.4	25.4	6.3	21.3	35.1	10.7	1.4	14.6	28.4	14.8	8.6
		✓	✓		18.3	31.7	15.5	5.2	16.2	28.0	9.4	1.5	10.4	20.4	12.4	8.4
		✓		✓	67.3	83.5	44.2	13.7	15.2	24.3	7.2	0.8	8.2	19.8	10.7	6.9
Ours (w/o U)	✓		✓		83.1	93.7	61.9	24.8	43.6	57.6	24.0	3.7	31.6	52.6	32.5	23.0
	✓			✓	68.3	84.9	45.9	14.6	28.4	42.3	9.1	1.2	13.8	25.8	14.8	9.0
		✓	✓		46.8	67.7	34.2	11.7	10.0	19.2	5.8	0.8	8.9	19.4	10.8	7.1
		✓		✓	75.8	88.9	52.4	18.3	14.7	23.6	6.0	0.5	14.3	28.5	15.0	8.9
Ours	✓		✓		**91.6**	**97.4**	**79.4**	**47.4**	**51.5**	**65.3**	**31.1**	**6.0**	**41.9**	**62.0**	**41.7**	**30.4**
	✓			✓	88.2	95.8	72.0	37.0	51.1	64.9	29.7	5.2	39.2	59.3	38.4	27.2
		✓	✓		82.5	93.6	67.5	36.0	50.5	64.7	30.5	5.7	35.3	55.4	35.5	25.4
		✓		✓	89.2	96.4	74.3	39.4	48.7	62.4	28.5	4.9	33.2	55.3	34.7	25.0
(b) Evaluation of pixelation.																
Base	✓		✓		87.4	96.1	73.4	39.5	25.0	38.3	15.3	2.6	28.5	50.0	31.5	22.9
	✓			✓	75.3	91.1	53.6	17.2	16.3	29.0	8.7	1.0	17.7	36.5	17.6	9.1
		✓	✓		70.9	86.4	54.7	24.1	14.6	24.7	9.0	1.6	15.1	29.5	17.7	12.3
		✓		✓	64.3	83.5	43.4	13.0	10.6	19.7	5.7	0.7	8.8	20.3	9.9	5.3
Ours (w/o U)	✓		✓		86.3	95.3	69.8	34.0	34.3	47.7	18.8	2.8	24.9	44.9	27.3	19.1
	✓			✓	80.1	92.6	57.4	18.6	26.2	40.4	12.3	1.3	24.2	44.8	23.3	13.6
		✓	✓		75.1	89.1	57.1	24.3	24.6	37.0	12.9	1.9	19.6	35.1	20.8	14.1
		✓		✓	73.2	89.1	49.7	15.7	20.5	32.4	9.3	1.0	12.1	25.9	13.2	7.5
Ours	✓		✓		**89.4**	**96.2**	**75.4**	**42.4**	48.6	63.2	**29.8**	**5.9**	**38.8**	**64.3**	**42.4**	31.6
	✓			✓	88.5	95.7	71.9	35.9	**49.1**	**63.6**	29.3	5.5	37.8	60.5	41.4	**32.2**
		✓	✓		86.8	94.9	72.3	39.1	48.5	63.6	29.8	5.7	30.4	50.6	30.6	20.7
		✓		✓	87.0	95.6	70.5	34.8	48.1	62.7	29.3	5.6	27.6	47.3	28.5	19.2
(c) Evaluation of Gaussian noise.																
Base	✓		✓		75.9	89.4	51.5	16.5	24.0	34.2	11.1	1.3	14.0	27.4	15.7	9.9
	✓			✓	50.8	70.3	30.2	5.8	20.4	32.4	8.6	0.9	9.1	19.4	9.9	5.4
		✓	✓		41.7	62.9	26.5	6.8	18.5	28.4	8.4	0.9	8.6	18.9	9.9	5.9
		✓		✓	68.7	85.9	43.2	11.9	18.2	27.6	7.8	0.8	8.1	18.7	10.2	5.9
Ours (w/o U)	✓		✓		90.4	96.7	75.9	41.9	41.9	55.6	24.0	4.2	28.1	47.6	30.2	21.8
	✓			✓	77.5	89.7	57.3	21.0	36.0	51.5	18.4	2.7	30.4	51.1	30.2	20.0
		✓	✓		67.5	82.0	50.8	20.7	29.4	41.4	15.7	2.5	30.4	50.4	29.9	20.3
		✓		✓	84.4	93.6	62.8	25.8	31.3	43.4	15.0	1.9	31.9	53.2	32.4	22.8
Ours	✓		✓		**91.7**	96.8	**78.2**	**45.4**	46.9	60.7	**27.6**	**4.9**	35.8	56.1	36.8	26.7
	✓			✓	83.5	92.8	68.0	33.6	**48.1**	**62.7**	27.3	4.4	36.4	57.3	36.3	25.6
		✓	✓		83.8	92.7	67.3	32.4	46.2	59.4	26.1	4.4	37.9	57.1	36.9	25.5
		✓		✓	91.2	**96.9**	77.0	44.3	46.0	59.4	26.0	4.4	**41.9**	**62.4**	**41.6**	**30.4**
Upper	✓		✓		95.7	98.4	88.6	66.7	68.6	79.7	49.8	15.0	67.3	82.8	65.8	54.6

5 CONCLUSION

This paper proposes a new reversible anonymization framework to explore the privacy-utility trade-off for pedestrian images from Re-ID perspective, which can reversibly generate full-body anonymous images with little performance degradation in Re-ID tasks. We further propose a progressive training strategy to improve the Re-ID performance. Extensive experiments further demonstrate the effectiveness of our method using anonymized pedestrian images for privacy protection, recovery, and person re-identification.

ACKNOWLEDGMENTS

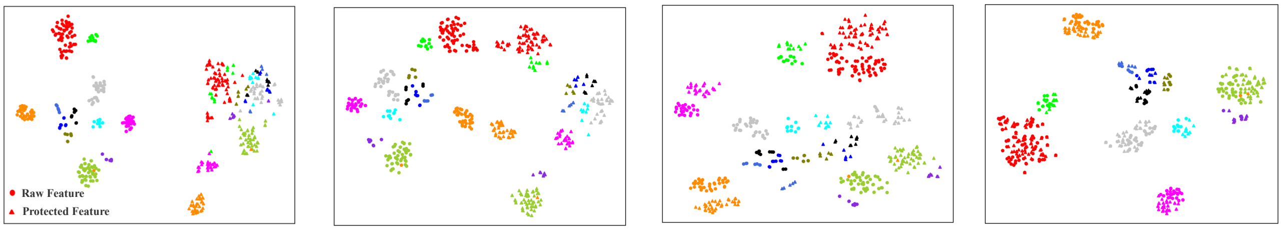
This work is supported by National Natural Science Foundation of China (62176188), Key Research and Development Program of Hubei Province (2021BAA187), Special Fund of Hubei Luojia Laboratory (220100015), Zhejiang Lab (NO.2022NF0AB01).

6 APPENDIX

A. Feature Distribution Comparison

In order to show more intuitively that our anonymized images are suitable as the public dataset for Re-ID research, we visualize the feature distribution of both raw and protected images in Fig. 9. The features are produced from the same batch of test samples and are extracted from the current non-protective Re-ID model, the baseline blurring-based model, and our blurring-based Re-ID model with/without supervision upgradation. If the protected feature distribution can be clustered distinctly by classes, and is similar to the original raw feature distribution, then the protected images should be suitable as a public Re-ID dataset.

a) Non-protective Model: Fig. 9(a) illustrates the feature distribution extracted from the current non-protective Re-ID model AGW



(a) Non-protective Model

(b) Baseline

(c) Ours(w/o U)

(d) Ours(w/ U)

Figure 9: Comparison of the feature distribution extracted by (a) non-protective Re-ID model, (b) the baseline blurring-based model, (c) ours (w/o U), and (d) ours (w/ U) on Market1501 dataset. Features with the same color are from images of the same person.

[36] which is trained on only raw images and tested on both raw and blurred images. It can be clearly seen that raw features are clustered by classes while protected features of different classes are mostly mixed together, indicating the Re-ID model trained on only raw images achieves poor performance when testing images contain protected images.

b) Baseline: The baseline Re-ID model is trained and tested on paired raw and blurred images. As illustrated in Fig. 9(b), compared to non-protective model, the protected features are clustered better and the distance between raw and protected feature distribution is narrowed. This indicates that the Re-ID performance is improved when the test set consists of protected images. However, the blurred images are not able to replace raw images as a Re-ID dataset due to the large deviation of feature distribution after blurring.

c) Ours (w/o U): Fig. 9(c) shows the feature distribution extracted from our model that is jointly trained without supervision upgradation and tested on raw and anonymized images. Compared to the baseline, the raw and protected feature distributions are significantly pulled in. However, there still exists observable misalignment between these two feature distributions.

d) Ours (w/ U): To further narrow the distance between these two feature distributions, we propose a training strategy, i.e., progressive supervision upgradation. As shown in Fig. 9(d), compared to baseline and our method w/o U, our method with supervision upgradation eliminates the deviation between these two types of feature distribution while retaining the distinction between classes. Compared to raw feature distribution from non-protective model, our method achieves comparable performance of forming clusters. The results indicate that the model trained on raw and our anonymized images can perform well under original, protected, and crossed settings and our anonymized images are suitable as the public Re-ID dataset.

B. Results of Privacy Protection and Recovery

In Fig. 6 and Fig. 7 of the main paper, we separately showed qualitative results of privacy protection and recovery. Besides, in Fig. 8 of the main paper, we performed comparison on supervision upgradation. In this part, we combine the three qualitative experiments and show more images in Fig. 10. It can be seen that our anonymized images achieve good visual obfuscation effect and our recovered images are visually similar to raw images.

REFERENCES

[1] Jia-Wei Chen, Li-Ju Chen, Chia-Mu Yu, and Chun-Shien Lu. 2021. Perceptual Indistinguishability-Net (PI-Net): Facial Image Obfuscation with Manipulable Semantics. In *CVPR*. 6478–6487.

[2] Deepfake. 2020. Deepfakes faceswap. https://github.com/deepfakes/faceswap.

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *CVPR*. 248–255.

[4] Julia Dietlmeier, Feiyan Hu, Frances Ryan, Noel E O'Connor, and Kevin McGuinness. 2022. Improving Person Re-Identification with Temporal Constraints. In *CVPR*. 540–549.

[5] Cynthia Dwork. 2008. Differential privacy: A survey of results. In *TAMC*. 1–19.

[6] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *TCC*. 265–284.

[7] Facebook. 2021. An Update On Our Use of Face Recognition. https://about.fb.com/news/2021/11/update-on-use-of-face-recognition.

[8] Oran Gafni, Lior Wolf, and Yaniv Taigman. 2019. Live face de-identification in video. In *ICCV*. 9378–9387.

[9] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *NIPS*.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *CVPR*. 770–778.

[11] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. 2019. Deepprivacy: A generative adversarial network for face anonymization. In *ISVC*. 565–578.

[12] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *CVPR*. 1125–1134.

[13] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. 2019. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977* (2019).

[14] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[15] Zhenzhong Kuang, Zhiqiang Guo, Jinglong Fang, Jun Yu, Noboru Babaguchi, and Jianping Fan. 2021. Unnoticeable synthetic face replacement for image privacy protection. *Neurocomputing* 457 (2021), 322–333.

[16] Zhenzhong Kuang, Huigui Liu, Jun Yu, Aikui Tian, Lei Wang, Jianping Fan, and Noboru Babaguchi. 2021. Effective De-identification Generative Adversarial Network for Face Anonymization. In *ACM MM*. 3182–3191.

[17] Tao Li and Lei Lin. 2019. Anonymousnet: Natural face de-identification with measurable privacy. In *CVPR*. 0–0.

[18] Hao Luo, Wei Jiang, Youzhi Gu, Fuxu Liu, Xingyu Liao, Shenqi Lai, and Jianyang Gu. 2019. A strong baseline and batch normalization neck for deep person re-identification. *ACM MM* 22, 10 (2019), 2597–2609.

[19] Maxim Maximov, Ismail Elezi, and Laura Leal-Taixé. 2020. Ciagan: Conditional identity anonymization generative adversarial networks. In *CVPR*. 5447–5456.

[20] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *AISTATS*. 1273–1282.

[21] Hugo Proença. 2020. The uu-net: Reversible face de-identification for visual surveillance video footage. *arXiv preprint arXiv:2007.04316* (2020).

[22] Zhongzheng Ren, Yong Jae Lee, and Michael S Ryoo. 2018. Learning to anonymize faces for privacy preserving action detection. In *ECCV*. 620–636.

[23] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. 2016. Performance measures and a data set for multi-target, multi-camera tracking. In *ECCV*. 17–35.

[24] Qianru Sun, Liqian Ma, Seong Joon Oh, Luc Van Gool, Bernt Schiele, and Mario Fritz. 2018. Natural and effective obfuscation by head inpainting. In *CVPR*. 5050–5059.

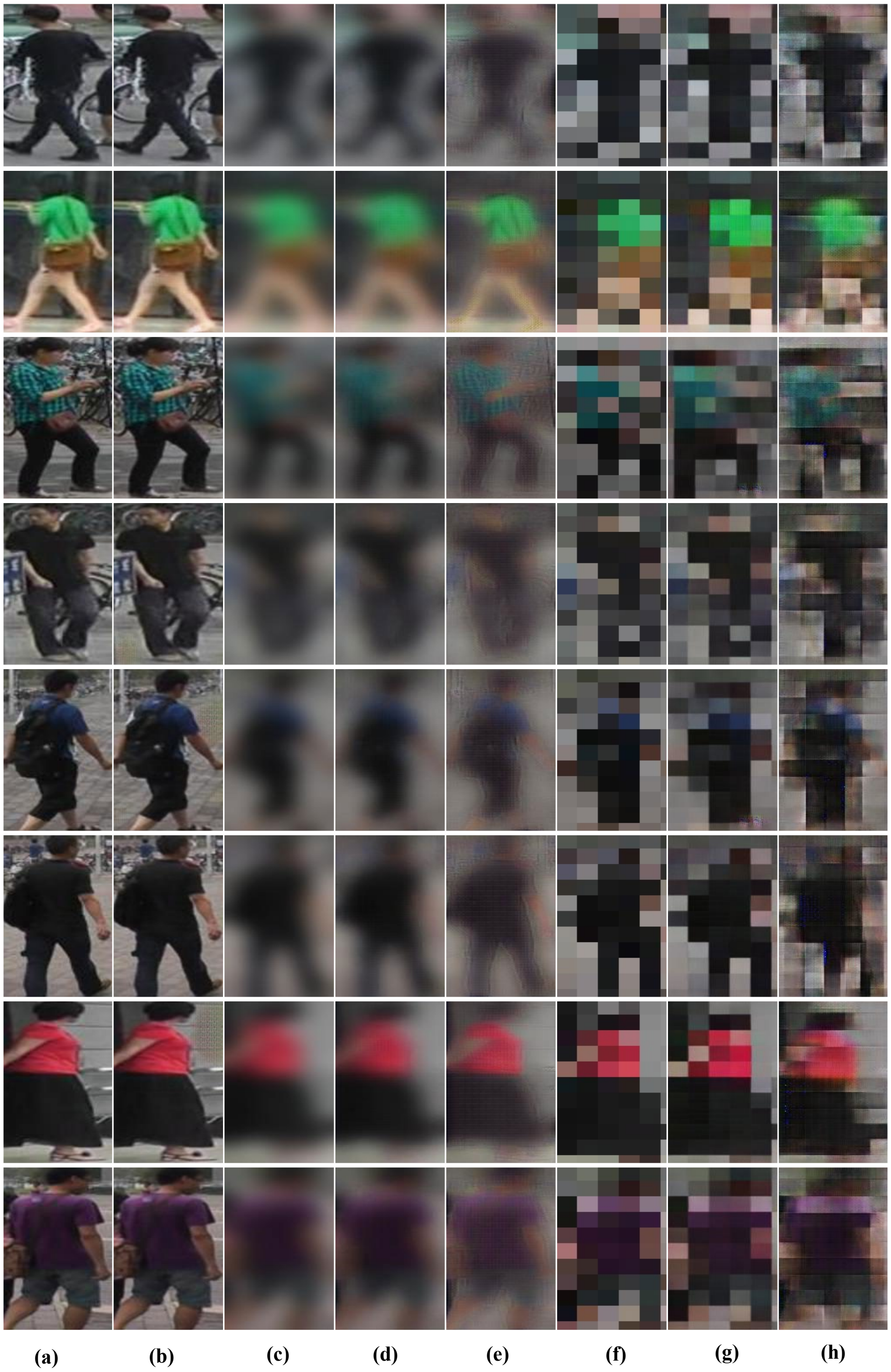


Figure 10: Qualitative comparison between baseline and our method. (a)/(b) denote raw/recovered images; (c) indicates blurred images; (d)/(e) represent anonymized images guided by blurred images without/with supervision upgradation; (f)/(g)/(h) are similar to (c)/(d)/(e) with blurred images being replaced by pixelated images.

[25] Qianru Sun, Ayush Tewari, Weipeng Xu, Mario Fritz, Christian Theobalt, and Bernt Schiele. 2018. A hybrid model for identity obfuscation by face replacement. In *ECCV*. 553–569.

[26] Antonio Torralba, Rob Fergus, and William T Freeman. 2008. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE TPAMI* 30, 11 (2008), 1958–1970.

[27] Guanshuo Wang, Yufeng Yuan, Xiong Chen, Jiwei Li, and Xi Zhou. 2018. Learning discriminative features with multiple granularities for person re-identification. In *ACM MM*. 274–282.

[28] Xiaogang Wang, Gianfranco Doretto, Thomas Sebastian, Jens Rittscher, and Peter Tu. 2007. Shape and appearance context modeling. In *ICCV*. 1–8.

[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE TIP* 13, 4 (2004), 600–612.

[30] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. 2018. Person transfer gan to bridge domain gap for person re-identification. In *CVPR*. 79–88.

[31] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. 2016. A discriminative feature learning approach for deep face recognition. In *ECCV*. 499–515.

[32] Yifan Wu, Fan Yang, and Haibin Ling. 2018. Privacy-protective-gan for face de-identification. *arXiv preprint arXiv:1806.08906* (2018).

[33] Kaiyu Yang, Jacqueline Yau, Li Fei-Fei, Jia Deng, and Olga Russakovsky. 2021. A study of face obfuscation in imagenet. *arXiv preprint arXiv:2103.06191* (2021).

[34] Mang Ye, Cuiqun Chen, Jianbing Shen, and Ling Shao. 2021. Dynamic tri-level relation mining with attentive graph for visible infrared re-identification. *IEEE TIFS* 17 (2021), 386–398.

[35] Mang Ye, He Li, Bo Du, Jianbing Shen, Ling Shao, and Steven CH Hoi. 2021. Collaborative refining for person re-identification with label noise. *IEEE TIP* 31 (2021), 379–391.

[36] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. 2021. Deep learning for person re-identification: A survey and outlook. *IEEE TPAMI* (2021).

[37] Mang Ye, Jianbing Shen, Xu Zhang, Pong C. Yuen, and Shih-Fu Chang. 2022. Augmentation Invariant and Instance Spreading Feature for Softmax Embedding. *IEEE TAPMI* 44, 2 (2022), 924–939.

[38] Zhengxin You, Sheng Li, Zhenxing Qian, and Xinpeng Zhang. 2021. Reversible Privacy-Preserving Recognition. In *ICME*. 1–6.

[39] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Scalable person re-identification: A benchmark. In *ICCV*. 1116–1124.

[40] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. 2021. Domain generalization: A survey. (2021).

[41] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. 2019. Omni-scale feature learning for person re-identification. In *CVPR*. 3702–3712.

[42] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. 2021. Learning generalisable omni-scale representations for person re-identification. *IEEE TPAMI* (2021).

[43] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. 2021. Domain generalization with mixstyle. *arXiv preprint arXiv:2104.02008* (2021).

[44] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*. 2223–2232.