2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Latent-RAG: Identity Retrieval-Guided Latent Augmentation for Privacy-Preserving Person Re-Identification

Seung-hyeok Back†, Eungi Lee†, Hyung-Il Kim* and Seok Bong Yoo*

***Abstract*— Person re-identification (re-ID) is crucial for security applications, including autonomous robots that monitor individuals via continuous image acquisition. Such data are transmitted to a database; however, if stored without adequate protection, they can be intercepted, posing privacy risks. In response, the existing methods balance privacy and accuracy, but protected images still reveal structural cues, such as silhouettes or edges. These methods rely on randomness to defend against recovery attacks, limiting the guarantee of complete protection. Thus, this work proposes latent retrieval-augmented generation (RAG), an identity retrieval-guided latent augmentation framework for privacy-preserving person re-ID that balances the re-ID performance with privacy protection. The proposed method generates augmented codes that distort appearance and disrupt mapping to the original input by retrieving identity-similar latent codes and applying inverse self-attention, enhancing its robustness to recovery attacks. Next, this approach employs gradient-based latent code manipulation to preserve identity vectors to maintain re-ID accuracy. The hierarchical latent codes are concurrently adjusted to eliminate structural cues that could threaten privacy. The experimental results demonstrate that Latent-RAG induces strong visual distortion, reliable re-ID accuracy and a robust defense against recovery attacks, even without additional training with a few frozen parameters in a pretrained generator. Our code is available at https://github.com/BACKAI/Latent-RAG.**

I. INTRODUCTION

Person re-identification (re-ID) is the task of identifying and matching the identity of an individual across images captured from various viewpoints. In robotics, re-ID is critical for mobile security and identification systems, such as robot navigation and autonomous security robots, which frequently collect images via onboard cameras [1]. The collected data are stored in the local database of the robot or are sent to a server and stored in the server database. If such collected and stored data are kept without protective measures, they could be exposed to unauthorized third parties, potentially enabling stalking or unauthorized peeking into individual's private information in the event of a malicious breach (see Fig. 1(a)). These attacks are primarily relevant to data that a robot directly captures and transmits during operation. In our setting, once the data are securely protected and transmitted without leakage, they are stored in a vector store hosted on a highly trusted server. Therefore, our goal is to protect the query data acquired on the robot side. To this end, a robotic system must visually obfuscate query images while preserving re-ID accuracy for authorized matching, which requires resolving a strict privacy–utility trade-off under practical compute and latency constraints. Accordingly, research on privacy-preserving person re-identification has been rapidly expanding to safeguard sensitive visual data in recent robotic applications.

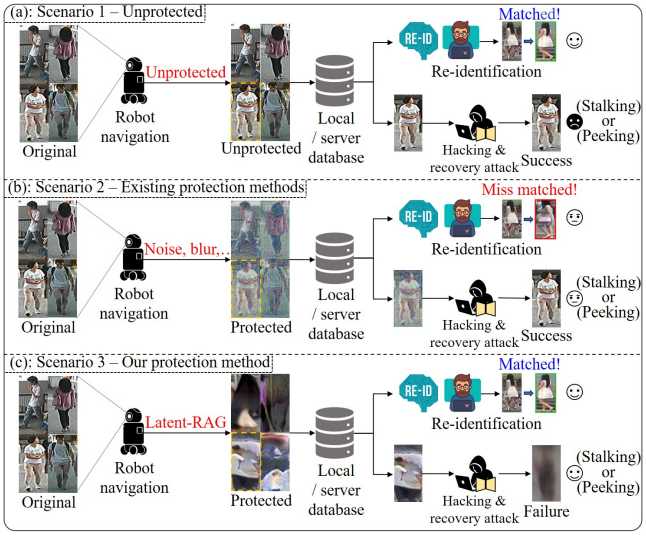

Fig. 1: Comparison of privacy protection in robot navigation scenarios. (a) Unprotected input: retains discriminative vectors for re-identification (re-ID) but exposes sensitive visual information. (b) Prior methods: provide partial obfuscation yet leak structural cues and degrade re-ID performance. (c) The proposed method: achieves visual obfuscation, recovery robustness and maintains re-ID discriminability.

The existing research on privacy-preserving person re-ID has been divided into two broad approaches. The first approach [2], [3] applies conventional image processing techniques, such as noise, blurring, or mosaic filters, to conceal visual information. Although these techniques offer visual protection by completely obscuring details, they also eliminate semantic features, decreasing re-ID performance. The second approach [4]–[14] applies deep learning to protect images while preserving the performance of a target re-ID model. This approach distorts images in a controlled manner and jointly trains the re-ID network, achieving visual privacy and high re-ID accuracy. However, if attackers can query the

This work was supported by the IITP grant funded by the Korea government (MSIT) (RS-2022-00156287, RS-2023-00256629, RS-2024-00437718).

The authors are with the Department of Artificial Intelligence Convergence, ({aiback856336, st0421, sbyoo}@jnu.ac.kr), and the School of Electronics and Computer Engineering, (hyungil.kim@jnu.ac.kr), Chonnam National University, Gwangju 61186, South Korea.

† These authors contributed equally to this work.

* Corresponding author.

protection model as a black box or know its architecture, they could train a recovery network to reconstruct the originals. Figure 1(b) illustrates this attack scenario.

Such recovery attacks pose a critical threat in real-world settings, making robustness against them a critical requirement for practical privacy-preserving systems. In this paper, we propose latent code retrieval-augmented generation (RAG), a privacy-preserving person re-ID framework that enhances robustness against recovery attacks while maintaining re-ID accuracy via identity retrieval-guided latent augmentation to address this threat. As illustrated in Fig. 1(c), the Latent-RAG balances visual privacy, re-ID accuracy and robustness to recovery attacks.

Latent-RAG integrates three core components to realize robust privacy-preserving person re-ID: identity-guided latent code retrieval, inverse self-attention-based augmentation and identity-aligned visual-divergence generation. Initially, the identity vector of the input image is extracted and applied to retrieve multiple semantically similar latent codes from a prebuilt vector store via cosine similarity. This identity-guided latent code retrieval decouples the input representation, disrupting any consistent mapping that an attacker might exploit to reconstruct original images. Then, the retrieved latent codes are fused using an inverse self-attention mechanism that calculates channel-wise importance weights to facilitate visual obfuscation. This augmentation maximizes distortion in the synthesized output, ensuring that sensitive visual details are concealed prior to image generation.

Next, the proposed method introduces an identity-aligned generation module to address the challenge that using retrieved latent codes rather than the original input may dilute identity information and influence re-ID accuracy. This approach generates an initial image by building on the augmented latent code and applying gradient-based latent manipulations to closely align the identity vectors of the protected image with those of the original. By operating directly in the latent space instead of performing pixel-level perturbations, the proposed method produces visually distinct yet identity-consistent outputs. This approach prevents recovery attacks via structural disentanglement and ensures strong and transferable re-ID performance across diverse models, setting a new standard in privacy-preserving person re-ID. The contributions are summarized below:

- This paper proposes a Latent-RAG that leverages the RAG framework to enhance visual privacy in the latent domain, improve recovery robustness and increase re-ID accuracy in privacy-preserving person re-ID through retrieval and augmentation.
- Latent-RAG retrieves latent codes using identity vectors similar to those of the original, decoupling the input representation to disrupt fixed mappings and enhance resistance to recovery attacks.
- The retrieved latent codes are augmented through an inverse self-attention mechanism to maximize visual distortion and conceal sensitive structural cues in the generated images.
- The augmented latent codes are manipulated through a loss that enables identity alignment and visual divergence, ensuring transferable re-ID performance while preserving privacy.

II. RELATED WORK

A. Privacy-Preserving Person Re-Identification

Person re-ID is a critical perception technology in robotics, enabling systems to recognize individuals across applications ranging from human–robot interaction to robot navigation [15]–[26]. However, as robots process and transmit increasing volumes of visual data in shared spaces, these systems raise privacy concerns [27]–[29]. Early work in privacy-preserving person re-ID has relied on visual obfuscation such as blur or noise [2], [3], to address privacy concerns. However, these methods often undermine the semantic features required for robotic perception.

Recently, deep learning-based anonymization [4]–[14] has enabled the generation of privacy-preserved features or reversible anonymized images that better balance individual privacy with the accuracy required for effective robot navigation and interaction. In addition to visual privacy, robust resistance to recovery attacks is crucial in privacy-preserving person re-ID. Conventional visual anonymization can be recovered by attackers who can reconstruct the original content from protected data, so recent studies have emphasized visual anonymization and recovery defense [30].

Encryption-based methods [7], [8], [10], [13], [31], [32] aim to prevent recovery attacks by transforming images into an encrypted space. However, their generalization is limited because they are effective only against the re-ID models used during training. Moreover, some recent approaches [9], [11] assume white-box access to the recovery model, which is often impractical in real-world scenarios. Other methods [6], [12], [14], [33] adopt randomization strategies to obfuscate mappings. Nevertheless, randomization alone fails to guarantee robust defense because it can be undermined via algorithmic analyses, repeated querying, statistical inference, or adaptive attacks [34], revealing a vulnerability in such defenses.

The RAG method replaces the input with those of highly semantically similar yet sufficiently distinct vectors retrieved from a vector store, making it infeasible to train a recovery model, even with paired supervision. This design preserves re-ID accuracy while ensuring robust visual privacy and robustness against recovery attacks.

B. Recovery Attack

This work adopts the commonly assumed recovery attack scenario [9], [11], [13], where an attacker aims to reconstruct original images from the protected outputs of a privacy model. In this setting, we further assume the attacker has black-box access to the privacy model and can obtain anonymized images for arbitrary input. By feeding large-scale public data into the privacy model, the attacker can collect input-output pairs and train a recovery network to minimize the L1 loss between the reconstructed and original images. Once trained, this network enables visual identity

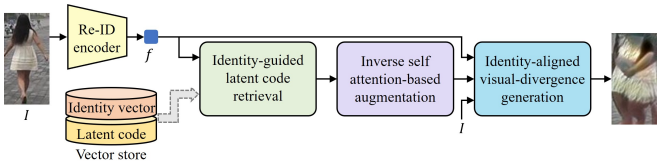

Fig. 2: Overall architecture of Latent-RAG.

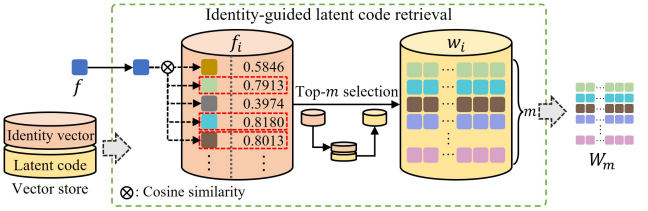

Fig. 3: Identity-guided latent code retrieval process.

recovery and defeats privacy protection, even without direct access to the re-ID system. While prior works employ a single recovery model based on a full-scale U-Net [35] or pix2pix [36], we adopt a more advanced reconstruction model, Restormer [37], to simulate a stronger attacker with enhanced reconstruction capability.

III. METHOD

A. Overview

The proposed approach was inspired by the RAG paradigm for privacy-preserving person re-ID via latent code retrieval and augmentation in the $\mathcal{W}^+$ space [38]–[40]. As illustrated in Fig. 2, the process extracts an identity vector from the image to be protected using a re-ID encoder [1] and compares it via cosine similarity with identity vectors stored in a prebuilt vector store. The vector store contains identity vectors for each individual and their corresponding latent codes extracted using E4E [38], and it is stored on a secure server isolated from malicious attackers.

The top several identity vectors are retrieved and replaced with their corresponding latent codes using similarity scores. The selected latent codes undergo inverse self-attention, emphasizing the lower channels and are averaged across all channels to produce an augmented latent code that induces pronounced visual distortion during image generation. Identity-guided loss is applied to enforce identity consistency to overcome the potential re-ID performance degradation caused by visual distortion. Additionally, the latent code underwent fine-grained, hierarchical manipulation to prevent the generated images from structurally resembling the originals due to this constraint.

B. Identity-Guided Latent Code Retrieval

The proposed method retrieves latent codes that are semantically similar in identity to the input image, as illustrated in Fig. 3. In this module, the vector f is extracted from an input image I using a re-ID encoder [1]. This vector is compared to identity vectors $\{f_i\}_{i=1}^N$ stored in a prebuilt vector store using cosine similarity:

$$c_i = \frac{f \cdot f_i}{\|f\|\|f_i\|}, \quad \forall i \in \{1, \ldots, N\}, \tag{1}$$

where $\cdot$ denotes the dot product and $\|\cdot\|$ the ℓ_2-norm. N is the total number of samples in the vector store. The resulting cosine similarity scores $C = \{c_i\}_{i=1}^N$ quantify the identity-level proximity between the input and each stored sample.

This approach employs FAISS [41], which indexes pre-extracted identity vectors for a fast approximate nearest neighbor search to efficiently retrieve the top-m most similar identities. Each retrieved identity vector f_i is paired with a precomputed latent code w_i obtained from the same reference image using the E4E encoder [38]. These latent codes reside in the $\mathcal{W}^+$ space and capture fine-grained visual attributes in a disentangled, multilayered representation. The latent codes corresponding to the top m retrieved identities were collected to form the latent code set:

$$W_m = \{w_i \mid i \in \text{argsort}(C)[:\, m]\}\,, \tag{2}$$

where $\text{argsort}(C)$ returns the indices of C sorted in descending order. Each w_i has the dimensions of 14×512, resulting in $W_m \in \mathbb{R}^{m \times 14 \times 512}$. These codes serve as the basis for latent fusion and visual obfuscation in the following steps.

The latent code set W_m does not include the latent code of the input image itself. This design enhances privacy by ensuring that no direct reconstruction path exists from the protected image to its original latent representation. The retrieved latent codes are selected based on the identity similarity; hence, the fused representation retains semantic consistency with the original identity, preserving re-ID performance to a reasonable extent. Hence, even if a malicious attacker attempts to train a recovery model to invert the protected images, the absence of consistent identity-latent pairs significantly hinders the ability of the model to reconstruct or identify individuals.

C. Inverse Self Attention-Based Augmentation

Prior to considering defenses against recovery attacks, protected images must preserve privacy through obfuscation. Therefore, the method applies self-attention inversely to the retrieved latent code set W_m. We term this mechanism inverse self-attention. In contrast to standard self-attention, it reverses the similarity measure so that dissimilar features receive higher scores. This mechanism in Fig. 4 emphasizes relationships between dissimilar components across identity-related codes, promoting visual privacy.

This method reorders the dimensions of W_m to enable channel-wise processing. This approach treats each of the 14 layers independently, with the dimensions $14 \times m \times 512$ corresponding to 14 channels, m retrieved codes and 512-dimensional vectors. For each layer, this method defines the corresponding $m \times 512$ matrix as the query (Q), key (K), and value (V) input to the self-attention mechanism. Attention scores are computed using matrix multiplication between Q and the transposed key matrix $K^\top$, were computed, yielding an attention matrix of $m \times m$. The scores were inverted by

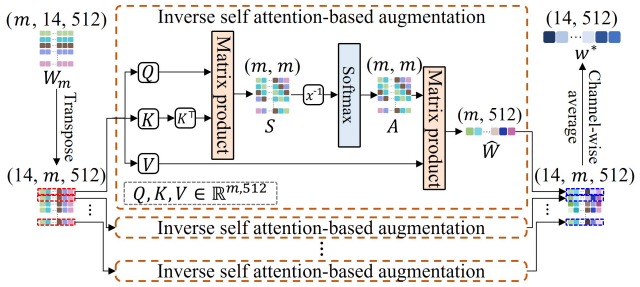

Fig. 4: Inverse self attention-based augmentation process.

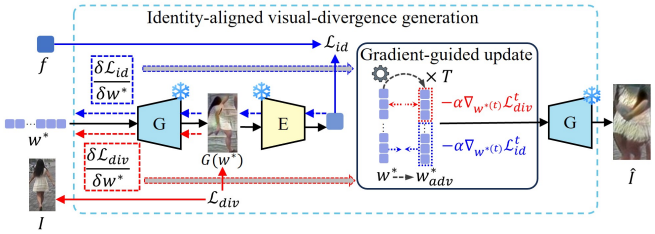

Fig. 5: Identity-aligned visual-divergence generation process.

applying the following element-wise reciprocal, so that visually dissimilar features are prioritized over visually similar ones:

$$S_{u,v} = \frac{1}{(S_{\text{attn}})_{u,v}}, \quad S_{\text{attn}} = (Q \times K^{\top}) \oslash D, \tag{3}$$

where $\times$ denotes matrix multiplication and $\oslash$ denotes element-wise division. D is an $m \times m$ matrix with every entry equal to $\sqrt{d}$ and d represents the dimensionality of Q and K. The scaling factor $\sqrt{d}$ mitigates magnitude inflating during the matrix-product attention. In this setting, S reflects the inverse similarity between the u-th and v-th latent codes in a given channel. If higher S scores are assigned to distinct features, those that are visually distant from the original will be more strongly emphasized, which can contribute to the visual distortion in the final output.

Next, the softmax function normalizes S along each row to obtain the inverse attention matrix A:

$$A_{u,v} = \frac{\exp(S_{u,v})}{\sum_{v=1}^{m} \exp(S_{u,v})}, \tag{4}$$

where A is applied as a weight to V, yielding weighted latent channels to suppress identity-redundant features and highlight diverse, less identity-dependent attributes:

$$\hat{W} = A \otimes V, \quad \hat{W} \in \mathbb{R}^{m \times 512}, \tag{5}$$

This process is repeated across all 14 channels, yielding a matrix of dimensions $14 \times m \times 512$. Each matrix is averaged across the m retrieved codes to produce one representative vector per channel. The final augmented latent code $w^* \in \mathbb{R}^{14 \times 512}$ is obtained by concatenating these 14 vectors.

D. Identity-Aligned Visual-Divergence Generation

Privacy-preserving transformations often degrade re-ID performance by distorting identity-related vectors. To address this issue, we propose the identity-aligned visual-divergence generation module, which aims to protect visual privacy while retaining discriminative cues essential for person re-ID. Unlike pixel-level adversarial perturbations that are designed as targeted attacks to produce similar vectors to the original image for a specific model, our approach perturbs semantic attributes in the latent space. Such perturbed attributes ensure re-ID performance even when no vectors corresponding to the original are present in the vector store during the retrieval process. This manipulation, as illustrated in Fig. 5, can produce perturbations that are less model-specific, improving transferability across re-ID models.

This work leverages the hierarchical latent space of $\mathcal{W}^+$, where coarse layers (0–2) encode high-level structural attributes (e.g., pose and layout) and fine layers (3–13) capture detailed visual vectors [42]. Our method applies layer-wise adaptive manipulation of the augmented latent code w^*. Coarse layers are modified to reduce structural resemblance to the original image, enhancing privacy. Fine layers are adjusted to preserve identity-discriminative information, maintaining re-ID effectiveness. Formally, privacy enhancement is guided by the divergence loss:

$$\mathcal{L}_{\text{div}} = -\|G(w^*) - I\|, \tag{6}$$

where I is the input image and generator (G) is the StyleGAN3 [43]. The gradient of $\mathcal{L}_{\text{div}}$ with respect to w^* is applied only to the coarse layers, encouraging the generator to disrupt structural information that may reveal visual content. Identity preservation is enforced via the identity loss:

$$\mathcal{L}_{\text{id}} = 1 - \frac{E(G(w^*)) \cdot f}{\|E(G(w^*))\| \|f\|}, \tag{7}$$

This loss is applied to the fine layers through gradient-guided updates, promoting identity similarity while avoiding structural cues. This identity-preserving loss enables us to maintain re-ID utility even when the vector store is small or lacks sufficiently diverse, highly similar samples for a given identity, since retrieval only needs to provide auxiliary guidance rather than perfectly matched latent candidates.

The latent code w^* is iteratively updated over T steps with step size α as follows:

$$w^{*(t+1)} = w^{*(t)} - \alpha \nabla_{w^{*(t)}} \mathcal{L}^{(t)}, \quad t = 0, \ldots, T, \tag{8}$$

where $\mathcal{L}^{(t)}$ includes $\mathcal{L}_{\text{div}}$ and $\mathcal{L}_{\text{id}}$, applied to coarse and fine layers with step sizes of 0.001 and 0.01, respectively. Each gradient is confined to its target layer, maintaining updates for privacy and identity objectives. After T iterations, we obtain the adversarial latent code:

$$w^*_{\text{adv}} = w^{*(T)}, \tag{9}$$

which is then used to generate the final protected image: $\hat{I} = G(w^*_{\text{adv}})$. This image obscures visual details while preserving identity-discriminative information, enabling robust re-ID performance under privacy constraints.

TABLE I: Privacy strength measured by the visual similarity between original and protected images. Lower scores indicate stronger visual privacy protection.

Protection Methods	Market-1501		MSMT17		CUHK03	
	PSNR↓	SSIM↓	PSNR↓	SSIM↓	PSNR↓	SSIM↓
–	∞	1.00	∞	1.00	∞	1.00
Gaussian Blur [1]	18.76	0.35	20.35	0.38	17.56	0.45
Mosaic [1]	14.69	0.21	16.34	0.22	13.55	0.20
PrivacyReID [7]	16.96	0.47	18.53	0.66	15.79	0.65
AVIH [9]	13.78	0.16	15.47	0.24	12.44	0.23
PixelFade [14]	12.33	0.08	14.02	0.13	11.12	0.12
SecureReID [13]	13.53	0.25	15.26	0.36	12.41	0.34
Ours	**10.43**	**0.07**	**12.17**	**0.11**	**9.17**	**0.06**

IV. EXPERIMENTS

A. Evaluation Setup

This work employs three benchmarks for person re-ID: Market-1501 [44], MSMT17 [45] and CUHK03 [46]. Market-1501 has 32,668 images of 1,501 identities from six cameras, MSMT17 is larger with 126,441 bounding boxes of 4,101 identities from 15 cameras across indoor and outdoor scenes and finally, CUHK03 comprises 14,097 detected bounding boxes associated with 1,467 identities.

This work employs a re-ID model trained on raw images for both query and gallery sets to assess the generalizability of privacy-preserving methods. Gallery vectors are pre-extracted and stored in a vector store; thus, protection is applied only to queries.

B. Evaluation Metrics

This work evaluates the proposed model using image quality and person re-ID metrics. For image quality, peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) are adopted to assess similarity between protected or recovered images and originals. Lower PSNR and SSIM values indicate stronger privacy protection, both in the protected images themselves and under recovery attacks.

This work employs Rank-1 matching accuracy and mean average precision (mAP) [44], to assess the effectiveness of re-ID. Higher scores in these metrics correspond to a better preservation of identity information and utility for pedestrian recognition. Combining these metrics enables a comprehensive measure of the visual privacy protection and utility of the protected images. In all tables, the bold values indicate the best results.

C. Implementation Details

This work applies the vector store using the $N = 32,668$ gallery images provided in the Market-1501 dataset. These images contain pairs of identity vectors extracted via a person re-ID model and latent codes computed by E4E [38]. Even when new data are added to the vector store, they are stored as vector units. As a result, incremental updates can be performed efficiently without storage concerns, making the approach scalable under memory constraints. We set the number of retrieved latent codes to 10 ($m = 10$). These latent codes are employed for image generation using StyleGAN3 [43]. In gradient-guided update, latent codes are perturbed using the iterative fast gradient sign method [49] with $T = 10$ iterations. All experiments were performed on a single Nvidia RTX 3080 graphics processing unit.

Fig. 6: Visual comparison of protected images (a) generated by various privacy-preserving methods and their corresponding reconstructed images (b) by recovery attacks. (1) Original image; (2) Gaussian Blur; (3) Mosaic; (4) PrivacyReID; (5) AVIH; (6) PixelFade; (7) SecureReID; (8) Latent-RAG.

For the baseline comparison, conventional methods such as Gaussian blur and mosaic, were implemented using the default configurations from PixelFade [14]. For deep learning-based privacy-preserving methods, including PrivacyReID [7], AVIH [9], PixelFade [14] and SecureReID [13], the open-source implementations were used, adopting their default settings. For SecureReID, the encryption–decryption module was excluded to ensure a fair comparison because it relies on a reversible recovery before re-ID.

D. Main Results

Figure 6(a) and Table I illustrate the privacy effectiveness of protection methods from qualitative and quantitative perspectives. Figure 6(a) shows that conventional techniques, such as Gaussian blur and mosaic, partially preserve the human silhouette, leaving visual clues that may reveal identity. PixelFade causes stronger corruption, yet high-frequency details like edge structures remain visible. Table I supports these observations: these methods yield relatively high PSNR and SSIM scores, indicating that much of the original visual information is retained. In contrast, the proposed Latent-RAG introduces more aggressive distortions that disrupt the human shape, as observed in the figure and achieves the lowest PSNR and SSIM scores, confirming it provides the strongest visual obfuscation among the compared approaches.

Table II compares re-ID performance of across Market-1501, MSMT17 and CUHK03. Conventional approaches, such as Gaussian blur and mosaic, substantially degrade

TABLE II: Comparison of re-ID performance on protected images across person re-ID datasets.

Re-ID Backbone	Protection Methods	Market-1501		MSMT17		CUHK03	
		Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
AGW [1]	Gaussian Blur	18.7	15.6	16.1	9.5	10.2	11.0
	Mosaic	70.4	54.2	14.0	9.1	15.2	14.7
	PrivacyReID	81.9	66.7	49.2	30.1	33.9	32.1
	AVIH	91.8	80.3	58.5	40.8	55.6	52.4
	PixelFade	93.6	85.4	62.8	45.7	59.0	57.2
	SecureReID	87.4	75.2	54.4	36.7	50.1	44.6
	Ours	**94.4**	**87.1**	**64.1**	**47.9**	**61.5**	**59.0**
	w/o protection	95.1	87.8	68.3	49.3	63.6	62.0
BagTricks [47]	Gaussian Blur	17.7	13.5	12.1	5.5	7.1	5.7
	Mosaic	69.5	52.4	10.2	4.9	10.1	9.3
	PrivacyReID	81.5	64.5	44.4	27.1	28.8	25.5
	AVIH	91.2	78.2	53.2	36.3	49.9	46.1
	PixelFade	93.1	83.1	57.9	41.8	53.4	51.7
	SecureReID	86.9	73.1	49.7	33.5	39.1	37.9
	Ours	**93.9**	**85.7**	**59.8**	**44.2**	**56.4**	**53.2**
	w/o protection	94.5	85.9	63.4	45.1	58.0	56.6
ABD-Net [48]	Gaussian Blur	19.2	17.4	17.8	10.7	12.1	11.8
	Mosaic	71.0	56.8	16.2	10.5	17.1	15.5
	PrivacyReID	82.1	68.5	49.9	31.6	34.8	32.8
	AVIH	92.6	82.1	59.4	43.1	56.8	53.1
	PixelFade	93.9	87.5	63.7	47.4	60.7	57.9
	SecureReID	87.6	71.4	55.0	41.2	41.8	44.6
	Ours	**95.1**	**88.9**	**64.8**	**49.5**	**62.3**	**59.4**
	w/o protection	95.6	89.3	68.8	50.7	64.5	62.6

TABLE III: Resistance to recovery attacks measured by the visual similarity between original and recovered images. Lower scores indicate stronger recovery protection.

Protection Methods	Market-1501		MSMT17		CUHK03	
	PSNR↓	SSIM↓	PSNR↓	SSIM↓	PSNR↓	SSIM↓
–	∞	1.00	∞	1.00	∞	1.00
Gaussian Blur	25.79	0.69	27.35	0.75	22.66	0.58
Mosaic	25.17	0.68	26.71	0.73	21.99	0.56
PrivacyReID	33.01	0.86	34.55	0.92	29.94	0.74
AVIH	25.51	0.70	27.01	0.76	22.42	0.59
PixelFade	23.64	0.60	25.21	0.66	20.50	0.49
SecureReID	30.35	0.84	31.86	0.91	27.19	0.71
Ours	**14.14**	**0.22**	**15.73**	**0.25**	**12.91**	**0.18**

re-ID accuracy by indiscriminately removing cues needed for identity recognition. Learning-based methods, such as PrivacyReID and SecureReID, perform better but require additional training on protected images, limiting practicality. AVIH and PixelFade notably improve re-ID performance through pixel-level adversarial perturbations, but their reliance on a specific target model (AGW) leads to poor generalization to unseen backbones. In contrast, Latent-RAG operates in feature space via identity-guided latent manipulation, enabling remarkably strong generalization across backbones and datasets. We adopt AGW [1], BagTricks [47] and ABD-Net [48], achieving the best performance in three widely used re-ID models.

E. Resistance to Recovery Attacks

We evaluate the robustness of privacy-preserving methods against recovery attacks, both qualitatively and quantitatively, as shown in Fig. 6(b) and Table III. Figure 6(b) indicates that these methods, including Gaussian blur and mosaic, preserve coarse human silhouettes, allowing attackers to reconstruct visually plausible figures. PixelFade and SecureReID introduce stronger corruption but still retain some structural or facial details in the reconstructed outputs. In contrast, images protected using the Latent-RAG method lead to degraded reconstructions, eliminating any identity-related visual cues. These observations are supported by the low PSNR and SSIM values in Table III, which indicate dissimilarity between the original and recovered images.

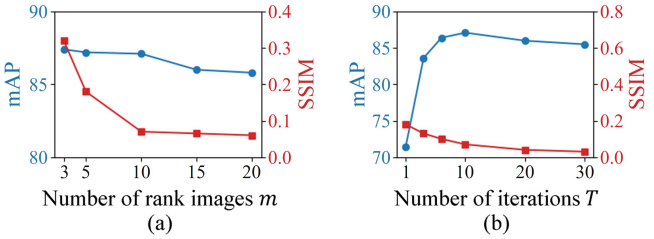

Fig. 7: Effects of (a) the number of retrieved rank images m and (b) the number of iterations T.

TABLE IV: Ablation results on the Market-1501 dataset.

Identity-guided latent code retrieval	Inverse self attention-based augmentation	Identity-aligned visual-divergence generation		SSIM↓ (Protected)	SSIM↓ (Recovered)	Rank-1↑ (Protected)
		L_{id}	L_{div}			
✓	✓	✓	✓	**0.07**	**0.22**	**94.4**
✓		✓	✓	0.18	0.25	93.2
		✓	✓	0.35	0.42	88.2
		✓		0.57	0.74	91.6
✓	✓			0.25	0.31	48.5
✓				0.31	0.33	46.0

F. Parameter Analysis of Latent-RAG

In this section, we evaluate the impact of two key hyperparameters of the proposed Latent-RAG through experiments on the Market-1501 dataset: the number of retrieved rank images m and the number of gradient update iterations T. Figure 7(a) shows that increasing m generally reduces the SSIM score between the protected and original images, indicating stronger visual privacy. However, excessively large m slightly decreases mAP, as excessive diversity may dilute identity-relevant vectors. The best trade-off is observed at $m = 10$, which achieves low SSIM while maintaining high mAP.

Figure 7(b) shows the effect of iteration count T in the identity-aligned visual-divergence stage. Increasing T initially improves mAP, but beyond $T = 10$, gains plateau and may slightly decline due to over-adjustments in the latent space. Meanwhile, SSIM consistently decreases with larger T, indicating stronger privacy. Thus, $T = 10$ is chosen as the optimal setting, balancing re-ID performance and privacy protection. This trend is well generalized and consistently observed in datasets such as MSMT17 and CUHK03.

G. Ablation Study

In this section, we conduct ablation studies on the Market-1501 dataset to evaluate each Latent-RAG component (Table IV). The baseline includes only identity-guided latent code retrieval. Identity-guided latent code retrieval disrupts latent mapping, blocking recovery attacks and enhancing visual privacy. Inverse self-attention augmentation reduces visual similarity, strengthening obfuscation. For identity-aligned visual-divergence generation, using only identity loss improves re-ID accuracy but may partially restore visual content, while adding divergence loss mitigates this, balancing accuracy and privacy. Combining latent code retrieval with identity-aligned generation recovers baseline-level accuracy

TABLE V: Complexity comparison of privacy-preserving methods.

Method	FLOPs(G)↓	Params(M)↓	Inference time(ms)↓
PrivacyReID	184.07	41.83	366.57
AVIH	121.79	114.38	853.01
PixelFade	89.16	23.51	1149.58
SecureReID	201.41	66.37	379.59
Latent-RAG	**86.65**	**0.79**	**347.93**
(R + A + G)	(5.42 + 0 + 81.23)	(0 + 0 + 0.79)	(11.63 + 0.78 + 335.52)

with better privacy and integrating all three achieves the best privacy-preserving trade-off, confirming their compatibility.

H. Computational Complexity

Table V compares the computational complexities of privacy-preserving person re-ID methods. PrivacyReID and SecureReID achieve fast inference by leveraging pix2pix-based image reconstruction. However, they require retraining whenever a new re-ID model is adopted, as their performance relies on end-to-end learning. In contrast, AVIH and PixelFade apply fine-grained, iterative pixel-level perturbations, which increase processing time. Latent-RAG, operating in the latent space, achieves strong privacy protection and fast inference without additional training. Our future research avoiding iterative procedures, such as universal adversarial perturbation, can significantly reduce inference time.

V. CONCLUSION

This paper proposes a method for privacy-preserving person re-ID that augments identity-guided latent codes to induce visual distortion, while preserving re-ID performance via latent manipulation. Our key idea is to decouple the protected output from the original input by retrieving identity-aligned latent candidates from a secure gallery and integrating them through inverse self-attention, which amplifies dissimilar components to suppress structural resemblance. To maintain authorized matching utility, we apply lightweight gradient-based updates in latent space to preserve identity embeddings while manipulating hierarchical latent codes to reduce appearance leakage. Experiments on Market-1501, MSMT17, and CUHK03 show a strong privacy–utility trade-off: the method lowers visual similarity to the originals, sustains competitive Rank-1 and mAP across multiple backbones, and improves robustness against recovery-based attackers. While practical with a pretrained generator and retrieval store, it still incurs iterative latent optimization cost; future work will target faster inference and more efficient retrieval/update policies.

REFERENCES

[1] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 6, pp. 2872–2893, 2021.

[2] T. Li and L. Lin, "Anonymousnet: Natural face de-identification with measurable privacy," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.

[3] A. Alshaibani and A. J. Quinn, "Pterodactyl: Two-step redaction of images for robust face deidentification," in *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 9, 2021, pp. 27–34.

[4] L. Rocher, J. M. Hendrickx, and Y.-A. De Montjoye, "Estimating the success of re-identifications in incomplete datasets using generative models," *Nature communications*, vol. 10, no. 1, p. 3069, 2019.

[5] J. Dietlmeier, J. Antony, K. McGuinness, and N. E. O'Connor, "How important are faces for person re-identification?" in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 6912–6919.

[6] Y. Wang, J. Liu, M. Luo, L. Yang, and L. Wang, "Privacy-preserving face recognition in the frequency domain," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 3, 2022, pp. 2558–2566.

[7] J. Zhang, M. Ye, and Y. Yang, "Learnable privacy-preserving anonymization for pedestrian images," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 7300–7308.

[8] B. Zhao, Y. Li, X. Liu, H. H. Pang, and R. H. Deng, "Freed: An efficient privacy-preserving solution for person re-identification," in *2022 IEEE Conference on Dependable and Secure Computing (DSC)*. IEEE, 2022, pp. 1–8.

[9] Z. Su, D. Zhou, N. Wang, D. Liu, Z. Wang, and X. Gao, "Hiding visual information via obfuscating adversarial perturbations," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4356–4366.

[10] B. Zhao, Y. Li, X. Liu, X. Li, H. H. Pang, and R. H. Deng, "Identifiable, but not visible: A privacy-preserving person reidentification scheme," *IEEE Transactions on Reliability*, vol. 72, no. 4, pp. 1295–1307, 2023.

[11] K. Kansal, Y. Wong, and M. Kankanhalli, "Privacy-enhancing person re-identification framework-a dual-stage approach," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 8543–8552.

[12] Y. Mi, Z. Zhong, Y. Huang, J. Ji, J. Xu, J. Wang, S. Wang, S. Ding, and S. Zhou, "Privacy-preserving face recognition using trainable feature subtraction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 297–307.

[13] M. Ye, W. Shen, J. Zhang, Y. Yang, and B. Du, "Securereid: Privacy-preserving anonymization for person re-identification," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 2840–2853, 2024.

[14] D. Zhang, Y.-X. Peng, X.-M. Wu, A. Wu, and W.-S. Zheng, "Pixelfade: Privacy-preserving person re-identification with noise-guided progressive replacement," in *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024, pp. 6326–6334.

[15] M. Munaro, S. Ghidoni, D. T. Dizmen, and E. Menegatti, "A feature-based approach to people re-identification using skeleton keypoints," in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 5644–5651.

[16] T. Wengefeld, M. Eisenbach, T. Q. Trinh, and H.-M. Gross, "May i be your personal coach? bringing together person tracking and visual re-identification on a mobile robot," in *Proceedings of ISR 2016: 47st International Symposium on Robotics*. VDE, 2016, pp. 1–8.

[17] M. Marras, P. A. Marín-Reyes, J. J. Lorenzo Navarro, M. F. Castrillón Santana, and G. Fenu, "Averobot: an audio-visual dataset for people re-identification and verification in human-robot interaction," *ICPRAM (Setúbal)*, 2019.

[18] Y. Wang, J. Shen, S. Petridis, and M. Pantic, "A real-time and unsupervised face re-identification system for human-robot interaction," *Pattern Recognition Letters*, vol. 128, pp. 559–568, 2019.

[19] S. Coşar and N. Bellotto, "Human re-identification with a robot thermal camera using entropy-based sampling," *Journal of Intelligent & Robotic Systems*, vol. 98, no. 1, pp. 85–102, 2020.

[20] K. De Langis and J. Sattar, "Realtime multi-diver tracking and re-identification for underwater human-robot collaboration," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 11 140–11 146.

[21] I. Condés, J. Fernández-Conde, E. Perdices, and J. M. Cañas, "Robust person identification and following in a mobile robot based on deep learning and optical tracking," *Electronics*, vol. 12, no. 21, p. 4424, 2023.

[22] M. Srouji, Y.-H. H. Tsai, H. Thomas, and J. Zhang, "Human following in mobile platforms with person re-identification," *arXiv preprint arXiv:2309.12479*, 2023.

[23] S. Banerjee, A. Kumar, and A. Shekhar, "Indoor surveillance robot with person following and re-identification," in *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2024, pp. 1–8.

[24] P. A. Carlsen, A. M. Taylor, D. M. Chan, M. Z. Uddin, L. Riek, and J. Torresen, "Real-time person re-identification to improve human-robot interaction," in *2024 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. IEEE, 2024, pp. 425–430.

[25] F. Rollo, A. Zunino, N. Tsagarakis, E. M. Hoffman, and A. Ajoudani, "Continuous adaptation in person re-identification for robotic assistance," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 425–431.

[26] H. Ye, J. Zhao, Y. Zhan, W. Chen, L. He, and H. Zhang, "Person re-identification for robot person following with online continual learning," *IEEE Robotics and Automation Letters*, 2024.

[27] L. Li, A. Bayuelo, L. Bobadilla, T. Alam, and D. A. Shell, "Coordinated multi-robot planning while preserving individual privacy," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 2188–2194.

[28] S. Eick and A. I. Antón, "Enhancing privacy in robotics via judicious sensor selection," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 7156–7165.

[29] M. Li, W. Ding, and D. Zhao, "Privacy risks in reinforcement learning for household robots," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 5148–5154.

[30] E. Lee, M. S. Lee, J. H. Yoon, and S. B. Yoo, "Intenspure: Attack intensity-aware secondary domain adaptive diffusion for adversarial purification." in *IJCAI*, 2024, pp. 956–964.

[31] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, and T. Toft, "Privacy-preserving face recognition," in *International symposium on privacy enhancing technologies symposium*. Springer, 2009, pp. 235–253.

[32] A.-R. Sadeghi, T. Schneider, and I. Wehrenberg, "Efficient privacy-preserving face recognition," in *International conference on information security and cryptology*. Springer, 2009, pp. 229–244.

[33] Y. Mi, Y. Huang, J. Ji, M. Zhao, J. Wu, X. Xu, S. Ding, and S. Zhou, "Privacy-preserving face recognition using random frequency components," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19 673–19 684.

[34] K. Lucas, M. Jagielski, F. Tramèr, L. Bauer, and N. Carlini, "Randomness in ml defenses helps persistent attackers and hinders evaluators," *CoRR*, vol. abs/2302.13464, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2302.13464

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[36] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[37] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739.

[38] O. Tov, Y. Alaluf, Y. Nitzan, O. Patashnik, and D. Cohen-Or, "Designing an encoder for stylegan image manipulation," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 4, pp. 1–14, 2021.

[39] I. Lee, E. Lee, and S. B. Yoo, "Latent-ofer: Detect, mask, and reconstruct with latent vectors for occluded facial expression recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 1536–1546.

[40] E. Lee, J. H. Yoon, and S. B. Yoo, "Scol: Style code orchestration in latent space for proactive face-swapping defense," in *Proceedings of the 33rd ACM International Conference on Multimedia*, 2025, pp. 11 472–11 481.

[41] M. Douze, A. Guzhva, C. Deng, J. Johnson, G. Szilvasy, P.-E. Mazaré, M. Lomeli, L. Hosseini, and H. Jégou, "The faiss library," *arXiv preprint arXiv:2401.08281*, 2024.

[42] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.

[43] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," *Advances in neural information processing systems*, vol. 34, pp. 852–863, 2021.

[44] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1116–1124.

[45] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer gan to bridge domain gap for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 79–88.

[46] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 152–159.

[47] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, "Bag of tricks and a strong baseline for deep person re-identification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 4321–4329.

[48] T. Chen, S. Ding, J. Xie, Y. Yuan, W. Chen, Y. Yang, Z. Ren, and Z. Wang, "Abd-net: Attentive but diverse person re-identification," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 8351–8361.

[49] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.