

Supplementary Materials: PixelFade: Privacy-preserving Person Re-identification with Noise-guided Progressive Replacement

Anonymous Authors

A HUMAN EVALUATION RESULTS

In this section, we follow [6] to perform human evaluation experiments to test whether an adversary can distinguish the identity of a protected pedestrian image by the human eye. We ask participants whether the image pair represents the same person, where the image pair is randomly sampled from the original or protected Market-1501 test set. For each method, we sample 100 image pairs which are evenly distributed over 10 participants. Optimal privacy is achieved when the privacy value reaches 50%, which corresponds to random guessing. As shown in Table S1, our PixelFade achieves the lowest privacy value, which is quite close to the result of random guessing, demonstrating that our method robustly protects the visual privacy of pedestrian images. At the same time, our method obtains the highest Rank-1, indicating that the protected images still provide high utility.

B FURTHER COMPARISON WITH AVIH

In our main paper, we reproduce and set the maximum number of steps of AVIH [4] to 100, which is the same as the default setting of our PixelFade for a fair comparison. The default number of steps for the AVIH method is 800. To demonstrate a comprehensive comparison, we also set different iteration steps up to 800 for both methods and show the trend of the Re-ID performance. Here we set the feature constraint ϵ of PixelFade to 0.01. As shown in Figure S1, Our PixelFade converges faster than AVIH since PixelFade utilizes coarse-grained information during the optimization process to suggest a better optimization direction for the Re-ID model.

Moreover, when the iteration number is set to 800, our method exceeds AVIH in both mAP and mNIP, being very close to “origin” which indicates the performance of the unprotected model. The reason AVIH achieves a lower mNIP and mAP is that AVIH only reduces the distance between the protected image and a specific pedestrian image. This approach results in “overfitting” to that particular image, consequently hindering its ability to match pedestrian samples from different viewpoints. In contrast, our PixelFade effectively captures identity-relative intrinsic features of pedestrian images and ensures their retention in the protected image, thereby achieving relatively high mNIP and mAP.

C CALCULATION OF AD VALUE

To measure the pixel chaos degree of an protected image, we follow [3] to perform an Anderson-Darling test, calculating how similar the image is to a normally distributed noise image. Specifically, for each protected image, we sample a random noise image from the standard normal distribution with the same shape as the protected image. After normalizing both protected image and noise image to a data range between 0 and 1, we use the “anderson” function from “scipy.stat” library to calculate the corresponding statistic (AD value), which indicates how closely the protected image’s pixel values conform to a normal distribution. As the AD value approaches

Table S1: Human evaluation results of different privacy-preserving person re-identification (PPPR) methods. Privacy Value(%) indicates verification accuracy by human eyes. A lower privacy value means better protection of visual privacy.

Image Pair Method	Original-Protected		Protected-Protected	
	Privacy Value↓	Re-ID Rank-1↑	Privacy Value↓	Re-ID Rank-1↑
Blur	79	40.1	82	67.3
Mosaic	71	75.3	75	64.3
PrivacyReID [6]	83	88.2	82	89.2
AVIH [4]	56	92.6	51	91.2
Ours	55	95.0	51	94.2

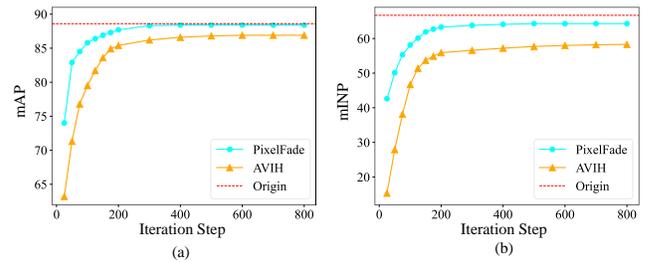


Figure S1: Further comparison between iterative method AVIH [4] and our PixelFade. (a) Comparison of mAP. (b) Comparison of mNIP.

0, it represents increasingly chaotic pixels in the protected image. In our paper, we computed the AD value for each image in the protected test set for every PPPR method. Subsequently, we calculated the average AD value for all images to determine the final AD value for each PPPR method.

D MORE VISUALIZATION RESULTS

We provide more visualization results in Figure S2. Previous methods (columns a-d) still expose some visual information (e.g., clothing color, contour). In comparison, our PixelFade (column e) effectively hides the visual information of pedestrians, making it difficult for malicious attackers to distinguish the identity.

E DISCUSSION OF THE REVERSIBILITY

Although the related work PrivacyReID [6] claims that pedestrian protected images should be reversible, meaning that they can be recovered to original images by authorized models to retain the utility of original images. However, such reversibility gives the adversary an opportunity to launch recovery attacks [1, 2, 5, 7] to invade privacy. Furthermore, preserving the reversibility of protected images is not necessary. It is feasible to store corresponding original images of protected images in another secure location to

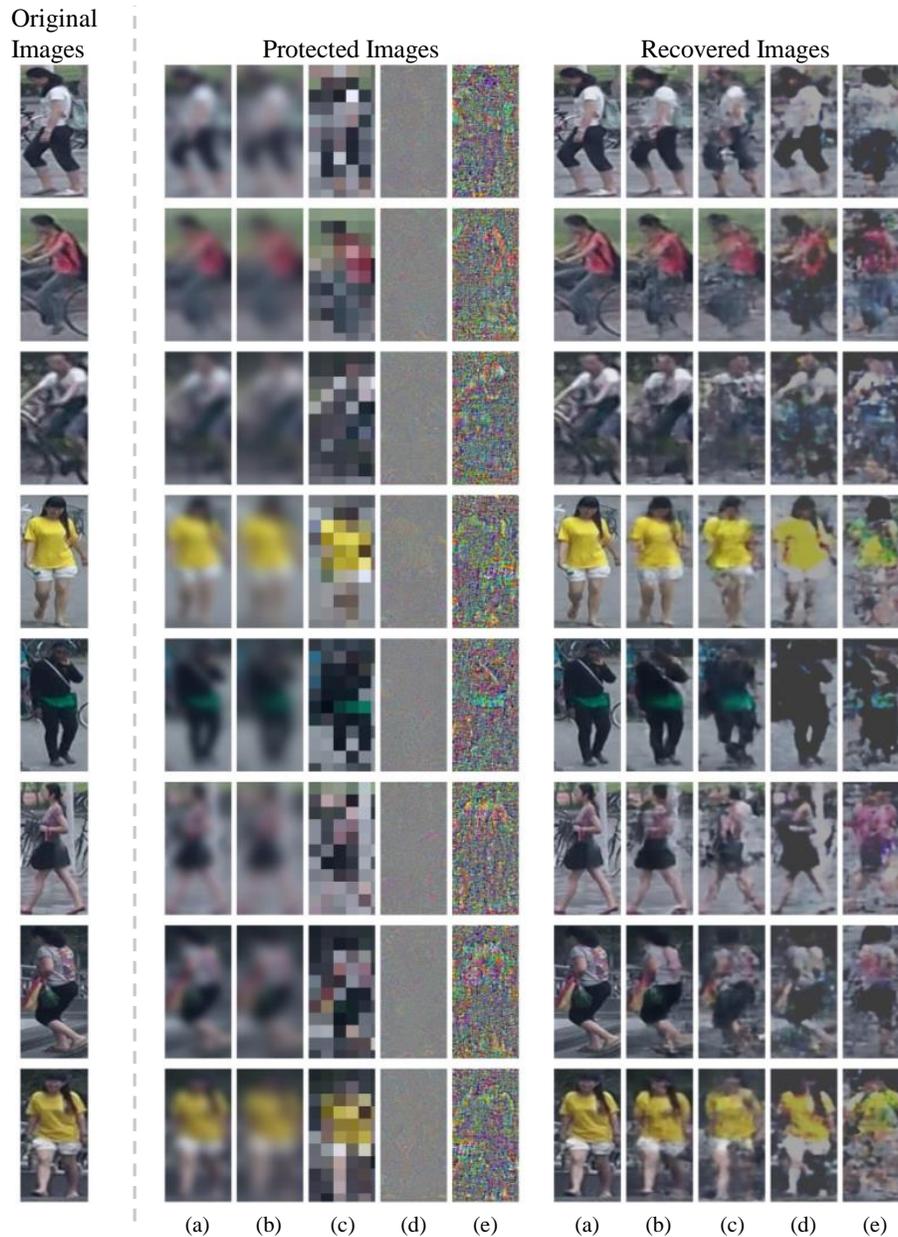


Figure S2: More Visualization of protected and recovered images from different PPPR methods. (a) PrivacyReID [6]; (b) Blurring; (c) Mosaic; (d) AVIH [4]; (e) Our PixelFade.

fully utilize their utility. Next, we give a concrete example of the application of our PixelFade.

In the scenario of tracking a suspect, images captured by the surveillance can be processed by PixelFade into two copies, original and protected images, each sharing a common *index* number. Original image can be stored locally at the trusted site, while protected image can be stored in the cloud, where Re-ID services are provided by an untrusted third party. When there is a need to track suspects across different cameras, the third party can use Re-ID

models to retrieve protected images. Then, they return the *indexes* of the retrieved images to the trusted local site, which uses these *indexes* to get the corresponding original images for further crime investigation.

In summary, our PixelFade provides a technique to protect pedestrian images and resist recovery attacks, which does not contradict the utility of original images. We hope that our novel approach will advance the development of PPPR task.

REFERENCES

- [1] Zecheng He, Tianwei Zhang, and Ruby B Lee. 2019. Model inversion attacks against collaborative inference. In *Proceedings of the 35th Annual Computer Security Applications Conference*. 148–162.
- [2] Guangcan Mai, Kai Cao, Pong C Yuen, and Anil K Jain. 2018. On the reconstruction of face images from deep face templates. *IEEE transactions on pattern analysis and machine intelligence* 41, 5 (2018), 1188–1202.
- [3] Nornadiah Mohd Razali, Yap Bee Wah, et al. 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics* 2, 1 (2011), 21–33.
- [4] Zhigang Su, Dawei Zhou, Nannan Wang, Decheng Liu, Zhen Wang, and Xinbo Gao. 2023. Hiding visual information via obfuscating adversarial perturbations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4356–4366.
- [5] Mang Ye, Wei Shen, Junwu Zhang, Yao Yang, and Bo Du. 2024. Securereid: Privacy-preserving anonymization for person re-identification. *IEEE Transactions on Information Forensics and Security* (2024).
- [6] Junwu Zhang, Mang Ye, and Yao Yang. 2022. Learnable privacy-preserving anonymization for pedestrian images. In *Proceedings of the 30th ACM International Conference on Multimedia*. 7300–7308.
- [7] Andrey Zhmoginov and Mark Sandler. 2016. Inverting face embeddings with convolutional neural networks. *arXiv preprint arXiv:1606.04189* (2016).