SANER: Switchable Adapter with Non-parametric Enhanced Routing for Person De-Reidentification (Supplemental Material)

Yimin Liu[1,4,5,6] Nan Pu[1*] Fengxiang Yang[2] Wenjing Li[1*] Zhihui Li[3] Zhun Zhong[1]
[1]School of Computer Science and Information Engineering, Hefei University of Technology, China
[2]vivo BlueImage Lab, vivo Mobile Communication Co., Ltd.
[3]School of Information Science and Technology, University of Science and Technology of China
[4]Jianghuai Advance Technology Center; [5]Anhui Provincial Key Laboratory of Humanoid Robots
[6]Anhui Provincial Industry Innovation Center of Humanoid Robots

Appendix Contents

A. Joint Optimization

Our proposed SANER is optimized using two complementary loss functions, the forgetting loss and the retaining loss, as described in Section 3.1 of the main text, which were originally introduced by [6]. These losses respectively optimize the F-Adapter and R-Adapter, ensuring that training samples are mapped into a decoupled feature space. The forgetting loss $\mathcal{L}_f$ guides the model to erase identity-specific information of forgotten persons. Specifically, the forgetting loss $\mathcal{L}_f$ is defined as:

$$\mathcal{L}_f = \mathcal{L}_p + \mathcal{L}_{TRC}, \tag{10}$$

where $\mathcal{L}_p$ is the view-specific loss that encourages the model to push apart the features of the same forgotten identity under different augmentations:

$$\mathcal{L}_p = \left[\sigma_c - \|f(x_i^t) - f(\hat{x}_i^t)\|_2^2\right]_+, \tag{11}$$

where x_i^t denotes each image of forgotten identities, $\hat{x}_i^t = T(x_i^t)$ denotes the corresponding augmented view generated using a data augmentation function $T(\cdot)$ [1], $[a]_+ = \max(a, 0)$, and σ_c is a margin parameter. $\mathcal{L}_{TRC}$ is the triplet relation constraint loss, which further separates forgotten identities from accessible ones. It is formulated as:

$$\mathcal{L}_{TRC} = \left[\|x_i^t - x_{i,K}^r\|_2^2 - \|x_i^t - x_h^t\|_2^2 + \sigma_r\right]_+, \tag{12}$$

where $x_{i,K}^r \in \mathcal{S}_R$ is the K-th nearest accessible image to x_i^t, x_h^t is another image of the same forgotten person, and σ_r is a margin. This constraint explicitly enforces the feature representations of forgotten and retaining identities to be pushed apart, enhancing the forgetting effect.

The retaining loss $\mathcal{L}_r$ helps preserve the discriminative consistency of accessible identities, it is defined as:

$$\mathcal{L}_r = \mathcal{L}_o + \mathcal{L}_{RCR}. \tag{13}$$

where $\mathcal{L}_o$ encourages the model to pull together the features of accessible identity and their corresponding augmented views. The loss is formulated as:

$$\mathcal{L}_o = \|f(x_j^r) - f(\hat{x}_j^r)\|_2^2, \tag{14}$$

where $x_j^r \in \mathcal{S}_R$ denotes an accessible identity image, an augmented view $\hat{x}_j^r = T(x_j^r)$ is generated using a data augmentation function $T(\cdot)$. To preserve discriminative relationships among accessible identities, the relation consistent regularization loss $\mathcal{L}_{RCR}$ is defined as:

$$\mathcal{L}_{RCR} = \|\text{sim}(f_p(x_j^r), f_p(x_h^r)) - \text{sim}(f(x_j^r), f(x_h^r))\|_2^2, \tag{15}$$

where $f_p(\cdot)$ is the pretrained model and $\text{sim}(\cdot,\cdot)$ denotes similarity in the embedding space. The whole procedure is shown in Algorithm 1.

*Corresponding author

Algorithm 1 Training and Testing Pipeline of SANER

Require: Training sets $\mathcal{S}_T$ (forgetting) and $\mathcal{S}_R$ (retaining); Frozen backbone $f_p(\cdot)$; F-Adapter and R-Adapter; Data augmentation $T(\cdot)$; Hyper-parameters.

procedure TRAINING
  for each mini-batch (x, y) **do**
    Assign routing variable $R \in \{0, 1\}$ for each sample
    Extract frozen features and augmented features using $f_p(\cdot)$
    Generate decoupled features $\mathbf{f}_f$ and $\mathbf{f}_r$ via our designed **Switchable Adapter** (SA).
    Forgetting branch ($R = 1$)**:**
    Compute $\mathcal{L}_p$ and $\mathcal{L}_{TRC}$ using Eq. 11 and Eq. 12
    Compute forgetting loss $\mathcal{L}_f$ using Eq. 10
    Retaining branch ($R = 0$)**:**
    Compute $\mathcal{L}_o$ and $\mathcal{L}_{RCR}$ using Eq. 14 and Eq. 15
    Compute retaining loss $\mathcal{L}_r$ using Eq. 13
    Compute joint loss $\mathcal{L} = \mathcal{L}_r + \mathcal{L}_f$
    Update adapter parameters (F-Adapter / R-Adapter) by gradient descent
  end for
end procedure

procedure TESTING
  Extract frozen backbone feature $\mathbf{f} = f_p(x_q)$
  Use **Non-parametric Enhanced Routing** (NER) to calculate prototype similarity and make routing decisions
  if route = retain **then**
    $\mathbf{f}_q = \mathbf{f} + \Delta\mathbf{f}_r$
  else
    $\mathbf{f}_q = \mathbf{f} + \Delta\mathbf{f}_f$
  end if
  Retrieve with mapped query feature $\mathbf{f}_q$
end procedure

B. More Implementation Details

We use ViT-B [2] as the backbone for both $f_p(\cdot)$ and $f(\cdot)$. The $f_p(\cdot)$ is trained on the pretraining subset of each dataset using cross-entropy and triplet losses, following PASS [9]. For De-ReID, we initialize $f(\cdot)$ with $f_p(\cdot)$ and fine-tune it using our proposed Switchable Adapter modules. These adapters are inserted into the feed-forward network (FFN) layers and the query and value projections of the multi-head attention module in the last 6 Transformer blocks. The Switchable Adapter rank is set to 8 for the query and value projections and 16 for the FFN layers. In our experiments, the method uses approximately 11GB of GPU memory on an NVIDIA RTX 4090.

C. More Experimental Results

In this section, we provide additional experiments to further analyze the effectiveness, robustness, and scalability of the proposed method.

C.1. Image Restoration from Forgotten Identity Prototypes

To assess the privacy leakage risk, we attempt to reconstruct images from the forgotten identity prototypes using D2D [8]. As shown in Fig. A(a), the reconstructed images are visually blurry and lack clear identity information, indicating that the proposed method effectively suppresses identity-specific features and provides a certain level of privacy preservation.

C.2. Impact of Routing Errors

We analyze the sensitivity of the proposed method to routing errors by intentionally varying the routing error rate in the non-parametric enhanced routing (NER) module. As shown in Fig. A(b), increasing the routing error rate leads to consistent performance degradation, demonstrating that accurate routing is critical for maintaining strong De-ReID performance. In practice, our NER achieves a low routing error rate of 2.3% on Market-1501, ensuring reliable behavior in real-world scenarios.

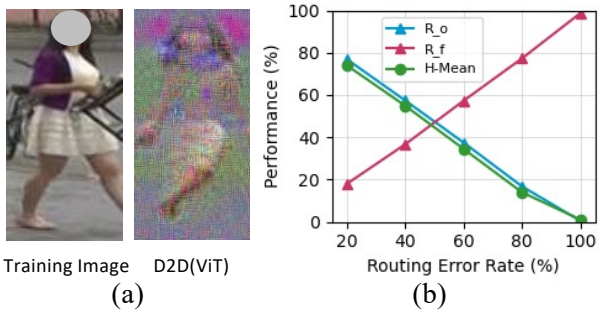

Figure A. (a) Image restoration from forgotten identity prototypes via D2D. (b) Impact of routing error rates on Market-1501.

C.3. Results with CNN-based Backbones

To verify the generalization ability across architectures, we further evaluate the proposed method using a ResNet50 backbone. As shown in Tab. A, our method consistently outperforms the baseline VIS under different settings, demonstrating its effectiveness beyond transformer-based models.

Table A. Results of CNN-based backbones (ResNet50) on Market.

Method	$M_T = 25$			$M_T = 50$		
	R-1$_T$ ↓	R-1$_O$ ↑	H ↑	R-1$_T$ ↓	R-1$_O$ ↑	H ↑
$f_p(\cdot)$	76.0	90.3	-	81.3	90.4	-
VIS [6]	32.0	63.1	51.8	40.0	58.3	48.3
Ours	**16.0**	**68.3**	**63.9**	**10.2**	**69.4**	**70.2**

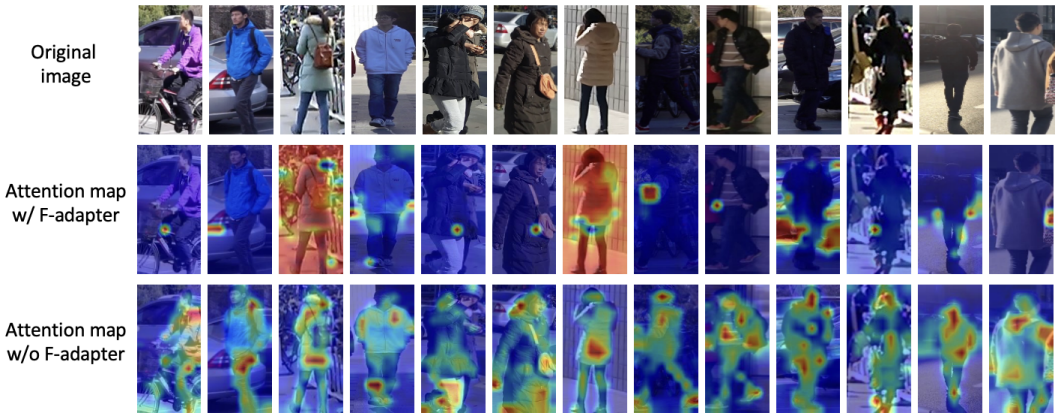

Figure B. Extended attention maps for forgetting identities on MSMT17 [7]. Each column shows one sample per identity, with rows displaying the original image, attention map with F-Adapter, and attention map without F-Adapter.

C.4. Comparison with Different PEFT Methods

We further evaluate our method with different parameter-efficient fine-tuning (PEFT) techniques, including LoRA [3], DoRA [4], and PiSSA [5]. As shown in Tab. B, all variants achieve consistently strong De-ReID performance, demonstrating the robustness and general applicability of our framework.

Table B. Comparison with different PEFT methods on Market.

Method	$M_T = 25$			$M_T = 50$		
	R-1$_T$ ↓	R-1$_O$ ↑	H ↑	R-1$_T$ ↓	R-1$_O$ ↑	H ↑
Ours w/ LoRA	7.3	95.7	88.3	6.7	95.3	91.1
Ours w/ DoRA	7.8	96.1	88.2	10.7	96.0	89.2
Ours w/ PiSSA	7.0	96.1	88.6	6.5	95.7	91.4

C.5. Computational Cost Analysis

We further analyze the computational cost of the proposed method on Market-1501. As shown in Tab. C, our method introduces only a small number of additional parameters and maintains comparable training time to VIS, while incurring a modest inference overhead. Despite this, it achieves a significant improvement in H-Mean, demonstrating a favorable trade-off between efficiency and performance.

Table C. Parameter efficiency, prototype overhead, and runtime comparison.

Method	#Params	Train Time	Mem. (MB)	Inf. Time (/img)	H
Full FT	86.5M	3h 02m	0	1.86 ms	56.7
VIS [6]	0.52M	2h 20m	0	1.86 ms	83.2
Ours	1.04M	1h 50m	0.146	2.53 ms	**91.1**

C.6. Scalability with Large Numbers of Forgotten Identities

To evaluate scalability, we increase the number of forgotten identities to 200 and 400. As shown in Tab. D, the proposed method maintains strong De-ReID performance even with a large number of forgotten identities, demonstrating its robustness and scalability.

Table D. Impact of the number of forgotten identities.

Method	$M_T = 200$			$M_T = 400$		
	R-1$_T$ ↓	R-1$_O$ ↑	H ↑	R-1$_T$ ↓	R-1$_O$ ↑	H ↑
Ours	1.0	98.3	97.8	1.3	98.4	97.7

D. Additional Attention Map Examples

We provide additional visualizations of attention maps for identities to be forgotten in the MSMT17 dataset. Compared with the examples shown in the main text, this extended figure includes more samples per identity, allowing a more comprehensive examination of the model's attention patterns. Each column shows different images of the same person, and the three rows display the original image, the attention map generated with the F-Adapter, and the attention map after removing the F-Adapter. Brighter regions indicate areas where the model focuses its attention.

The visualizations show that when the F-Adapter is applied, the model's attention tends to concentrate on identity-irrelevant regions or is more widely dispersed across the image. This indicates that the F-Adapter effectively guides the model to suppress identity-specific features, achieving the forgetting objective. Conversely, when the F-Adapter is removed and only the R-Adapter remains, the attention clearly focuses on identity-relevant regions, supporting improved re-identification performance.

Further analysis demonstrates the complementary roles of the F-Adapter and R-Adapter in decoupling features. The F-Adapter weakens identity information, enabling selective forgetting of specific identities, while the R-Adapter ensures that key identity features are preserved for correct re-identification. This controllable attention distribution shows that the model can reduce the influence of sensitive information in forgetting tasks while maintaining discriminative ability in retaining tasks, highlighting the effectiveness of the proposed adapter mechanism in multi-task optimization.

References

[1] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 702–703, 2020. 1

[2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021. 2

[3] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022. 3

[4] Shih-Yang Liu, Chien-Yi Wang, Hongxu Yin, Pavlo Molchanov, Yu-Chiang Frank Wang, Kwang-Ting Cheng, and Min-Hung Chen. Dora: Weight-decomposed low-rank adaptation. In *International Conference on Machine Learning (ICML)*, pages 32100–32121, 2024. 3

[5] Fanxu Meng, Zhaohui Wang, and Muhan Zhang. Pissa: Principal singular values and singular vectors adaptation of large language models. In *Advances in Neural Information Processing Systems*, 2024. 3

[6] Yixing Peng, Yuming Tang, Kunyu Lin, Qize Yang, Jingke Meng, Xihan Wei, and Weishi Zheng. Person de-reidentification: A variation-guided identity shift modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 29331–29341, 2025. 1, 2, 3

[7] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 79–88, 2018. 3

[8] Hongxu Yin, Pavlo Molchanov, Zhizhong Li, Jose M. Alvarez, Arun Mallya, Derek Hoiem, and Niraj K. Jha. Dreaming to distill: Data-free knowledge transfer via deepinversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8715–8724, 2020. 2

[9] Kuan Zhu, Haiyun Guo, Tianyi Yan, Yousong Zhu, Jinqiao Wang, and Ming Tang. Pass: Part-aware self-supervised pre-training for person re-identification. In *European Conference on Computer Vision*, pages 198–214. Springer, 2022. 2