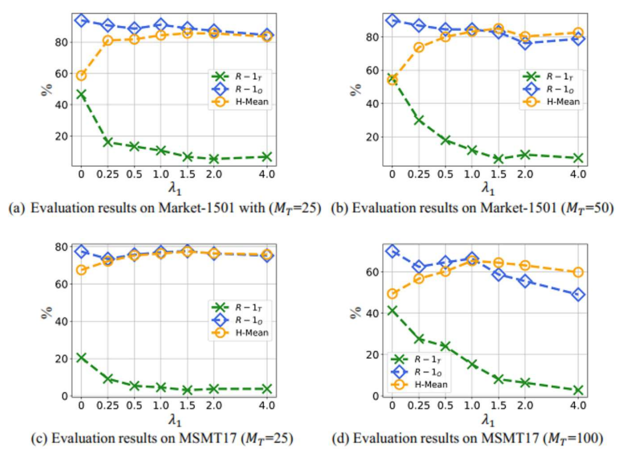
Person De-reidentification: A Variation-guided Identity Shift Modeling

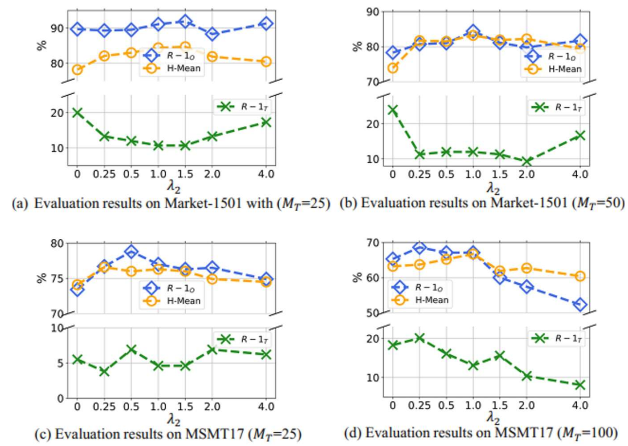
Supplementary Material



(a) Evaluation results on Market-1501 with (M_T=25) (b) Evaluation results on Market-1501 (M_T=50)

(c) Evaluation results on MSMT17 (M_T=25) (d) Evaluation results on MSMT17 (M_T=100)

Figure A1. Evaluation under different λ_1.



(a) Evaluation results on Market-1501 with (M_T=25) (b) Evaluation results on Market-1501 (M_T=50)

(c) Evaluation results on MSMT17 (M_T=25) (d) Evaluation results on MSMT17 (M_T=100)

Figure A2. Evaluation under different λ_2.

Our Appendix includes Sec. A1 for more experimental results in MSMT17 and Market-1501, Sec. A2 for more implementation details, Sec. A3 for dataset statistics and Sec. A4 for experimental results in Occ-Duke and SYSU-MM01.

A1. More Experimental Results in MSMT17 and Market-1501

We report the ablation study results in the cases of $M_T \in \{50, 75\}$ in MSMT17 dataset in Table A1. As analyzed in our main manuscript, each component of our method is effective and clearly improves the H-Mean.

More evaluation results using different trade-off parameters. Besides, we report the detailed performance of our method under different trade-off parameters λ_1 and λ_2 in Fig. A1 and Fig. A2. λ_1 controls the importance of the proposed variation-based unlearning, and λ_2 controls the importance of relation regularization. In the range of $[1.0, 2.0]$. the two trade-off parameters work effectively in different M_T and datasets when they are in the range of $[1.0, 2.0]$.

Ablation study on data augmentation function $\mathcal{T}$. In our experiments, to introduce abundant variations for our variation-guided identity shift, our data augmentation includes random crop, random flip, and *AutoContrast*, *Brightness*, *Color*, *Contrast* as well as *Equalize* using in RandAug [1]. Notably, apart from the five augmentations mentioned above, the augmentation pool of RandAug contains more augmentations such as *Posterize*. However, we empirically choose the augmentations that can properly simulate the natural cross-view variations. We ablation the effectiveness of different augmentations in Table A2.

A2. More Implementation Details

We choose ViT-B [2] as our backbone of $f_p(\cdot)$ and $f(\cdot)$. The $f_p(\cdot)$ is trained in the pretraining subset of each dataset using cross-entropy loss and triplet loss following PASS [5]. When learning De-ReID, we initialize $f(\cdot)$ as $f_p(\cdot)$. We adopt LoRA modules [3] for fine-tuning $f(\cdot)$. We insert LoRA modules into the FFN layers, query, and value projection in the multi-head attention module in the last 6 transformer blocks of $f(\cdot)$. The LoRA rank is set to 8 for query and value projection and is set to 16 for FFN layers. In experiments, our method requires about 10GB GPU memory and costs less than 3 hours.

Table A1. Ablation study in MSMT17. "W/o" means "without", and other notations are the same as Table 2 in the main manuscript. $f_p(\cdot)$ is the initial model for De-ReID learning. The full model achieves the best H-Mean. Notably, without L_{RCR}, the model forgets ReID knowledge and collapses. "SD" denotes the "self-augmented discrimination". Please refer to the text in Sec.4.4 in the main manuscript for more details.

Method	MSMT17					
	$M_T = 50$			$M_T = 75$		
	R-1_T	R-1_O	H-Mean	R-1_T	R-1_O	H-Mean
$f_p(\cdot)$	77.6	85.4	—	79.7	85.4	—
Components in Relation regularization						
W/o L_{TRC}	12.9	67.5	66.1	6.9	59.1	65.2
W/o L_{RCR}	—	—	—	—	—	—
Components in Variation-guided Identity Shift						
W/o SD	6.2	65.5	68.3	7.4	58.6	64.7
W/o L^o_{VIS}	14.4	65.6	64.4	17.4	69.7	65.8
W/o L_{VIS}	25.6	68.3	59.0	30.4	72.1	58.6
Full model	**10.8**	**72.9**	**69.7**	**12.4**	**69.8**	**68.5**

Table A2. Ablation study on the data augmentation function. “RC” means random crop and “RF” means the horizontal random flip. Our data augmentation is derived from the RandAug including *AutoContrast*, *Brightness*, *Color*, *Contrast* and *Equalize*.

Augmentation	Market-1501						MSMT17					
	$M_T = 25$			$M_T = 50$			$M_T = 25$			$M_T = 100$		
	R-1_T	R-1_O	H	R-1_T	R-1_O	H	R-1_T	R-1_O	H	R-1_T	R-1_O	H
RC, RF	10.7	89.0	83.4	11.3	78.2	80.4	6.2	73.3	73.3	7.9	52.9	60.8
RandAug	**6.7**	**88.8**	**85.6**	13.3	83.6	82.1	3.1	74.1	75.2	8.4	62.5	66.5
Ours	10.7	91.1	84.4	**12.0**	**84.4**	**83.2**	**4.6**	**77.0**	**76.3**	**13.1**	**67.1**	**66.7**

Table A3. Statistics of the datasets in the case of $M_T \in \{50, 75\}$ in MSMT17. Notably, the accessible persons in training are different from those in testing. For the unlearned persons, the images in training and those in testing are from different cameras. We keep the accessible persons unchanged when varying the number of M_T in the same dataset.

	MSMT17							
	M_T	$\lvert S_T\rvert$	M_O	$\lvert S_O\rvert$	M_T	$\lvert S_T\rvert$	M_O	$\lvert S_O\rvert$
Train	50	2100	141	5061	75	3207	141	5061
Query	50	259	2960	10706	75	404	2960	10706
Gallery	50	2095	2960	74165	75	3168	2960	74165

A3. Dataset Statistics

We report the statistics of the dataset in Table A4 and Table A3.

A4. More Experimental Results in Occ-Duke and SYSU-MM01

We report additional results in Table A5 & A6. In briefly, (1) our method significantly outperforms other methods in the datasets, and (2) each component is effective. Since Occ-Duke has a similar size with Market, we evaluate with $M_T \in \{25, 50\}$, where M_T is the number of unlearned persons. Since SYSU-MM01 only has 96 persons in testing set, we set M_T=16 and evaluate our method in all-search mode as in [4].

References

[1] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. Randaugment: Practical automated data augmentation with a reduced search space. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 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*, 2021. 1

[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*, 2022. 1

[4] Ancong Wu, Wei-Shi Zheng, Hong-Xing Yu, Shaogang Gong, and Jianhuang Lai. Rgb-infrared cross-modality person re-identification. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017. 2

[5] Kuan Zhu, Haiyun Guo, Tianyi Yan, Yousong Zhu, Jinqiao Wang, and Ming Tang. Pass: Part-aware self-supervised pre-training for person re-identification. *arXiv preprint arXiv:2203.03931*, 2022. 1

Table A4. Statistics of the datasets in the case of $M_T = 25$ in MSMT17 and $M_T = 25$ in Market1501. M_T is the number of unlearned persons for De-ReID. Notably, the accessible persons in training are different from those in testing. For the unlearned persons, the images in training and those in testing are from different cameras.

	MSMT17				Market-1501			
	M_T	$\|S_T\|$	M_O	$\|S_O\|$	M_T	$\|S_T\|$	M_O	$\|S_O\|$
Train	25	1126	141	5061	25	481	151	2284
Query	25	130	2960	10706	25	75	700	3068
Gallery	25	1064	2960	74165	25	523	700	18033

Method	Occ-Duke						SYSU-MM01		
	$M_T = 25$			$M_T = 50$			$M_T = 16$		
	R-1$_T$ ↓	R-1$_O$ ↑	H ↑	R-1$_T$ ↓	R-1$_O$ ↑	H ↑	R-1$_T$ ↓	R-1$_O$ ↑	H ↑
LabelAug	54.6	57.7	26.4	54.6	57.7	18.5	9.0	33.7	36.2
BS	30.0	44.2	42.9	30.3	35.4	35.3	27.0	49.3	29.6
SCRUB	52.0	63.8	30.1	55.5	63.2	17.4	30.8	38.9	23.9
LIRF	17.2	57.8	56.1	20.2	51.3	48.2	36.6	40.8	17.9
LIRF*	16.8	57.1	55.9	21.3	51.6	47.6	43.8	50.7	7.9
GS-LoRA	30.3	60.2	49.1	32.6	46.4	38.6	37.9	47.6	16.8
Ours	6.1	60.0	62.7	7.8	58.5	58.2	7.5	41.2	40.9

Table A5. Comparison results in the Occ-Duke and SYSU-MM01. The notations are consistent with tables in paper.

Table A6. Ablation study in Occ-Duke and SYSU-MM01.

Method	SYSU-MM01			Occ-Duke		
	$M_T = 16$			$M_T = 50$		
	R-1$_T$ ↓	R-1$_O$ ↑	H ↑	R-1$_T$ ↓	R-1$_O$ ↑	H ↑
$f_p(\cdot)$	48.1	52.6	—	65.6	69.8	—
Components in Relation regularization						
W/o L_{TRC}	9.9	28.4	32.6	10.6	57.6	56.2
W/o L_{RCR}	—	—	—	—	—	—
Components in Variation-guided Identity Shift						
W/o SD	7.4	40.1	40.2	9.0	55.8	56.2
W/o L_{VIS}^o	8.7	37.8	38.6	8.3	56.8	57.0
W/o L_{VIS}	11.0	38.0	37.5	35.3	56.5	39.4
Full model	**7.5**	**41.2**	**40.9**	**7.8**	**58.5**	**58.2**