Distribution-aware Knowledge Prototyping for Non-exemplar Lifelong Person Re-identification

Kunlun Xu[1], Xu Zou[2], Yuxin Peng[1], Jiahuan Zhou[1*]

[1] Wangxuan Institute of Computer Technology, Peking University, Beijing 100871, China

[2] School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China

xkl@stu.pku.edu.cn, zoux@hust.edu.cn, {pengyuxin, jiahuanzhou}@pku.edu.cn

In our supplementary materials, we present an additional set of ablation studies on different hyperparameters. Moreover, we present a comprehensive proof of Eq. (4) and Eq. (5) in the original paper. Additionally, we further introduce the datasets employed in the LReID benchmark with comprehensive details. Finally, we visualize the person retrieval results to show the effectiveness of our model compared to the state-of-the-art method PatchKD.

1. More Ablation on Hyperparameters

1.1. Influence of Temperature Parameter

In our original paper, λ_1 and λ_2 in Eq. (6) and Eq. (7) are the temperature parameters to scale the matrix values. In this section, we conduct experiments to explore the impact of different values for λ_1 and λ_2 on the model performance. When adjusting λ_1 or λ_2, all other hyperparameters are set to the default values. The corresponding results are shown in Fig. 1 (a) and Fig. 1 (b) respectively. We observe that the model achieves the best performance when $\lambda_1 = 0.1$ and $\lambda_2 = 0.1$, which is adopted as the default temperature parameters setting in our original paper. Besides, Fig. 1 (a) illustrates that the model performance is relatively sensitive to λ_1. This sensitivity arises because λ_1 significantly affects the relative coordinates of the new features from the perspective of the sampled prototypes. Specifically, a small λ_1 can cause each coordinate vector $\mathbf{C}_{p,i} \in \mathbb{R}^{N_{t-1}}$, *i.e.*, each row of $\mathbf{C}_p \in \mathbb{R}^{d \times N_{t-1}}$ which is obtained by Eq. (7) in the original paper, to concentrate on a specific dimension, while a large λ_1 can cause the values of each dimension in $\mathbf{C}_{p,i}$ to approach uniformity. Consequently, both directions can result in a loss of the diverse information encoded within different dimensions compared to the optimal value. In contrast, Fig. 1 (b) demonstrates that the model exhibits greater robustness to the changes of λ_2, with only minor performance degradation when deviating from the optimal value. This can be attributed to the inherent alignment between the relative coordinates of the new features and the original new features which reduce the influence of the matrix value distortions caused by different λ_2.

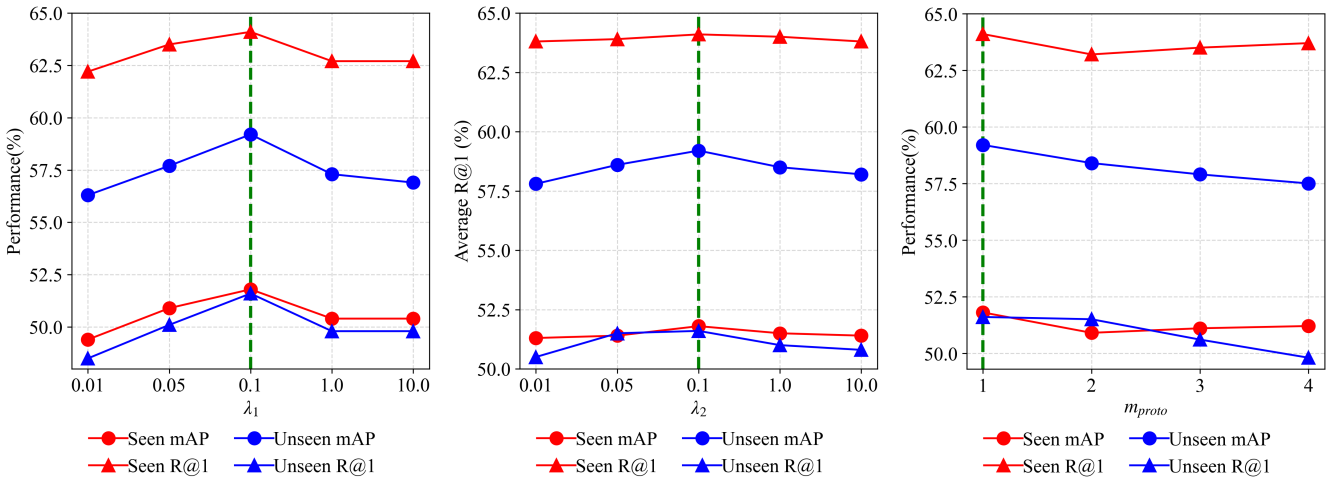
1.2. Influence of Prototype sample number

In the original paper, for Prototype-based Knowledge Transfer (PKT), given a set of prototypes $\mathcal{P}_{t-1} = \{(\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i^2)\}_{i=1}^{N_{t-1}}$, we sample N_{t-1} prototype features $\mathcal{F}_p = \{\boldsymbol{g}_i\}_{i=1}^{N_{t-1}}$. This indicates that each prototype is sampled once. Here, we also provide additional experiments to show the influence of the prototype sampling number on the model performance. For the convenience of expression, we denote the sampling number of each prototype as m_{proto}. Specifically, we generate m_{proto} groups of prototype samples and calculate the Prototype Knowledge Transfer loss for each group according to Equation (8) in the original paper. The average loss of these groups is adopted as $\mathcal{L}_{proto-d}$ to replace the counterpart in the original paper. As shown in Figure 1 (c), we observe that increasing the prototype sampling number has a minor influence on the performance of the seen domains, but it significantly degrades the performance of the unseen domains. This indicates that $m_{proto} = 1$ is sufficient for accomplishing knowledge transfer, and larger values of m_{proto} can cause the model to overfit on the seen domains. Therefore, by default, we only sample each prototype once in the original paper.

2. Proof of Distribution-oriented Prototype Generation

According to the notations in the original paper, for an identity k with n_k instances $\{x_i^k\}_{i=1}^{n_k}$ in the training set (For simplicity, these n_k instances are denoted as $\{x_i\}_{i=1}^{n_k}$), we assume that the distribution of identity k is the mix of the distributions of $\{x_i\}_{i=1}^{n_k}$. Note that the distribution function

*Corresponding author



(a) Temperature parameter λ_1 (b) Temperature parameter λ_2 (c) Prototype sample number m_{proto}

Figure 1. Ablation studies on additional hyperparameters. The values marked by the dashed lines are adopted by our proposed method. m_{proto} denotes the sample number for each prototype.

denoted by $\mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k^2)$ is represented as:

$$\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k) = \frac{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)}}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_k|^{1/2}}, \tag{1}$$

where $\boldsymbol{\Sigma}_k$ is a diagonal matrix with diagonal elements formed as a vector equal to $\boldsymbol{\sigma}_k^2$.

Additionally, given $\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{c}_i, \boldsymbol{v}_i)$ denoted by $\mathcal{N}(\boldsymbol{c}_i, \boldsymbol{v}_i^2)$, we can obtain:

$$\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k) = \frac{1}{n_k}\sum_{i=1}^{n_k} \boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{c}_i, \boldsymbol{v}_i), \tag{2}$$

where $\boldsymbol{\mu}_k$ and $\boldsymbol{\sigma}_k^2$ are the mean and variance of the identity k. $\boldsymbol{c}_i$ and $\boldsymbol{v}_i^2$ are the learned enter and variance of the instance x_i. Besides, according to probability theory [9], we obtain $\boldsymbol{c}_i = \int \boldsymbol{x}\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{c}_i, \boldsymbol{v}_i)d\boldsymbol{x}$ and $\boldsymbol{v}_i^2 = \int \boldsymbol{x}^2\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{c}_i, \boldsymbol{v}_i)d\boldsymbol{x} - (\int \boldsymbol{x}\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{c}_i, \boldsymbol{v}_i)d\boldsymbol{x})^2$. Moreover, the identity distribution center $\boldsymbol{\mu}_k$ could be obtained as follows:

$$\begin{aligned}
\boldsymbol{\mu}_k &= \int \boldsymbol{x}\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k)d\boldsymbol{x} \\
&= \int \frac{1}{n_k}\sum_{i=1}^{n_k} \boldsymbol{x}\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{c}_i, \boldsymbol{v}_i)d\boldsymbol{x} \\
&= \frac{1}{n_k}\sum_{i=1}^{n_k} \int \boldsymbol{x}\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{c}_i, \boldsymbol{v}_i)d\boldsymbol{x} \\
&= \frac{1}{n_k}\sum_{i=1}^{n_k} \boldsymbol{c}_i,
\end{aligned} \tag{3}$$

Similarly, the identity distribution variance $\boldsymbol{\sigma}_k^2$ could be obtained by:

$$\begin{aligned}
\boldsymbol{\sigma}_k^2 &= \int \boldsymbol{x}^2\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k)d\boldsymbol{x} - (\int \boldsymbol{x}\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k)d\boldsymbol{x})^2 \\
&= \int \boldsymbol{x}^2 \frac{1}{n_k}\sum_{i=1}^{n_k} \boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{c}_i, \boldsymbol{v}_i)d\boldsymbol{x} - \boldsymbol{\mu}_k^2 \\
&= \frac{1}{n_k}\sum_{i=1}^{n_k} \int \boldsymbol{x}^2\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{c}_i, \boldsymbol{v}_i)d\boldsymbol{x} - \boldsymbol{\mu}_k^2 \\
&= \frac{1}{n_k}\sum_{i=1}^{n_k}(\boldsymbol{c}_i^2 + \boldsymbol{v}_i^2) - (\frac{1}{n_k}\sum_{i=1}^{n_k} \boldsymbol{c}_i)^2
\end{aligned} \tag{4}$$

3. Datasets Details

We conducted all the experiments on the LReID benchmark [8] which consists of 12 existing ReID datasets, the detailed statistics of which are shown in Table 1. 'Original Identities' denotes the identity numbers in the original datasets and 'LReID Identities' denotes the selected identity numbers in the LReID benchmark. Note that CUHK-SYSU was initially proposed for the person search task and we preprocess it for the ReID task following [8] where the ground-truth person bounding box annotation is used to crop individual-level images and a subset in which each identity contains at least 4 bounding boxes is selected.

4. Visualization of Retrieval Results

In this section, we visualize the person retrieval results of our model and the state-of-the-art PatchKD [11] on different seen and unseen datasets to show the superiority of our method. The results on seen datasets CUHK-SYSU (Fig. 2), MSMT17-V2 (Fig. 3) are visualized to show the acquisition and anti-forgetting capability of our model. Besides, the results on unseen datasets CUHK01 (Fig. 4) and SenseReID (Fig. 5) are visualized to show

Type	Datasets Name	Original Identities			LReID Identities		
		Train	Query	Gallery	Train	Query	Gallery
Seen	CUHK03 [6]	767	700	700	500	700	700
	Market-1501 [15]	751	750	751	500	751	751
	DukeMTMC-ReID [10]	702	702	1110	500	702	1110
	CUHK-SYSU [13]	942	2900	2900	500	2900	2900
	MSMT17-V2 [12]	1041	3060	3060	500	3060	3060
Unseen	i-LIDS [1]	243	60	60	-	60	60
	VIPR [2]	316	316	316	-	316	316
	GRID [7]	125	125	126	-	125	126
	PRID [3]	100	100	649	-	100	649
	CUHK01 [5]	485	486	486	-	486	486
	CUHK02 [4]	1677	239	239	-	239	239
	SenseReID [14]	1718	521	1718	-	521	1718

Table 1. The statistics of datasets used in the LReID benchmark. '-' indicates the datasets only serve as test domains.

the generalization of our model. The images in black represent query images, and the ones in the green and red boxes represent the correct and false retrievals respectively. The retrieval results are arranged in descending order from left to right based on the matching scores.

In many cases with prominent changes in human pose, viewpoint, and scenery, our model could extract reliable features to retrieve more correct matchings. Besides, it often occurs that different persons with similar clothes appear in the datasets, which frequently leads to false retrievals in PatchKD, while our model could utilize more accurate information to measure the affinity for persons in query and gallery sets, thus achieving better performance. These results show that the abundant knowledge consolidated by our method could guide the model to extract more discriminative information.

References

[1] Home Office Scientific Development Branch. Imagery library for intelligent detection systems (i-lids). In *2006 IET Conference on Crime and Security*, pages 445–448. IET, 2006. 3

[2] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, pages 262–275. Springer, 2008. 3

[3] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof. Person re-identification by descriptive and discriminative classification. In *Image Analysis*, pages 91–102. Springer, 2011. 3

[4] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *CVPR*, pages 3594–3601. IEEE, 2013. 3

[5] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *ACCV*, pages 31–44. Springer, 2012. 3

[6] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, pages 152–159. IEEE, 2014. 3

[7] Chen Change Loy, Tao Xiang, and Shaogang Gong. Time-delayed correlation analysis for multi-camera activity understanding. *IJCV*, 90(1):106–129, 2010. 3

[8] Nan Pu, Wei Chen, Yu Liu, Erwin M Bakker, and Michael S Lew. Lifelong person re-identification via adaptive knowledge accumulation. In *CVPR*, pages 7897–7906. IEEE, 2021. 2

[9] Alfréd Rényi. *Probability theory*. Courier Corporation, 2007. 2

[10] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *ECCV*, pages 17–35. Springer, 2016. 3

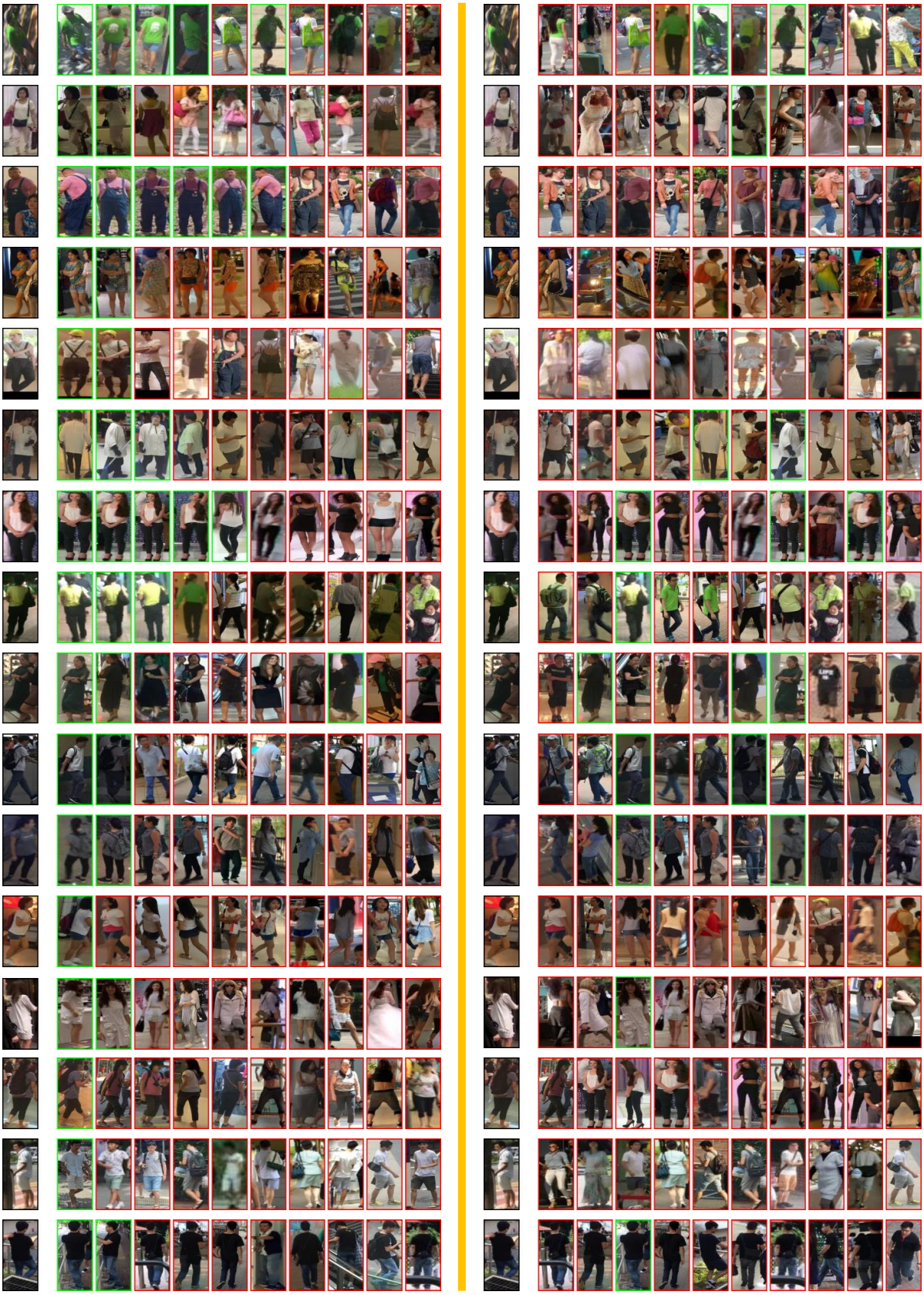
[11] Zhicheng Sun and Yadong Mu. Patch-based knowledge distillation for lifelong person re-identification. In *ACMMM*, pages 696–707, 2022. 2

[12] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *CVPR*, pages 79–88. IEEE, 2018. 3

[13] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. End-to-end deep learning for person search. *arXiv:1604.01850*, 2(2):4, 2016. 3

[14] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *CVPR*, pages 907–915. IEEE, 2017. 3

[15] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *ICCV*, pages 1116–1124. IEEE, 2015. 3



(a) Ours (b) PatchKD

Figure 2. Person retrieval results of on seen domain CUHK-SYSU.

(a) Ours (b) PatchKD

Figure 3. Person retrieval results of on seen domain MSMT17-V2.

(a) Ours (b) PatchKD

Figure 4. Person retrieval results of on seen domain CHUK01.

(a) Ours (b) PatchKD

Figure 5. Person retrieval results of on seen domain SenseReID.