Patch-based Discriminative Feature Learning for UnsupervisedPerson Re-identification

Qize Yang, Hong-Xing Yu, Ancong Wu, and Wei-Shi Zheng

Reference: https://kovenyu.com/papers/2019_CVPR_PEDAL.pdf

Intelligent Information Fusion Research Group

Figure 1. Some image samples of MSMT17 [40] and Market-1501 [53]. It is easier to find that if two images are similar, then their patches would probably also be similar. And the gap of the similar patches would be smaller than the similar images

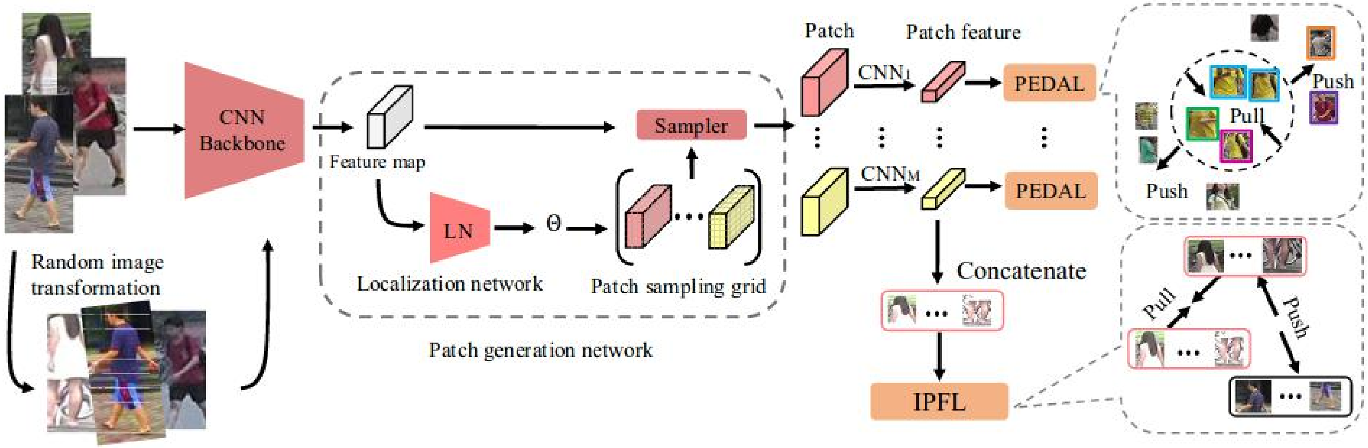
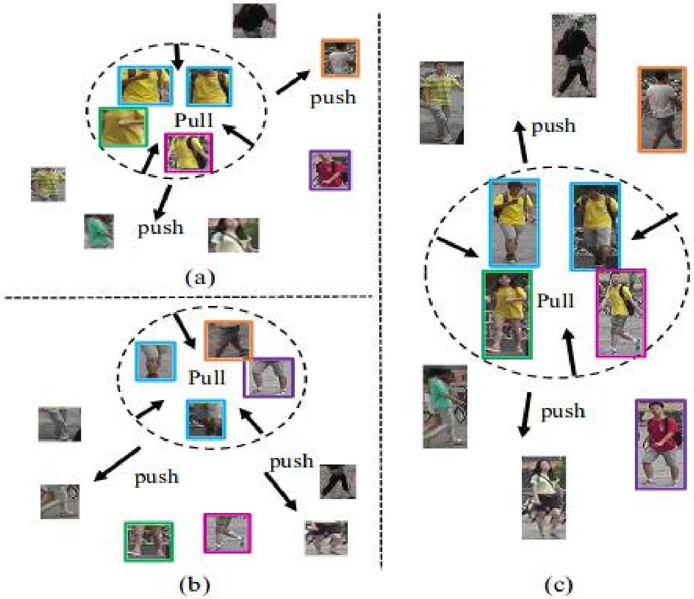


Figure 2. An illustration of the PAUL. The PatchNet is mainly composed of a CNN backbone and the patch generation network. First, we generate surrogate positive samples by using random image transformations. Next, we generate M patches for each feature map by using patch generation network (PGN) which can be split into three parts including the localization network (LN), the patch sampling grid, and the sampler. The PEDAL is designed to pull similar patches together and push the dissimilar patches. The IPFL is designed to pull the real sample and the surrogate positive samples together while pushing hard negative samples away.



$$\mathbf{w}_{j,t}^{m} = \begin{cases} (1-l) \times \mathbf{w}_{j,t-1}^{m} + l \times \mathbf{x}_{j,t}^{m}, & t > 0, \\ \mathbf{x}_{j,t}^{m}, & t = 0, \end{cases}$$

$$\mathcal{L}_{c}^{m} = -\log \frac{\sum_{\mathbf{w}_j^m \in \mathcal{K}_i^m} e^{-\frac{s}{2}\left\|\mathbf{x}_i^m - \mathbf{w}_j^m\right\|_2^2}}{\sum_{j=1, j\neq i}^{N} e^{-\frac{s}{2}\left\|\mathbf{x}_i^m - \mathbf{w}_j^m\right\|_2^2}},$$

Figure 4. Illustration of the cyclic ranking. We compute the ranking list $\mathcal{N}_i$ for the target image $\mathbf{x}_i$, Then we traverse the ranking list $\mathcal{N}_i$ in order, and we compute the ranking list for $\mathbf{x}_j \in \mathcal{N}_i$ until we find a hard negative sample. (Best viewed in color)

Table 1. Performance (%) comparison on Market-1501 dataset.

Methods	Rank-1	Rank-5	Rank-10	mAP
LOMO [20]	27.2	41.6	49.1	8.0
Bow [53]	35.8	52.4	60.3	14.8
UMDL [25]	34.5	52.6	59.6	12.4
PUL [7]	45.5	60.7	66.7	20.5
CAMEL [46]	54.5	-	-	26.3
PTGAN [40]	38.6	-	66.1	-
SPGAN + LMP [5]	57.7	75.8	82.4	26.7
TJ-AIDL [37]	58.2	74.8	81.1	26.5
HHL [55]	62.2	78.8	84.0	31.4
DECAMEL [47]	60.2	76.0.	81.1	32.4
SyRI [1]	65.7	-	-	-
PAUL (Ours)	**68.5**	**82.4**	**87.4**	**40.1**

Table 2. Performance (%) comparison on DukeMTMC dataset.

Methods	Rank-1	Rank-5	Rank-10	mAP
LOMO [20]	12.3	21.3	26.6	4.8
Bow [53]	17.1	28.8	34.9	8.3
UMDL [25]	18.5	31.4	37.6	7.3
PUL [7]	30.0	43.4	48.5	16.4
PTGAN [40]	27.4	-	50.7	-
SPGAN + LMP [5]	46.4	62.3	68.0	26.2
TJ-AIDL [37]	44.3	59.6	65.0	23.0
HHL [55]	46.9	61.0	66.7	27.2
PAUL (Ours)	**72.0**	**82.7**	**86.0**	**53.2**

Table 3. Ablation study (%). The results of "ResNet-50" and "PatchNet" mean we directly test the model (pre-trained on MSMT17) without training on unlabeled datasets.

Methods	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
ResNet-50 [11] †	46.6	22.7	52.6	33.1
ResNet-50 [11] + $\mathcal{L}_c$ ‡	25.6	11.8	29.5	15.4
PatchNet (Baseline)	59.3	31.0	65.7	45.6
PatchNet+$\mathcal{L}_c$	66.2	38.0	70.6	52.1
PatchNet+$\mathcal{L}_v$	65.4	37.6	67.1	48.0
PatchNet+$\mathcal{L}_v + \mathcal{L}_c$ (PAUL)	**68.5**	**40.1**	**72.0**	**53.2**

† Image-level feature learning.

‡ Here, $\mathcal{L}_c$ means pulling the features of the similar images together.

Table 4. Analysis on the proposed method with different patch generation schemes (%). For each patch generation schemes, we train the PatchNet on the MSMT17 and then perform unsupervised training with PEDAL and IPFL.

Generation schemes	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
Randomly	24.1	14.8	16.0	10.9
Equally	66.6	38.5	56.2	39.4
PGN	**68.5**	**40.1**	**72.0**	**53.2**
4 patches	67.3	39.5	70.4	50.6
6 patches	**68.5**	**40.1**	**72.0**	**53.2**
8 patches	66.7	37.2	70.6	51.8

Figure 6. A visualization of the nearest patches in the patch-based discriminative feature learning loss (Eq. (2)). We also show the whole images corresponding to the patches. Blue bounding box indicates the same identity. Red bounding box indicates the location of the patch.