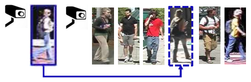
Convolutional Feature Transfer via Camera-Specific Discriminative Pooling for Person Re-Identification

Tetsu Matsukawa and Einoshin Suzuki
Kyushu University, Japan

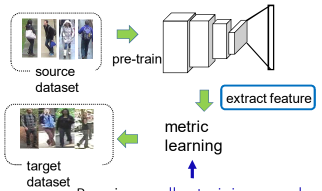
Approach

Practical issue

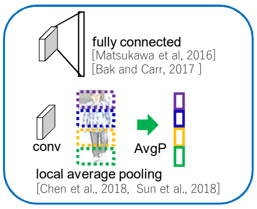
- Lack of training person IDs
- Require GPU for fine-tuning



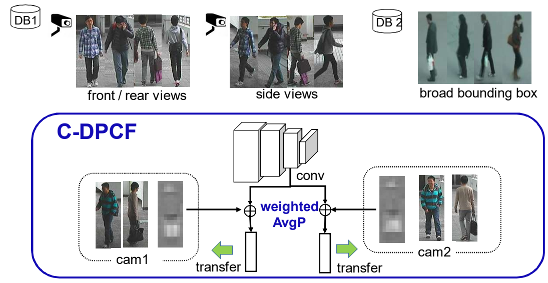
CNN-feature transfer



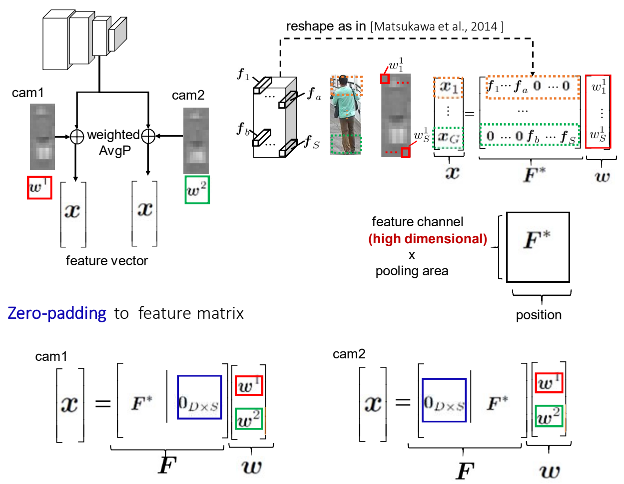
Require smaller training samples



Conventional features are less transferable to different camera/datasets due to spatial bias



Camera-specific weight local AvgP



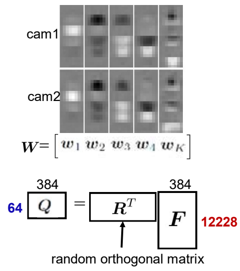
Weight map learning

- Given a training data $\{F_i, p_i, c_i\}_{i=1}^{N}$ (Person ID, Cam ID)
- Optimize sum of distances of K-weight map pairs
  - Random projection distance

$$\delta^2_{w_k}(i,j) = \|R^T x_{k,i} - R^T x_{k,j}\|_2^2 = \|R^T F_i w_k - R^T F_j w_k\|_2^2 = \|Q_i w_k - Q_j w_k\|_2^2$$

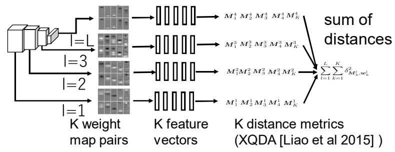
- Optimization problem

$$\max_{W} \; J(W) = \underbrace{\mathrm{Tr}\left[W^T \Sigma_D W\right]}_{\text{avg. distance of different persons}} - \underbrace{\mathrm{Tr}\left[W^T \Sigma_S W\right]}_{\text{avg. distances of same person}} \quad s.t. \quad W^T W = I_K$$



Distance for re-id

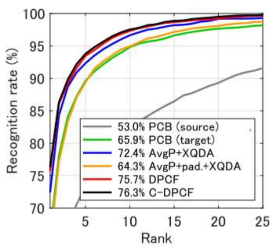
- C-DPCF is applied to L layers



Experiments

Comparison with baselines

Baseline: PCB [Sun et al., 2018] trained on (source/target) dataset

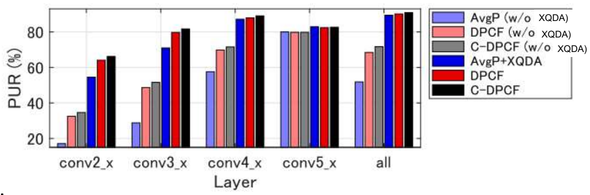


Source dataset: MSMT17
Target datasets: VIPeR

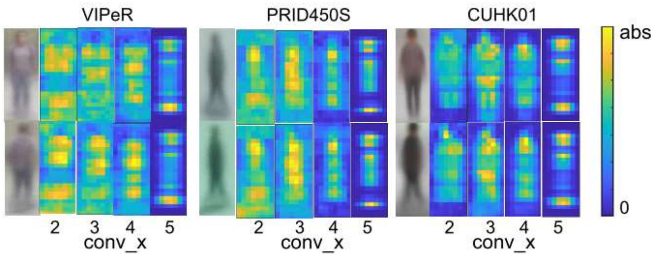
Training time

C-DPCF	PCB(target)
42.5 sec (CPU)	312 sec (GPU)

Convolutional layers



Weight maps

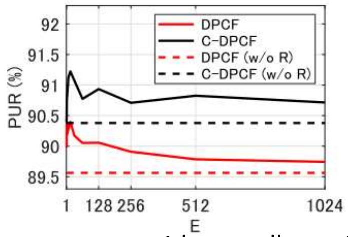


Rank-1 rates of SOTA

S: Supervised
U: Unsupervised
DG: Domain Generalization

		Type	VIPeR	GRID	PRID	CUHK01
	CMDL[PAMI18]	S	66.4	30.9	52.0	78.2
	HGD[PAMI20]	S	52.8	28.2	-	-
CNN features	Synthesis[ECCV18]	U	43.0	-	-	54.9
CNN features	One-shot [CVPR17]	U+S	34.3	-	-	45.6
CNN features	CRAFT [PAMI18]	S	50.3	22.4	-	-
CNN features	C-DPCF [ours]	S	76.3	34.8	79.4	89.1
Mobilenet -V2	DIMN [CVPR19]	DG	51.2	29.3	-	-
Mobilenet -V2	DN [BMVC19]	DG	58.8	39.7	73.6	-
Mobilenet -V2	DN + ours	DG+S	73.9	42.3	84.1	-

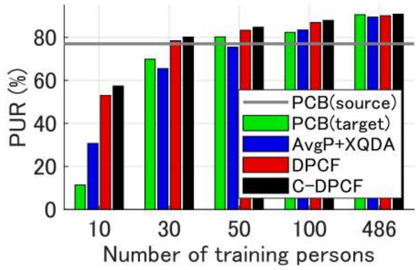
Random projection dim



Training time (CUHK01)

C-DPCF	(E=64)
w/o R	1684.4 sec
w/ R	49.0 sec

Accurate even with a smaller training data



- C-DPCF improves PCB(source) with 30 persons
- Camera-specific weight maps always outperforms common weight maps