One-Shot Metric Learning for Person Re-identification

Slawomir Bak, Peter Carr

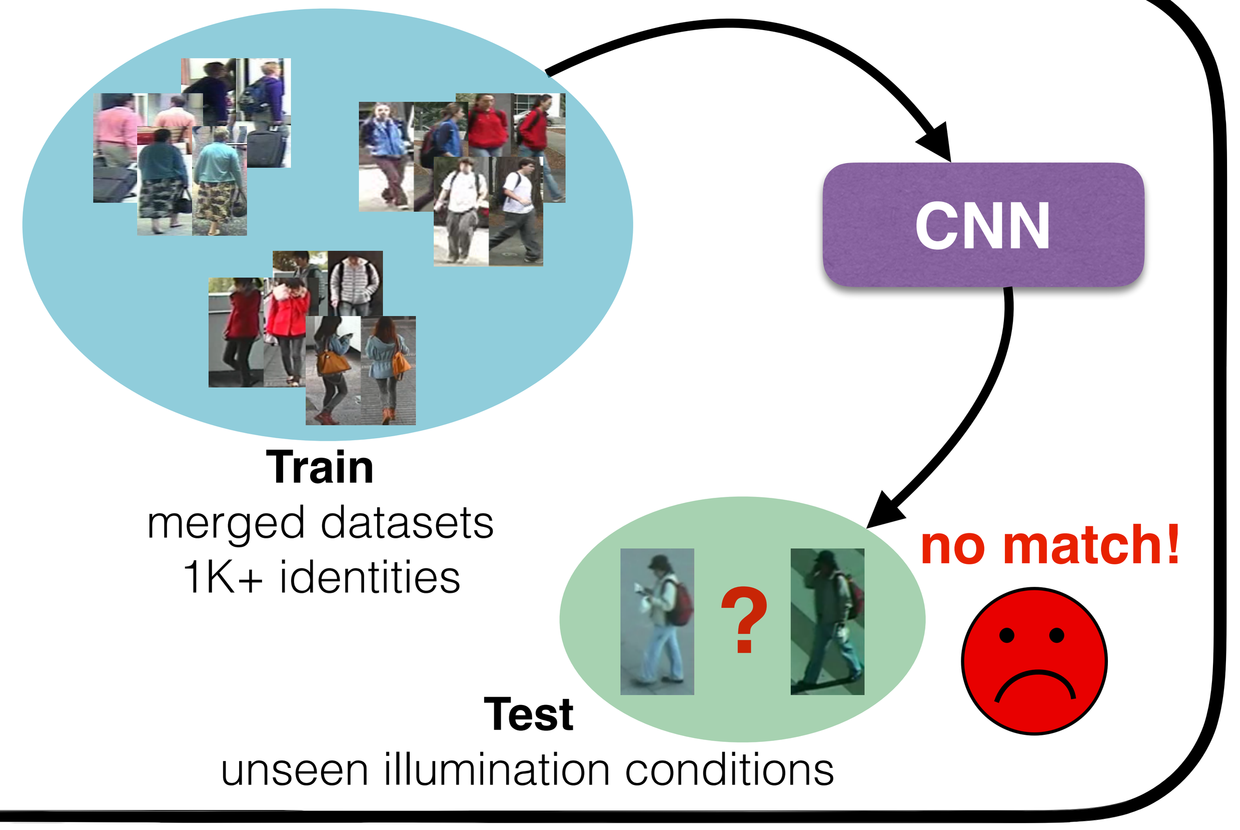
Disney Research

IEEE 2017 Conference on Computer Vision and Pattern Recognition

CVPR 2017, July 21-26, Honolulu

Motivation

- Supervised re-identification **does not scale** to large camera networks
  - **Poor generalization** properties to unseen camera conditions
  - **Requires fine-tuning** - often **hundreds** of image pairs



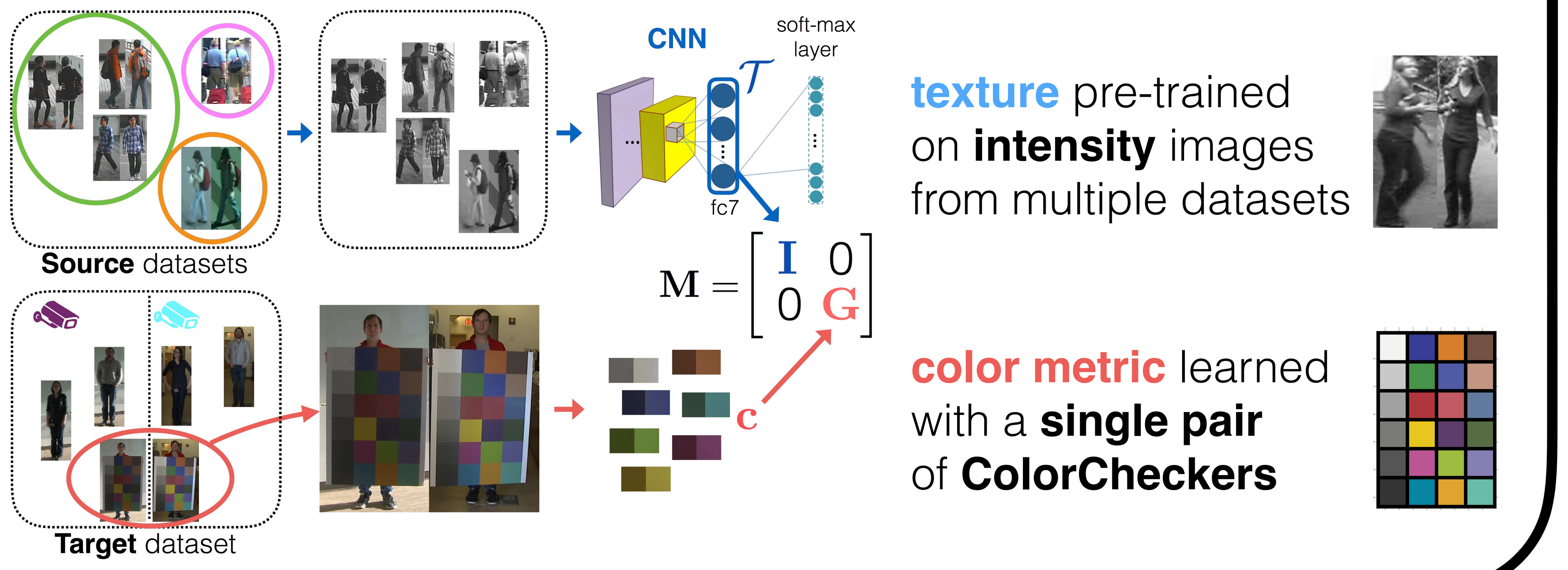
Approach: One-Shot Metric Learning

We learn a Mahalanobis metric

$$d^2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j).$$

, where $\mathbf{M}$ is split into **texture** and **color** components

$$\mathbf{M} = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{G} \end{bmatrix}$$



$\mathbf{M} = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{G} \end{bmatrix}$

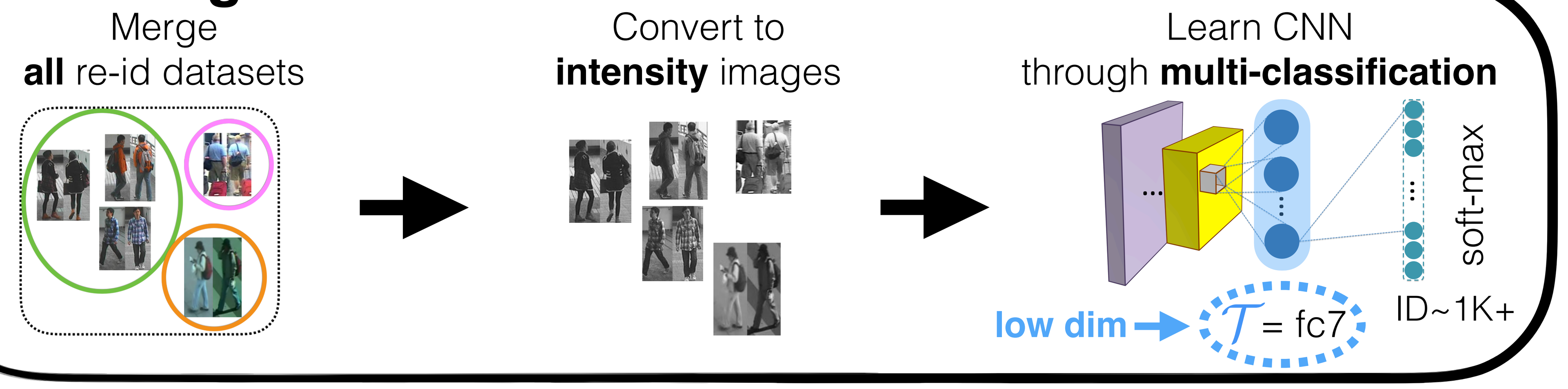
texture pre-trained on **intensity** images from multiple datasets

color metric learned with a **single pair** of **ColorCheckers**

Contributions

- A metric **split into texture and color** components
- Color invariant **deep texture** learned with only intensity images
- **One-shot metric learning** based on patches of a ColorChecker chart
- Spatial variations are handled by **explicitly modeling background distortions**

Learning Texture $\mathcal{T}$



Learning Color Metric $\mathbf{G}$

- **KISS** ML

$$\mathbf{G} = (\Sigma^+)^{-1} - (\Sigma^-)^{-1}$$

Covariance of **positive** pairwise differences

Easy: random sampling of people patches from both cameras

Covariance of **negative** pairwise differences

- **One-shot** ML

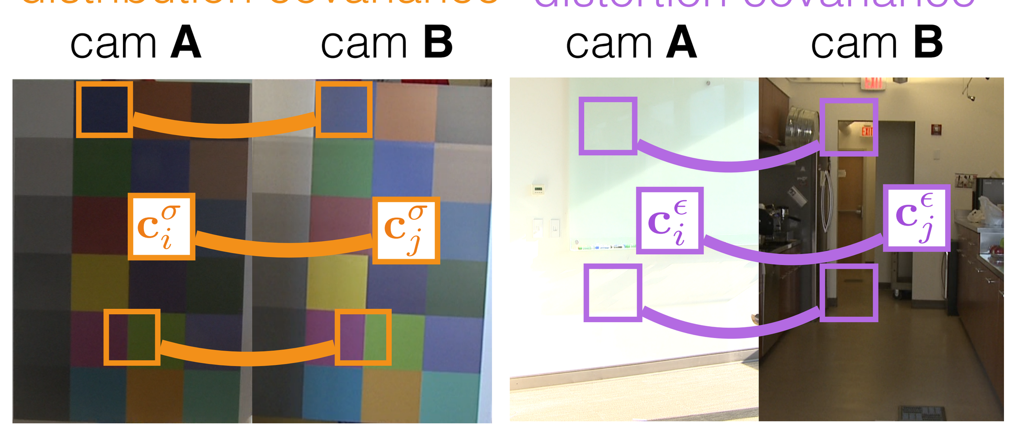
$$\Sigma^+ = \Sigma_\sigma^+ + \Sigma_\epsilon^+$$

$$\Sigma_\sigma^+ = (\mathbf{c}_i^\sigma - \mathbf{c}_j^\sigma)(\mathbf{c}_i^\sigma - \mathbf{c}_j^\sigma)^T$$

$$\Sigma_\epsilon^+ = (\mathbf{c}_i^\epsilon - \mathbf{c}_j^\epsilon)(\mathbf{c}_i^\epsilon - \mathbf{c}_j^\epsilon)^T$$

Cross-camera-color distribution covariance

Background distortion covariance



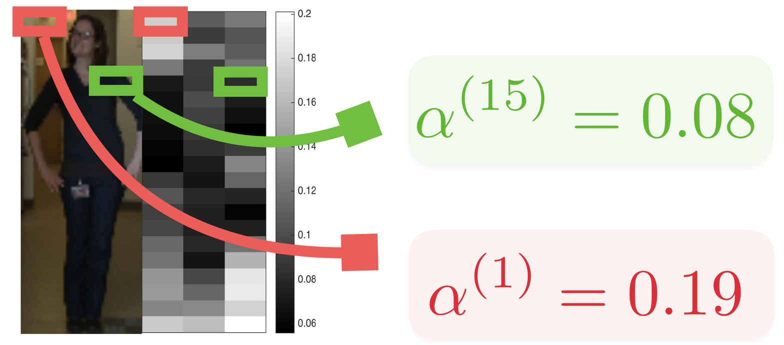
Spatial Variations

- Patches at different locations have **different amounts** of background pixels

$$\Sigma^+ = \Sigma_\sigma^+ + \Sigma_\epsilon^+$$

Learn using auxiliary dataset

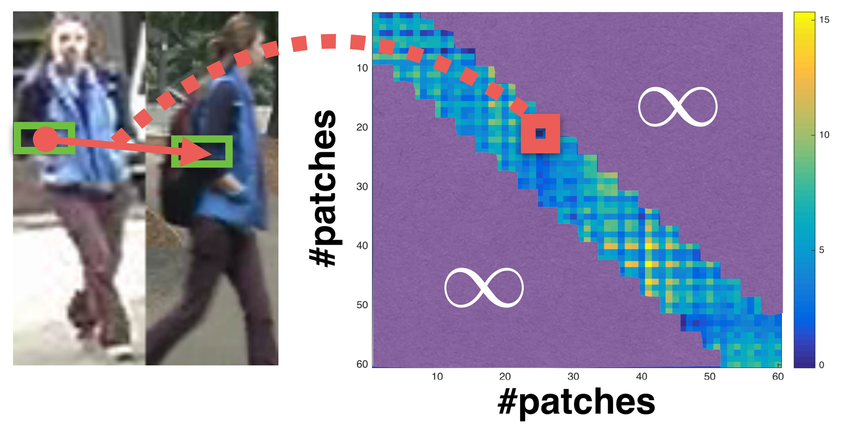
$$\Sigma^{+(n)} = \Sigma_\sigma^+ + \alpha^{(n)} \Sigma_\epsilon^+$$



$\alpha^{(15)} = 0.08$

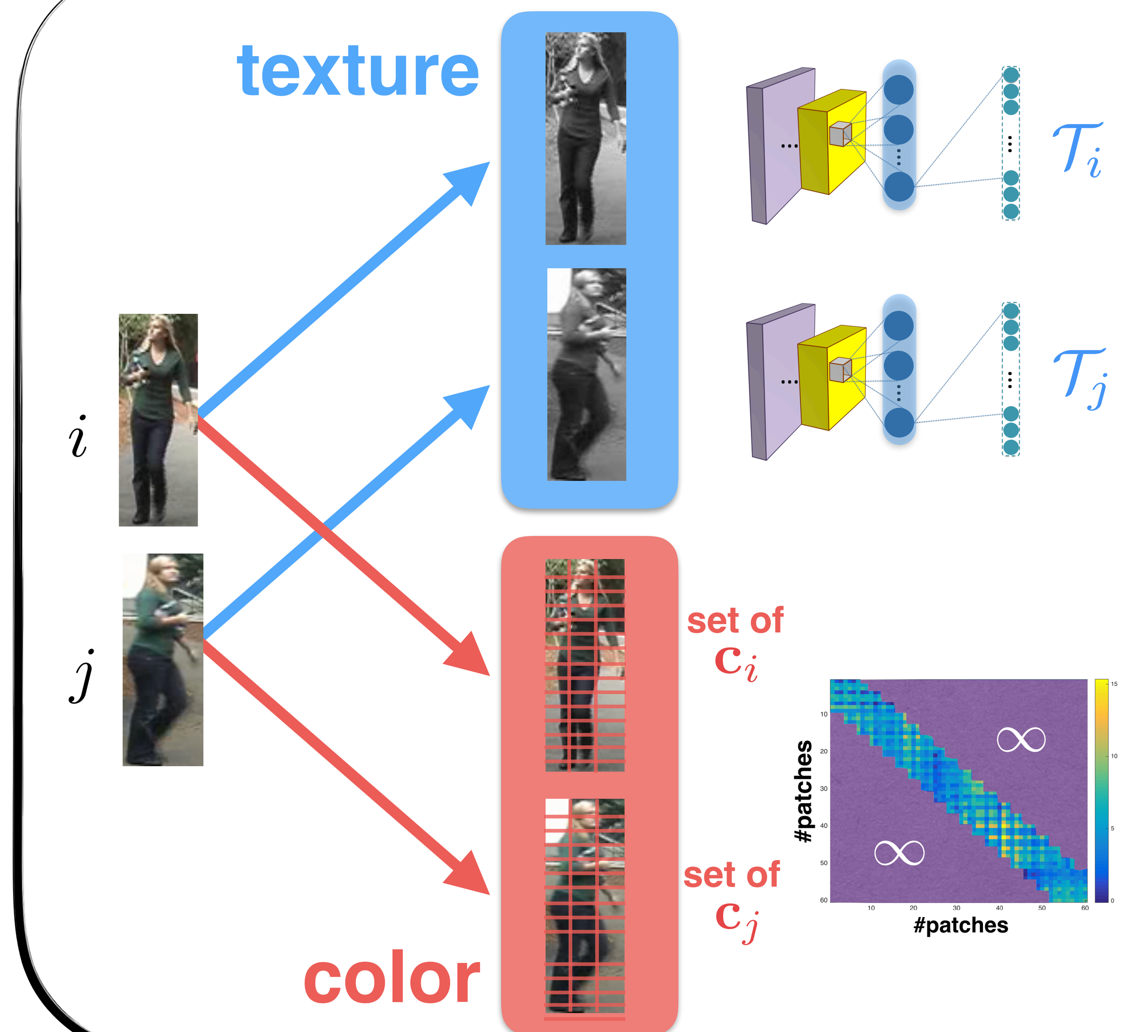
$\alpha^{(1)} = 0.19$

- Features extracted on a fixed grid may not correspond due to **pose changes**



Define a linear patch assignment problem to accommodate pose misalignments - solved by the **Hungarian** algorithm.

Distance Computation



$$\mathbf{T} = |\mathcal{T}_i - \mathcal{T}_j|_2$$

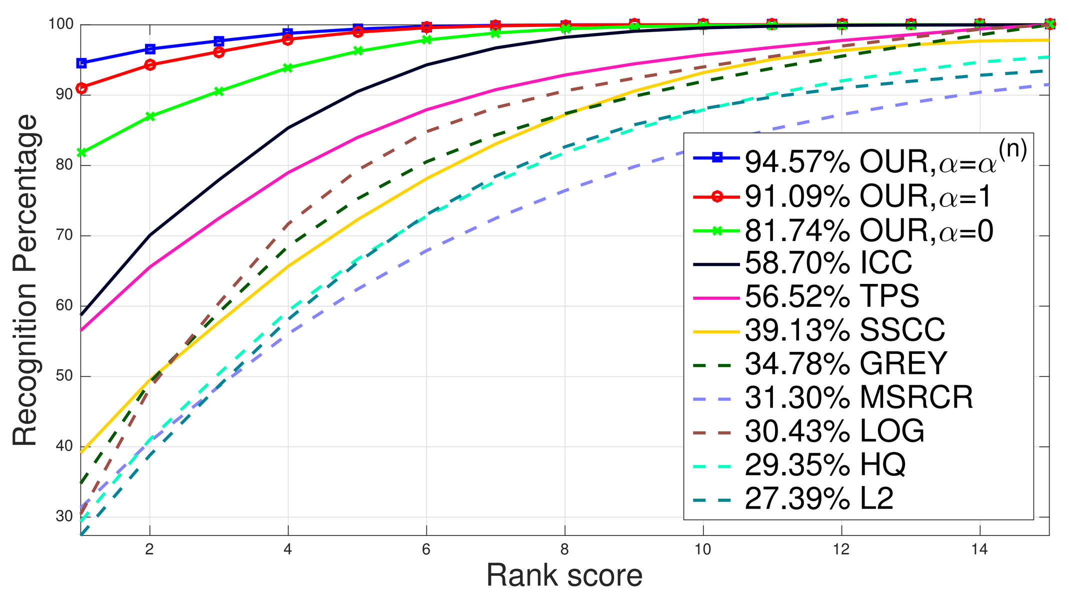
$$d^2(i,j) = (1-\gamma)\mathbf{T} + \gamma \mathbf{C}$$

$$\mathbf{C} = \sum_{n=1}^{N} \Phi^2(\mathbf{c}_i^{\Omega_{ij}^*(n)}, \mathbf{c}_j^n; \mathbf{G}^{(n)})$$

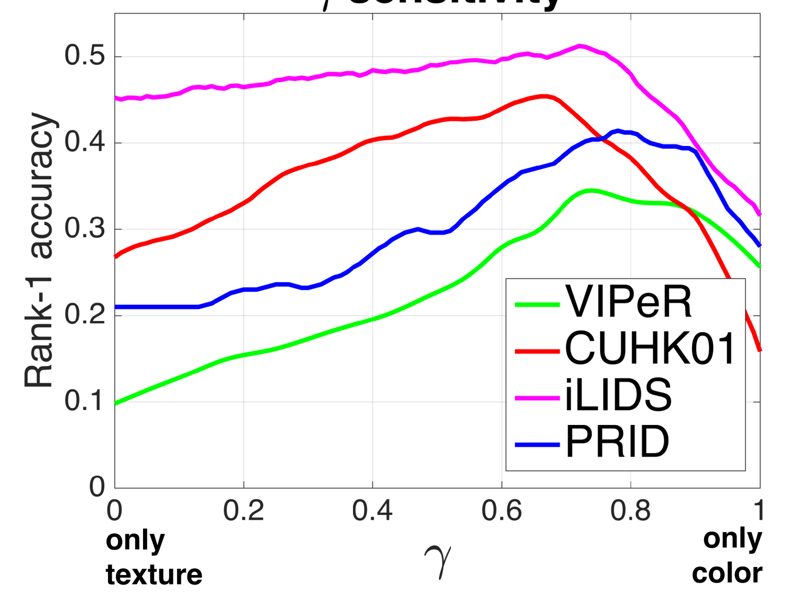
γ controls relative importance of color

Experimental Results

Color Calibration on CCH dataset



γ sensitivity



ColorCheckers

	Method	#IDs	VIPeR	CUHK01	iLIDS	PRID
semi/unsupervised	**Our**, $\alpha = \alpha^{(n)}$	1	**34.3**	**45.6**	**51.2**	**41.4**
	Our, $\alpha = 0$	1	30.1	39.6	49.9	31.9
	$\text{JSTL}^{I}_{\text{LOO}}$ [CVPR16]	0	9.8	26.8	44.0	21.0
	JSTL_{LOO} [CVPR16]	0	20.9	37.1	43.5	2.0
	GL [ECCV16]	0	33.5	41.0	-	25.0
	TL-semi [CVPR16]	80	34.1	32.1	50.3	25.3
supervised	FT-JSTL+DGD [CVPR16]	2629	38.6	**66.6**	**64.6**	**64.0**
	LOMO+XQDA [CVPR15]	240	40.0	63.2	-	26.7
	Ensembles [CVPR15]	240	45.9	53.4	50.3	17.9
	Null Space [CVPR16]	240	42.2	64.9	-	29.8
	Triplet Loss [CVPR16]	240	47.8	53.7	60.4	22.0
	Gaussian+XQDA [CVPR16]	240	**49.7**	57.8	-	-