Person Re-identification

Wei-Shi Zheng (郑伟诗)

http://isee.sysu.edu.cn/~zhwshi

Sun Yat-sen University

机器智能与先进计算
教育部重点实验室

研究领域

研究方面一：

跨场景行为分析

- 解决行人碎片化轨迹问题
- 行人重识别

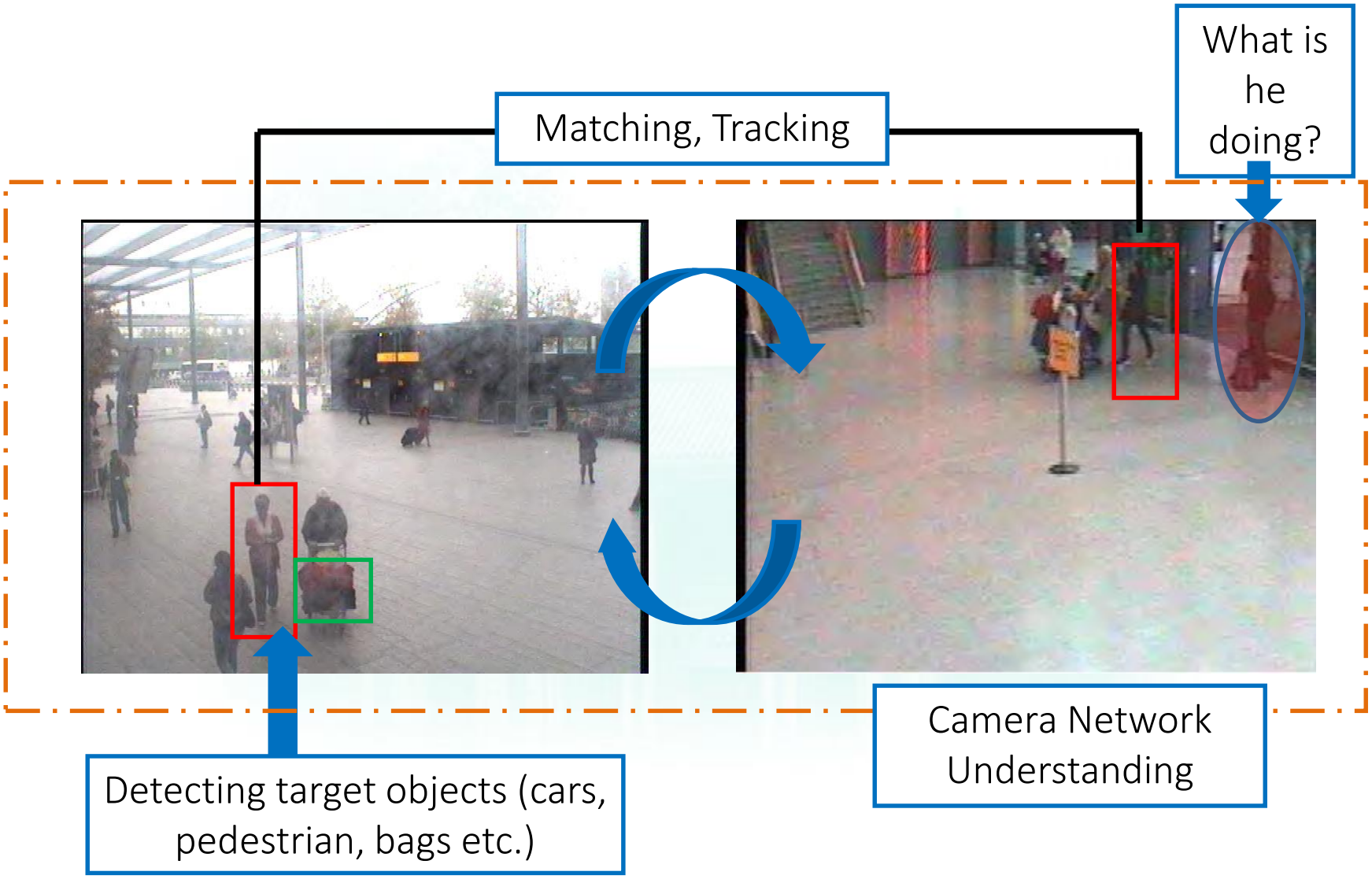
研究方面二：

行为意图识别

- 解决复杂交互行为建模
- 解决多通道信息融合问题
- 解决中、长短间隔预测

跨场景、多人、长时间行为识别

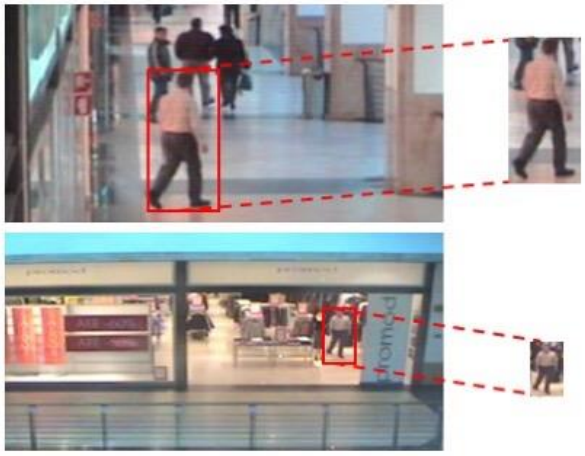
Person Re-identification



Person Re-identification: Challenges

Person Re-identification: Challenges

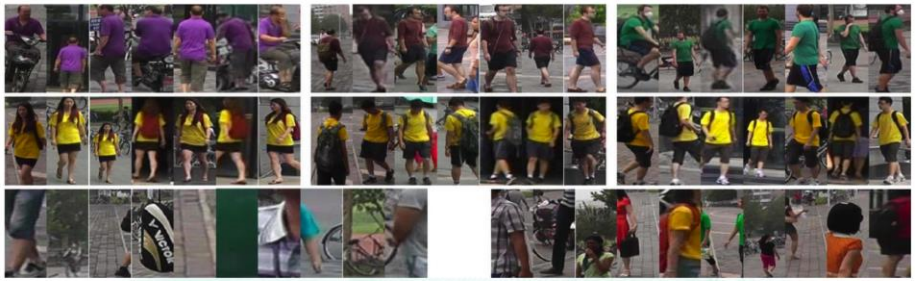
- **Some Main Variations**



View | Lighting | Occlusion | Low Resolution | Clothing Change

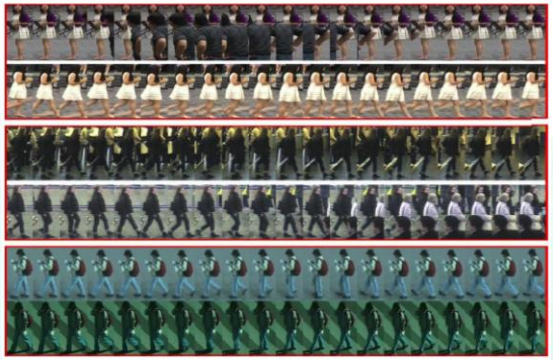
Recent Development & Question

- **Pose-guided, Local, Attention-based, GAN-based, ……**
  [a ppt: https://share.weiyun.com/5VPtcZa]
- **What should we do? I would guess we will soon have 99% matching rate this year or early next year on benchmarks**



Dataset	Reference	Rank-1	mAP	
Market-1501	Harmonious Attention Network for Person Re-Identification	91.2	75.7	Single query
		93.8	82.8	Multiple queries
DukeMTMC	Learning Discriminative Features with Multiple Granularities for Person Re-Identification	88.7	78.4	

- Video-based RE-ID datasets: MARS, iLIDS-VID, PRID 2011.



- The state-of-the-art performance (Rank-1):

	MARS	iLIDS-VID	PRID 2011
CVPR 2018	82.3 (65.8)	80.2	93.2

- **Have we already solved it?**

My Today's Focus

- **Tell less about performance**

- **Aim to tell something of my understanding about Re-ID**

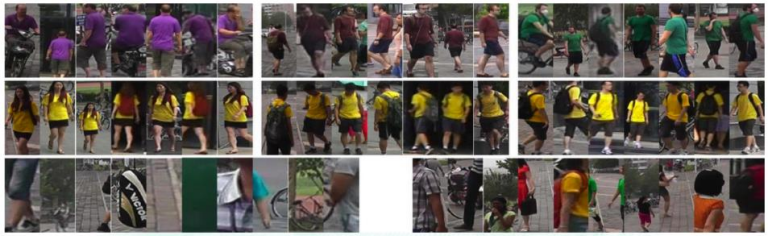
1. Unsupervised Person Re-id?

Unsupervised Re-ID

- **Existing works mainly focus on supervised models**
- **Scalability of supervised models**
  - Require quantities of labelled data across views
  - Manually labelling not very reliable
  - Prohibitive time and money cost
  - …
- **Unsupervised Re-ID to help solve this problem**
  - Feature Representation Learning
    - Hand-crafted
    - Sparse/dictionary learning
    - CCA
    - transfer

How to explicitly quantify cross-view matching?

How to perform effective deep learning for unsupervised re-id?

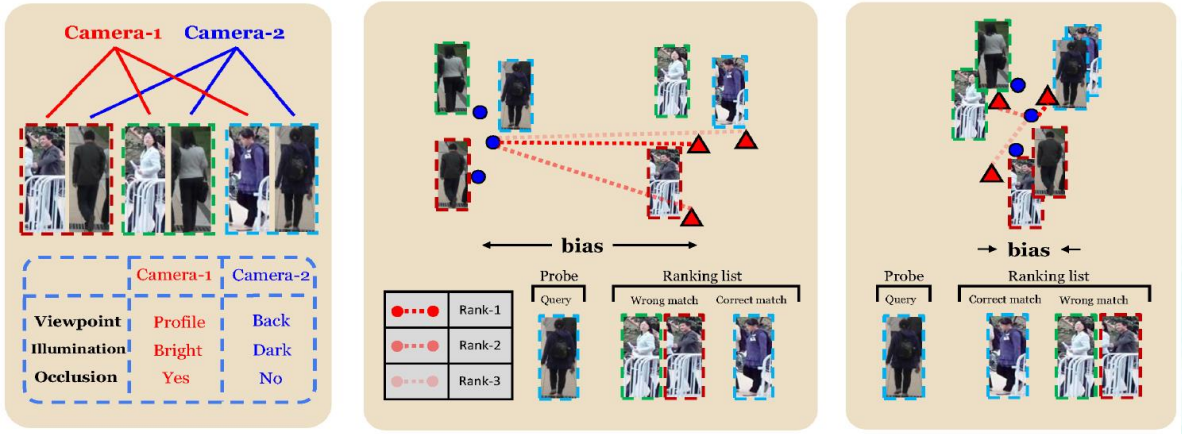


Dataset	Reference	Rank-1	mAP	
Market-1501	Harmonious Attention Network for Person Re-Identification	91.2	75.7	Single query
		93.8	82.8	Multiple queries
DukeMTMC	Learning Discriminative Features with Multiple Granularities for Person Re-Identification	88.7	78.4	

Deep Unsupervised Person Re-id

Unsupervised Deep Learning

- DEep Clustering-based Asymmetric MEtric Learning (DECAMEL)



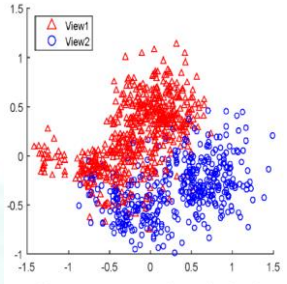
$$f_{loss} = f_{intra} + \lambda f_{consistency} + \gamma f_{constraint}$$

$$= \frac{1}{N}\sum_{k=1}^{K}\sum_{i\in\mathcal{C}_k}\|\boldsymbol{U}_{v_i}^{\mathrm{T}}\mathbf{x}_i - \mathbf{c}_k\|^2 + \lambda\sum_{u,w}\|\boldsymbol{U}_v - \boldsymbol{U}_w\|_F^2$$

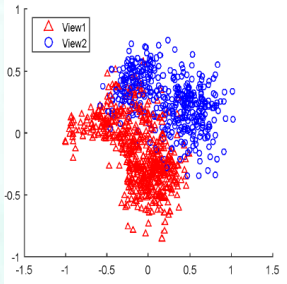
$$+ \gamma\sum_{v=1}^{V}\|\boldsymbol{U}_v^{\mathrm{T}}\boldsymbol{\Sigma}_v\boldsymbol{U}_v - \boldsymbol{I}\|_F^2,$$

Asymmetric metric clustering

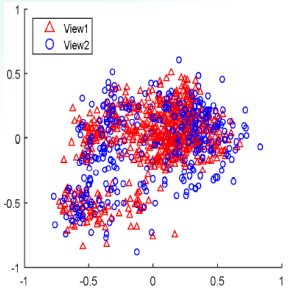
Cross-view consistency regularization



(a) Feature representation distribution



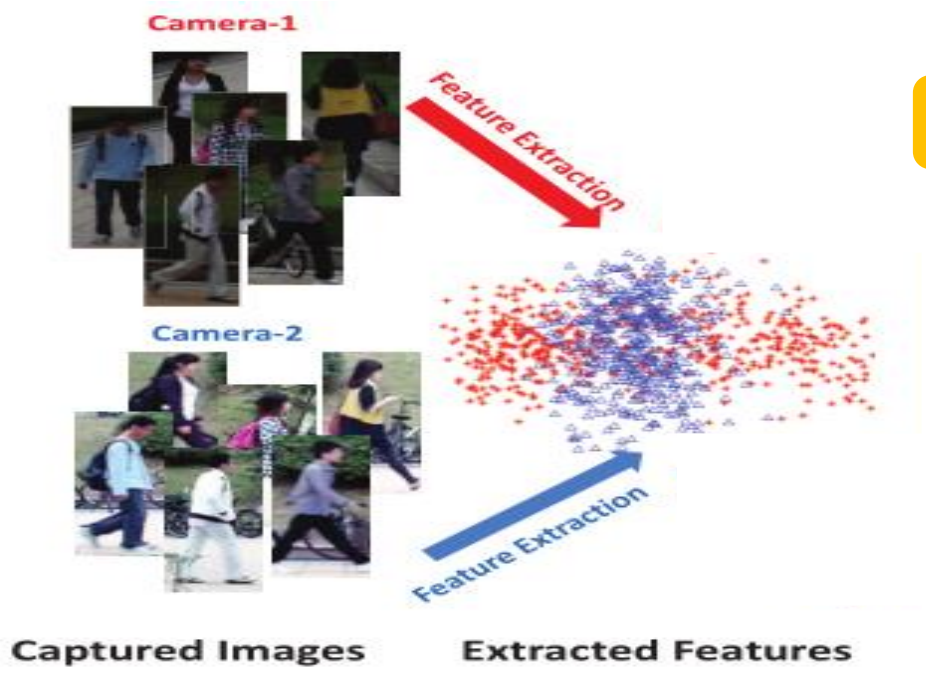
(b) Using symmetric metric



(c) Using asymmetric metric

Hong-Xing Yu, Ancong Wu, Wei-Shi Zheng(PI)*. Unsupervised Person Re-identification by Asymmetric Metric Embedding. In IEEE IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2019.

Asymmetric Metric for Re-ID: Unsupervised



Asymmetric Metric Learning

$$d(\{x_i^p, p\}, \{x_j^q, q\}) = \|U^{pT} x_i^p - U^{qT} x_j^q\|_2$$

$$U^p \neq U^q$$

$$min \quad f = f_{\text{cross}} + \eta f_{\text{intra}}$$

$$f_{\text{cross}} = \sum_{p=1}^{N-1} \sum_{q=p+1}^{N} \sum_{i=1}^{n^p} \sum_{j=1}^{n^q} W_{ij}^{p,q} \|U^{pT} x_i^p - U^{qT} x_j^q\|_2^2$$

$$f_{\text{intra}} = \sum_{p=1}^{N} \sum_{i=1}^{n^p} \sum_{j=1}^{n^p} W_{i,j}^{p,p} \|U^{pT} x_i^p - U^{pT} x_j^p\|_2^2.$$

Learning universal feature transformation → Learning view-specific feature transformation

Yingcong Chen, Xiatian Zhu, Wei-Shi Zheng*, and Jian-Huang Lai. Person Re-Identification by Camera Correlation Aware Feature Augmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 2017.

Science in China 2018

Asymmetric Metric for Re-ID

$$d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \mathbf{M} (\boldsymbol{x}_i - \boldsymbol{x}_j)}$$
$$= ||\boldsymbol{U}^T \boldsymbol{x}_i - \boldsymbol{U}^T \boldsymbol{x}_j||_2,$$

Learn different feature transformation for different camera views

Pseudometric

$$d(\{\boldsymbol{x}_i^p, p\}, \{\boldsymbol{x}_j^q, q\}) = ||\boldsymbol{U}^{pT} \boldsymbol{x}_i^p - \boldsymbol{U}^{qT} \boldsymbol{x}_j^q||_2$$
$$\boldsymbol{U}^p \neq \boldsymbol{U}^q$$

Non-negativity **Symmetry**

$$\begin{aligned} d(\{\boldsymbol{x}_i^p, p\}, \{\boldsymbol{x}_j^q, q\}) &= ||\boldsymbol{U}^{pT} \boldsymbol{x}_i^p - \boldsymbol{U}^{qT} \boldsymbol{x}_j^q||_2 \\ &= ||\boldsymbol{U}^{qT} \boldsymbol{x}_j^q - \boldsymbol{U}^{pT} \boldsymbol{x}_i^p||_2 \\ &= d(\{\boldsymbol{x}_j^q, q\}, \{\boldsymbol{x}_i^p, p\}), \end{aligned}$$

Triangle Inequality

$$||\boldsymbol{U}^{rT} \boldsymbol{x}_k^r - \boldsymbol{U}^{qT} \boldsymbol{x}_j^q||_2 \leq$$
$$||\boldsymbol{U}^{rT} \boldsymbol{x}_k^r - \boldsymbol{U}^{pT} \boldsymbol{x}_i^p||_2 + ||\boldsymbol{U}^{pT} \boldsymbol{x}_i^p - \boldsymbol{U}^{qT} \boldsymbol{x}_j^q||_2.$$

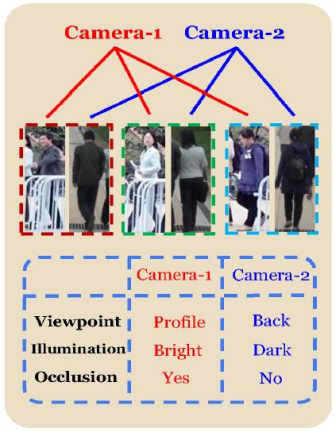
Coincidence

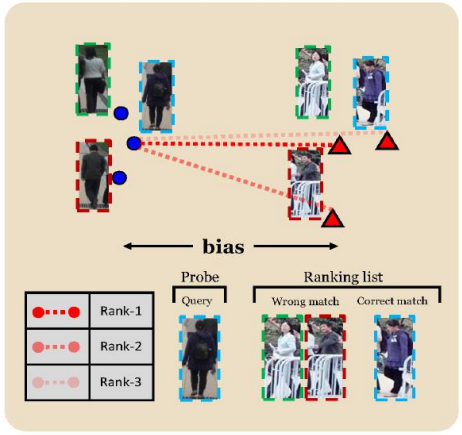
$$d(\{\boldsymbol{x}^p, p\}, \{\boldsymbol{x}^q, q\}) = 0$$
$$\boldsymbol{U}^{pT} \boldsymbol{x}^p = \boldsymbol{U}^{qT} \boldsymbol{x}^q \quad \Leftarrow \quad \boldsymbol{U}^{pT} \boldsymbol{x}^p = \boldsymbol{U}^{qT} \boldsymbol{x}^q$$
$$||\boldsymbol{U}^p - \boldsymbol{U}^q||_F^2$$

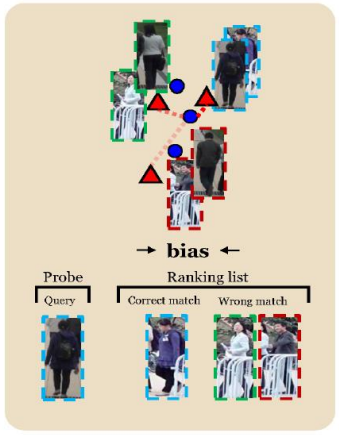
Deep Unsupervised Person Re-id

Deep Asymmetric Clustering

- DEep Clustering-based Asymmetric MEtric Learning (DECAMEL)







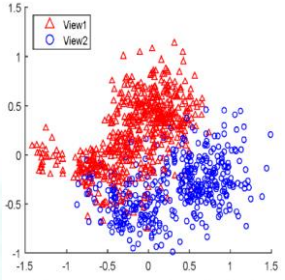
$$f_{loss} = f_{intra} + \lambda f_{consistency} + \gamma f_{constraint}$$

$$= \frac{1}{N}\sum_{k=1}^{K}\sum_{i\in\mathcal{C}_k}\|\boldsymbol{U}_{v_i}^{\mathrm{T}}\mathbf{x}_i - \mathbf{c}_k\|^2 + \lambda\sum_{u,w}\|\boldsymbol{U}_v - \boldsymbol{U}_w\|_F^2$$

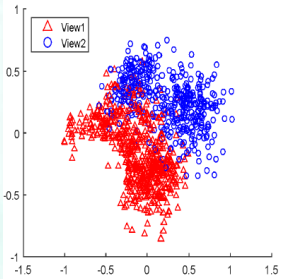
$$+ \gamma\sum_{v=1}^{V}\|\boldsymbol{U}_v^{\mathrm{T}}\boldsymbol{\Sigma}_v\boldsymbol{U}_v - \boldsymbol{I}\|_F^2,$$

Asymmetric metric clustering

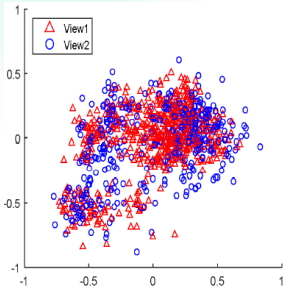
Cross-view consistency regularization



(a) Feature representation distribution



(b) Using symmetric metric



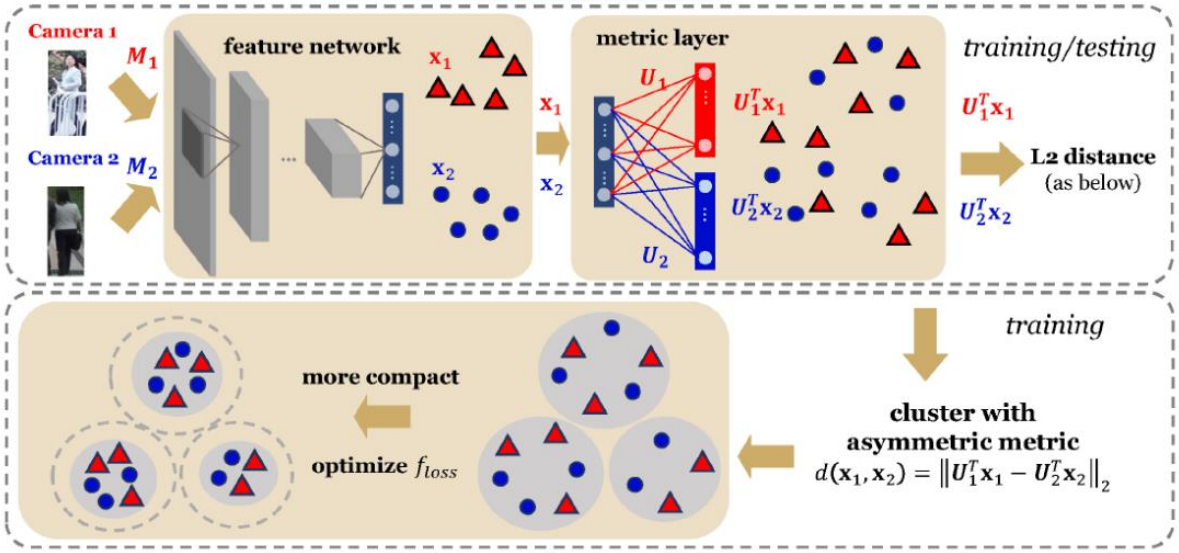
(c) Using asymmetric metric

Hong-Xing Yu, Ancong Wu, Wei-Shi Zheng(PI)*. Unsupervised Person Re-identification by Asymmetric Metric Embedding. In IEEE IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2019.

Deep Unsupervised Person Re-id

Unsupervised Deep Learning

- Challenge: How to optimize unsupervised deep model?



Algorithm 1: DECAMEL

Input : The training images $\mathcal{M}$, the deep feature extractor $f(\cdot;\Theta)$
Training:
Metric initialization:
Extract feature representations using f to obtain the initial feature set $\mathcal{X}$.
Conduct k-means clustering in $\mathcal{X}$ to obtain $\{\mathbf{c}_k\}_{k=1}^K$ and to initialize $\boldsymbol{H}$ according to Eq. (12) and (13).
Fix $\boldsymbol{H}$ and solve the eigen-decomposition problem described by Eq. (23) and (24) to construct $\widetilde{\boldsymbol{U}}$.
$t \leftarrow 1$ where t denotes each step in the following loop.
while $\{f_{obj}^t\}$ *not converged* **do**
- Alternate fixing $\widetilde{\boldsymbol{U}}$ and $\boldsymbol{H}$ while optimizing the other.
- $t \leftarrow t+1$.

end
Decompose $\widetilde{\boldsymbol{U}}$ to obtain $\{\boldsymbol{U}_v\}_{v=1}^V$.
Initialize the deep framework $g(\cdot,\cdot;\Theta,\{\boldsymbol{U}_v\}_{v=1}^V)$ using Θ and $\{\boldsymbol{U}_v\}_{v=1}^V$.
End-to-end joint learning:
Update $\{\mathbf{c}_k\}_{k=1}^K$ from $\boldsymbol{H}$ according to Eq. (12) and (13).
while *maximum iteration not reached* **do**
- Update Θ and $\{\boldsymbol{U}_v\}_{v=1}^V$ by performing SGD using the gradients in Eq. (25), (26) and (27).
- Update $\{\mathbf{c}_k\}_{k=1}^K$ while fixing Θ and $\{\boldsymbol{U}_v\}_{v=1}^V$.

end
Testing:
Given two testing images $\{\mathbf{M}_i, v_i\}$ and $\{\mathbf{M}_j, v_j\}$, the distance/dissimilarity is computed by $||g(\mathbf{M}_i, v_i) - g(\mathbf{M}_j, v_j)||_2$.

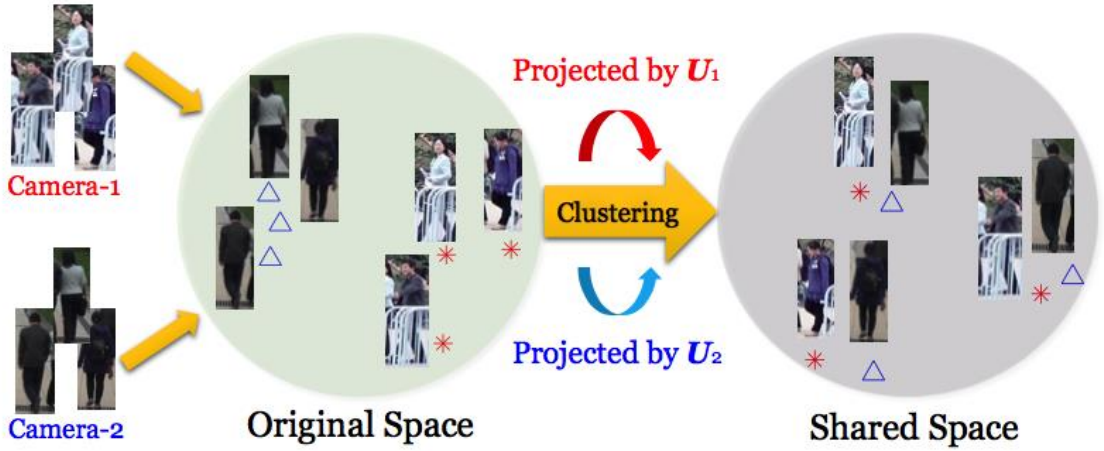
(1) **Metric initialization** by the linear Clustering-based Asymmetric Metric Learning (**CAMEL**)

(2) **End-to-end** joint feature-metric learning by any gradient based method e.g. SGD

Deep Unsupervised Person Re-id

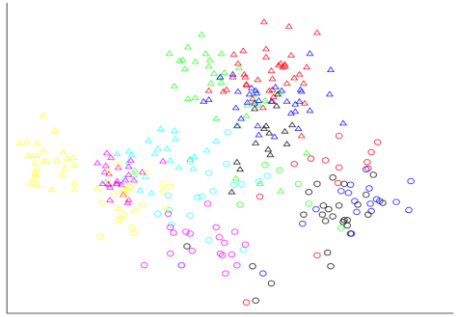
CAMEL: Metric initialization

- Clustering-based Asymmetric MEtric Learning (CAMEL)

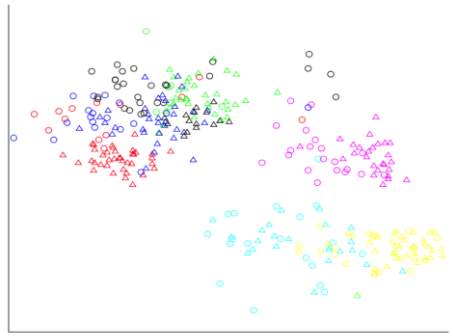


$$\min_{\boldsymbol{U}^1,\cdots,\boldsymbol{U}^V} \mathcal{F}_{obj} = \frac{1}{N}\sum_{k=1}^{K}\sum_{i\in C_k}\|\boldsymbol{U}^{p\mathrm{T}}\mathbf{x}_i^p - \mathbf{c}_k\|^2 + \lambda\sum_{p\neq q}\|\boldsymbol{U}^p - \boldsymbol{U}^q\|_F^2$$

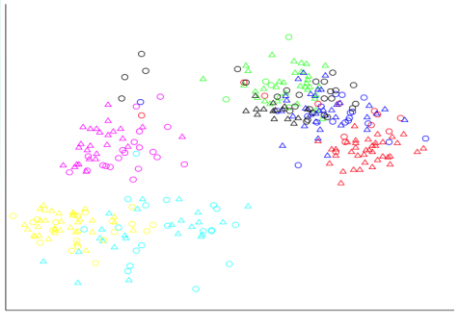
$$s.t. \quad \boldsymbol{U}^{p\mathrm{T}}\boldsymbol{\Sigma}^p\boldsymbol{U}^p = \boldsymbol{I} \quad (p = 1,\cdots,V),$$



(a) Original Feature Distribution



(b) Middle Stage of CAMEL



(c) Convergence Stage of CAMEL

Hong-Xing Yu, Ancong Wu, Wei-Shi Zheng*. Cross-view Asymmetric Metric Learning for Unsupervised Person Re-identification. In IEEE Conf. on Computer Vision (ICCV), 2017.

Deep Unsupervised Person Re-id

❑ Theoretical insight

Cross-view consistency regularization ⇨ **Bound the Coincidence Discrepancy**

End-to-end learning with the asymmetric metric ⇨ **Share know-ledge for the feature learning**

Remark 1: Cross-view Consistency Regularization for the Metric. We note that although asymmetric metric learning has been successfully applied in supervised Re-ID [3], [28], it is a pseudo metric rather than a strict metric [28], because it may not meet the coincidence property: given two identical feature vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ ($\mathbf{x}_i = \mathbf{x}_j$) from different camera views v_i and v_j, the asymmetric metric may not guarantee $d(\mathbf{x}_i, \mathbf{x}_j) = 0$. In this aspect, the cross-view consistency regularization plays a role to control an upper bound of coincidence discrepancy. In fact, according to the Cauchy Inequality, we have

$$\begin{aligned} d(\mathbf{x}_i, \mathbf{x}_j) &= \|\boldsymbol{U}_{v_i}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{U}_{v_j}^{\mathrm{T}}\mathbf{x}_j\|_2 = \|\boldsymbol{U}_{v_i}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{U}_{v_j}^{\mathrm{T}}\mathbf{x}_i\|_2 \\ &\leq \|\mathbf{x}_i\|_2 \cdot \|\boldsymbol{U}_{v_i} - \boldsymbol{U}_{v_j}\|_F. \end{aligned} \tag{8}$$

The cross-view consistency regularization controls $\|\boldsymbol{U}_{v_i} - \boldsymbol{U}_{v_j}\|_F$ which is a scaled upper bound of the coincidence discrepancy. Thus, it makes the learned asymmetric metric more mathematically principled and rigorous.

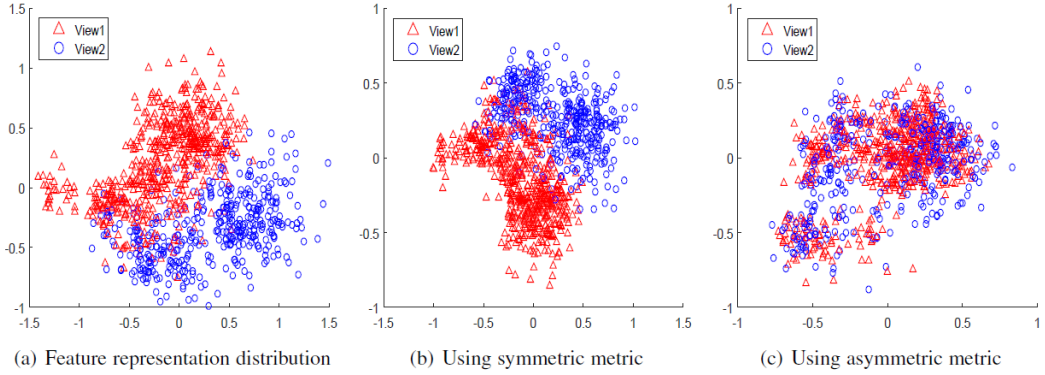
Remark 2: Explanation for Deep Metric Embedding. We can see from Eq. (25), (26) and (27) that, the sample gradient flows over the whole network, while the metric gradient only flows to the metric. However, the metric is actually embedded into the sample gradient: according to the chain rule, the sample gradient for the feature extractor parameter $\boldsymbol{\Theta}$ is

$$\begin{aligned} \frac{\partial f_{loss}}{\partial \boldsymbol{\Theta}} &= \frac{\partial f_{loss}}{\partial f(\mathbf{M};\boldsymbol{\Theta})}\frac{\partial f(\mathbf{M};\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} = \frac{\partial f_{loss}}{\partial \mathbf{x}}\frac{\partial f(\mathbf{M};\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} \\ &= \frac{\partial f_{loss}}{\partial \boldsymbol{U}_v^{\mathrm{T}}\mathbf{x}}\frac{\partial \boldsymbol{U}_v^{\mathrm{T}}\mathbf{x}}{\partial \mathbf{x}}\frac{\partial f(\mathbf{M};\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} \\ &= 2(\boldsymbol{U}_v^{\mathrm{T}}\mathbf{x} - \mathbf{c}_k)\boldsymbol{U}_v^{\mathrm{T}} \cdot \frac{\partial f(\mathbf{M};\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}}. \end{aligned} \tag{28}$$

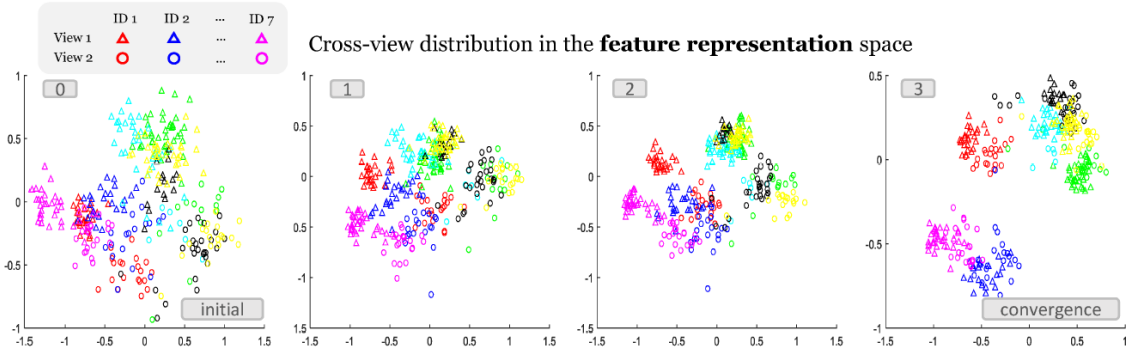
Thus, the metric $\boldsymbol{U}_v^{\mathrm{T}}$ is back-propagated to the whole network. As we will see in Sec. 4.2, the jointly learned feature bears resemblance to the metric. Furthermore, we will also see in Sec. 5.4.3 that this improves the cross-view discriminability of the feature. These observations seem like the metric is being "embedded" into the feature, and thus we refer to it as the "deep metric embedding".

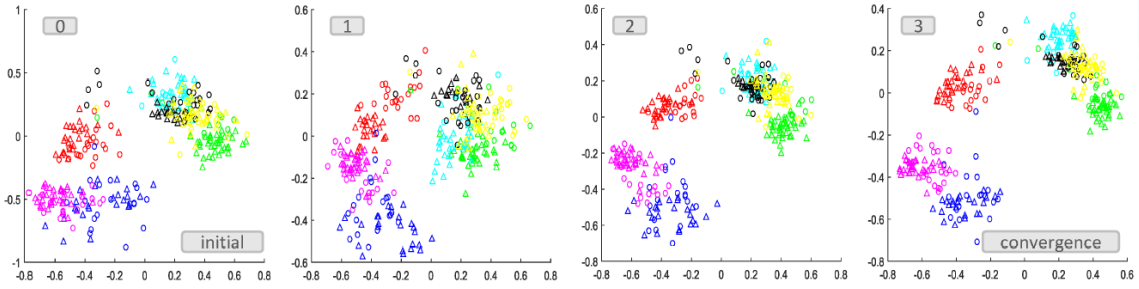
Deep Unsupervised Person Re-id

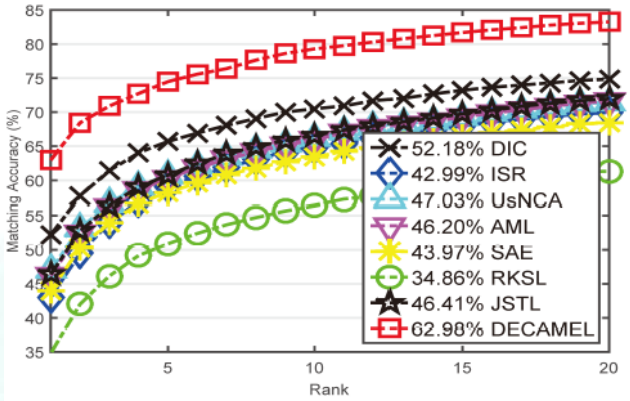
❑ Superiority of asymmetric metric



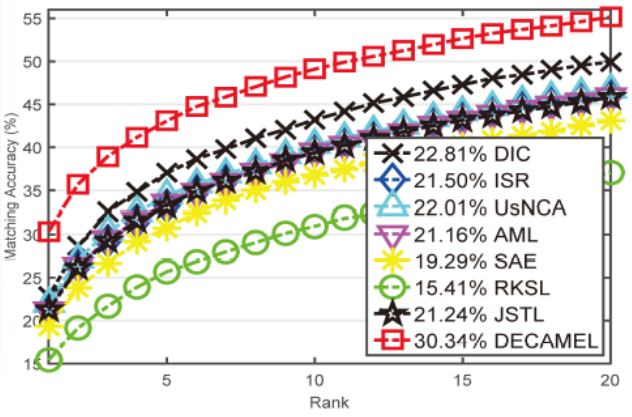
(a) Feature representation distribution (b) Using symmetric metric (c) Using asymmetric metric







(e) ExMarket



(f) MSMT17

Deep Unsupervised Person Re-id

- **Superiority of asymmetric metric: Comparison to the state of the art**

TABLE 2
Comparison with related unsupervised models: single-shot ("single") and multi-shot ("multi") rank-1 matching rate and MAP in percentage. In each column, the best is indicated in **red** and the second in **blue**.

Dataset	VIPeR	CUHK01	CUHK01	CUHK03	CUHK03	SYSU	SYSU	Market	ExMarket	MSMT17
Measure	single	single	multi	single	multi	single	multi	multi(MAP)	multi(MAP)	multi(MAP)
DIC [48]	**29.94**	49.31	52.85	27.38	36.51	21.28	28.56	**50.21(22.68)**	**52.18(21.19)**	**22.81(7.01)**
ISR [46]	27.53	**53.17**	**55.66**	**31.13**	**38.50**	**23.16**	**33.77**	40.32(14.27)	42.99(15.74)	21.50(6.10)
RKSL [47]	25.76	45.41	50.13	25.79	34.75	17.64	23.01	33.97(11.03)	34.86(10.40)	15.41(4.30)
SAE [83]	20.70	45.33	49.94	21.18	30.51	18.02	24.15	42.40(16.23)	43.97(15.10)	19.29(5.50)
JSTL [32]	25.73	46.26	50.61	24.66	33.15	19.92	25.59	44.69(18.36)	46.41(16.68)	21.24(6.05)
AML [54]	23.10	46.78	51.14	22.19	31.41	20.88	26.39	44.71(18.36)	46.20(16.22)	21.16(6.08)
UsNCA [55]	24.27	47.01	51.70	19.76	29.59	21.07	27.18	45.22(18.91)	47.03(16.91)	22.01(6.53)
DECAMEL	**34.15**	**65.81**	**69.00**	**38.27**	**45.82**	**36.14**	**43.90**	**60.24(32.44)**	**62.98(33.28)**	**30.34(11.13)**

- **Superiority of asymmetric metric: Comparison to the symmetric model**

Dataset	CUHK01	CUHK03	SYSU	Market	ExMarket	MSMT17
Measure	single	single	single	multi(MAP)	multi(MAP)	multi(MAP)
DECMEL	55.95	27.86	25.38	49.94(23.15)	53.00(26.53)	23.88(8.01)
DECAMEL	**65.81**	**38.27**	**36.14**	**60.24(32.44)**	**62.98(33.28)**	**30.34(11.13)**

Deep Unsupervised Person Re-id

- **Challenge for asymmetric learning**
  - Generalization to unseen views
  - Computation grows quickly with more views
- **View Clustering (VC)**

Algorithm 2: DECAMEL with View Clustering

Input: The training images $\mathcal{M}$, the deep feature extractor $f(\cdot;\Theta)$

Training:

Compute the view representations $\{\mathbf{w}_v\}_{v=1}^{V}$ by Eq. (30) .

Conduct k-means clustering in $\{\mathbf{w}_v\}_{v=1}^{V}$ to obtain the cluster separation $\mathcal{B}_j = \{v|\mathbf{w}_v \in j\text{-th cluster}\}$.

For each image M_i in the training set, reassign a view label $v_i' \leftarrow j$ where $v_i \in \mathcal{B}_j$ to it. So we now have $1 \leq v_i' \leq J$ (the number of view clusters).

Feed $\mathcal{M}' = \{\mathbf{M}_i, v_i'\}$ to Algorithm 1 to train a deep framework $g(\cdot,\cdot;\Theta,\{\boldsymbol{U}_j\}_{j=1}^{J})$.

Use the learned feature extractor $g(\cdot;\Theta)$ to compute view prototypes/centroids $\{\mathbf{b}_j\}_{j=1}^{J}$.

Testing for an unseen view u:

Extract the view representation $\mathbf{w}_u$ using the learned feature extractor $g(\cdot;\Theta)$.

Assign this view to a view prototype j where $j = \arg\min_j \|\mathbf{w}_u - \mathbf{b}_j\|_2$.

Assign a view label j to all testing images from this view.

Follow the testing procedure in Algorithm 1.

View distance

$$d_V(View_u, View_v)^2 = \frac{1}{2}(\|\mathbf{m}_u - \mathbf{m}_v\|_2^2 + \|\boldsymbol{\sigma}_u - \boldsymbol{\sigma}_v\|_2^2)$$

View representation

$$\mathbf{w}_v = [\mathbf{m}_v^{\mathrm{T}}, \boldsymbol{\sigma}_v^{\mathrm{T}}]^{\mathrm{T}}$$

View clustering

$$\min_{\{\mathbf{b}_j\}_{j=1}^{J}} f_{vc} = \frac{1}{V}\sum_{j=1}^{J}\sum_{v\in\mathcal{B}_j}\|\mathbf{w}_v - \mathbf{b}_j\|_2^2,$$

Deep Unsupervised Person Re-id

View-extendable setting for Re-ID

TABLE 10
Comparative results in the view-extendable setting. The performances of Rank-1 accuracy (MAP) are ***only*** of all the ***unseen*** views (see the text in Sec. 5.5).

Method	AML	Dic	DECAMEL$_{VC}$
Market-1501	41.57(15.01)	43.82(18.19)	56.50(29.99)
MSMT17	19.77(5.87)	21.04(6.00)	26.42(8.75)

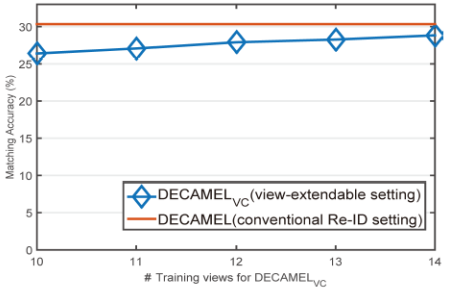


Fig. 10. Performances of DECAMEL$_{VC}$ in the view-extendable setting in the MSMT17 dataset. We fix $J = 10$.

Computation-efficient and also effective

Complexity of metric initialization:
w/o VC: O((Vd)^3), V the # views
w/ VC: O((Jd)^3), J user-defined
(i.e. constant)

Method	Dic [48]	ISR [46]	DECAMEL
Training/Testing Time	35.2h/0.02s	-/98.0s	5.6h/0.02s

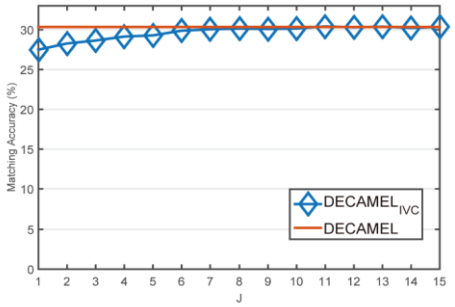


Figure S 1. DECAMEL Initialized with View Clustering (DECAMEL$_{IVC}$) affecting the performance in the MSMT17 dataset. J denotes the number of view clusters.

2. How to match heterogeneous person images across camera views?

Person Re-ID vs. Cross-Modality

- **Matching between Heterogeneous Images**

RGB-Infrared Re-ID

Cross-Modality Learning: RGB-IR Re-ID

- Deep zero-padding

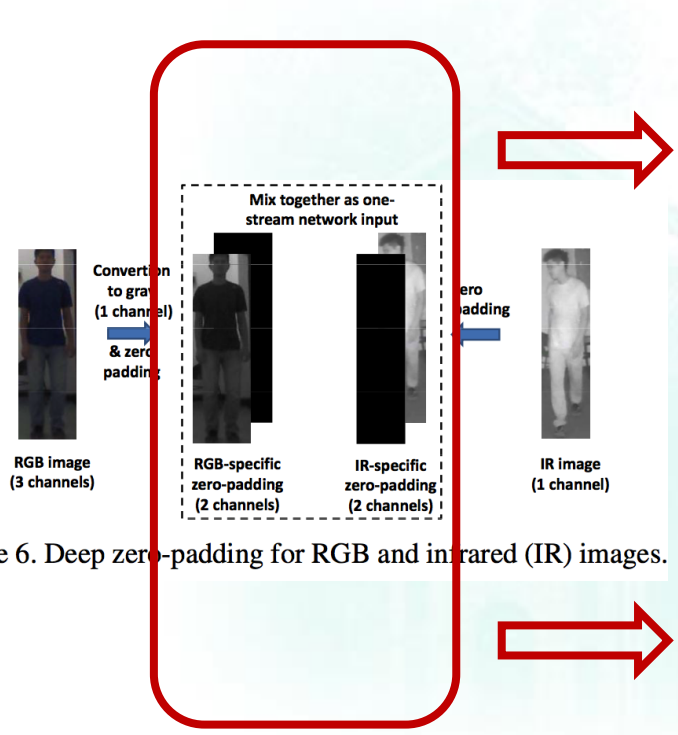
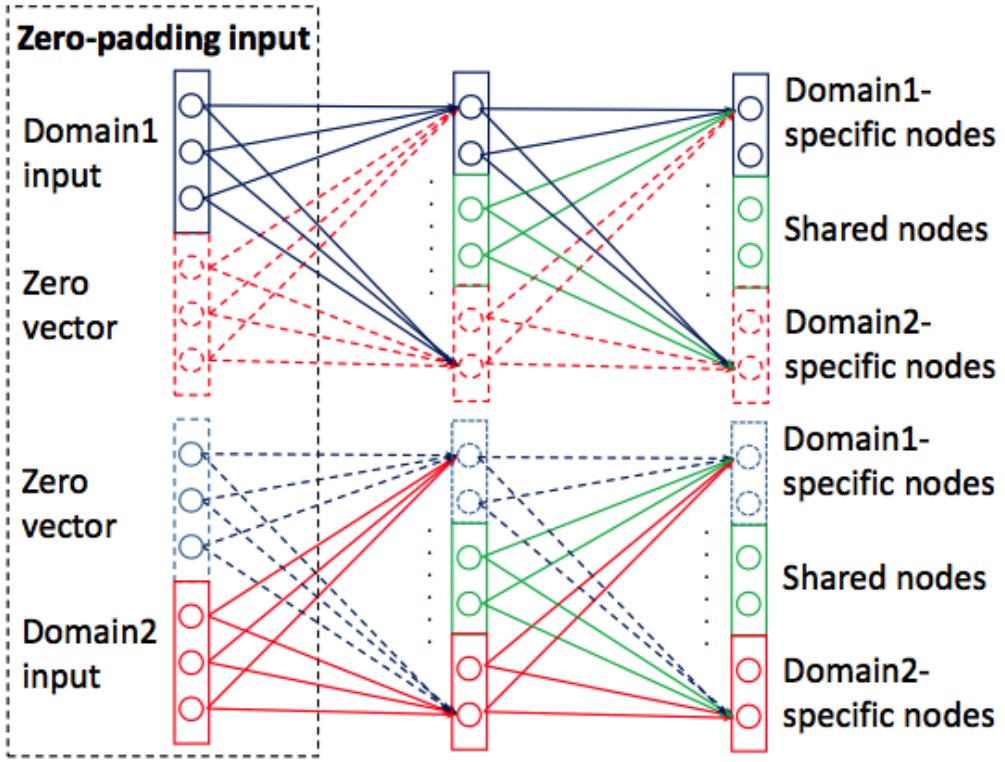


Figure 6. Deep zero-padding for RGB and infrared (IR) images.



Ancong Wu, Wei-Shi Zheng*(PI), Hong-Xing Yu, Shaogang Gong, Jianhuang Lai. RGB-Infrared Cross-Modality Person Re-Identification. In IEEE Conf. on Computer Vision (ICCV), 2017.

RGB-Infrared Re-ID

Cross-Modality Learning: RGB-IR Re-ID

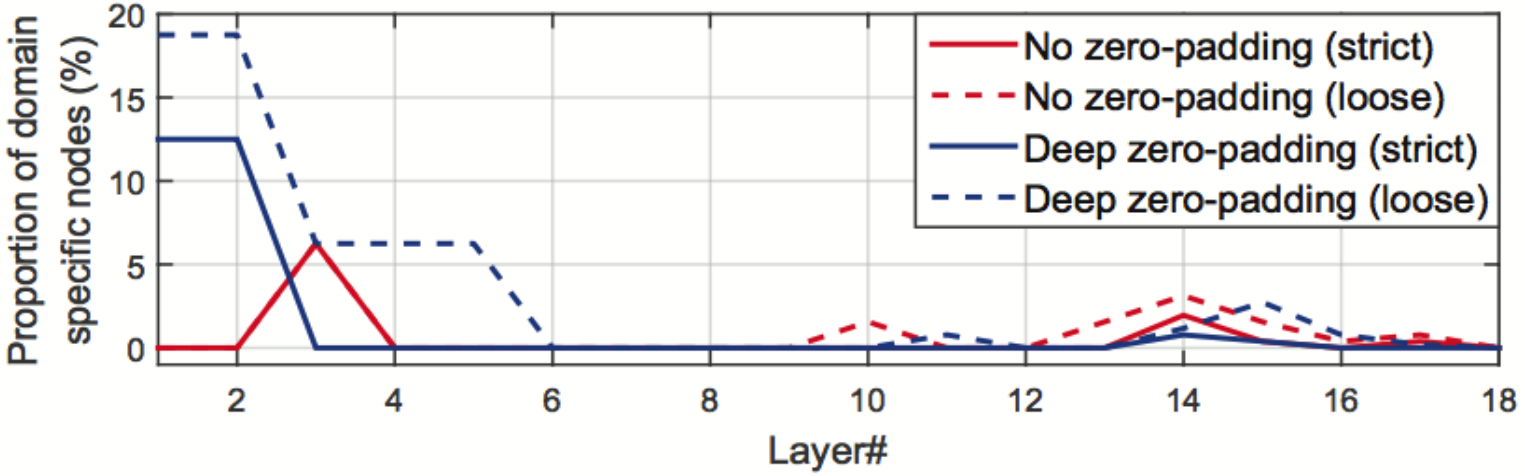


Figure 8. Relation between proportion of domain-specific nodes and layer depth. The x-axis denotes layer depth from bottom to top of the network, and the y-axis denotes the proportion of domain-specific nodes. The strict threshold is $T = 0.01\ std(x_i^{(l)})$ and the loose threshold is $T = 0.05\ std(x_i^{(l)})$ ($std(x_i^{(l)})$ is the standard deviation of the output of the i-th node in layer l). Generally, the proportion of domain-specific nodes using deep zero-padding is higher than that without zero-padding.

RGB-Infrared Re-ID

Cross-Modality Learning: RGB-IR Re-ID

- SYSU RGB-IR Re-ID Dataset

Feature	Metric	All-search								Indoor-search							
		Single-shot				Multi-shot				Single-shot				Multi-shot			
		r1	r10	r20	mAP	r1	r10	r20	mAP	r1	r10	r20	mAP	r1	r10	r20	mAP
One-stream network (deep zero-padding)	Euclidean	**14.80**	**54.12**	**71.33**	**15.95**	**19.13**	**61.40**	**78.41**	**10.89**	**20.58**	**68.38**	**85.79**	**26.92**	**24.43**	**75.86**	**91.32**	**18.64**
One-stream network	Euclidean	12.04	49.68	66.74	13.67	16.26	58.14	75.05	8.59	16.94	63.55	82.10	22.95	22.62	71.74	87.82	15.04
Asymmetric FC layer network	Euclidean	9.30	43.26	60.38	10.82	13.06	52.11	69.52	6.68	14.59	57.94	78.68	20.33	20.09	69.37	85.80	13.04
Lin's	GSM	5.29	33.71	52.95	8.00	6.19	37.15	55.66	4.38	9.46	48.98	72.06	15.57	11.36	51.34	73.41	9.03
HIPHOP	CRAFT	1.80	14.56	26.29	3.40	1.92	16.00	28.31	1.77	2.86	23.40	41.94	7.16	3.01	25.53	44.97	3.43
HOG	Euclidean	2.76	18.25	31.91	4.24	3.82	22.77	37.63	2.16	3.22	24.68	44.52	7.25	4.75	29.06	49.38	3.51
	KISSME	2.12	16.21	29.13	3.53	2.79	18.23	31.25	1.96	3.11	25.47	46.47	7.43	4.10	29.32	50.59	3.61
	LFDA	2.33	18.58	33.38	4.35	3.82	20.48	35.84	2.20	2.44	24.13	45.50	6.87	3.42	25.27	45.11	3.19
	CCA	2.74	18.91	32.51	4.28	3.25	21.82	36.51	2.04	4.38	29.96	50.43	8.70	4.62	34.22	56.28	3.87
	CDFE	2.09	16.68	30.51	3.75	2.47	19.11	34.11	1.86	2.80	23.39	44.46	6.91	3.28	27.31	48.61	3.24
	GMA	1.07	10.42	20.91	2.52	1.03	10.29	20.73	1.39	1.84	17.97	36.14	5.64	1.80	18.10	35.79	2.63
	SCM	1.86	15.16	28.27	3.57	2.40	17.45	31.22	1.66	3.30	25.82	46.23	7.52	3.90	28.84	51.64	3.22
	CRAFT	2.59	17.93	31.50	4.24	3.58	22.90	38.59	2.06	3.03	24.07	42.89	7.07	4.16	27.75	47.16	3.17
LOMO	Euclidean	1.75	14.14	26.63	3.48	1.96	15.06	27.30	1.85	2.24	22.53	41.53	6.64	2.24	22.79	41.80	3.31
	KISSME	2.23	18.95	32.67	4.05	2.65	20.36	34.78	2.45	3.83	31.09	52.86	8.94	4.46	34.35	58.43	4.93
	LFDA	2.98	21.11	35.36	4.81	3.86	24.01	40.54	2.61	4.81	32.16	52.50	9.56	6.27	36.29	58.11	5.15
	CCA	2.42	18.22	32.45	4.19	2.63	19.68	34.82	2.15	4.11	30.60	52.54	8.83	4.86	34.40	57.30	4.47
	CDFE	3.64	23.18	37.28	4.53	4.70	28.23	43.05	2.28	5.75	34.35	54.90	10.19	7.36	40.38	60.33	5.64
	GMA	1.04	10.45	20.81	2.54	0.99	10.50	21.06	1.47	1.79	17.90	36.01	5.63	1.71	18.11	36.17	2.88
	SCM	1.54	14.12	26.27	3.34	1.66	15.17	28.41	1.57	2.86	24.34	44.53	7.06	2.89	25.81	48.33	3.02
	CRAFT	2.34	18.70	32.93	4.22	3.03	21.70	37.05	2.13	3.89	27.55	48.16	8.37	2.45	20.20	38.15	2.69

When the input is not image?

Attribute-Image Person Re-ID

- Match person images with **specific attribute**

Zhou Yin, Wei-Shi Zheng*(PI), et al. Adversarial Attribute-Image Person Re-identification, IJCAI 2018

Attribute-Image Person Re-ID

Method	Market				Duke				PETA			
	rank1	rank5	rank10	mAP	rank1	rank5	rank10	mAP	rank1	rank5	rank10	mAP
DeepCCAE [Wang *et al.*, 2015]	8.12	23.97	34.55	9.72	33.28	59.35	67.64	14.95	14.24	22.09	29.94	14.45
DeepCCA [Andrew *et al.*, 2013]	29.94	**50.70**	**58.14**	17.47	36.71	58.79	65.11	13.53	14.44	20.77	26.31	11.49
2WayNet [Eisenschtat and Wolf, 2017]	11.29	24.38	31.47	7.76	25.24	39.88	45.92	10.19	23.73	38.53	41.93	15.38
DeepMAR [Li *et al.*, 2015]	13.15	24.87	32.90	8.86	36.60	57.70	67.00	14.34	17.80	25.59	31.06	12.67
CMCE [Li *et al.*, 2017]	35.04	**50.99**	56.47	**22.80**	39.75	56.39	62.79	15.40	31.72	39.18	48.35	26.23
ours w/o adv	33.83	48.17	53.48	17.82	39.30	55.88	62.50	15.17	36.34	48.48	53.03	25.35
ours w/o sc	2.08	4.80	4.80	1.00	5.26	9.37	10.87	1.56	3.43	4.15	4.15	5.80
ours w/o adv+MMD	34.15	47.96	57.20	18.90	41.77	**62.32**	**68.61**	14.23	**39.31**	48.28	**54.88**	**31.54**
ours w/o adv+DeepCoral	**36.56**	47.61	55.92	20.08	**46.09**	**61.02**	68.15	**17.10**	35.62	**48.65**	53.75	27.58
ours	**40.26**	49.21	**58.61**	**20.67**	**46.60**	59.64	**69.07**	**15.67**	**39.00**	**53.62**	**62.20**	**27.86**

Our model:

Wrong samples

- Outperforms traditional cross modality retrieval methods (DeepCCAE, DeepCCA, 2WayNet, CMCE)
-
-

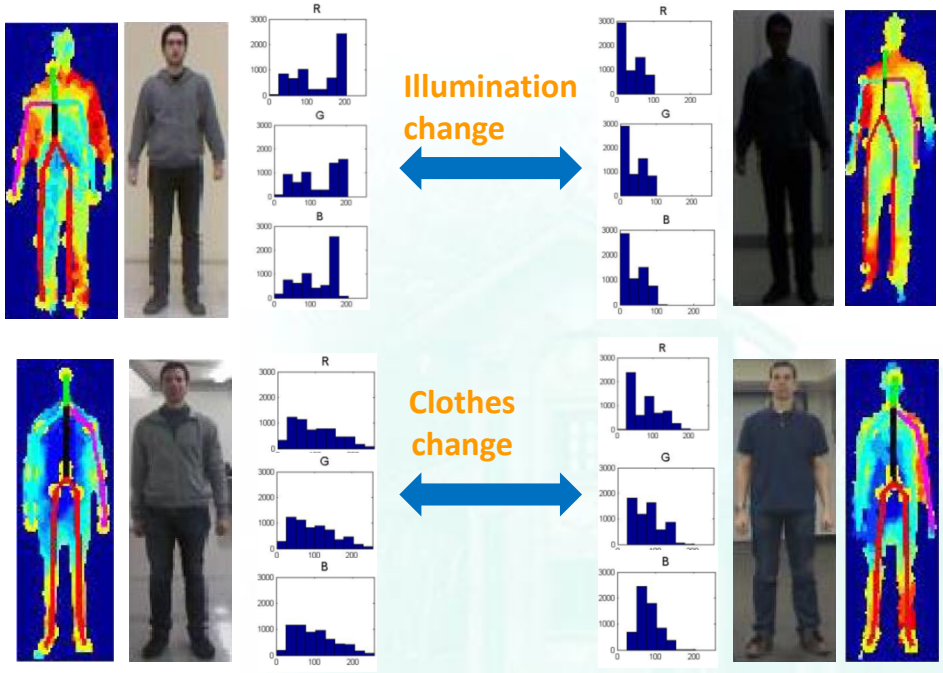
ibutions
al).

Part of query attributes	Teenager	Backpack	Bag	DownBlue	UpWhite	UpPink
	☑	☒	☒	☑	☑	☒

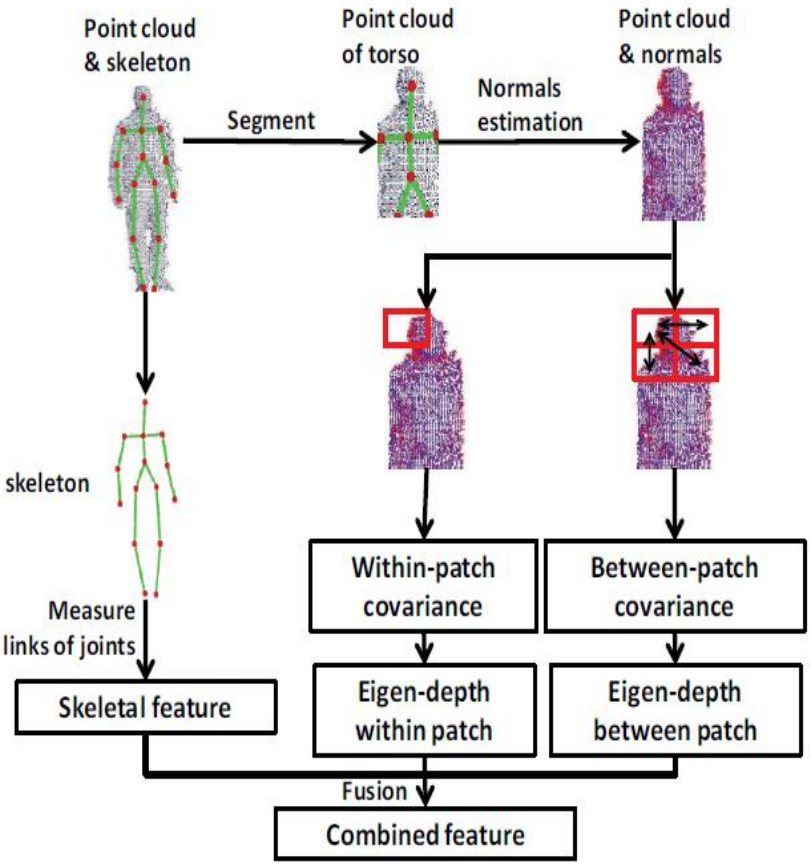
When dressing differently?

Depth Re-ID

- **Something to see**



In these cases, appearance cues are not reliable.



Depth Re-ID

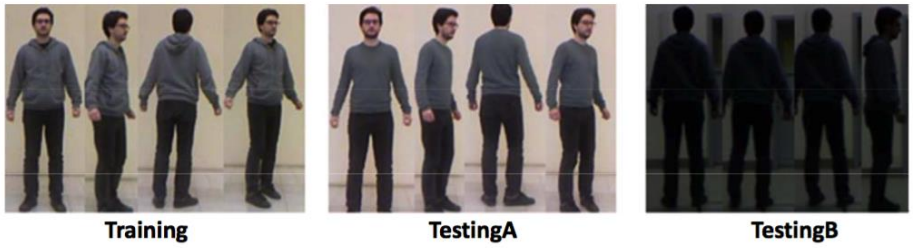


Fig. 10. Examples of images in IAS-Lab RGBD-ID. All samples were captured from multiple views. Compared to "Training", samples in "TestingA" changed clothes and some samples in "TestingB" were captured in dark environment.

Fig. 6. Examples of images in "Walking1" and "Walking2" in PAVIS. Most persons in "Walking2" dressed different clothes from "Walking1".

TABLE III

PAVIS DATASET: RANK-1 AND RANK-5 ACCURACIES (%), INCLUDING RESULTS OF OUR PROPOSED METHODS AND COMPARISONS WITH RGB-BASED APPEARANCE FEATURES AND DEPTH-BASED FEATURES

Setting	Single-shot		Multi-shot	
Rank	1	5	1	5
RGB-based appearance features				
LOMO [12]	12.05	35.03	19.74	44.36
ELF18 [50]	52.15	77.85	52.62	78.26
Color Hist [11]	47.90	74.97	48.92	74.82
HOG [13]	45.03	73.49	45.33	73.95
LBP [80]	42.92	71.33	45.64	72.36
Depth-based features				
RIFT2M [5]	7.13	22.77	8.77	27.69
Fehr's [6]	24.26	51.64	30.56	58.67
Skeleton [2]	33.13	67.85	37.33	71.13
Proposed				
DVCov (depth voxel covariance)	61.49	81.23	66.00	82.92
DVCov+SKL	**67.64**	**87.33**	**71.74**	**88.46**
ED (Eigen-depth feature)	44.67	72.10	51.59	76.15
ED+SKL	55.95	84.77	61.23	87.64

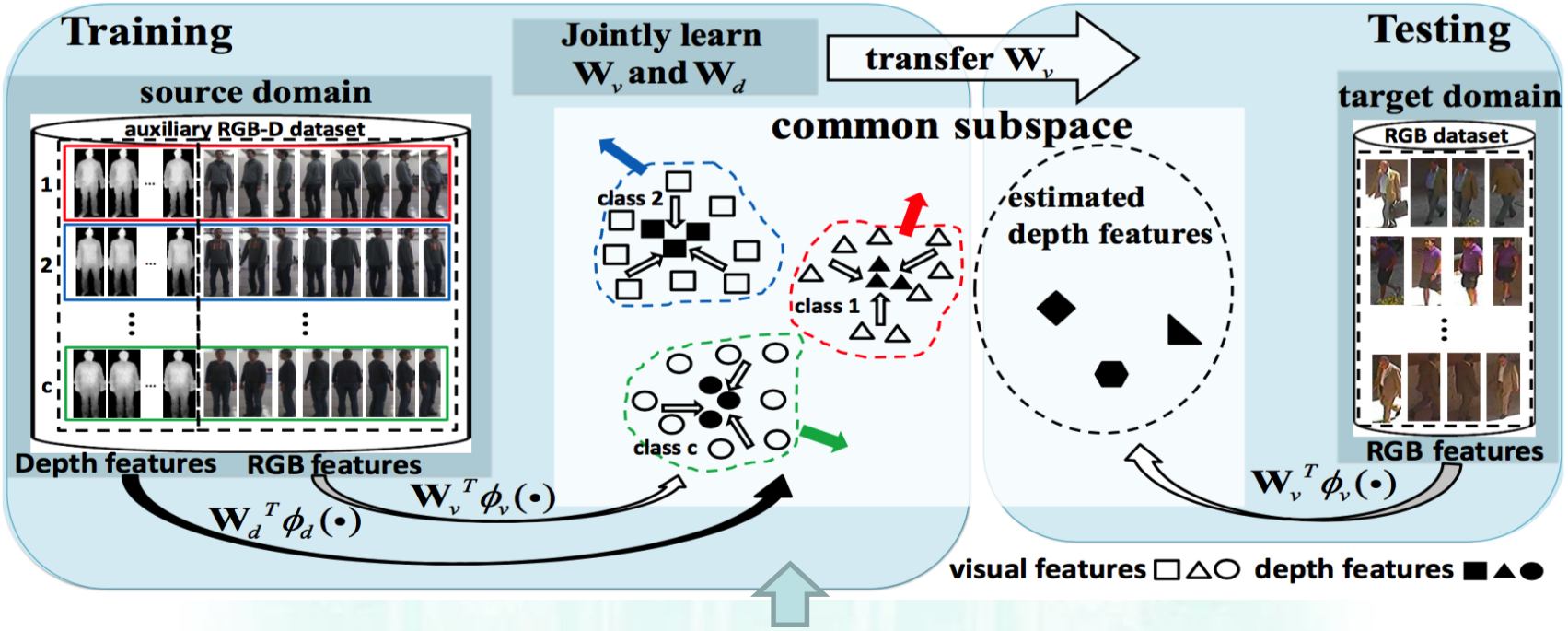
TABLE IV

BIWI RGBD-ID DATASET "STILL" AND "WALKING": RANK-1 AND RANK-5 ACCURACIES (%), INCLUDING RESULTS OF OUR PROPOSED METHODS AND COMPARISONS WITH RGB-BASED APPEARANCE FEATURES AND DEPTH-BASED FEATURES

Probe	Still				Walking			
Setting	Single-shot		Multi-shot		Single-shot		Multi-shot	
Rank	1	5	1	5	1	5	1	5
RGB-based appearance features								
LOMO [12]	9.07	28.21	18.17	35.47	8.74	23.33	10.31	25.39
ELF18 [50]	2.79	18.18	4.11	19.13	1.32	16.03	1.50	16.77
Color Hist [11]	7.02	25.47	10.61	31.92	5.43	19.56	5.86	21.70
HOG [13]	8.42	25.69	12.35	30.39	6.38	21.00	6.94	23.29
LBP [80]	7.37	26.04	10.87	33.57	4.87	20.04	5.34	23.31
Depth-based features								
RIFT2M [5]	4.04	19.52	4.34	20.78	3.25	17.46	3.75	18.31
Fehr's [6]	12.08	38.17	14.06	43.78	9.33	32.39	12.09	39.60
Skeleton [2]	21.34	53.32	26.55	62.73	14.52	42.36	16.94	47.18
Proposed								
DVCov	16.32	45.93	23.07	58.89	12.58	39.22	17.24	45.93
DVCov+SKL	23.49	57.06	34.37	72.77	16.59	46.67	21.40	54.12
ED	28.98	61.85	36.22	**73.11**	20.90	51.98	28.71	63.85
ED+SKL	**30.52**	**67.86**	**39.38**	72.13	**24.47**	**60.63**	**29.96**	**65.18**

Depth Re-ID

Learning & Transferring Depth

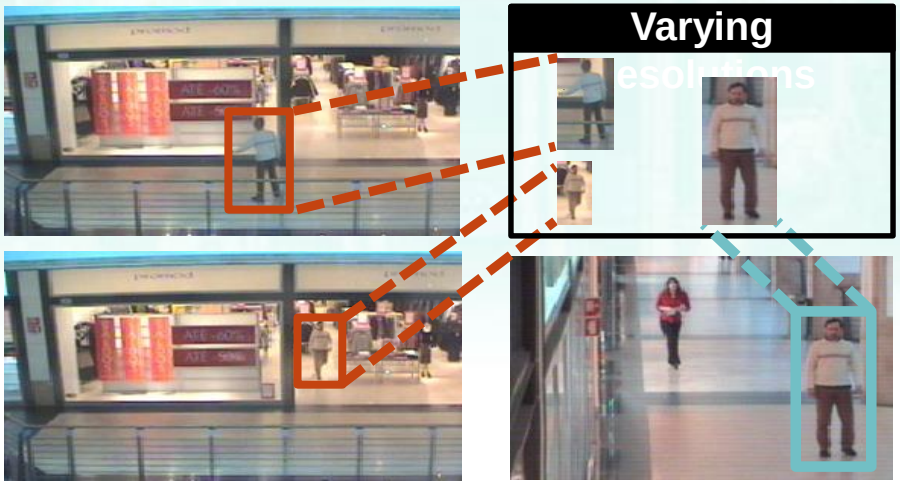


Learning relation between RGB features and depth features

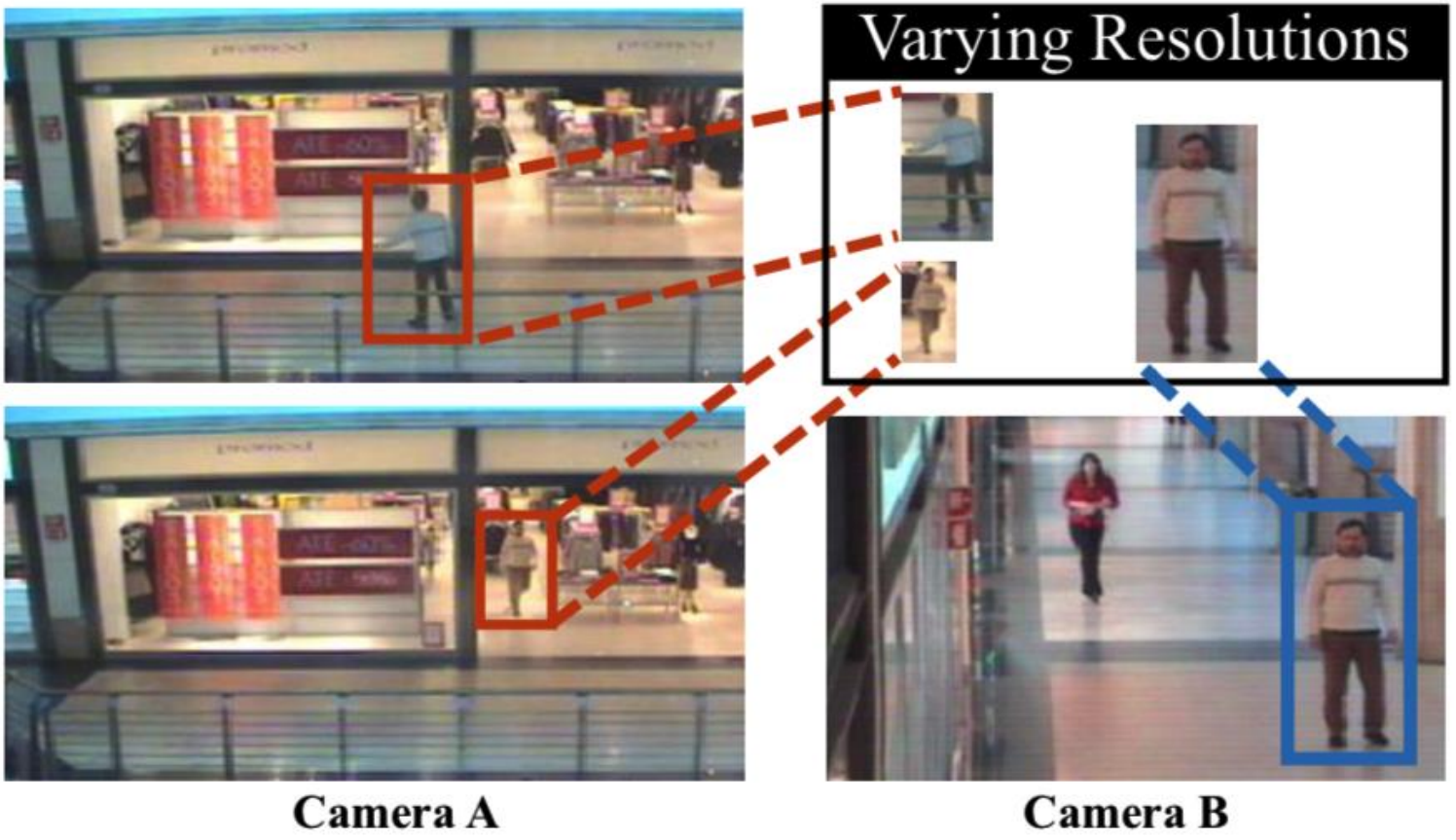
Dataset	Probe	ED	ED+SKL	3D RAM [4]	PCM [3]	PCM+SKL [3]	SKL [3]
PAVIS	Walking2	54.4	**57.0**	41.3	-	-	28.6
IAS-Lab	TestingA	44.0	**49.9**	48.3	28.6	25.6	22.5
RGBD-ID	TestingB	55.5	**66.6**	63.7	43.7	63.3	55.5

Ancong Wu, **Wei-Shi Zheng*(PI),** and Jian-Huang Lai. Robust Depth-based Person Re-identification. IEEE Transactions on Image Processing, 2017

3. Low-resolution Person Re-identification



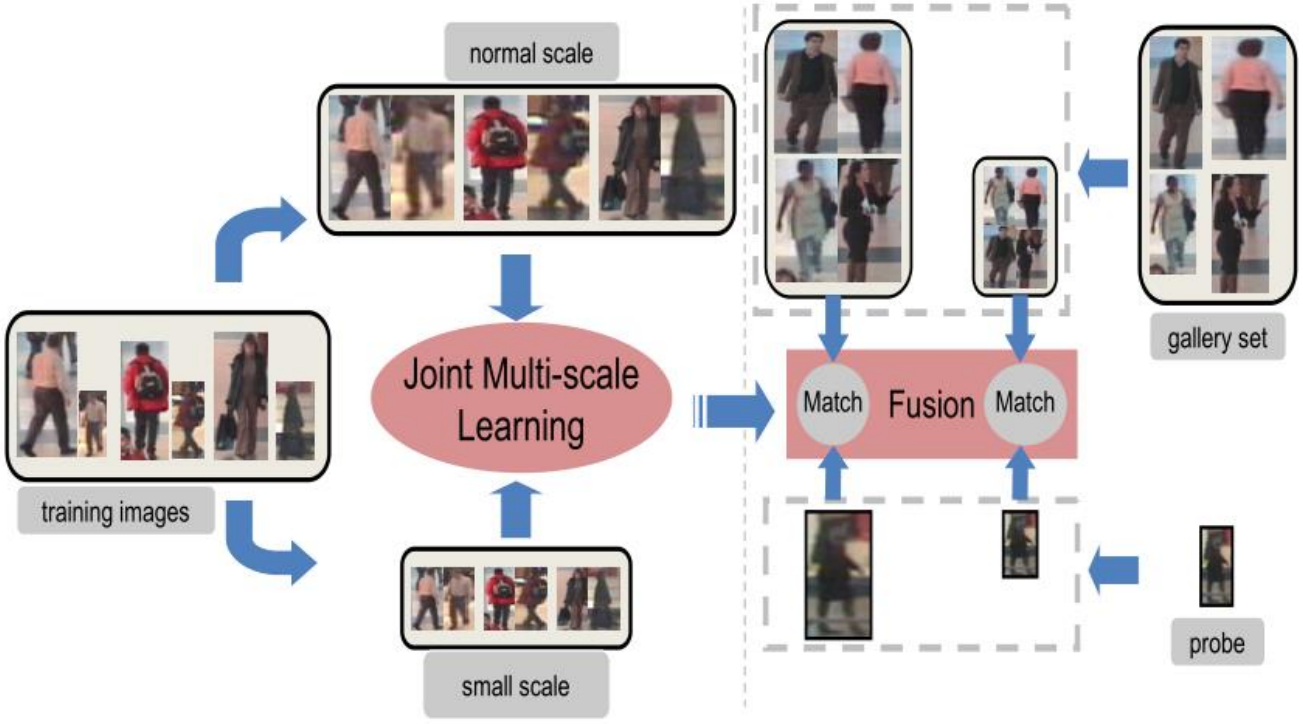
Low-resolution Re-ID



Low-resolution Re-ID

Low-resolution Re-ID

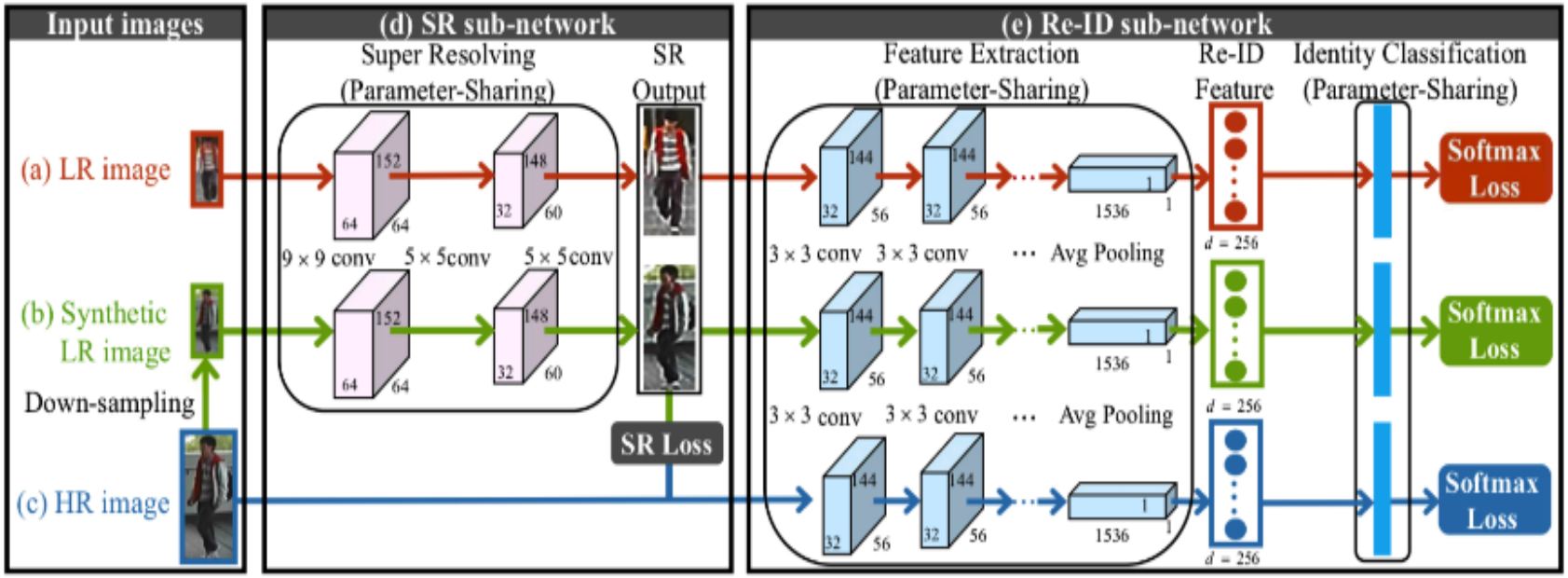
- JUDEA : joint multi-scale discriminant component analysis



Xiang Li, Wei-Shi Zheng*(PI), Xiaojuan Wang, Tao Xiang, Shaogang Gong. Multi-scale Learning for Low-resolution Person Re-identification. IEEE Conf. on Computer Vision (ICCV), 2015.

Low-resolution Re-ID

- **Super- resolution and Identity joiNt learninG (SING)**



Jiening Jiao, Wei-Shi Zheng*(PI), Ancong Wu, Xiatian Zhu, and Shaogang Gong. Deep Low-resolution Person Re-identification. AAAI 2018

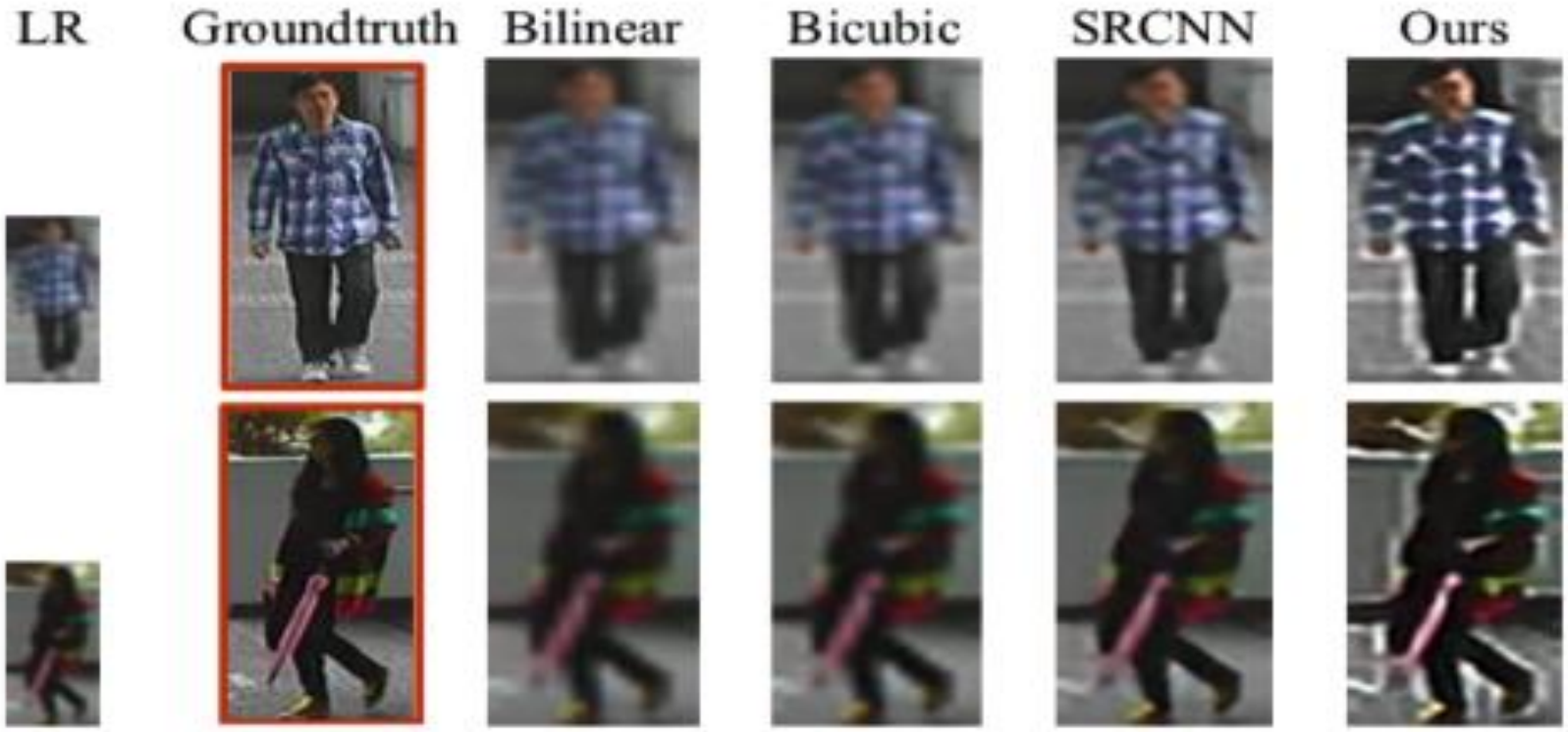
Low-resolution Re-ID

- Super- resolution and Identity joiNt learninG (SING)

Low-resolution Re-ID

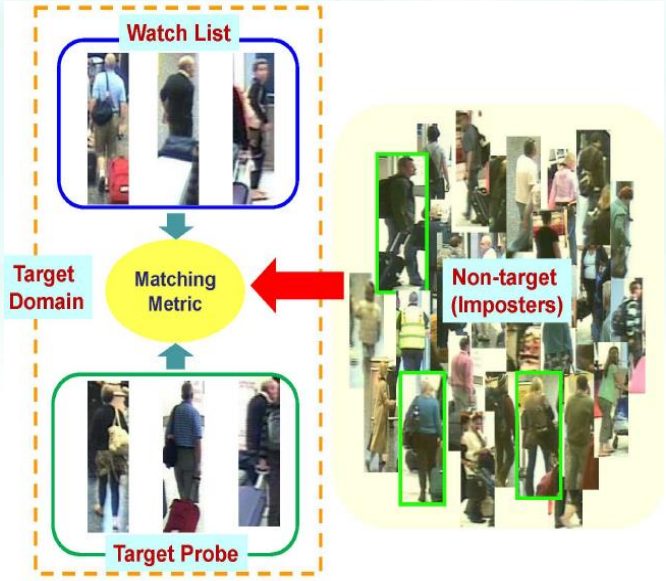
Results

Table 1: Comparing state-of-the-art LR re-id methods (%).
The 1st /2nd best results are indicated in red/blue



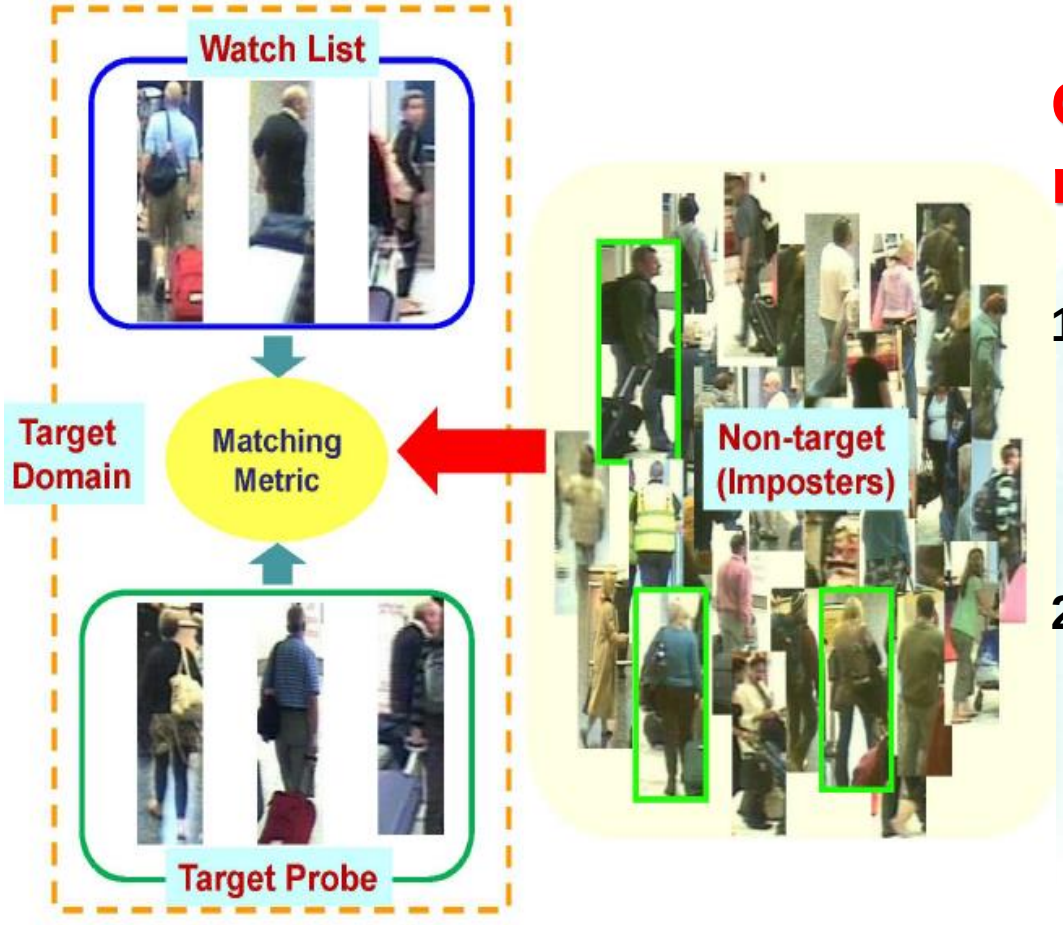
SDF	9.32	38.1	32.4	68.0
SING	33.5	57.0	66.5	76.6

4. Open-world Person Re-identification



One-Shot Open-World Group-based Re-id

Motivation



Open-world person re-identification setting

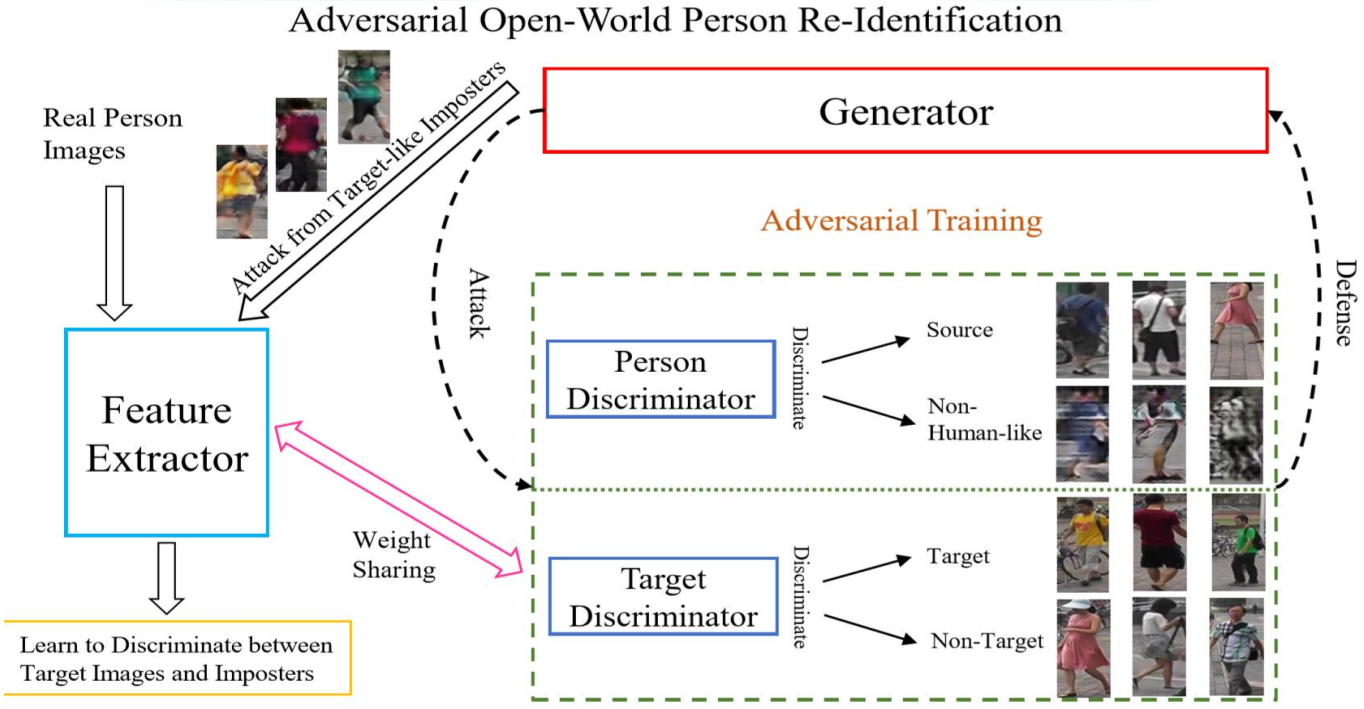
1) A large amount of non-target imposters captured along with the target people on the watch list.

2) Their images will also appear in the probe set and some of them will look visually similar to the target people

Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Towards Open-World Person Re-Identification by One-Shot Group-based Verification. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 38, no. 3, pp. 591-606, 2016.

Adversarial Open-World Re-id

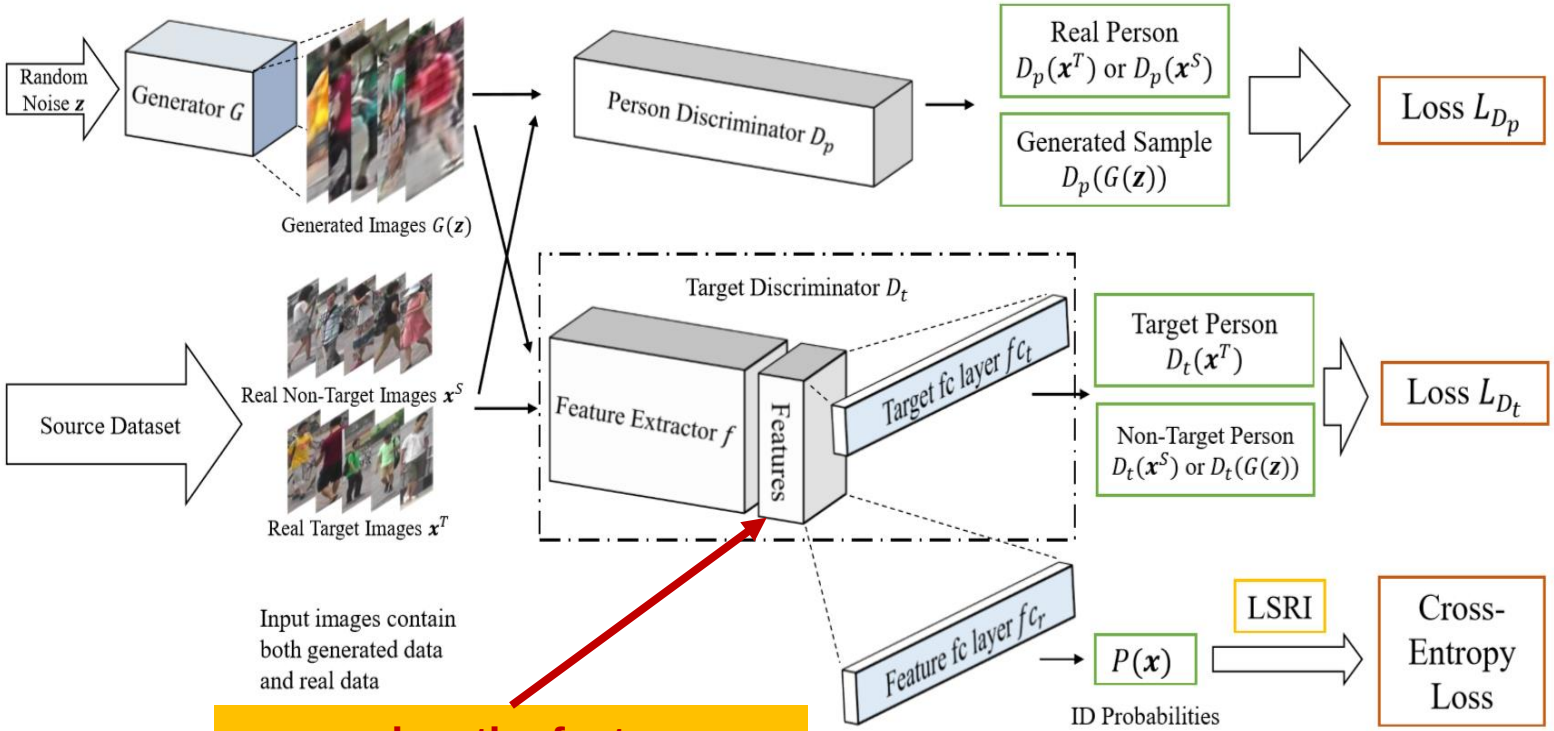
- Learn to attack feature extractor on the target people using GAN to generate very target-like images (imposters
- Make the feature extractor learn to tolerate the attack by discriminative learning so as to realize group-based verification.



Xiang Li, Ancong Wu, Wei-Shi Zheng*(PI). Adversarial Open-World Person Re-Identification. In European Conference on Computer Vision (ECCV), 2018.

Adversarial Open-World Re-id

- Jointly learns a generator, a person discriminator, a target discriminator and a feature extractor
- The feature extractor and target discriminator share the same weights



Adversarial Open-World Re-id

Adversarial Open-World Re-id

Results

Table 1: Comparison with typical person re-identification: TTR (%) against FTR

Dataset	Market-1501						CHUK01						CUHK03					
FTR	0.1%	1%	5%	10%	20%	30%	0.1%	1%	5%	10%	20%	30%	0.1%	1%	5%	10%	20%	30%
Evaluation	Set Verification																	
t-LRDC [42]	3.00	18.88	42.06	51.07	65.24	75.54	5.56	5.56	38.89	50.00	66.67	83.33	8.87	20.16	32.66	37.90	49.60	58.06
XICE [46]	6.77	21.69	45.17	58.88	73.68	81.80	11.11	33.33	44.44	55.56	72.22	83.33	3.14	13.03	31.65	44.36	59.70	69.98
GOG+XQDA [22]	0.43	2.15	9.01	17.17	28.76	39.06	0	5.56	33.33	38.89	66.67	72.22	10.48	16.53	27.02	36.69	53.63	60.48
LOMO+XQDA [22]	4.72	14.16	35.62	46.35	58.80	65.67	0	5.56	38.89	44.44	72.22	**88.89**	25.81	38.31	51.61	61.69	71.77	83.47
hiphop+CRAFT [6]	2.15	9.44	27.04	38.63	48.07	55.36	11.11	22.22	38.89	55.56	**77.78**	83.33	23.79	33.87	42.74	47.58	55.65	59.68
JSTL-DGD [38]	26.92	61.54	80.00	88.46	92.31	94.61	33.33	33.33	33.33	55.56	55.56	66.67	38.10	59.52	71.43	76.19	88.10	92.86
ResNet-50 [12]	34.62	80.00	93.85	96.92	98.46	99.23	44.44	**55.56**	**55.56**	55.56	**77.78**	77.78	61.90	73.81	90.48	**95.24**	**95.24**	**95.24**
DCGAN+LSRO [43]	36.15	78.46	94.62	96.15	**99.23**	99.23	44.44	**55.56**	**55.56**	55.56	55.56	77.78	64.29	71.43	88.10	90.48	92.86	**95.24**
DeepFool [28]	34.62	78.46	94.63	96.92	96.92	99.23	44.44	**55.56**	**55.56**	55.56	66.67	77.78	64.29	76.19	90.48	**95.24**	**95.24**	**95.24**
APN	**43.85**	**82.31**	**96.92**	**98.46**	**99.23**	**100**	**55.56**	**55.56**	**55.56**	**66.67**	**77.78**	77.78	**66.67**	**78.57**	**92.86**	**95.24**	**95.24**	**95.24**
Evaluation	Individual Verification																	
t-LRDC [42]	15.54	39.89	51.44	68.49	78.10	87.63	15.23	32.15	51.82	67.56	73.54	89.13	16.57	37.40	48.98	58.83	70.76	90.17
XICE [46]	34.64	61.58	84.87	90.68	96.86	97.21	33.33	36.11	55.56	55.56	72.22	88.89	18.63	48.94	71.89	81.64	89.71	97.43
GOG+XQDA [22]	10.49	30.60	51.83	62.77	77.29	86.14	25.00	55.56	88.89	91.67	97.22	**100**	33.93	45.73	64.93	77.85	87.40	91.99
LOMO+XQDA [22]	25.32	59.10	81.98	86.96	92.96	94.84	5.56	36.11	80.56	88.89	88.89	88.89	40.72	57.79	77.78	86.90	94.44	96.03
hiphop+CRAFT [6]	31.75	62.59	84.09	91.42	93.55	96.30	**50.00**	**72.22**	**100**	**100**	**100**	**100**	42.39	63.27	77.38	89.58	95.68	**99.18**
JSTL-DGD [38]	47.23	63.85	86.92	93.73	93.73	97.53	33.33	48.15	59.26	72.84	72.84	82.72	53.74	78.18	81.67	92.15	92.15	94.87
ResNet-50 [12]	82.26	95.86	98.54	99.38	**99.58**	**99.58**	44.44	50.00	72.22	77.78	83.33	83.33	76.19	91.67	**95.24**	**95.24**	95.24	95.24
DCGAN+LSRO [43]	81.71	95.36	98.33	98.96	99.17	**99.58**	44.44	61.11	72.22	77.78	83.33	88.83	73.81	90.48	**95.24**	**95.24**	95.24	95.24
DeepFool [28]	82.26	95.86	95.86	98.96	99.17	**99.58**	44.44	61.11	72.22	77.78	83.33	83.33	75.61	91.67	**95.24**	**95.24**	95.24	95.24
APN	**84.00**	**96.72**	**98.69**	**99.58**	**99.58**	**99.58**	44.44	61.11	77.78	77.78	83.33	88.89	**79.54**	**94.05**	**95.24**	**95.24**	**97.15**	97.15

More ……

Hash Re-ID for Fast Search

- **FAST Re-ID on Numbers of Cameras**
  - Learning view-specific hash code for each camera

$$f_p(\boldsymbol{x}_i^p) = \boldsymbol{x}_i^p \boldsymbol{W}_p, \qquad f_g(\boldsymbol{x}_j^g) = \boldsymbol{x}_j^g \boldsymbol{W}_g$$

$$\boldsymbol{B}_p = \text{sign}(\boldsymbol{X}_p \boldsymbol{W}_p) \in \{-1, 1\}^{n_p \times c},$$
$$\boldsymbol{B}_g = \text{sign}(\boldsymbol{X}_g \boldsymbol{W}_g) \in \{-1, 1\}^{n_g \times c},$$

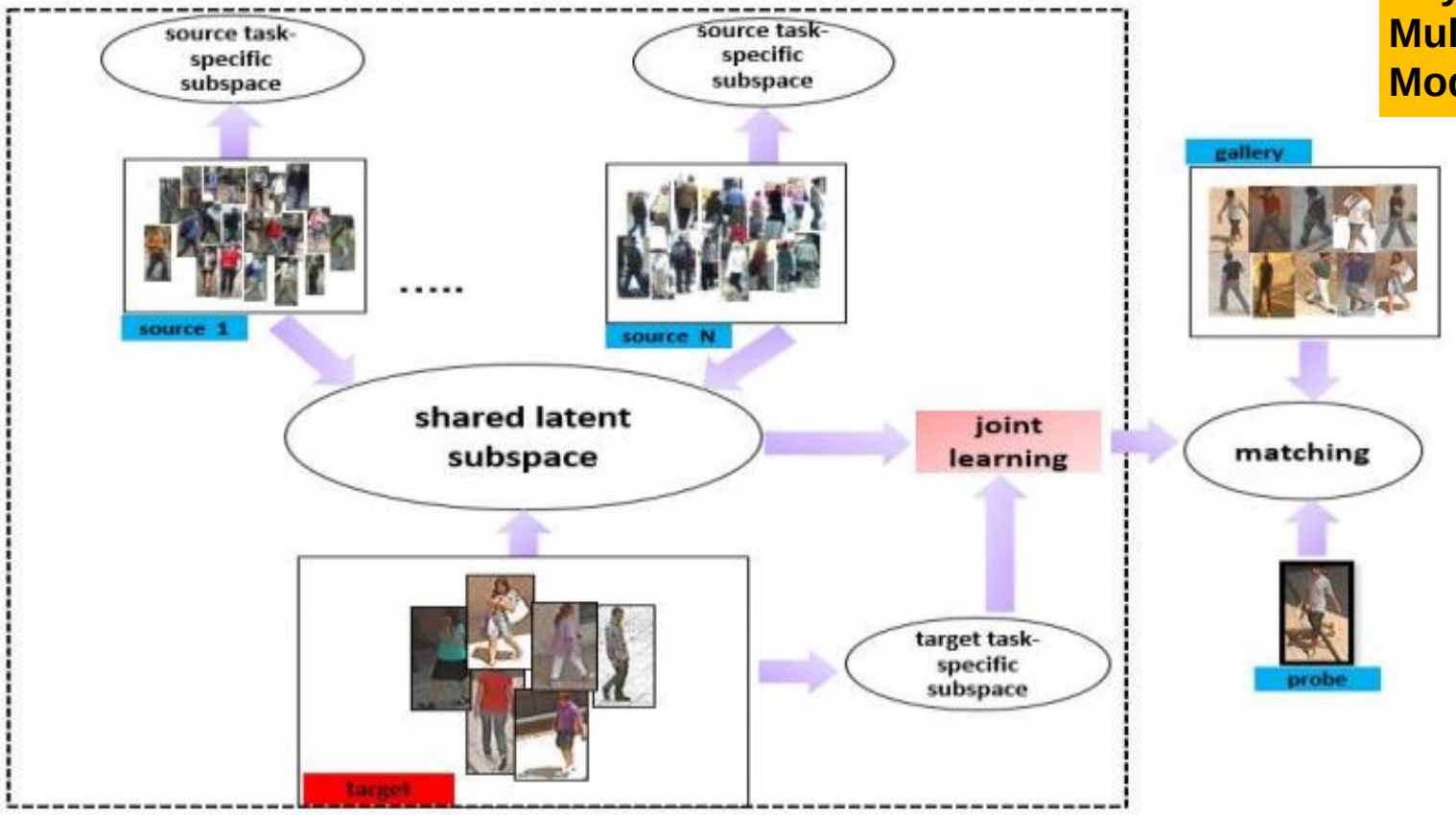
Xiatian Zhu, Botong Wu, Dongcheng Huang, Wei-Shi Zheng*(PI). Fast Open-World Person Re-Identification. IEEE Transactions on Image Processing, 2018.

Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Towards Open-World Person Re-Identification by One-Shot Group-based Verification. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 38, no. 3, pp. 591-606, 2016.

Cross-scenario Re-ID

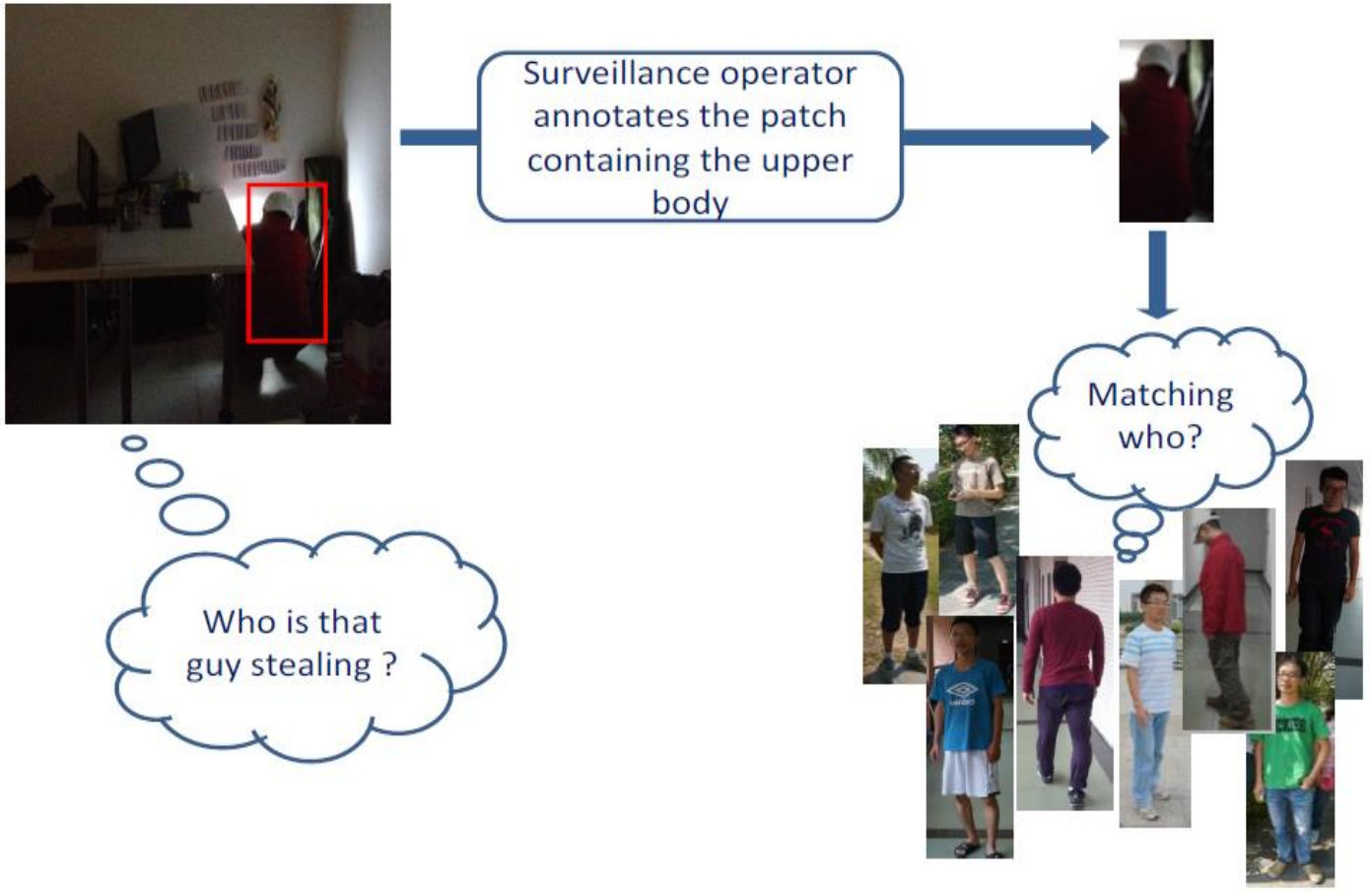
❑ Transferring between sets

An Asymmetric Multi-task Modelling



Xiaojuan Wang, Wei-Shi Zheng*(PI), Xiang Li, and Jianguo Zhang. Cross-scenario Transfer Person Re-identification. IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 8, pp. 1447-1460, 2016.

Partial Re-ID



Wei-Shi Zheng, Xiang Li, Tao Xiang, Shengcai Liao, JianHuang Lai, Shaogang Gong. Partial Person Re-identification. ICCV, 2015.

More

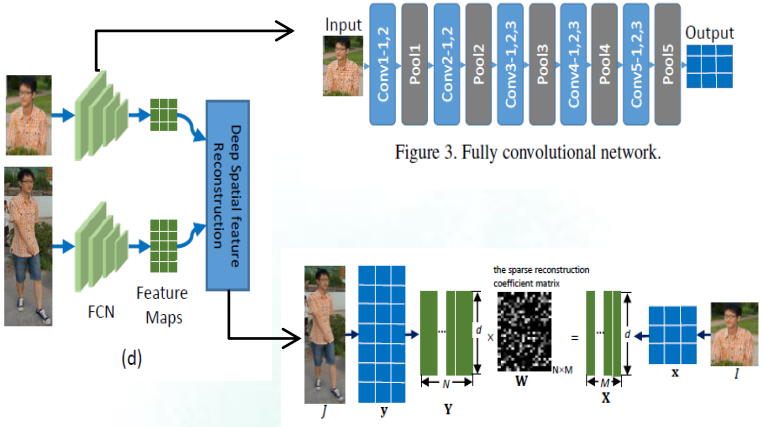
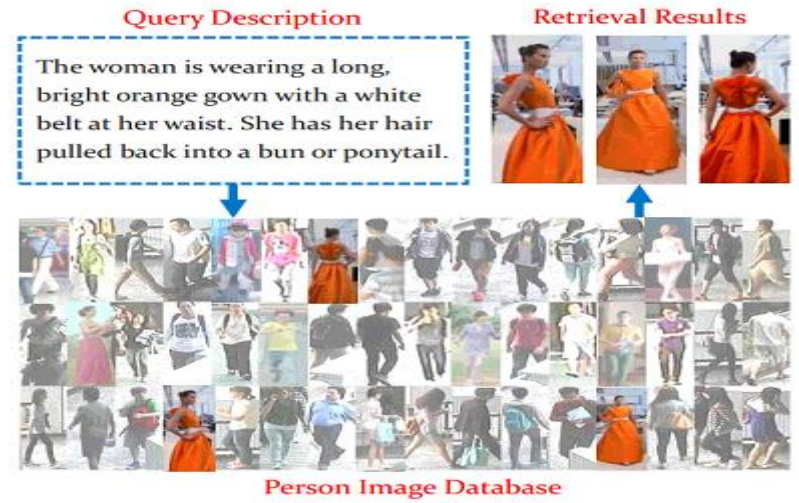


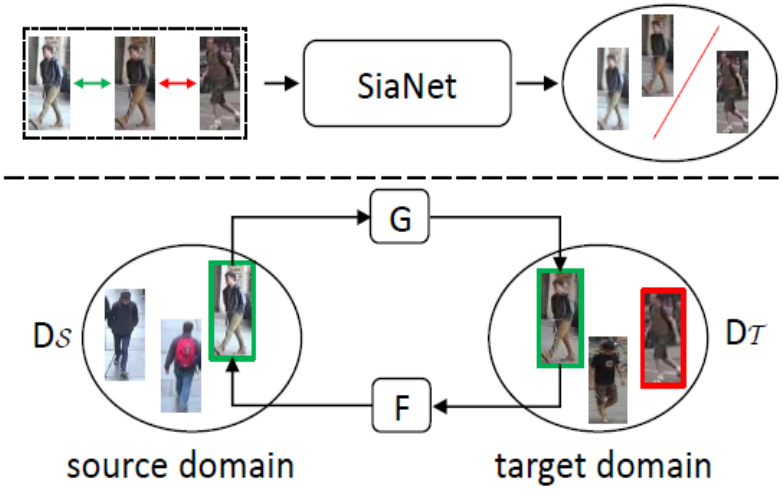
Figure 3. Fully convolutional network.

Figure 4. Deep Spatial feature Reconstruction.

Lingxiao He, et al., CVPR 2018.



S.Li et al., CVPR 2017



Weijian Deng, et al,. CVPR, 2018.

Longhui Wei, et al,. CVPR, 2018.

Take Home Message

- Person Re-id is far from being solved
- Unsupervised Learning….
- Illumination, Occlusion, Low-resolution
- Cross-modal Searching
- Open-world, ….
- More?

Intelligence Science & System Lab

智能科学与系统实验室

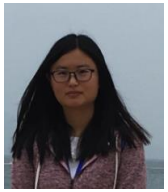
Thanks to my students









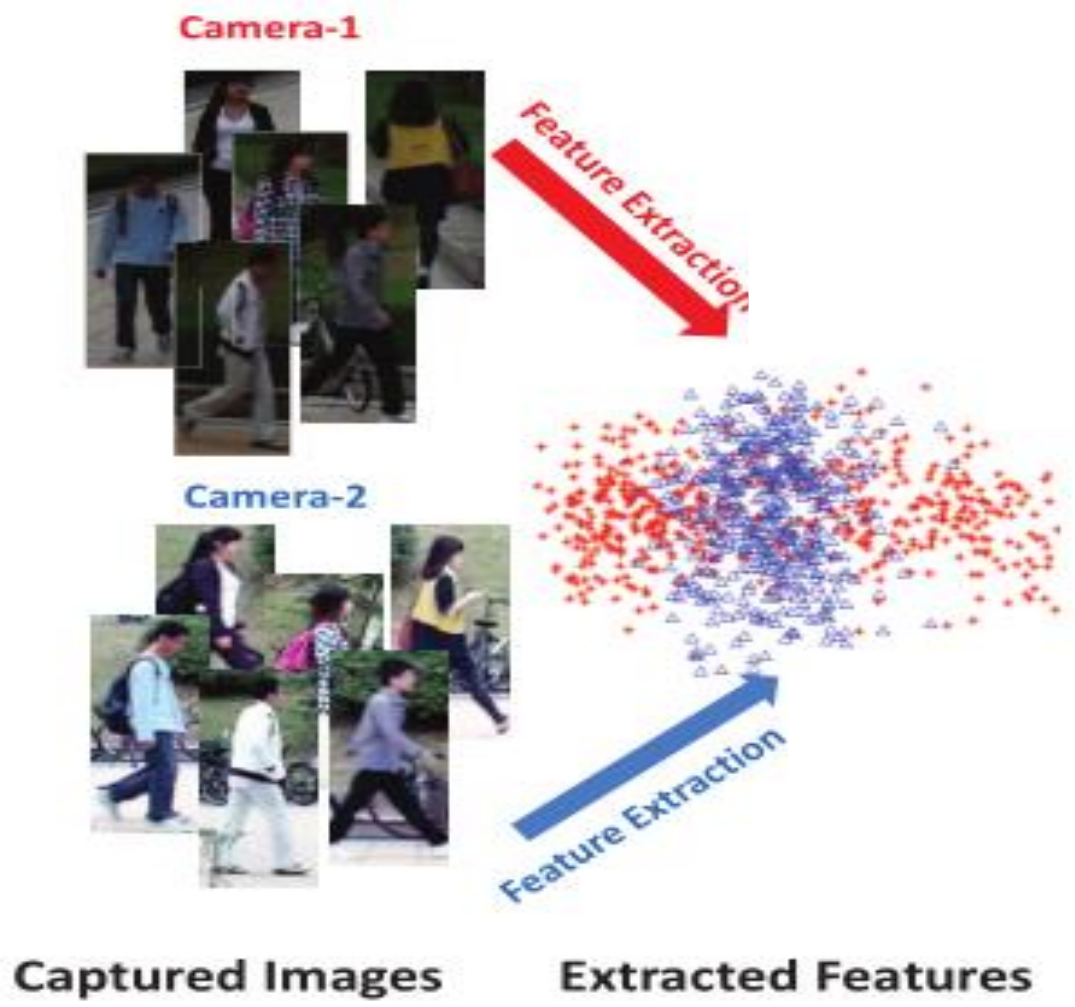


http://isee.sysu.edu.cn/~zhwshi

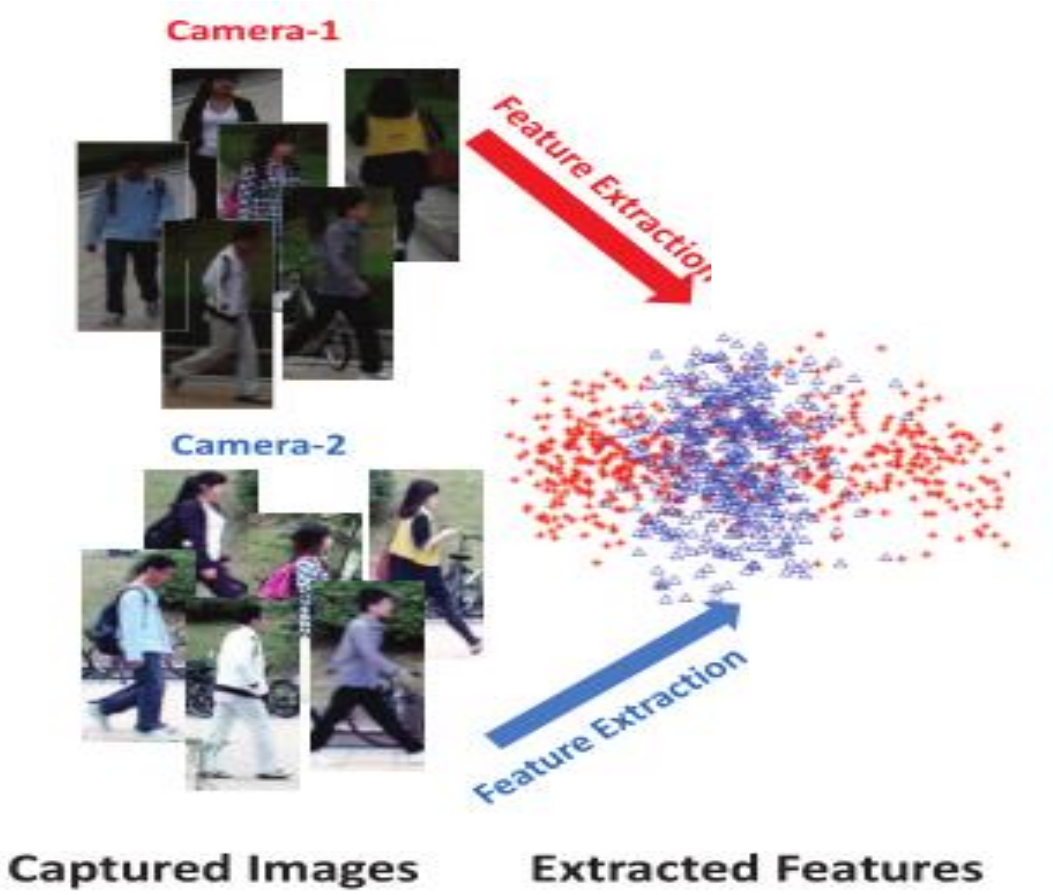
EMAIL ME：wszheng@ieee.org

Person Re-ID vs. Cross-Modality

- **View Bias**



Asymmetric Metric for Re-ID



Learning universal feature transformation

Learning view-specific feature transformation

Asymmetric Metric for Re-ID

$$d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \mathbf{M} (\boldsymbol{x}_i - \boldsymbol{x}_j)}$$
$$= ||\boldsymbol{U}^T \boldsymbol{x}_i - \boldsymbol{U}^T \boldsymbol{x}_j||_2,$$

Learn different feature transformation for different camera views

Pseudometric

$$d(\{\boldsymbol{x}_i^p, p\}, \{\boldsymbol{x}_j^q, q\}) = ||\boldsymbol{U}^{pT} \boldsymbol{x}_i^p - \boldsymbol{U}^{qT} \boldsymbol{x}_j^q||_2$$
$$\boldsymbol{U}^p \neq \boldsymbol{U}^q$$

Non-negativity **Symmetry**

$$\begin{aligned} d(\{\boldsymbol{x}_i^p, p\}, \{\boldsymbol{x}_j^q, q\}) &= ||\boldsymbol{U}^{pT} \boldsymbol{x}_i^p - \boldsymbol{U}^{qT} \boldsymbol{x}_j^q||_2 \\ &= ||\boldsymbol{U}^{qT} \boldsymbol{x}_j^q - \boldsymbol{U}^{pT} \boldsymbol{x}_i^p||_2 \\ &= d(\{\boldsymbol{x}_j^q, q\}, \{\boldsymbol{x}_i^p, p\}), \end{aligned}$$

Triangle Inequality

$$||\boldsymbol{U}^{rT} \boldsymbol{x}_k^r - \boldsymbol{U}^{qT} \boldsymbol{x}_j^q||_2 \leq$$
$$||\boldsymbol{U}^{rT} \boldsymbol{x}_k^r - \boldsymbol{U}^{pT} \boldsymbol{x}_i^p||_2 + ||\boldsymbol{U}^{pT} \boldsymbol{x}_i^p - \boldsymbol{U}^{qT} \boldsymbol{x}_j^q||_2.$$

Coincidence

$$d(\{\boldsymbol{x}^p, p\}, \{\boldsymbol{x}^q, q\}) = 0$$
$$\boldsymbol{U}^{pT} \boldsymbol{x}^p = \boldsymbol{U}^{qT} \boldsymbol{x}^q \quad \Leftarrow \quad \boldsymbol{U}^{pT} \boldsymbol{x}^p = \boldsymbol{U}^{qT} \boldsymbol{x}^q$$
$$||\boldsymbol{U}^p - \boldsymbol{U}^q||_F^2$$

Asymmetric Metric for Re-ID

- Re-ID Reformulation by Augmentation

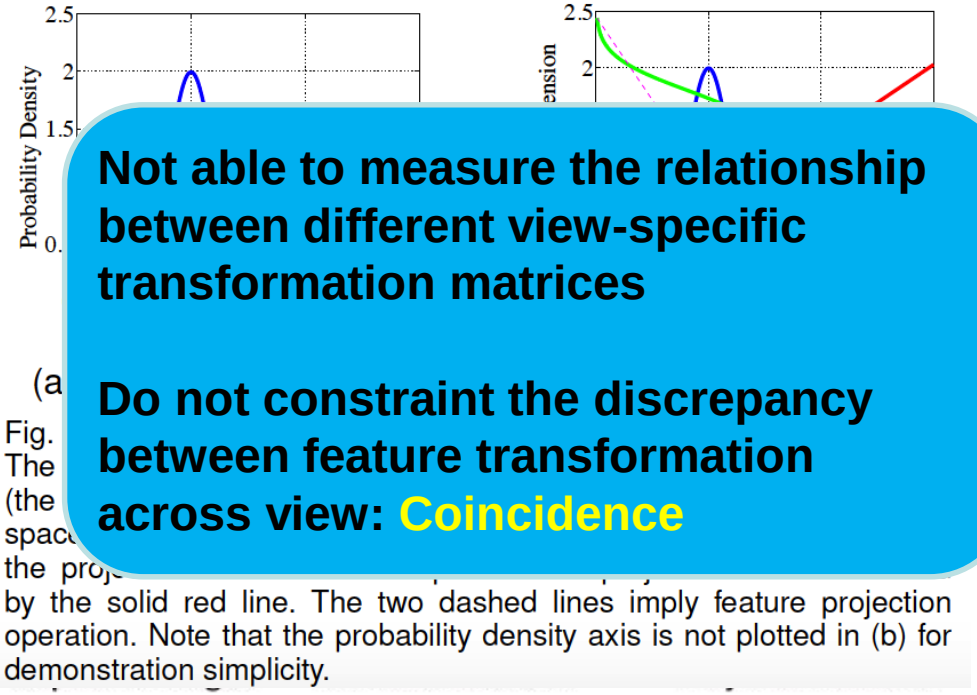
$$\tilde{\boldsymbol{X}}^a_{\text{zp}} = \begin{bmatrix} \boldsymbol{I}_{d\times d} \\ \boldsymbol{O}_{d\times d} \end{bmatrix} \boldsymbol{X}^a, \qquad \tilde{\boldsymbol{X}}^b_{\text{zp}} = \begin{bmatrix} \boldsymbol{O}_{d\times d} \\ \boldsymbol{I}_{d\times d} \end{bmatrix} \boldsymbol{X}^b$$

$$\hat{\boldsymbol{W}} = \min_{\boldsymbol{W}} f_{\text{obj}}(\boldsymbol{W}^\top \tilde{\boldsymbol{X}}_{\text{zp}})$$

$$\hat{\boldsymbol{W}} = [(\hat{\boldsymbol{W}}^a)^\top, (\hat{\boldsymbol{W}}^b)^\top]^\top$$

View-specific transformation

$$\hat{\boldsymbol{W}}^\top \tilde{\boldsymbol{X}}^a_{\text{zp}} = (\hat{\boldsymbol{W}}^a)^\top \boldsymbol{X}^a$$
$$\hat{\boldsymbol{W}}^\top \tilde{\boldsymbol{X}}^b_{\text{zp}} = (\hat{\boldsymbol{W}}^b)^\top \boldsymbol{X}^b$$



Not able to measure the relationship between different view-specific transformation matrices

Do not constraint the discrepancy between feature transformation across view: Coincidence

Yingcong Chen, Xiatian Zhu, Wei-Shi Zheng*, and Jian-Huang Lai. Person Re-Identification by Camera Correlation Aware Feature Augmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 2017.

Asymmetric Metric for Re-ID

Adaptive feature augmentation

$$\tilde{\boldsymbol{X}}^a_{\text{zp}} = \begin{bmatrix} \boldsymbol{I}_{d\times d} \\ \boldsymbol{O}_{d\times d} \end{bmatrix} \boldsymbol{X}^a, \qquad \tilde{\boldsymbol{X}}^b_{\text{zp}} = \begin{bmatrix} \boldsymbol{O}_{d\times d} \\ \boldsymbol{I}_{d\times d} \end{bmatrix} \boldsymbol{X}^b$$

$$\tilde{\boldsymbol{X}}^a_{\text{craft}} = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{M} \end{bmatrix} \boldsymbol{X}^a, \qquad \tilde{\boldsymbol{X}}^b_{\text{craft}} = \begin{bmatrix} \boldsymbol{M} \\ \boldsymbol{R} \end{bmatrix} \boldsymbol{X}^b$$

generalised

$$f_a(\tilde{\boldsymbol{X}}^a_{\text{craft}}) = \boldsymbol{W}^\top \tilde{\boldsymbol{X}}^a_{\text{craft}} = (\boldsymbol{R}^\top \boldsymbol{W}^a + \boldsymbol{M}^\top \boldsymbol{W}^b)^\top \boldsymbol{X}^a$$

$$f_b(\tilde{\boldsymbol{X}}^b_{\text{craft}}) = \boldsymbol{W}^\top \tilde{\boldsymbol{X}}^b_{\text{craft}} = (\boldsymbol{M}^\top \boldsymbol{W}^a + \boldsymbol{R}^\top \boldsymbol{W}^b)^\top \boldsymbol{X}^b$$

control the discrepancy **Between** f_a **and** f_b

Yingcong Chen, Xiatian Zhu, Wei-Shi Zheng*, and Jian-Huang Lai. Person Re-Identification by Camera Correlation Aware Feature Augmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 2017.

Asymmetric Metric for Re-ID

- **Learning:**

  Camera coRrelation Aware Feature augmenTation (CRAFT)

$$\hat{\boldsymbol{W}} = \arg\min_{\boldsymbol{W}} f_{\text{obj}}(\boldsymbol{W}^\top \tilde{\boldsymbol{X}}_{\text{craft}}) + \lambda \text{tr}(\boldsymbol{W}^\top \boldsymbol{C}\boldsymbol{W})$$

Generalize any symmetric metric learning models to asymmetric ones: e.g. MFA

$$\min_{\boldsymbol{H}} \sum_{i \neq j} \boldsymbol{A}^c_{ij} ||\boldsymbol{H}^\top (\ddot{\boldsymbol{x}}_i - \ddot{\boldsymbol{x}}_j)||_2^2 + \lambda \text{tr}(\boldsymbol{H}^\top \boldsymbol{H})$$

$$\text{s.t.} \sum_{i \neq j} \boldsymbol{A}^p_{ij} ||\boldsymbol{H}^\top (\ddot{\boldsymbol{x}}_i - \ddot{\boldsymbol{x}}_j)||_2^2 = 1,$$

$$\boldsymbol{A}^c_{ij} = \begin{cases} 1 & \text{if } i \in N^+_{k_1}(j) \text{ or } j \in N^+_{k_1}(i) \\ 0 & \text{otherwise,} \end{cases}$$

$$\boldsymbol{A}^p_{ij} = \begin{cases} 1 & \text{if } (i,j) \in P_{k_2}(y_i) \text{ or } (i,j) \in P_{k_2}(y_j) \\ 0 & \text{otherwise,} \end{cases}$$