Deep association learning for unsupervised video person re-identification

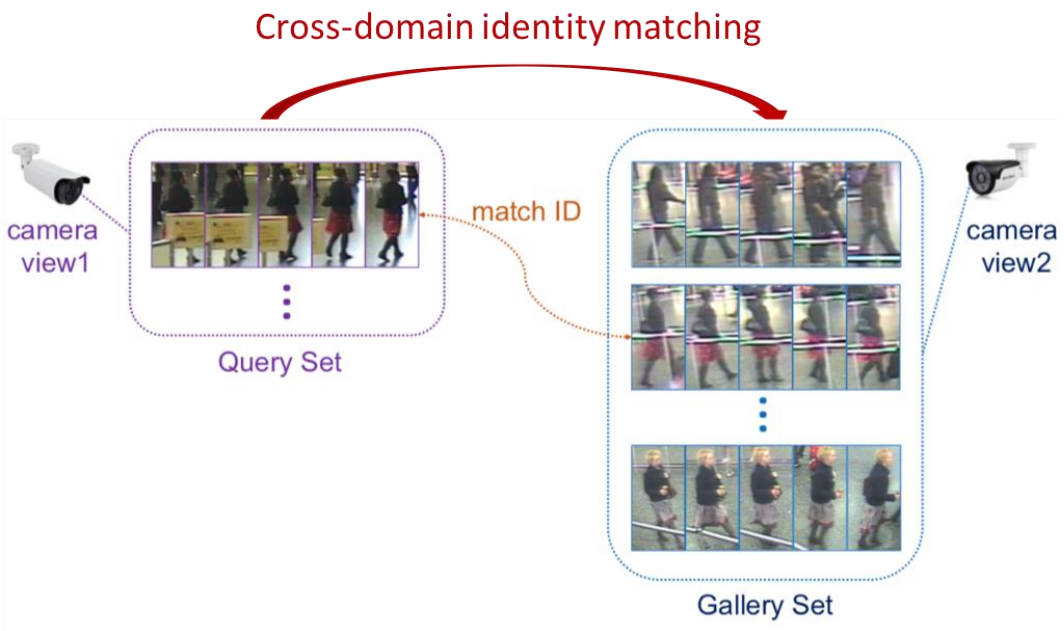
Yanbei Chen, Xiatian Zhu, Shaogang Gong

Published in:
British Machine Vision Conference (BMVC) 2018

Centre for Intelligent Sensing
Queen Mary University of London

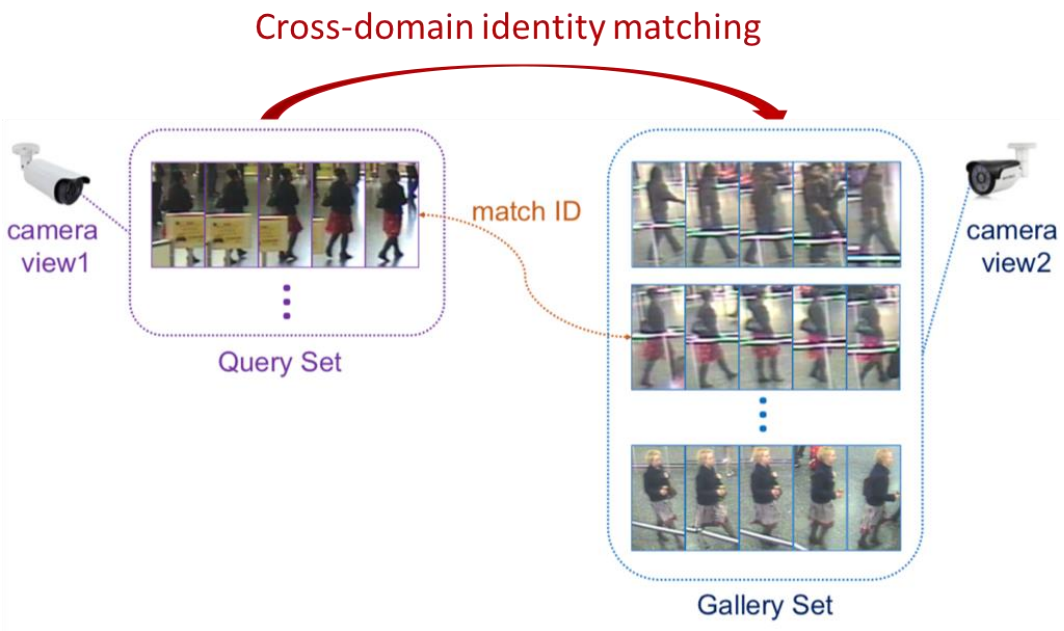
Problem

- **Video Person Re-Identification (ReID)**
- A task to match person identities in the video tracklets captured from disjoint surveillance camera views.



Problem

- **Unsupervised Video Person ReID**
- The key is to formulate supervision signals without utilising any pairwise ID matching information to guide model learning

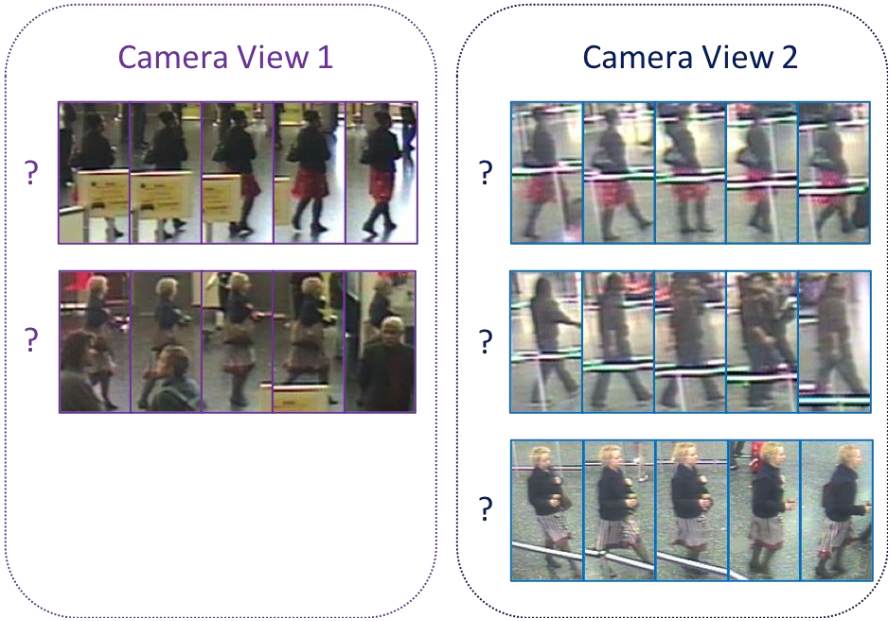


Challenges

- **Unsupervised Video Person ReID**

1) Unsupervised learning problem

Missing pairwise identity labels



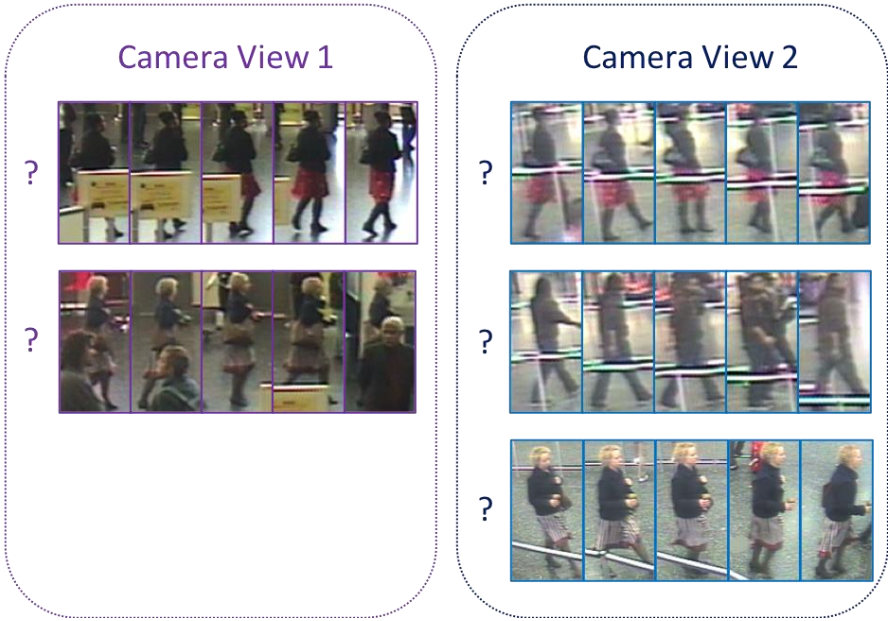
Challenges

- **Unsupervised Video Person ReID**

2) Cross-domain image retrieval challenge

Large cross-camera variations:

illumination, viewpoints, resolution, occlusion, background clutter.

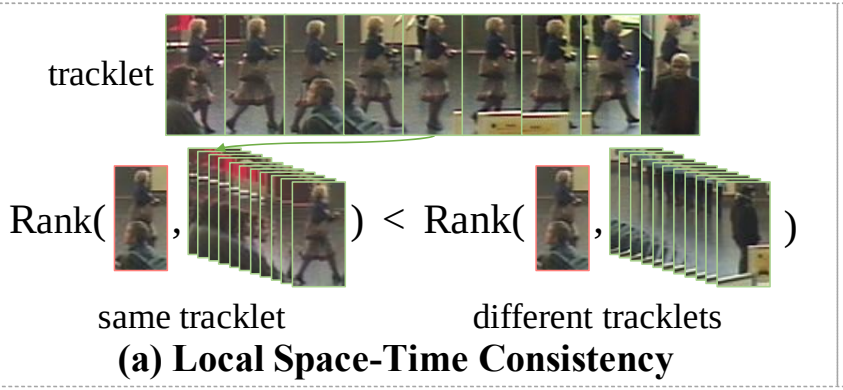


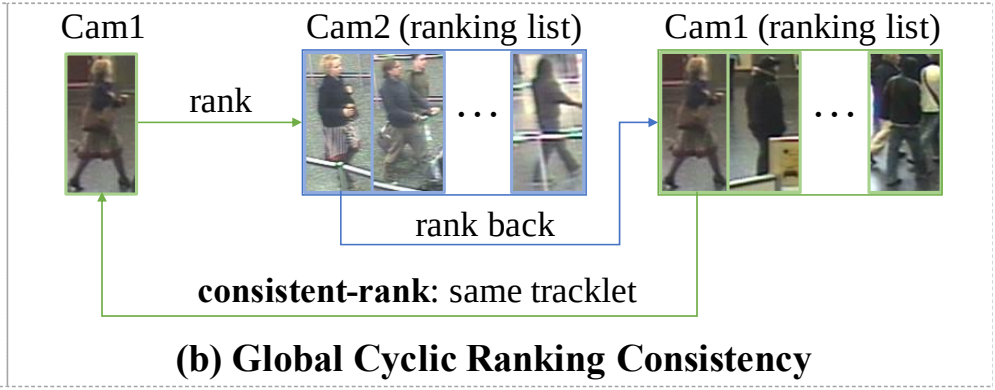
Methodology

- **Deep Association Learning**

(1) intra-camera image-to-tracklet association

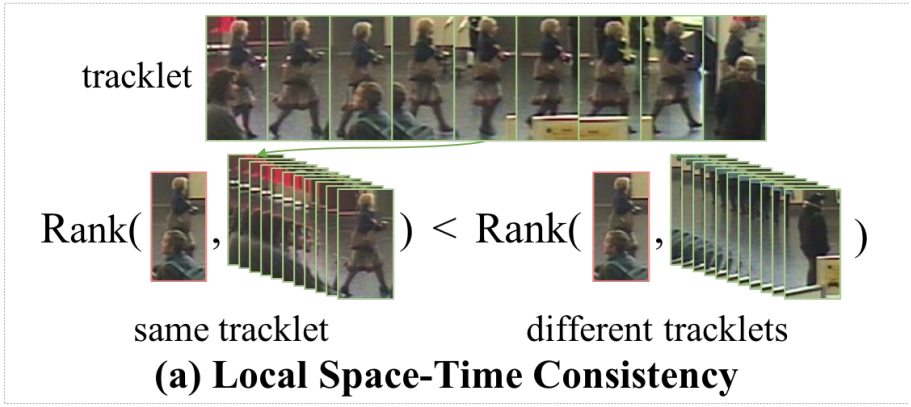
(2) cross-camera tracklet-to-tracklet association

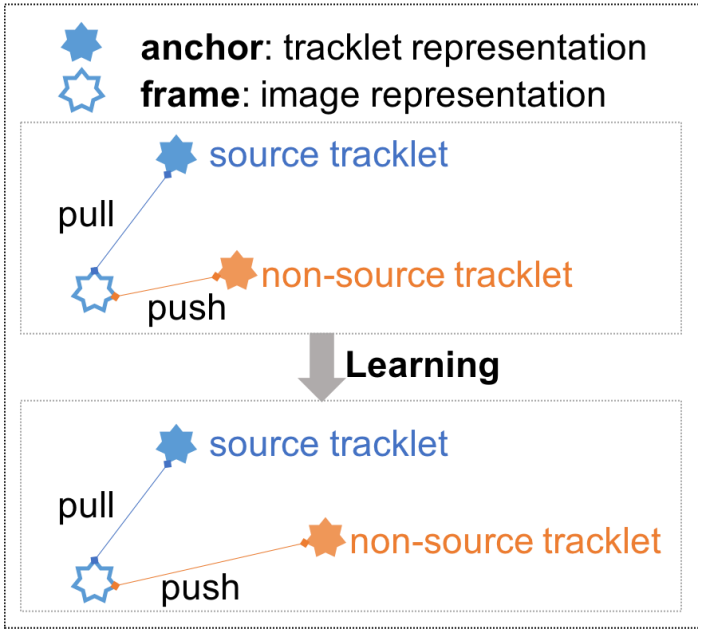




Methodology

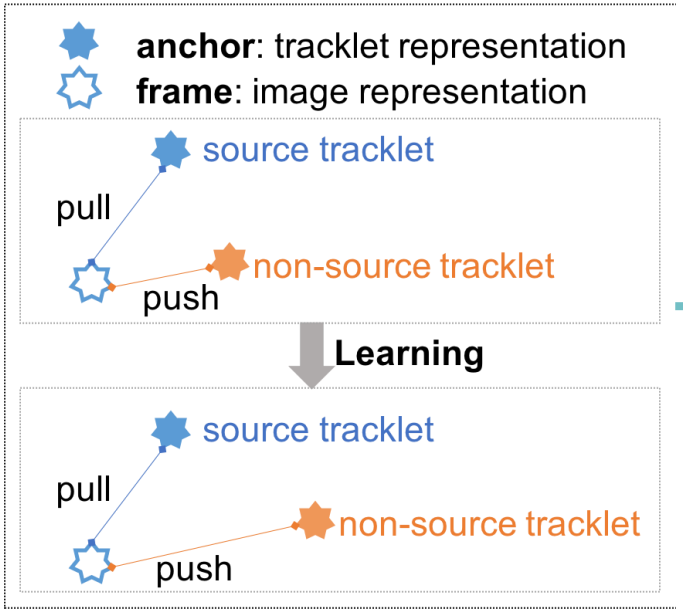
- **Intra-Camera Association Learning**
- Learning constraint on image-to-tracklet association under the same camera view



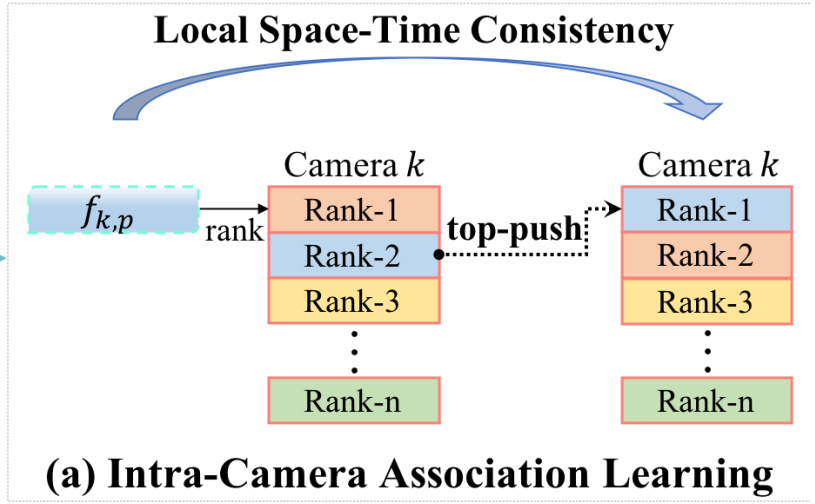


Methodology

- **Intra-Camera Association Learning**
- Learning constraint on image-to-tracklet association under the same camera view

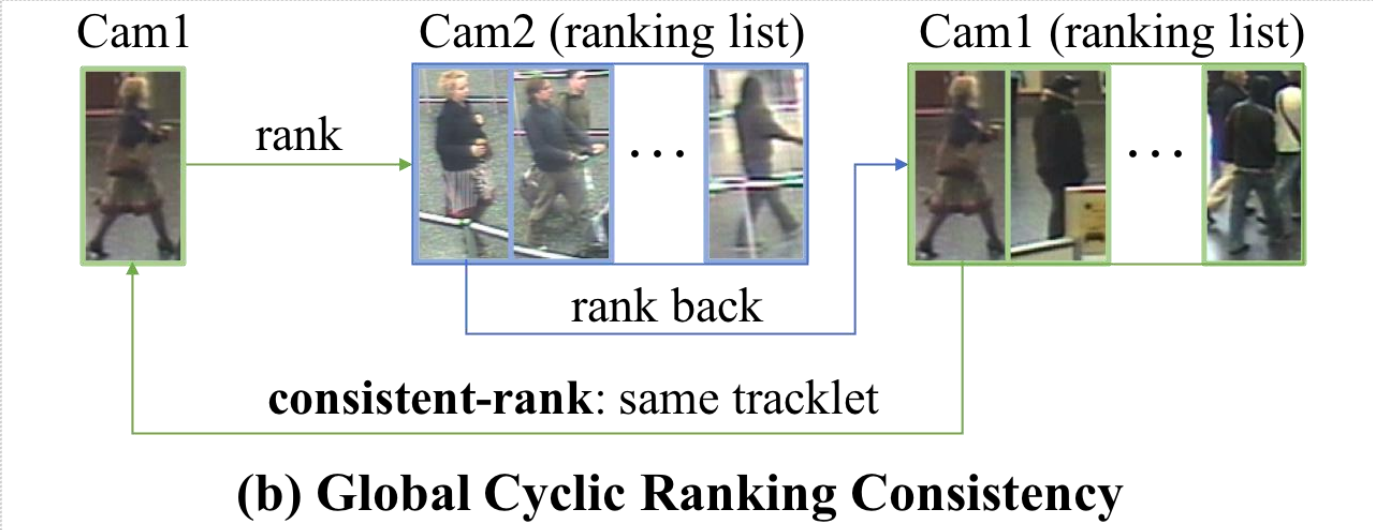


Tracklet representation: Learnable anchor



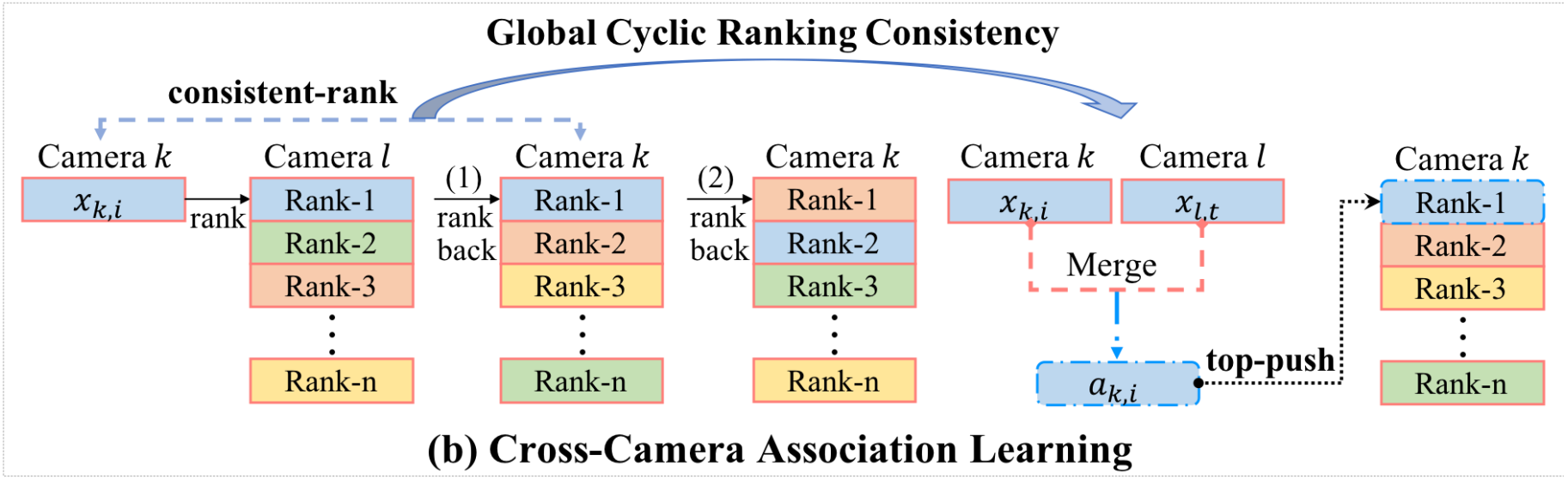
Methodology

- **Cross-Camera Association Learning**
- Learning constraint on tracklet-to-tracklet association across disjoint camera views



Methodology

- **Cross-Camera Association Learning**
- Learning constraint on tracklet-to-tracklet association across disjoint camera views



(b) Cross-Camera Association Learning

Methodology

- **Algorithmic overview**

Algorithm 1 Deep Association Learning.

Input: Unlabelled video tracklets captured from different cameras.
Output: A deep CNN model for re-id matching.
for $t = 1$ **to** *max_iter* **do**
    Randomly sample a mini-batch of image frames.
    Network forward propagation.
    Tracklet association ranking on the *intra-camera anchors* (Eq. (2)).
    Compute two margin-based association loss terms (Eq. (3), (6)).
    Update the corresponding *intra-camera anchors* based on the EMA strategy (Eq. (1)).
    Update the corresponding *cross-camera anchors* based on cyclic ranking (Eq. (4), (5)).
    Network update by back-propagation (Eq. (7)).
end for

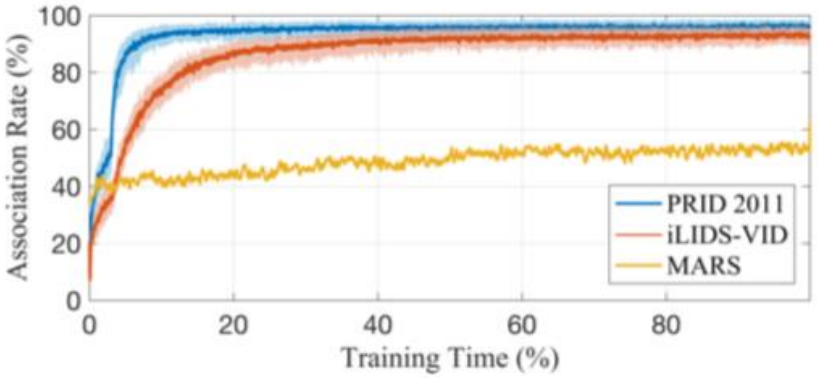
Experiments

- **Comparison to state-of-the-art performance**

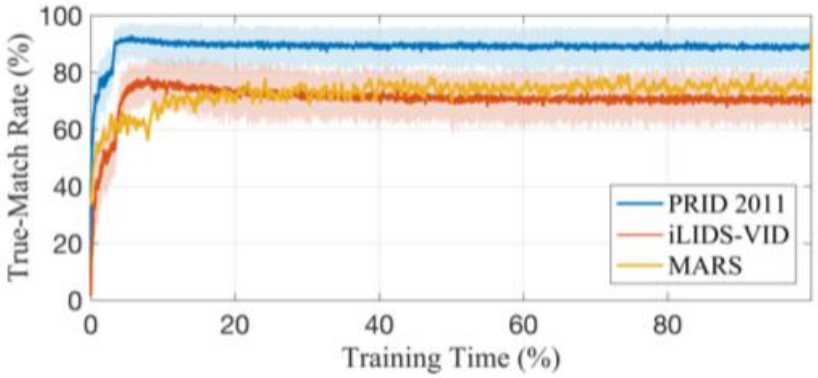
Datasets	PRID 2011				iLIDS-VID				MARS				
Rank@k	1	5	10	20	1	5	10	20	1	5	10	20	mAP
DVDL [18]	40.6	69.7	77.8	85.6	25.9	48.2	57.3	68.9	-	-	-	-	-
STFV3D [25]	42.1	71.9	84.4	91.6	37.0	64.3	77.0	86.9	-	-	-	-	-
MDTS-DTW [27]	41.7	67.1	79.4	90.1	31.5	62.1	72.8	82.4	-	-	-	-	-
UnKISS [19]	59.2	81.7	90.6	96.1	38.2	65.7	75.9	84.1	-	-	-	-	-
DGM+IDE [42]	56.4	81.3	88.0	96.4	36.2	62.8	73.6	82.7	36.8	54.0	61.6	68.5	21.3
Stepwise [26]	80.9	95.6	98.8	99.4	41.7	66.3	74.1	80.7	23.6	35.8	-	44.9	10.5
DAL (ResNet50)	85.3	97.0	98.8	99.6	56.9	80.6	87.3	91.9	46.8	63.9	71.6	77.5	21.4
DAL (MobileNet)	84.6	96.3	98.4	99.1	52.8	76.7	83.4	91.6	49.3	65.9	72.2	77.9	23.0

Experiments

- **Evolution of cross-camera tracklet associations**



(a) Evolution on association rate.



(b) Evolution on true-match rate.

Experiments

- **The first end-to-end optimized unsupervised deep learning method for video person re-id.**

- **A novel association learning method for unsupervised cross-domain image matching.**

Thank you