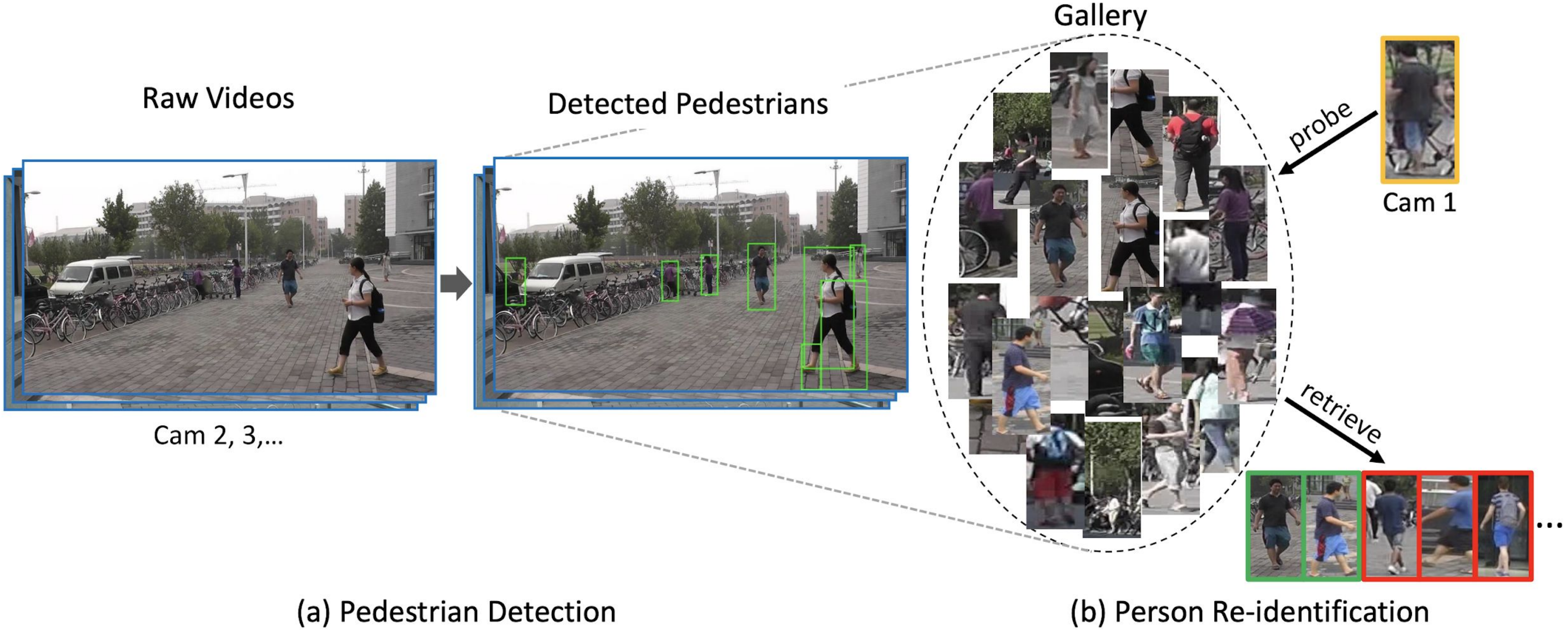
Mutual Mean-Teaching:

Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification

Yixiao Ge, Dapeng Chen, Hongsheng Li
Multimedia Laboratory,
The Chinese University of Hong Kong

Person re-identification

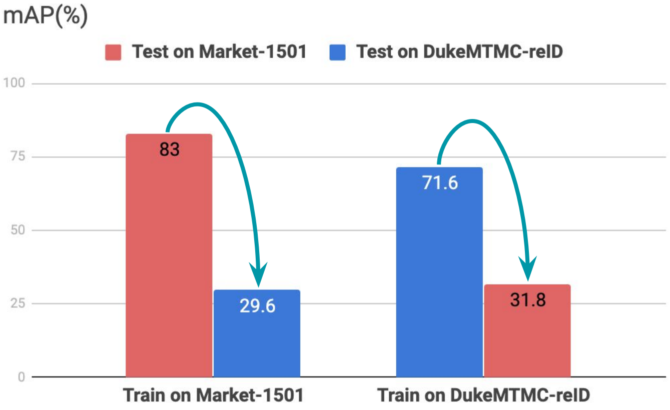


[1] Zheng L, et al. Person re-identification: Past, present and future[J]. arXiv preprint arXiv:1610.02984, 2016.

Single domain (dataset) *vs* Direct transfer

Market-1501[2]

Captured in Tsinghua University



DukeMTMC-reID[3]

Captured in Duke University

[2] Zheng L, et al. Scalable person re-identification: A benchmark[C]. CVPR, 2015: 1116-1124.
[3] Ristani E, et al. Performance measures and a data set for multi-target, multi-camera tracking[C]. ECCV, 2016: 17-35.

Unsupervised domain adaptation (UDA)

e.g. Market-1501

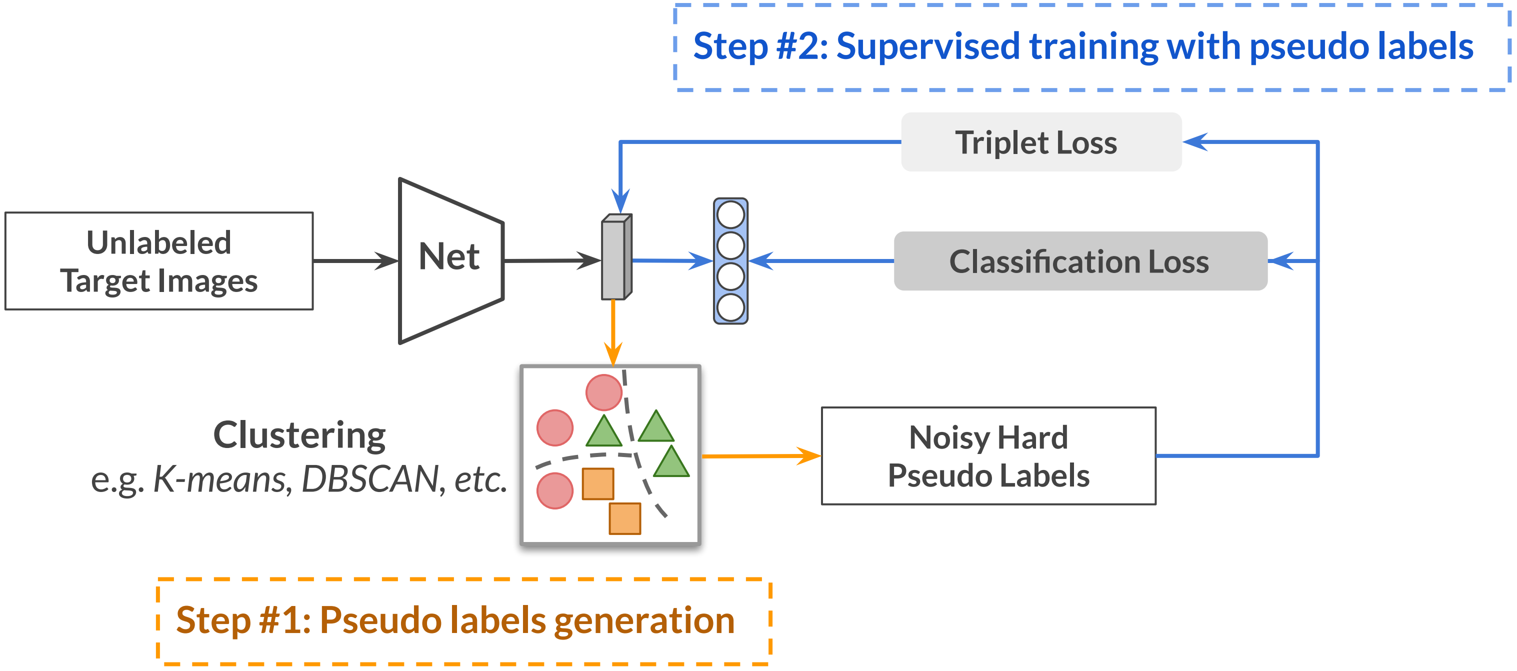
Source domain (labeled)

Adaptation

e.g. DukeMTMC-reID

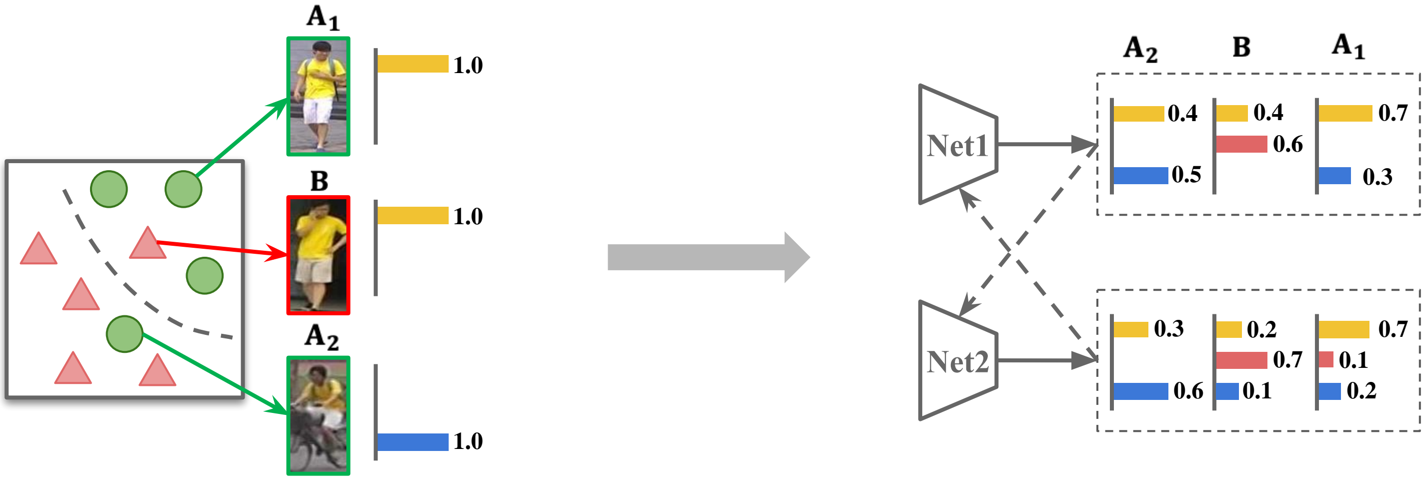
Target domain (unlabeled)

Clustering-based UDA Pipeline

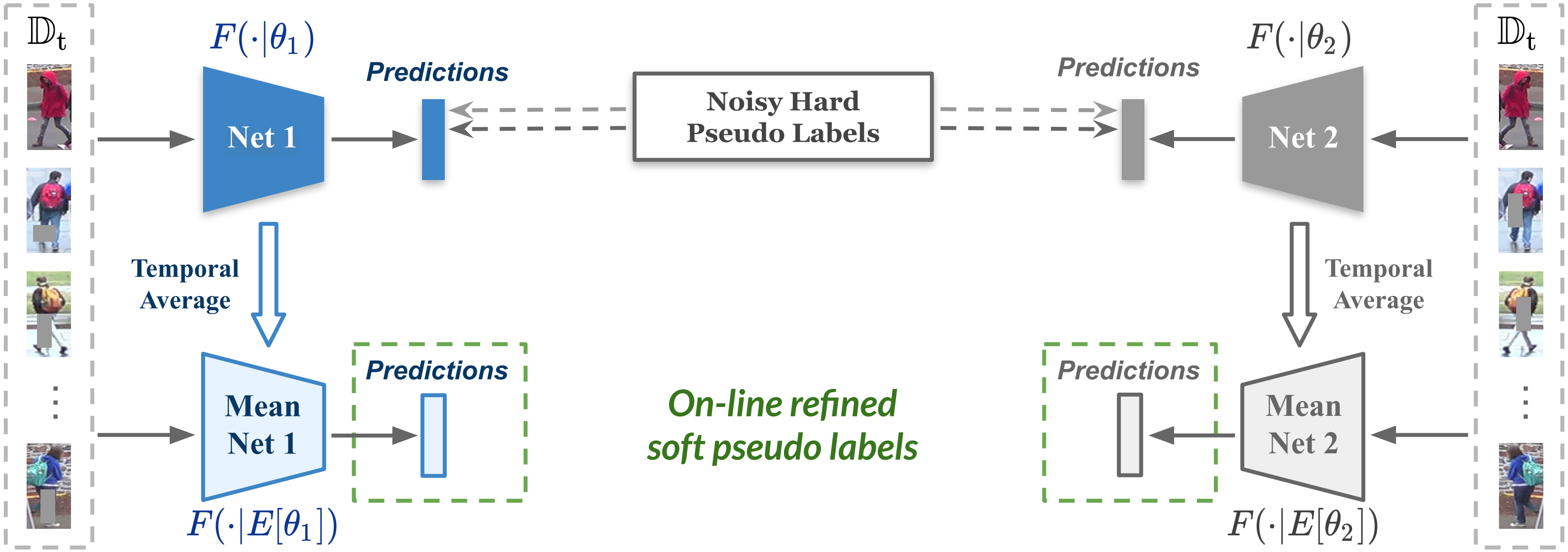


Issue: noisy hard labels

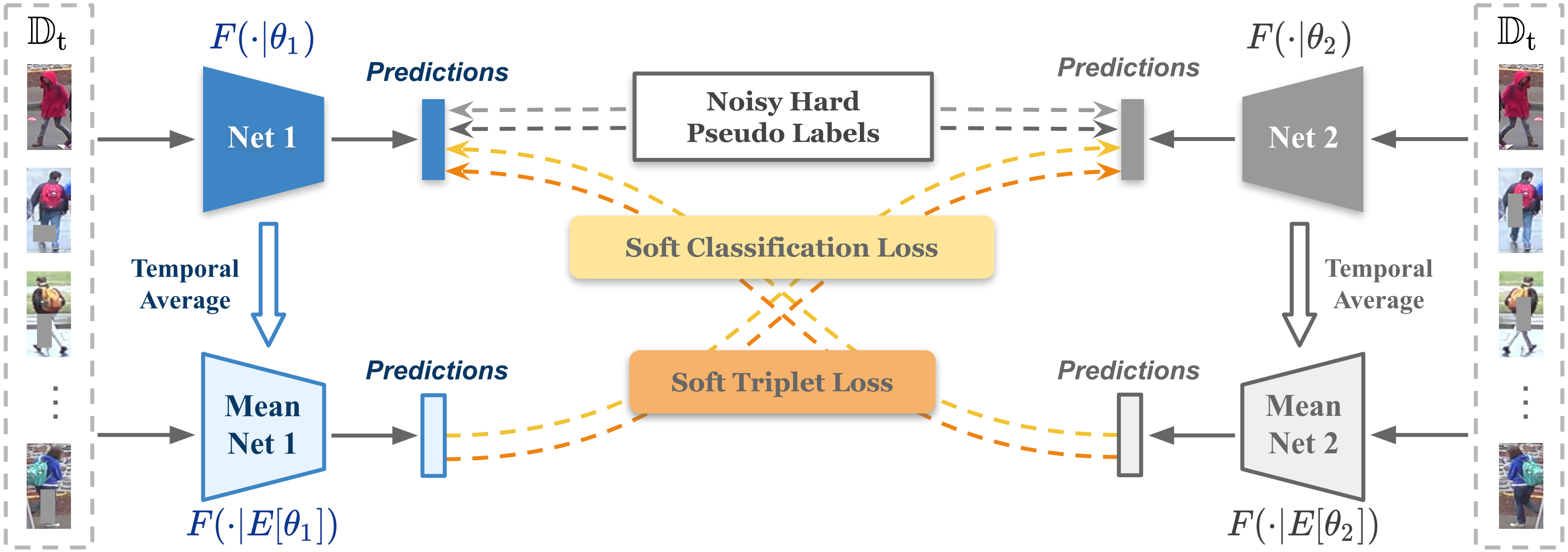
Solution: robust soft labels



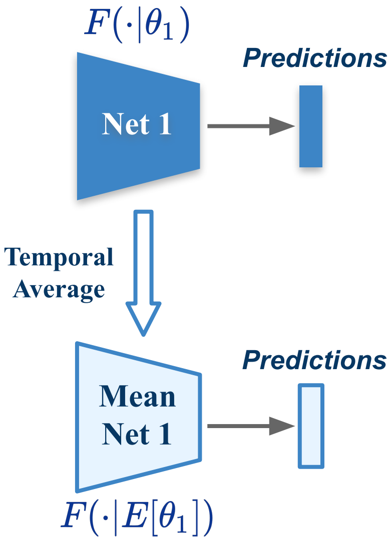
Mutual Mean-Teaching (MMT)



Mutual Mean-Teaching (MMT)



Mean Net



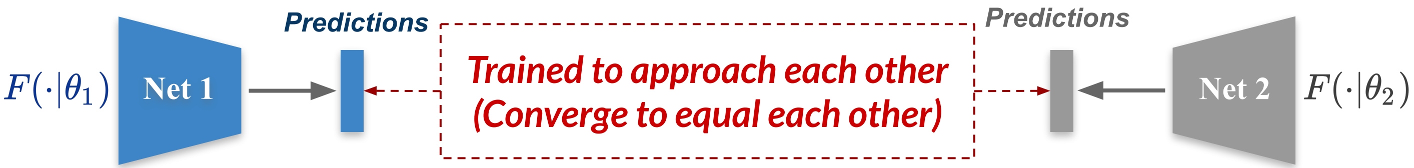
At iteration 0,

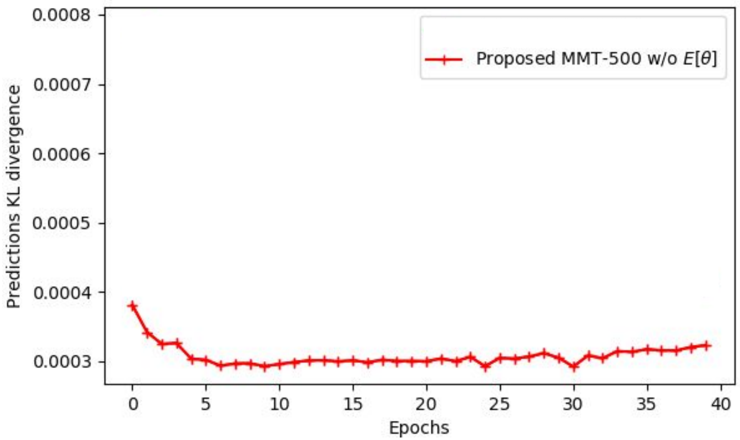
$$E^{(0)}[\boldsymbol{\theta}_1] = \boldsymbol{\theta}_1, E^{(0)}[\boldsymbol{\theta}_2] = \boldsymbol{\theta}_2$$

At iteration T (T>0),

$$E^{(T)}[\boldsymbol{\theta}_1] = \alpha E^{(T-1)}[\boldsymbol{\theta}_1] + (1-\alpha)\boldsymbol{\theta}_1$$
$$E^{(T)}[\boldsymbol{\theta}_2] = \alpha E^{(T-1)}[\boldsymbol{\theta}_2] + (1-\alpha)\boldsymbol{\theta}_2$$

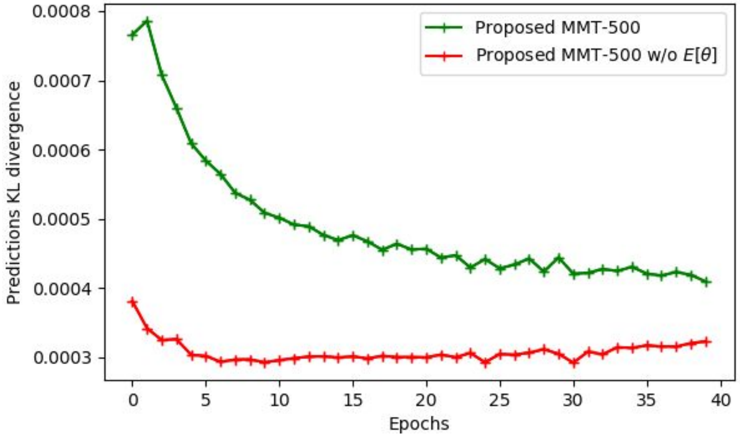
Why mean-teaching? *--- One option:*





[4] Han B, et al. Co-teaching: Robust training of deep neural networks with extremely noisy labels. NIPS, 2018: 8527-8537.
[5] Zhang Y, et al. Deep mutual learning. CVPR, 2018: 4320-4328.

Why mean-teaching?



Soft classification loss

$$\mathcal{L}_{sid}^{t}(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) = -\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\boxed{C_2^t(F(\boldsymbol{x}'^{t}_{i}|E^{(T)}[\boldsymbol{\theta}_2]))} \cdot \log C_1^t(F(\boldsymbol{x}_i^t|\boldsymbol{\theta}_1))\right)$$

$$\mathcal{L}_{sid}^{t}(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1) = -\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\boxed{C_1^t(F(\boldsymbol{x}_i^t|E^{(T)}[\boldsymbol{\theta}_1]))} \cdot \log C_2^t(F(\boldsymbol{x}'^{t}_{i}|\boldsymbol{\theta}_2))\right)$$

replace one-hot labels in cross-entropy loss

[6] Hinton G, et al. Distilling the knowledge in a neural network[J]. arXiv preprint arXiv:1503.02531, 2015.

Soft triplet loss --- *hard-version softmax-triplet*

Softmax-triplet: $$\mathcal{T}_i(\boldsymbol{\theta}_1) = \frac{\exp(\|F(\boldsymbol{x}_i^t|\boldsymbol{\theta}_1) - F(\boldsymbol{x}_{i,n}^t|\boldsymbol{\theta}_1)\|)}{\exp(\|F(\boldsymbol{x}_i^t|\boldsymbol{\theta}_1) - F(\boldsymbol{x}_{i,p}^t|\boldsymbol{\theta}_1)\|) + \exp(\|F(\boldsymbol{x}_i^t|\boldsymbol{\theta}_1) - F(\boldsymbol{x}_{i,n}^t|\boldsymbol{\theta}_1)\|)}$$

***Hard** Softmax-triplet loss:* $$\mathcal{L}_{tri}^t(\boldsymbol{\theta}_1) = \frac{1}{N_t}\sum_{i=1}^{N_t} \mathcal{L}_{bce}\Big(\mathcal{T}_i(\boldsymbol{\theta}_1), \boxed{\mathbf{1}}\Big)$$

The sample should be closer to its (potential) positive than its (potential) negative.

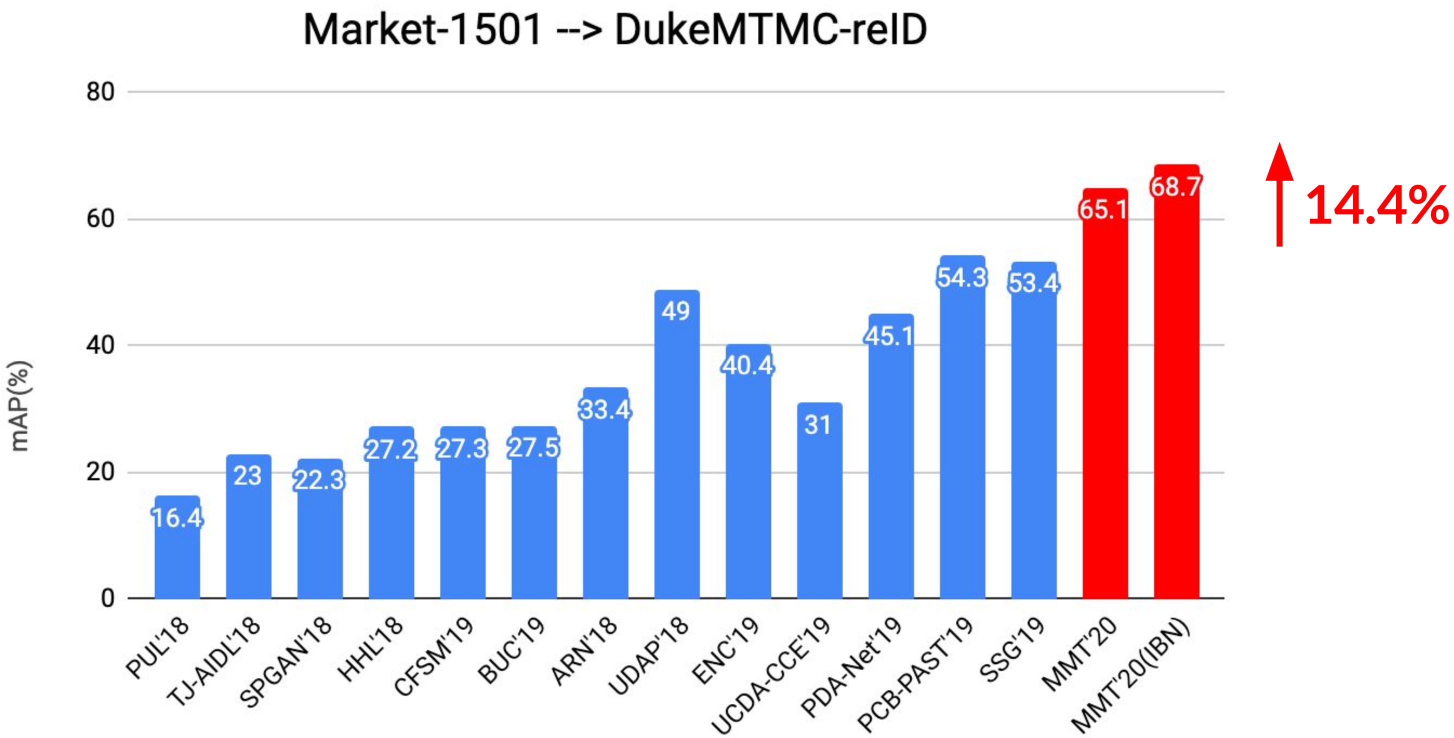
Soft triplet loss --- *soft-version softmax-triplet*

Softmax-triplet: $$\mathcal{T}_i(\boldsymbol{\theta}_1) = \frac{\exp(\|F(\boldsymbol{x}_i^t|\boldsymbol{\theta}_1) - F(\boldsymbol{x}_{i,n}^t|\boldsymbol{\theta}_1)\|)}{\exp(\|F(\boldsymbol{x}_i^t|\boldsymbol{\theta}_1) - F(\boldsymbol{x}_{i,p}^t|\boldsymbol{\theta}_1)\|) + \exp(\|F(\boldsymbol{x}_i^t|\boldsymbol{\theta}_1) - F(\boldsymbol{x}_{i,n}^t|\boldsymbol{\theta}_1)\|)}$$

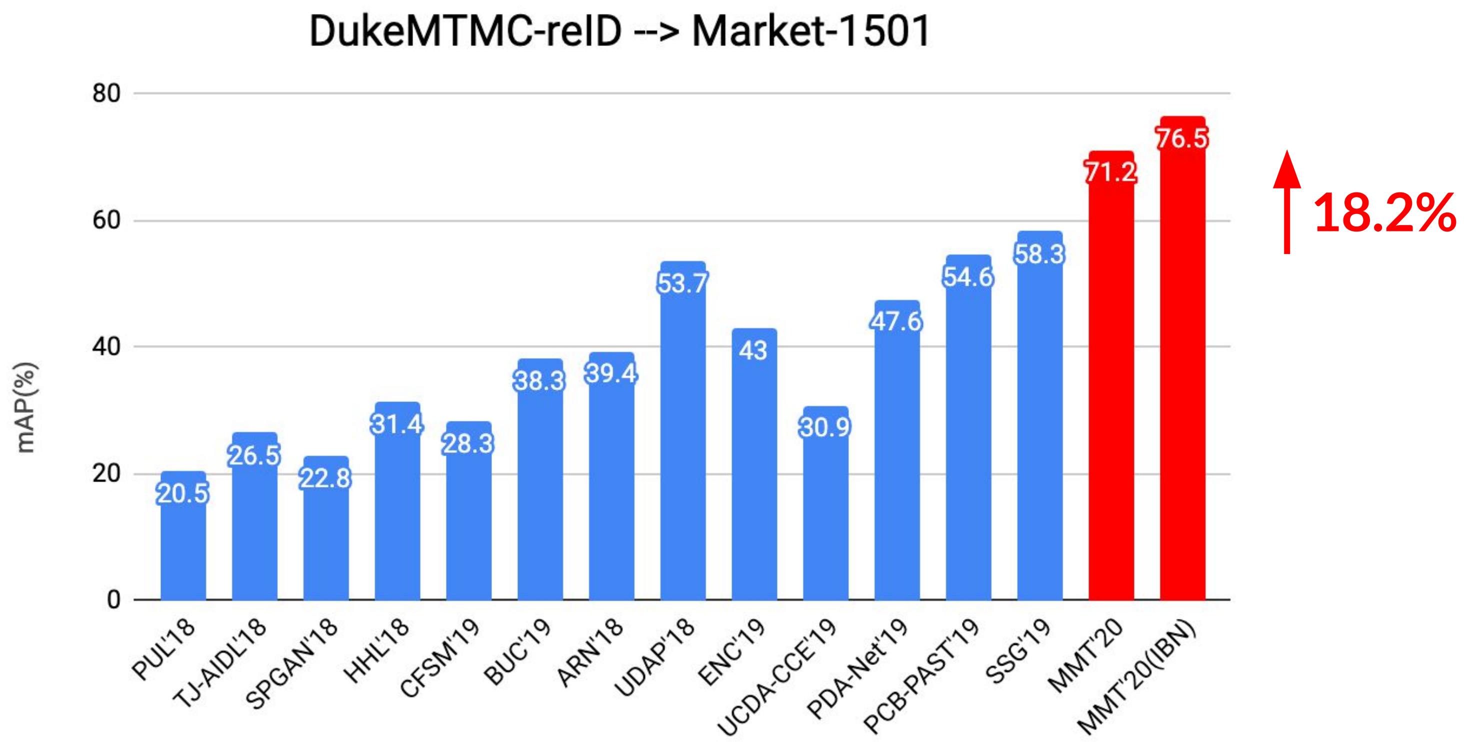
***Soft** Softmax-triplet loss:*
$$\mathcal{L}_{stri}^t(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) = \frac{1}{N_t}\sum_{i=1}^{N_t}\mathcal{L}_{bce}\left(\mathcal{T}_i(\boldsymbol{\theta}_1), \boxed{\mathcal{T}_i\left(E^{(T)}[\boldsymbol{\theta}_2]\right)}\right)$$
$$\mathcal{L}_{stri}^t(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1) = \frac{1}{N_t}\sum_{i=1}^{N_t}\mathcal{L}_{bce}\left(\mathcal{T}_i(\boldsymbol{\theta}_2), \boxed{\mathcal{T}_i\left(E^{(T)}[\boldsymbol{\theta}_1]\right)}\right)$$

replace hard label "1"

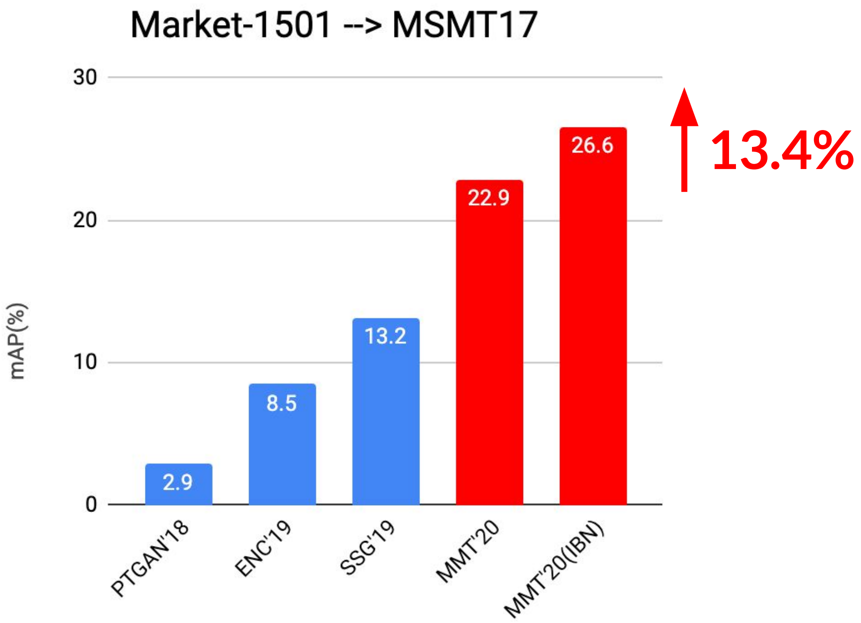
MMT *vs* state-of-the-arts

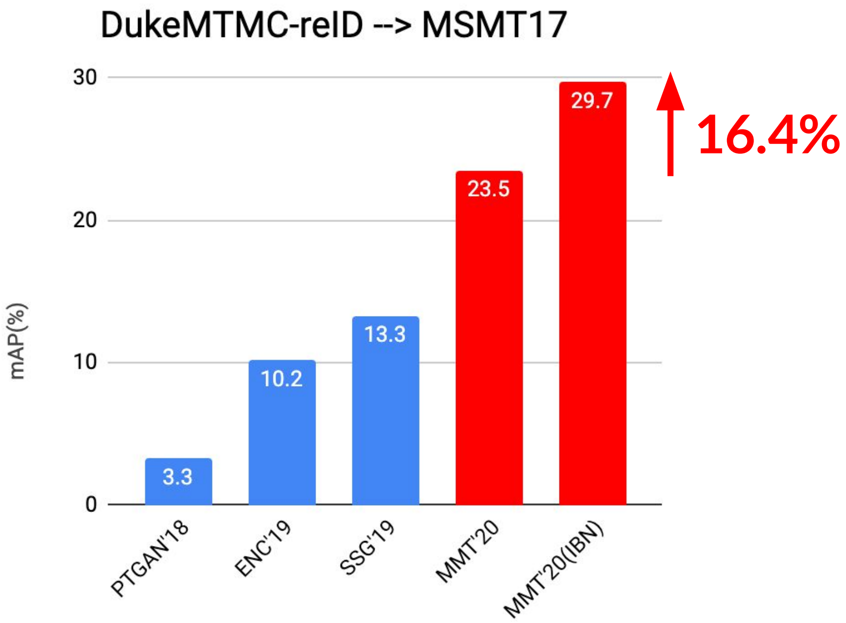


MMT *vs* state-of-the-arts



MMT *vs* state-of-the-arts





Mutual Mean-Teaching:

Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification

Yixiao Ge, Dapeng Chen, Hongsheng Li
Multimedia Laboratory, The Chinese University of Hong Kong

Code available at

https://github.com/yxgeee/MMT