Unsupervised Person Re-identification by Soft Multilabel Learning

CVPR2019

Hong-Xing Yu1 , Wei-Shi Zheng1,4∗ , Ancong Wu1 , Xiaowei Guo2 , Shaogang Gong3 , and Jian-Huang Lai1

1Sun Yat-sen University, China ; 2YouTu Lab, Tencent; 3Queen Mary University of London, UK

4Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China

汇报人：鲍涟漪

学号：19210240244

- Introduction
- Deep Soft Multilabel Reference Learning (MAR)
- Experiments
- Conclusion

Introduction

Propose a novel **soft multilabel** learning to mine the potential label information in the unlabeled RE-ID data

- Soft multilabel-guided hard negative mining
- Cross-view consistent soft multilabel learning
- Reference agent learning

Deep Soft Multilabel Reference Learning (MAR)

What is Soft Multilabel?

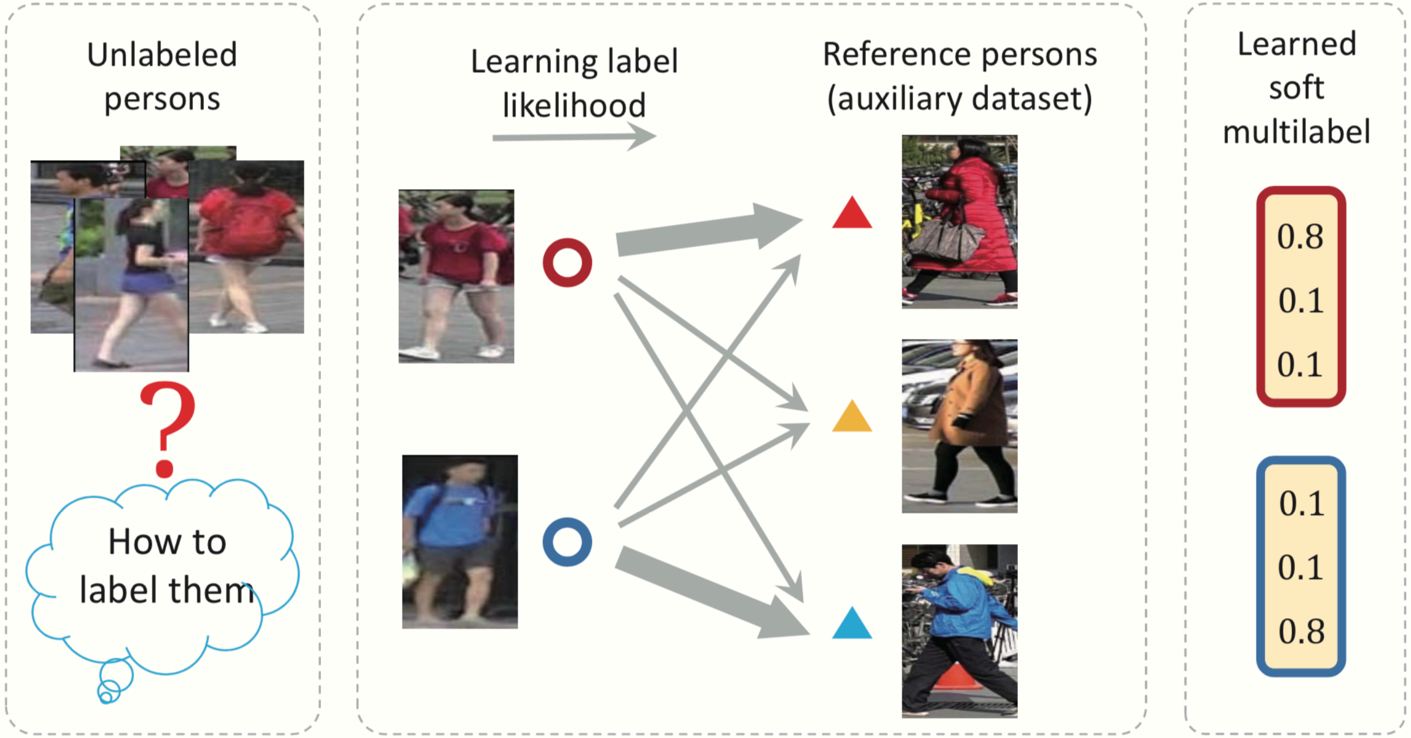

Figure 1. Illustration of our soft multilabel concept. We learn a soft multilabel (real-valued label vector) for each unlabeled person by comparing to a set of known auxiliary reference persons (thicker arrowline indicates higher label likelihood). Best viewed in color.

Deep Soft Multilabel Reference Learning (MAR)

Problem formulation and Overview

$\mathcal{X} = \{x_i\}_{i=1}^{N_u}$ an unlabeled target RE-ID dataset

$\mathcal{Z} = \{z_i, w_i\}_{i=1}^{N_a}$ an auxiliary RE-ID dataset

$f(\cdot)$ a discriminative deep feature embedding

$\{a_i\}_{i=1}^{N_p}$ reference person feature

$y = l(f(x), \{a_i\}_{i=1}^{N_p})$ The soft multilabel function

Soft multilabel-guided hard negative mining

- soft multilabel function
$$y^{(k)} = l(f(x), \{a_i\}_{i=1}^{N_p})^{(k)} = \frac{\exp(a_k^{\mathrm{T}} f(x))}{\Sigma_i \exp(a_i^{\mathrm{T}} f(x))} \tag{1}$$

- **Assumption 1**. If a pair of unlabeled person images x_i , x_j has high feature similarity $f(x_i)^{\mathrm{T}} f(x_j)$, we call the pair a similar pair. If a similar pair has highly similar comparative characteristics, it is probably a positive pair. Otherwise, it is probably a hard negative pair.

Soft multilabel-guided hard negative mining

- soft multilabel agreement $A(\cdot, \cdot)$

$$A(y_i, y_j) = y_i \wedge y_j = \Sigma_k \min(y_i^{(k)}, y_j^{(k)}) = 1 - \frac{||y_i - y_j||_1}{2}, \quad (2)$$

- mine the hard negative pairs

$$\begin{aligned} \mathcal{P} &= \{(i,j) | f(x_i)^{\mathrm{T}} f(x_j) \geq S, A(y_i, y_j) \geq T\} \\ \mathcal{N} &= \{(k,l) | f(x_k)^{\mathrm{T}} f(x_l) \geq S, A(y_k, y_l) < T\} \end{aligned} \quad (3)$$

We define the similar pairs in Assumption 1 as the pM pairs that have highest feature similarities among all the M = $N_u \times (N_u - 1)/2$ pairs within the unlabeled target dataset $\mathcal{X}$.

Soft Multilabel- guided Discriminative embedding Learning

$$L_{MDL} = -\log \frac{\overline{P}}{\overline{P} + \overline{N}}, \tag{4}$$

$$\overline{P} = \frac{1}{|\mathcal{P}|} \Sigma_{(i,j)\in\mathcal{P}} \exp(-||f(z_i) - f(z_j)||_2^2),$$

$$\overline{N} = \frac{1}{|\mathcal{N}|} \Sigma_{(k,l)\in\mathcal{N}} \exp(-||f(z_k) - f(z_l)||_2^2).$$

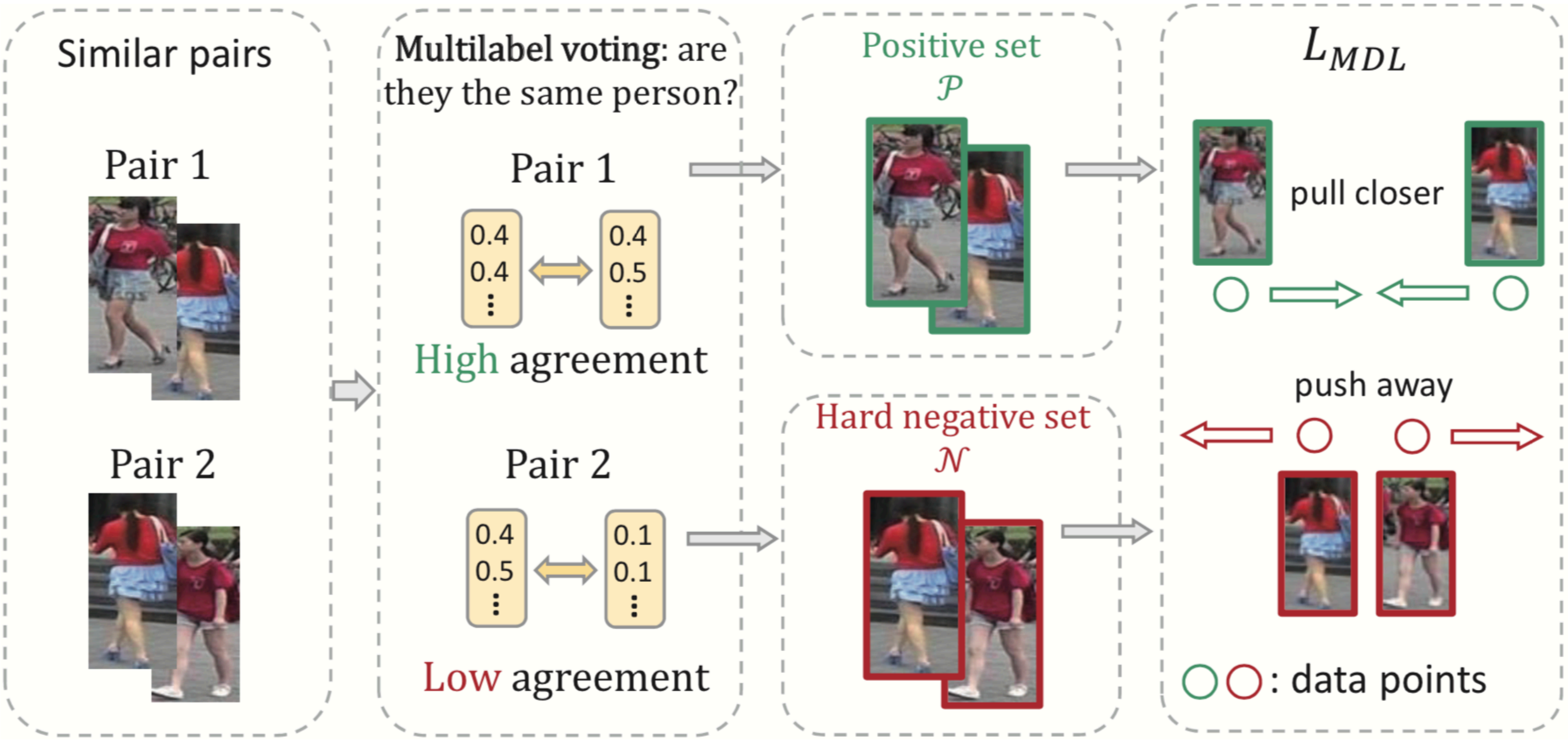

Figure 3. Illustration of the soft multilabel-guided hard negative mining. Best viewed in color.

Cross-view consistent soft multilabel learning

$$L_{CML} = \Sigma_v d(\mathbb{P}_v(y), \mathbb{P}(y))^2 \tag{5}$$

$$\Downarrow$$

$$L_{CML} = \Sigma_v ||\mu_v - \mu||_2^2 + ||\sigma_v - \sigma||_2^2 \tag{6}$$

Reference agent learning

- Agent Learning loss

$$L_{AL} = \Sigma_k - \log l(f(z_k), \{a_i\})^{(w_k)} = \Sigma_k - \log \frac{\exp(a_{w_k}^{\mathrm{T}} f(z_k))}{\Sigma_j \exp(a_j^{\mathrm{T}} f(z_k))} \tag{7}$$

- the Reference agent-based Joint embedding learning loss

$$L_{RJ} = \Sigma_i \Sigma_{j \in \mathcal{M}_i} \Sigma_{k \in \mathcal{W}_i} [m - \|a_i - f(x_j)\|_2^2]_+ + \|a_i - f(z_k)\|_2^2 \tag{8}$$

$$\Downarrow$$

$$L_{RAL} = L_{AL} + \beta L_{RJ} \tag{9}$$

Model training and testing

$$L_{MAR} = L_{MDL} + \lambda_1 L_{CML} + \lambda_2 L_{RAL} \tag{10}$$

where λ1 and λ2 are hyperparameters to control the relative importance of the cross-view consistent soft multilabel learning and the reference agent learning, respectively. We train our model end to end by the Stochastic Gradient Descent (SGD). For testing, we compute the cosine feature similarity of each probe(query)-gallery pair, and obtain the ranking list of the probe image against the gallery images.

Experiments

Datasets

Evaluation benchmarks

- Market-1501
- DukeMTMC-reID

Auxiliary dataset

- MSMT17

Comparison

Table 1. Comparison to the state-of-the-art unsupervised results in the Market-1501 dataset. **Red** indicates the best and **Blue** the second best. Measured by %.

Methods	Reference	Market-1501		
		rank-1	rank-5	mAP
LOMO [20]	CVPR'15	27.2	41.6	8.0
BoW [59]	ICCV'15	35.8	52.4	14.8
DIC [16]	BMVC'15	50.2	68.8	22.7
ISR [21]	TPAMI'15	40.3	62.2	14.3
UDML [29]	CVPR'16	34.5	52.6	12.4
CAMEL [52]	ICCV'17	54.5	73.1	26.3
PUL [8]	ToMM'18	45.5	60.7	20.5
TJ-AIDL [48]	CVPR'18	58.2	74.8	26.5
PTGAN [50]	CVPR'18	38.6	57.3	15.7
SPGAN [7]	CVPR'18	51.5	70.1	27.1
HHL [62]	ECCV'18	**62.2**	**78.8**	31.4
DECAMEL [53]	TPAMI'19	60.2	76.0	**32.4**
MAR	This work	**67.7**	**81.9**	**40.0**

Comparison

Table 2. Comparison to the state-of-the-art unsupervised results in the DukeMTMC-reID dataset. Measured by %.

Methods	Reference	DukeMTMC-reID		
		rank-1	rank-5	mAP
LOMO [20]	CVPR'15	12.3	21.3	4.8
BoW [59]	ICCV'15	17.1	28.8	8.3
UDML [29]	CVPR'16	18.5	31.4	7.3
CAMEL [52]	ICCV'17	40.3	57.6	19.8
PUL [8]	ToMM'18	30.0	43.4	16.4
TJ-AIDL [48]	CVPR'18	44.3	59.6	23.0
PTGAN [50]	CVPR'18	27.4	43.6	13.5
SPGAN [7]	CVPR'18	41.1	56.6	22.3
HHL [62]	ECCV'18	**46.9**	**61.0**	**27.2**
MAR	This work	**67.1**	**79.8**	**48.0**

Ablation study

Table 3. Ablation study. Please refer to the text in Sec. 4.4.

Methods	Market-1501			
	rank-1	rank-5	rank-10	mAP
Pretrained (source only)	46.2	64.4	71.3	24.6
Baseline (feature-guided)	44.4	62.5	69.8	21.5
MAR w/o L_{CML}	60.0	75.9	81.9	34.6
MAR w/o L_{CML}&L_{RAL}	53.9	71.5	77.7	28.2
MAR w/o L_{RAL}	59.2	76.4	82.3	30.8
MAR	**67.7**	**81.9**	**87.3**	**40.0**

Methods	DukeMTMC-reID			
	rank-1	rank-5	rank-10	mAP
Pretrained (source only)	43.1	59.2	65.7	28.8
Baseline (feature-guided)	50.0	66.4	71.7	31.7
MAR w/o L_{CML}	63.2	77.2	82.5	44.9
MAR w/o L_{CML}&L_{RAL}	60.1	73.0	78.4	40.4
MAR w/o L_{RAL}	57.9	72.6	77.8	37.1
MAR	**67.1**	**79.8**	**84.2**	**48.0**

Conclusion

- We address the unsupervised RE-ID problem by a novel **soft multilabel reference learning** method, in which we mine the potential label information latent in the unlabeled RE-ID data by exploiting the auxiliary source dataset for reference comparison.

- We formulate a novel deep model named deep soft multilabel reference learning (**MAR**),which enables simultaneously **the soft multilabel-guided hard negative mining**, **the cross-view consistent soft multilabel learning** and **the reference agent learning in a unified model**.

Thanks !