Learning to Reduce Dual-level Discrepancy for Infrared-Visible Person Re-identification

Zhixiang Wang[1] Zheng Wang[2](✉) Yinqiang Zheng[2] Yung-Yu Chuang[1] Shin'ichi Satoh[2,3]
[1]National Taiwan University [2]National Institute of Informatics [3]The University of Tokyo

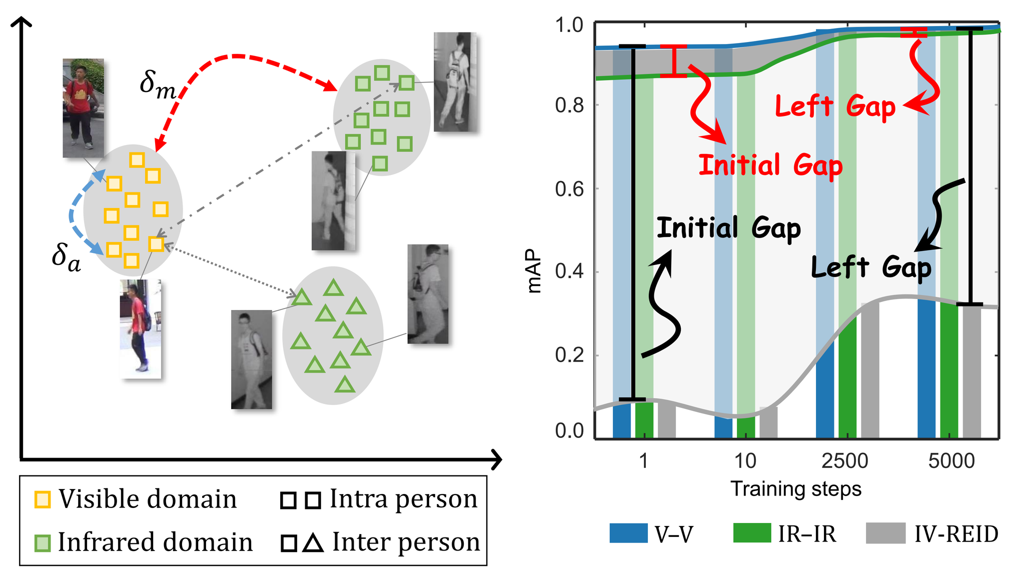
CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Problem and Difficulty

Problem

- Given a visible (or infrared) image of a specific person, the goal is to find the corresponding infrared (or visible) images of the person captured by other spectrum cameras. This cross-modality image matching task is named Infrared-Visible person RE-IDentification (IV-REID).

Difficulty

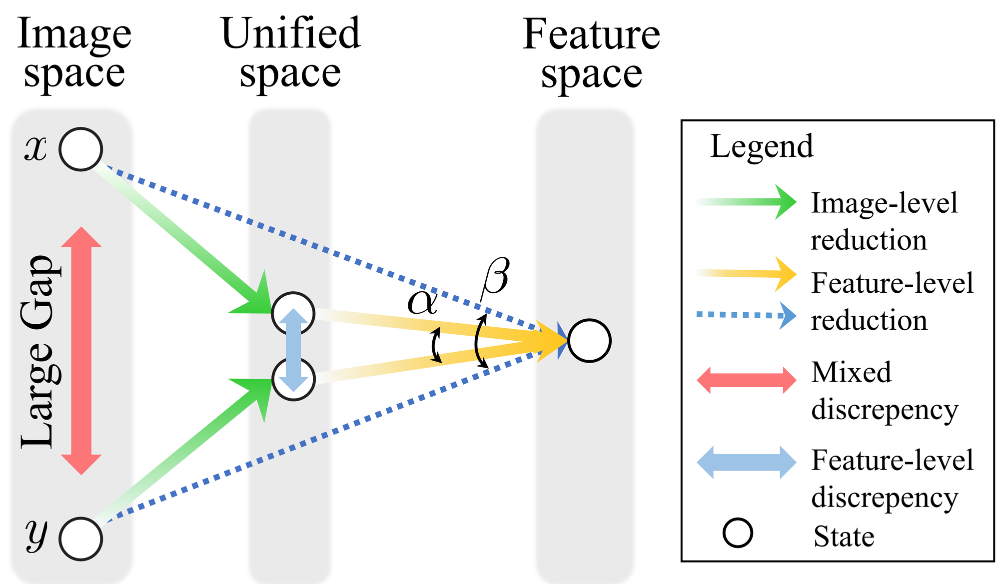


Contributions and Motivations

Contributions

- A novel dual-level discrepancy reduction learning scheme, which decompose the mixed discrepancies and handling them separately.
- Our end-to-end scheme enforces these two sub-networks benefit each other.
- Extensive experiments on two datasets demonstrate the superior performance.

Motivations



Method

Framework

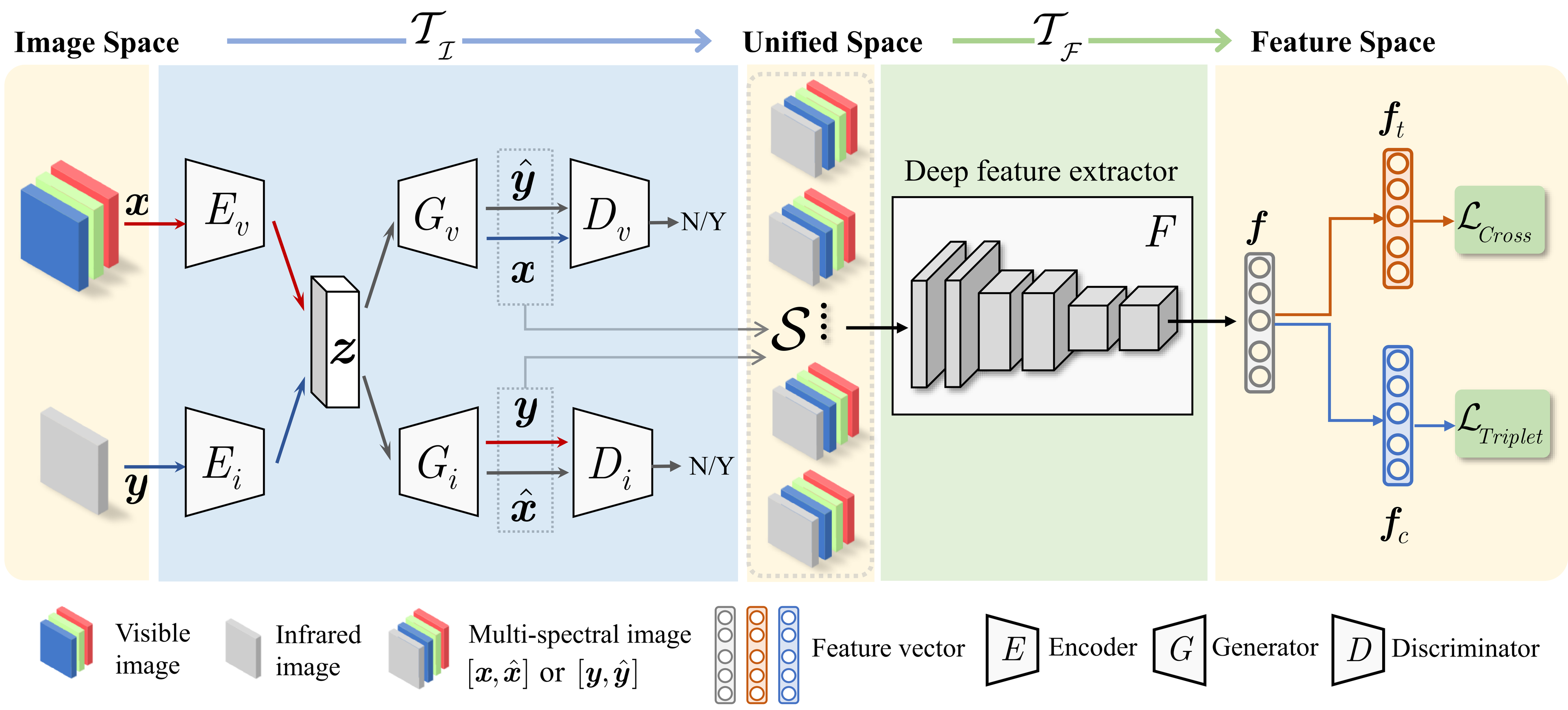


Image-level discrepancy reduction $\mathcal{T}_\mathcal{I}$

$$\mathcal{L}_\mathcal{I} = \mathcal{L}_{\mathrm{VAE}_v} + \mathcal{L}_{\mathrm{VAE}_i} + \mathcal{L}_{\mathrm{GAN}_v} + \mathcal{L}_{\mathrm{GAN}_i} + \mathcal{L}_{\mathrm{CC}_v} + \mathcal{L}_{\mathrm{CC}_i}$$

- $\mathcal{L}_{\mathrm{VAE}_v}, \mathcal{L}_{\mathrm{VAE}_i}$: **Style Disentanglement Loss**
- $\mathcal{L}_{\mathrm{GAN}_v}, \mathcal{L}_{\mathrm{GAN}_i}$: **Image Generation Loss**
- $\mathcal{L}_{\mathrm{CC}_v}, \mathcal{L}_{\mathrm{CC}_i}$: **Cycle-consistency Loss**
- $\mathcal{L}_\mathcal{I}$: **Objective for $\mathcal{T}_\mathcal{I}$**

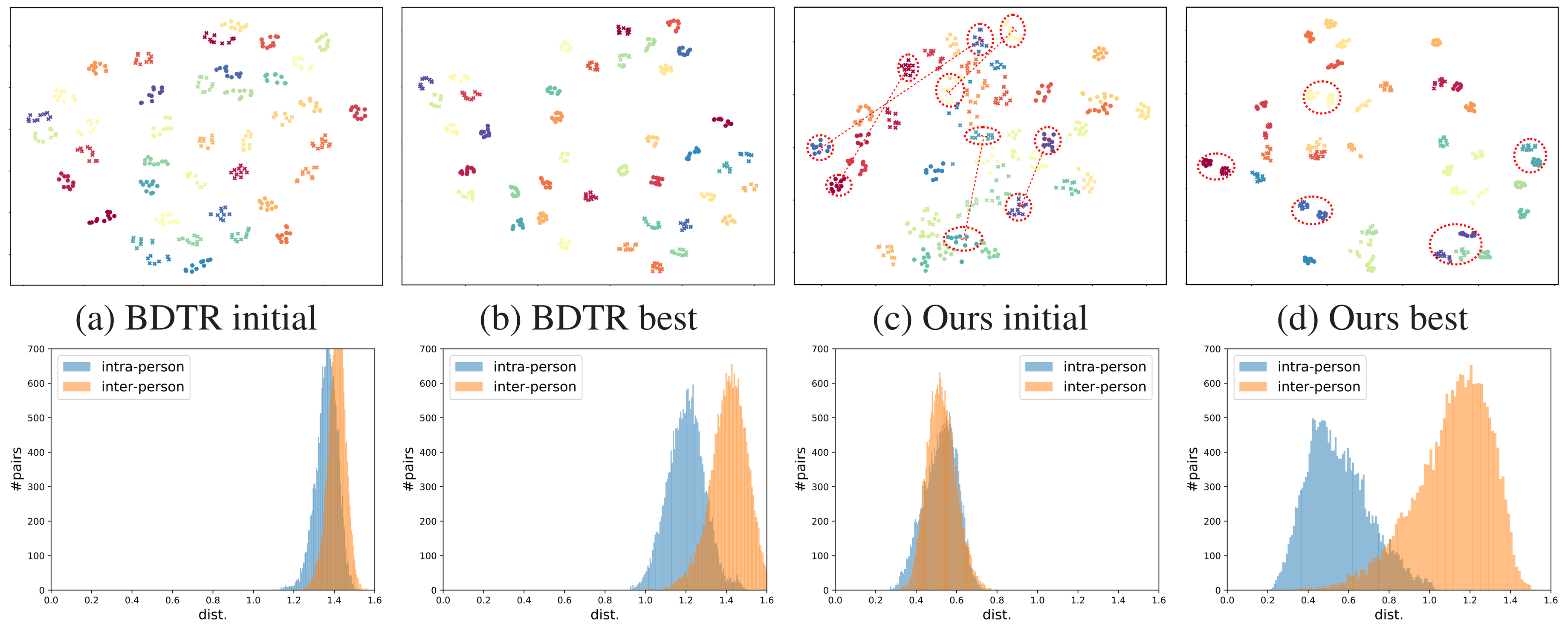
Feature-level discrepancy reduction $\mathcal{T}_\mathcal{F}$

$$\mathcal{L}_\mathcal{F} = \lambda_4 \mathcal{L}_\mathcal{F}^C + \lambda_5 \mathcal{L}_\mathcal{F}^T$$

- $\mathcal{L}_\mathcal{F}^T$: **Triplet Loss**
- $\mathcal{L}_\mathcal{F}^C$: **Cross-entropy Loss**
- $\mathcal{L}_\mathcal{F}$: **Objective for $\mathcal{T}_\mathcal{F}$**

Experiments & Results

Why separately reduce discrepancy?



(a) BDTR initial (b) BDTR best (c) Ours initial (d) Ours best
(e) BDTR initial (f) BDTR best (g) Ours initial (h) Ours best

Which modality to unify?

Metrics (%)	RegDB mAP	RegDB CMC-1	SYSU-MM01 mAP	SYSU-MM01 CMC-1
D^2RL(v)	36.4	39.1	28.4	28.1
D^2RL(i)	43.6	42.9	27.8	27.4
D^2RL	**44.1**	**43.4**	**29.2**	**28.9**

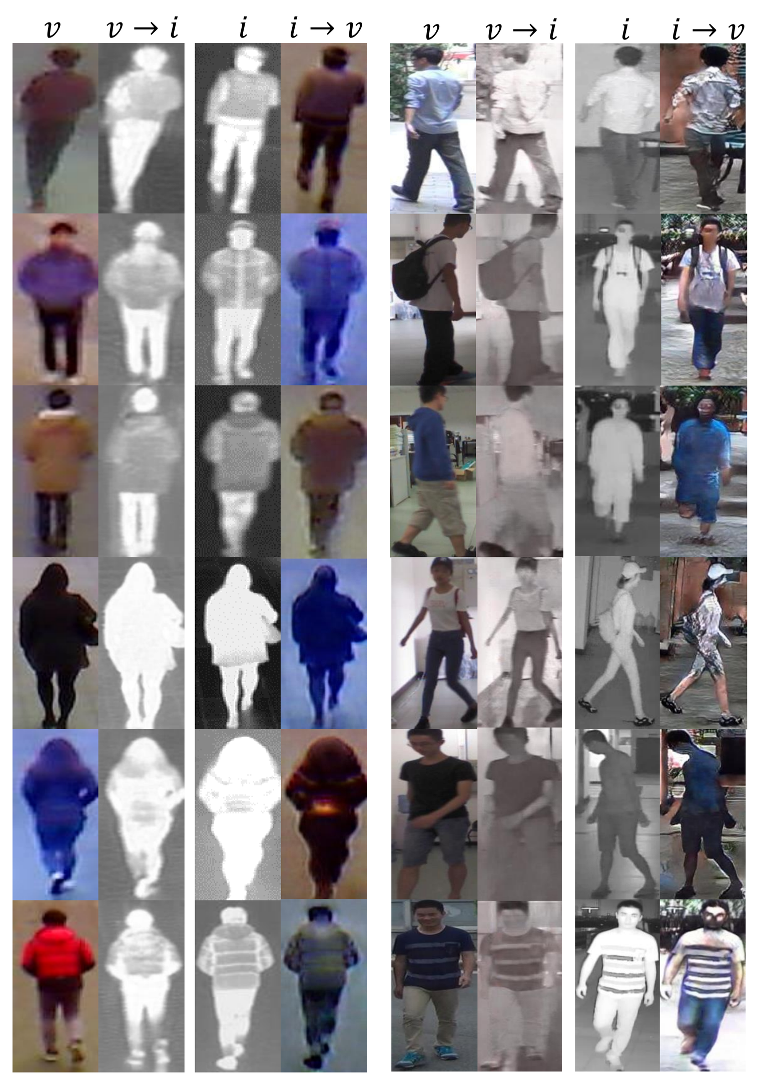
Why joint training?

Metrics (%)	RegDB mAP	RegDB CMC-1	SYSU-MM01 mAP	SYSU-MM01 CMC-1
Separate	40.7	39.9	25.7	26.1
Joint	**44.1**	**43.4**	**29.2**	**28.9**

Comparison with the state-of-the-art methods

Approach	Constraints Feature-level	Constraints Image-level	RegDB CMC-1	RegDB CMC-10	RegDB CMC-20	RegDB mAP	SYSU-MM01 CMC-1	SYSU-MM01 CMC-10	SYSU-MM01 CMC-20	SYSU-MM01 mAP
LOMO [1]	✗	✗	0.85	2.47	4.10	2.28	1.75	14.14	26.63	3.48
MLBP [2]	✗	✗	2.02	7.33	10.90	6.77	2.12	16.23	28.32	3.86
HOG [3]	✗	✗	13.49	33.22	43.66	10.31	2.76	18.25	31.91	4.24
GSM [4]	✗	✗	17.28	34.47	45.26	15.06	5.29	33.71	52.95	8.00
One-stream [5]	✓	✗	13.11	32.98	42.51	14.02	12.04	49.68	66.74	13.67
Two-stream [5]	✓	✗	12.43	30.36	40.96	13.42	11.65	47.99	65.50	12.85
Zero-Padding [5]	✓	✗	17.75	34.21	44.35	18.90	14.80	54.12	71.33	15.95
TONE [6]	✓	✗	16.87	34.03	44.10	14.92	12.52	50.72	68.60	14.42
HCML [6]	✓	✗	24.44	47.53	56.78	20.80	14.32	53.16	69.17	16.16
BDTR [7]	✓	✗	33.47	58.42	67.52	31.83	17.01	55.43	71.96	19.66
*cm*GAN [8]	✓	✗	–	–	–	–	26.97	67.51	80.56	27.80
Proposed D^2RL	✓	✓	**43.4**	**66.1**	**76.3**	**44.1**	**28.9**	**70.6**	**82.4**	**29.2**

Visualization results

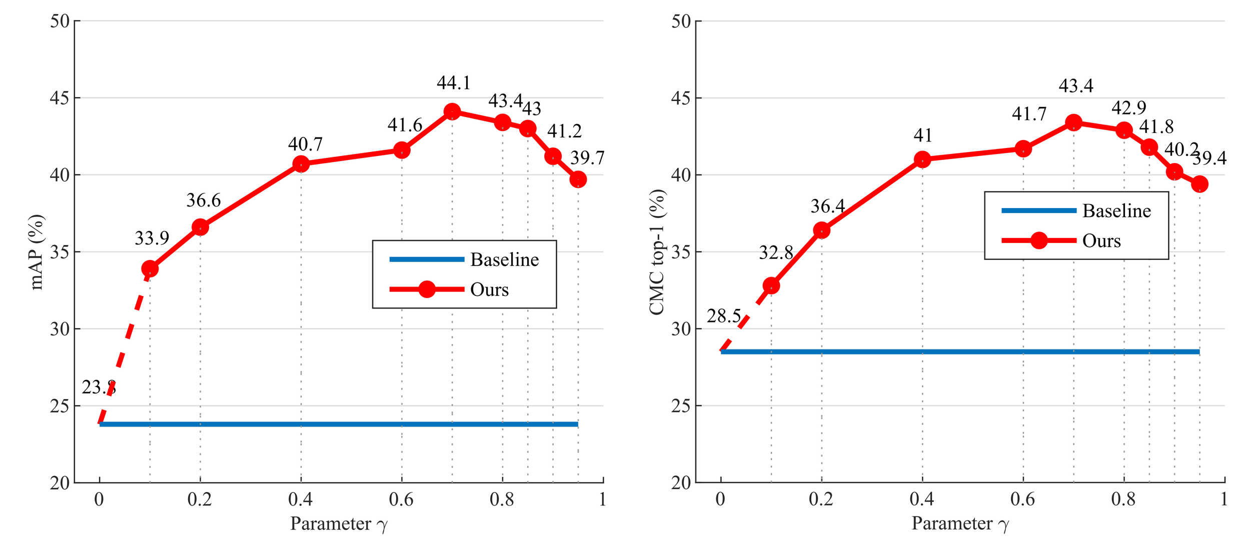


(a) RegDB (b) SYSU-MM01

Ablation study

Method	VAE	CC	CE	triplet	RegDB CMC-1 (%)	RegDB mAP (%)
Baseline	✓	✓	✗	✗	28.5	23.8
D^2RL (no VAE)	✗	✓	✓	✓	34.8	31.3
D^2RL (no CC)	✓	✗	✓	✓	33.7	29.9
D^2RL (no CE)	✓	✓	✗	✓	41.7	40.6
D^2RL (no triplet)	✓	✓	✓	✗	39.5	37.4
D^2RL	✓	✓	✓	✓	**43.4**	**44.1**

How to balance sub-networks $\mathcal{T}_\mathcal{I}$ and $\mathcal{T}_\mathcal{F}$?



Failure cases

References

[1] LOMO [Liao *et al.*, CVPR15]
[2] MLBP [Liao *et al.*, ICCV15]
[3] HOG [Dalal *et al.*, CVPR05]
[4] GSM [Lin *et al.*, TPAMI17]
[5] Zero-Padding [Wu *et al.*, ICCV17]
[6] TONE+HCML [Ye *et al.*, AAAI18]
[7] BDTR [Ye *et al.*, IJCAI18]
[8] *cm*GAN [Dai *et al.*, IJCAI18]