Learning with Twin Noisy Labels for Visible-Infrared Person Re-Identification

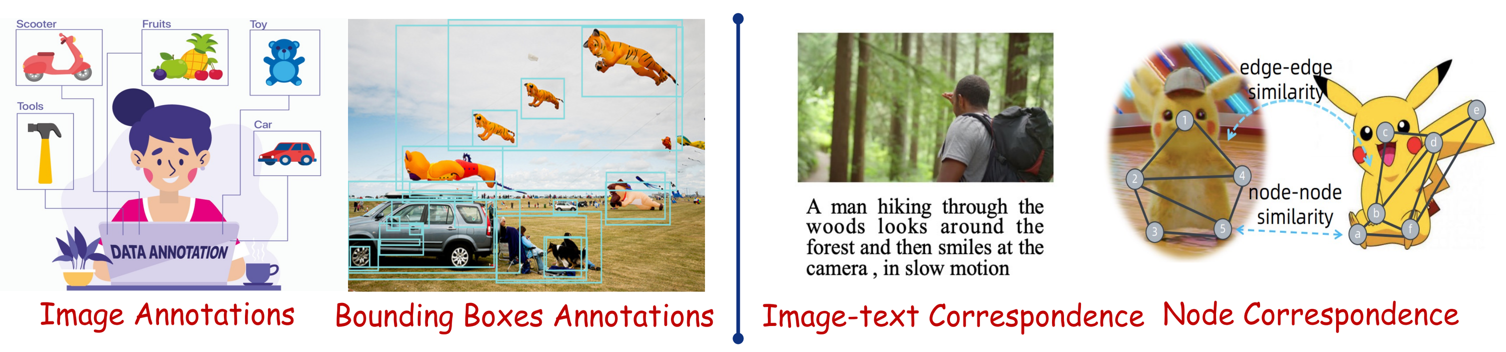
Mouxing Yang[1], Zhenyu Huang[1], Peng Hu[1], Taihao Li[2], Jiancheng Lv[1], Xi Peng[1]*

[1] Sichuan University [2] Zhejiang Lab, China *Corresponding Author

SICHUAN UNIVERSITY 四川大學 · 之江实验室 ZHEJIANG LAB · CVPR JUNE 19-24 2022 NEW ORLEANS LOUISIANA

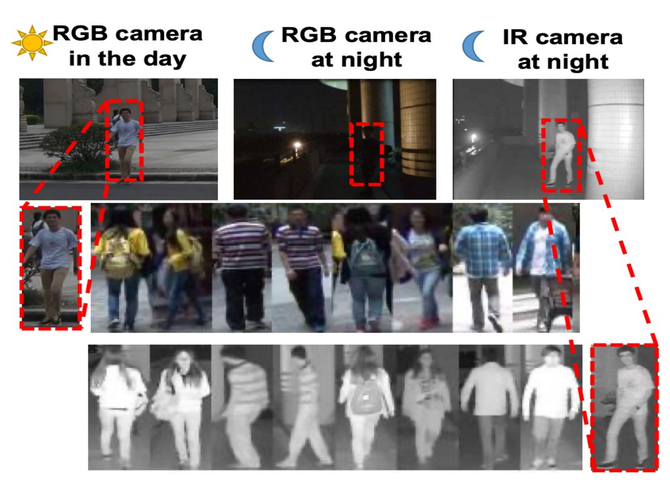
Observations & Motivations

Annotation and Correspondence: The data annotation aims at labeling the data in various formats like images and regions. The data correspondence refers to as the relationships between different samples such as images, texts and nodes.

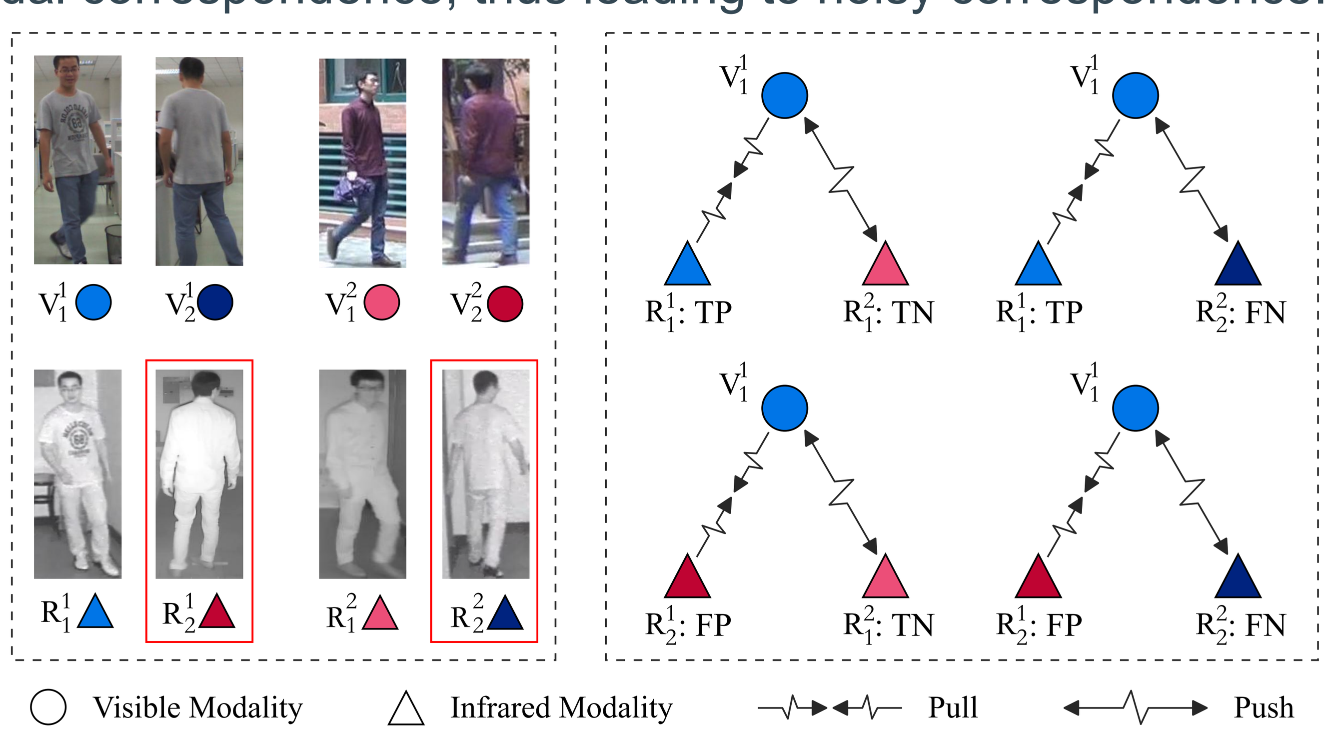


Examples of different types of data annotations and correspondences

VI-ReID: The task aims at finding the corresponding identities across visible and infrared modalities by simultaneously using the identity annotation (IA) and the cross-modal correspondence established by IA.



Twin Noisy Labels: In VI-ReID, some persons may be annotated with the wrong identity due to the poor recognizability in the infrared modality, which will eventually contaminate the cross-modal correspondence, thus leading to noisy correspondence.



(a) Noisy Annotations (b) Noisy Correspondence

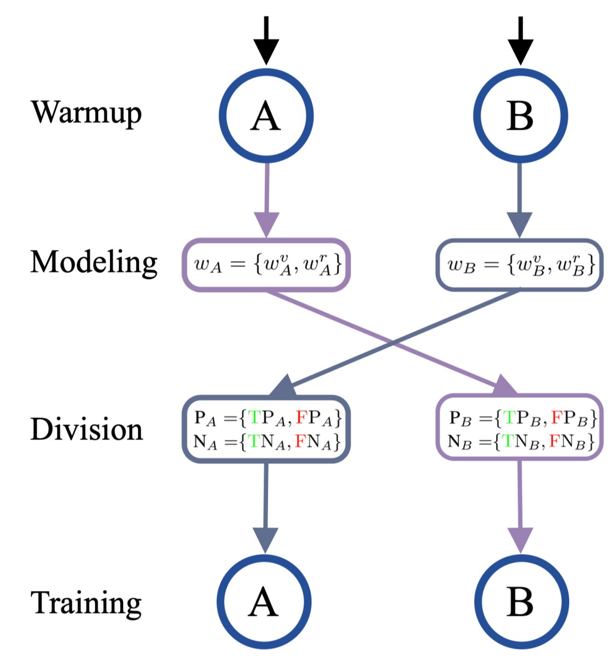
The twin noisy labels in VI-ReID. In the figure, V_i^j/R_i^j denotes sample i with the annotated identity j from the visible/infrared modality, the color indicates the latent correct identity, and R_1^2 and R_2^2 are noisy annotations.

Highlights & Contributions:

- We reveal a new problem for VI-ReID, termed twin noisy labels (TNL), which could be a new paradigm for noisy labels;
- To achieve robust VI-ReID, we propose a novel method for learning with TNL, termed dually robust training (DART), which could be the first successful solution towards TNL.

Method

Overview of the proposed method.



DART first computes the confidence of annotations by resorting to the memorization effect of DNNs. Then, DART divides the data into four groups, and rectifies the noisy correspondence with the estimated confidence. Finally, DART employs a novel dually robust loss consisting of a soft identification loss and an adaptive quadruplet loss to achieve robustness on the noisy annotation (NA) and noisy correspondence (NC).

➢ Co-modeling

Based on the memorization effect of DNNs, DART computes the correctly annotated confidence of each sample at each epoch.

1. Warmup two networks individually using the vanilla identification loss.

$$\mathcal{L}^{id}(\boldsymbol{x}_i^t, \boldsymbol{y}_i^t) = -\log P\left(\boldsymbol{y}_i^t \mid C^t\left(F^t\left(\boldsymbol{x}_i^t\right)\right)\right)$$

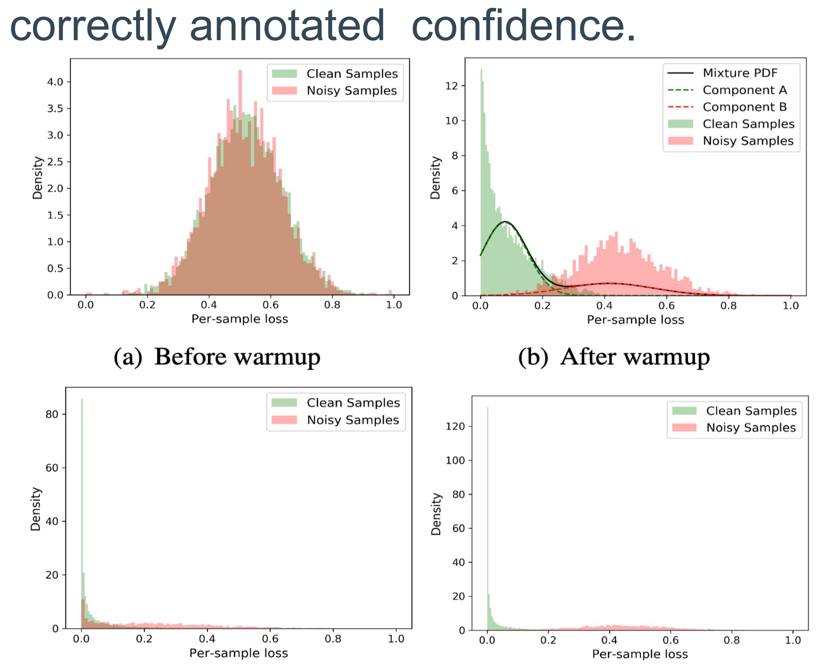
2. GMMs are used to estimate the correctly annotated confidence.

Gaussian mixture model:

$$p(\ell^{id} \mid \theta_t) = \sum_{k=1}^{K} \gamma_k \phi(\ell^{id} \mid k)$$

Clean confidence:

$$w_i = p(\kappa \mid \ell_i^{id})$$



(a) Before warmup (b) After warmup (c) Without $\mathcal{L}^{sid}$ (d) With $\mathcal{L}^{sid}$

➢ Pair Division & Rectification

DART divides the constructed pairs into four subsets, *i.e.*, true positive (TP), true negative (TN), false positive (FP), and false negative (FN) pairs, and rectifies their correspondences.

1. Division:

$$\mathcal{S}^c = \{(\boldsymbol{x}_i^{t_1}, \boldsymbol{x}_j^{t_2}), y_{ij}^p \mid w_i > \eta, w_j > \eta\}$$
$$\mathcal{S}^n = \{(\boldsymbol{x}_i^{t_1}, \boldsymbol{x}_j^{t_2}), y_{ij}^p \mid w_i > \eta, w_j \leq \eta\}$$

2. Rectification:

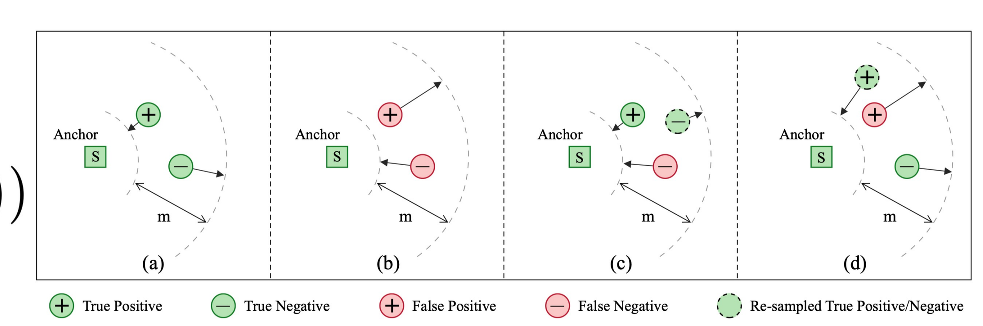
$$\hat{y}_{ij}^p = \mathbb{I}(y_{ij}^p \in \mathcal{S}^c) \odot y_{ij}^p$$
$$\hat{y}_{ij}^p = \mathbb{I}(C^t(F^t(\boldsymbol{x}_i^{t_1})) = C^t(F^t(\boldsymbol{x}_j^{t_2}))), \forall(\boldsymbol{x}_i^{t_1}, \boldsymbol{x}_j^{t_2}) \in \mathcal{S}^n$$

➢ Dually Robust Loss

$$\mathcal{L} = \mathcal{L}^{sid} + \mathcal{L}^{qdr}$$

$$\mathcal{L}^{sid} = -w_i \log P\left(\boldsymbol{y}_i^t \mid C^t\left(F^t\left(\boldsymbol{x}_i^t\right)\right)\right)$$

$$\mathcal{L}^{qdr} = \mathcal{L}^{tri} + \mathcal{L}^{qdt}$$



$$\mathcal{L}^{tri} = m + \frac{(-1)^{(\hat{y}_{ij}^p \otimes \hat{y}_{ik}^p)(1-\hat{y}_{ij}^p)} d_{ij} + (-1)^{(\hat{y}_{ij}^p \otimes \hat{y}_{ik}^p)(1-\hat{y}_{ik}^p)} d_{ik}}{2^{\hat{y}_{ij}^p \odot \hat{y}_{ik}^p}} \quad \mathcal{L}^{qdt} = (-1)^{\hat{y}_{ij}^p \hat{y}_{ik}^p} (\hat{y}_{ij}^p \odot \hat{y}_{ij}^p) d_{is}$$

Experiment

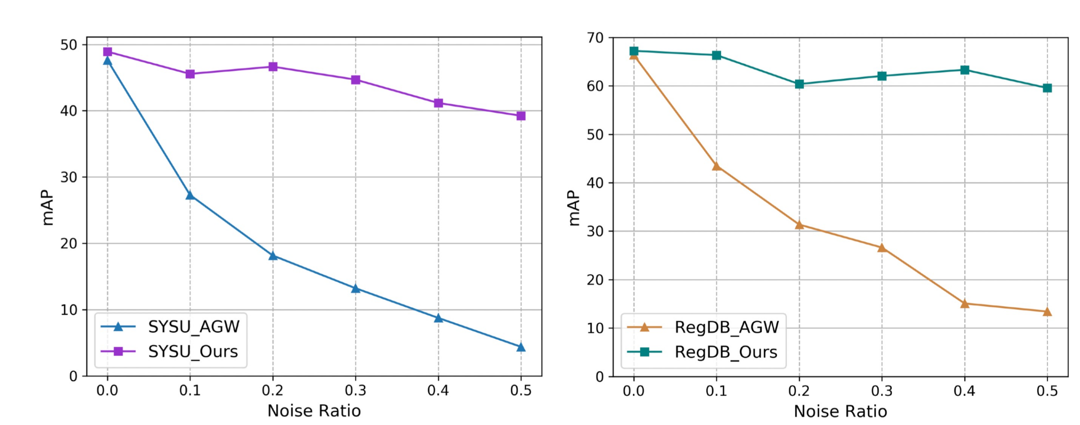
➢ Comparison on SYSU-MM01 under various noise rates

Noise	Methods	All-Search Rank-1	All-Search Rank-10	All-Search Rank-20	All-Search mAP	All-Search mINP	Indoor-Search Rank-1	Indoor-Search Rank-10	Indoor-Search Rank-20	Indoor-Search mAP	Indoor-Search mINP
0%	AGW (TPAMI2021)	47.50	84.39	92.14	47.65	35.30	54.17	91.14	95.98	62.97	59.23
	DDAG (ECCV2020)	54.75	90.39	95.81	53.02	39.62-	61.02	94.06	98.41	67.98	62.61
	LbA (ICCV2021)	55.41	–	–	54.14	–	58.46	–	–	66.33	–
	MPANet (CVPR2021)	**70.58**	96.21	98.8	**68.24**	–	**76.74**	**98.21**	**99.57**	**80.95**	–
	ADP (ICCV2021)	69.88	95.71	98.46	66.89	**53.61**	76.26	97.88	99.49	80.37	**76.79**
	DART (Ours)	68.72	**96.39**	**98.96**	66.29	53.26	72.52	97.84	99.46	78.17	74.94
20%	AGW (TPAMI2021)	17.68	56.80	72.45	18.15	8.55	20.83	65.01	82.43	29.80	25.31
	DDAG (ECCV2020)	14.55	46.58	61.81	13.99	5.56	15.13	50.68	69.33	22.37	18.34
	LbA (ICCV2021)	9.86	39.47	55.85	10.23	3.84	10.10	44.06	64.45	17.39	13.97
	MPANet (CVPR2021)	21.59	63.58	78.71	21.21	–	23.80	70.18	86.44	33.17	–
	ADP (ICCV2021)	25.44	67.55	80.88	23.71	11.05	26.61	70.68	85.19	34.97	29.61
	ADP-C (ICCV2021)	63.67	94.13	97.78	61.57	48.02	68.52	96.13	98.73	73.82	69.66
	DART (Ours)	**66.31**	**95.31**	**98.38**	**64.13**	**50.69**	**70.52**	**97.08**	**99.03**	**75.94**	**72.30**
50%	AGW (TPAMI2021)	7.93	37.56	55.78	9.75	4.38	9.61	47.87	70.47	18.14	15.22
	DDAG (ECCV2020)	6.68	28.95	43.77	7.52	2.93	8.39	37.87	57.86	15.12	12.33
	LbA (ICCV2021)	2.67	17.78	30.27	4.15	1.85	4.87	29.39	48.97	10.96	8.63
	MPANet (CVPR2021)	6.98	32.75	49.16	8.20	–	8.47	40.71	61.37	15.85	–
	ADP (ICCV2021)	8.00	42.55	62.14	10.83	5.21	11.49	52.99	76.77	20.81	17.53
	ADP-C (ICCV2021)	59.17	92.52	97.28	56.49	41.80	62.99	94.84	98.08	69.05	64.29
	DART (Ours)	**60.27**	**93.41**	**97.47**	**58.69**	**45.33**	**65.74**	**95.04**	**98.23**	**71.77**	**68.14**

➢ Ablation studies on SYSU-MM01 with noise ratio of 20%

Method	SYSU-MM01 under All-search Evaluation Rank-1	Rank-10	Rank-20	mAP	mINP
B	25.44	67.55	80.88	23.71	11.05
B + $\mathcal{L}^{sid}$	49.24	89.14	95.66	46.78	31.32
B + $\mathcal{L}^{sid}$ + $\mathcal{L}^{tri}$	65.44	95.01	98.13	63.15	50.35
B + $\mathcal{L}^{sid}$ + $\mathcal{L}^{tri}$ + $\mathcal{L}^{qdt}$	**66.31**	**95.31**	**98.38**	**64.13**	**50.69**

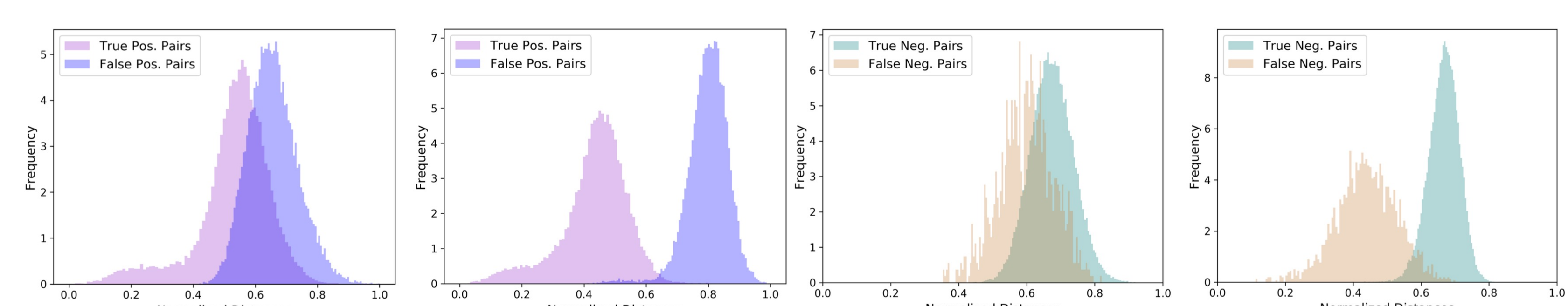
➢ Study on Robustness and Generalizability



The performance comparisons between AGW and our DART on various noise ratios from 0% to 50% with an interval of 10% show the robustness and generalizability ability of RART.

➢ Visualization on the Robustness

The baseline cannot handle NC so that different kinds of pairs are mixed up. In contrast, DART will prevent NC from dominating the network optimization.



(a) TP-FP distributions of ADP (b) TP-FP distributions of DART (c) TN-FN distributions of ADP (d) TN-FN distributions of DART

The code is available at https://github.com/XLearning-SCU/2022-CVPR-DART

Feel free to drop me an email.

yangmouxing@gmail.com