Supplementary Material for
Learning with Twin Noisy Labels for Visible-Infrared Person Re-Identication

Mouxing Yang[1], Zhenyu Huang[1], Peng Hu[1], Taihao Li[2], Jiancheng Lv[1], Xi Peng[1]*
[1] College of Computer Science, Sichuan University.
[2] Zhejiang Lab, China.
{yangmouxing, zyhuang.gm, penghu.ml, pengx.gm}@gmail.com;
lith@zhejianglab.com; lvjiancheng@scu.edu.cn

1. Introduction

In the supplementary material, we present additive experiments by comparing DART with a noisy annotation (NA) oriented method as discussed in Introduction. In addition, more experiments results are presented to further verify the effectiveness of DART.

2. Comparison with NA-oriented Method

In the Introduction, we elaborate on the necessity and effectiveness of this work. Namely, it is intractable to adopt existing NA-oriented methods to rectify the noisy annotations in VI-ReID so that the twin noisy label (TNL) problem is solved. To support such a claim, here, we compare DART with DivideMix [2] which is the state-of-the-art method on learning with noisy annotations. In brief, DivideMix first groups the training data into clean and noisy portions with the help of a co-divide module. Then the labels in clean and noisy portions are refined and re-generated with the prediction of two models. After that, MixMatch [1] is applied on the divided portions with revised labels to learn with noisy annotations. For a comprehensive comparison, we adopt DivideMix to preprocess the noisy annotations on each modality with the following two settings:

- DivideMix-divide: the data of each modality is divided into clean and noisy portions where the former is used for further training and the latter is discarded.

- DivideMix-full: the full pipeline of DivideMix is applied on each modality with carefully tuning to get the revised labels.

In the experiment, we first report the dividing and revised results of DivideMix-divide and DivideMix-full in Table 1-2. After that, we pass the data processed by these two baselines through ADP [3] for an extensive comparison. Notably, for the case where the processed class numbers between two modalities are inconsistent, we use the intersection of the same class across modalities for training. The results are summarized in Table 3-4. From the above results, one could have the following observations:

- It is challenging to distinguish the clean portion from the noisy one so that the influence of noisy labels is minimized. Specifically, as shown in Table 1, the capacity of the co-divide module is limited which leads to the inaccurate division and even the incomplete

Table 1. Performance of DivideMix-divide on the visible and infrared modalities of SYSU-MM01 and RegDB datasets. In the table, "Divide Acc." refers to the accuracy of correct division on clean portion and "Class Num." refers to the total class number of the divided clean portion.

Dataset	Modality	Noise(%)	Divide Acc.(%)	Class Num.
SYSU-MM01	Visible	20	94.17	395
		50	82.77	377
	Infrared	20	94.26	395
		50	77.39	383
RegDB	Visible	20	95.57	205
		50	80.87	189
	Infrared	20	91.69	206
		50	62.43	170

Table 2. Performance of DivideMix-full on the visible and infrared modalities of SYSU-MM01 and RegDB datasets. In the table, "Revised Acc." refers to the accuracy of revised labels on the divided clean and noisy portion and "Class Num." refers to the total class number of the revised labels.

Dataset	Modality	Noise(%)	Revised Acc.(%)	Class Num.
SYSU-MM01	Visible	20	85.80	354
		50	85.63	339
	Infrared	20	95.84	382
		50	75.74	344
RegDB	Visible	20	98.56	206
		50	77.18	176
	Infrared	20	98.85	206
		50	77.34	178

*Corresponding author

Table 3. Additional comparisons with the NA-oriented method on the SYSU-MM01 dataset under the noise ratio of 20% and 50%, respectively. The best and second best results are highlight in **bold** and underline.

Noise	Methods	All-Search					Indoor-Search				
		Rank-1	Rank-10	Rank-20	mAP	mINP	Rank-1	Rank-10	Rank-20	mAP	mINP
20%	ADP (ICCV2021)	25.44	67.55	80.88	23.71	11.05	26.61	70.68	85.19	34.97	29.61
	DivideMix-divide	61.83	93.23	97.13	58.36	43.36	67.02	95.4	98.70	72.51	68.17
	DivideMix-full	60.15	92.12	96.81	56.84	41.9	66.89	95.88	98.41	72.78	68.82
	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%	ADP (ICCV2021)	8.00	42.55	62.14	10.83	5.21	11.49	52.99	76.77	20.81	17.53
	DivideMix-divide	8.20	37.29	54.75	9.82	4.08	10.24	42.57	62.00	17.75	15.11
	DivideMix-full	53.84	90.15	96.12	52.03	38.05	59.6	93.52	97.28	67.27	63.84
	ADP-C (ICCV2021)	59.17	92.52	97.28	56.49	41.80	62.99	94.84	98.80	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**

Table 4. Additional comparisons with the NA-oriented method on the RegDB dataset under the noise ratio of 20% and 50%, respectively. The best and second best results are highlight in **bold** and underline.

Noise	Methods	Visible to Thermal					Thermal to Visible				
		Rank-1	Rank-10	Rank-20	mAP	mINP	Rank-1	Rank-10	Rank-20	mAP	mINP
20%	ADP (ICCV2021)	50.71	78.97	85.33	35.92	14.12	49.98	77.46	84.35	34.75	12.62
	DivideMix-divide	74.32	90.44	94.22	62.60	41.84	71.31	87.67	91.65	60.37	39.5
	DivideMix-full	80.27	91.15	95.28	72.97	57.55	78.94	92.39	94.96	69.53	55.61
	ADP-C (ICCV2021)	78.39	90.75	94.47	70.02	51.80	71.81	90.92	94.35	68.95	51.19
	DART (Ours)	**82.04**	**93.39**	**96.00**	**74.18**	**57.89**	**79.48**	**92.55**	**95.28**	**71.72**	**54.47**
50%	ADP (ICCV2021)	17.04	41.17	51.46	11.25	3.55	20.28	47.30	58.43	12.31	3.24
	DivideMix-divide	16.46	38.83	48.88	11.61	4.21	17.52	40.00	50.83	11.81	3.43
	DivideMix-full	63.98	83.79	89.51	55.86	38.23	63.54	84.42	89.81	53.68	34.51
	ADP-C (ICCV2021)	77.43	90.84	93.89	66.75	47.25	74.89	90.15	94.38	63.05	41.83
	DART (Ours)	**78.23**	**91.16**	**95.11**	**67.04**	**48.36**	**75.04**	**91.02**	**94.56**	**64.38**	**43.62**

class number phenomenon. As a result, its performance is inferior to ADP-C (ADP using only clean data) and our DART as depicted in Table 3-4. Besides, ADP-C, which is suboptimal to DART, could be regarded as the upper limit of DivideMix-divide. This also verifies that the DART could well utilize the noisy data discarded by ADP-C.

- It is intractable to rectify the noisy annotation ideally so that the TNL problem is solved. Specifically, DivideMix cannot revise the label of clean and noisy portions perfectly, especially on the challenging SYSU-MM01 dataset or under a high noise ratio. The incomplete class number phenomenon is also inevitable. As a result, its performance is also suboptimal compared to our DART.

Clearly, the above observations could further verify the effectiveness of DART and support our claim that it is suboptimal to adopt the existing NA-oriented method to combat the twin noisy labels in VI-ReID.

3. Additional Experiment Results

In this section, we present additional experiment results to verify the effectiveness of our DART. In brief, we first present the full results of DART on RegDB and then further study the robustness of DART.

3.1. Full Results on RegDB

Due to the limited space of the manuscript, we only present the main results of DART on RegDB. In the subsection, we present the full results on RegDB for a comprehensive comparison. The results are shown in Table 5 where one could observe that DART still achieves a remarkable performance superiority comparing with the baselines.

3.2. Robustness of DART

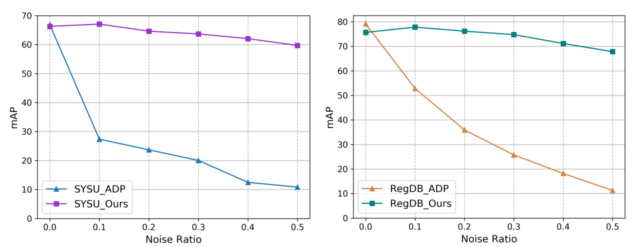
In the manuscript, we perform DART under the noise ratio of $0\%, 20\%$, and 50%, respectively. In this section, to further verify the robustness, we compare DART with ADP under the noise ratio increasing from 0% to 50% with an internal of 10%. The results are shown in Fig 1 which shows the robustness of DART.

4. Broader Impact.

This work could be one of the first works to reveal the importance of the twin noisy label problem in the VI-ReID task. Solving this problem could improve the tolerance for the errors of annotation and the accompanying noisy correspondence, which might benefit the practitioners in the

Table 5. Comparisons with state-of-the-art methods on the RegDB dataset under the noise ratio of 0%, 20% and 50%, respectively. The best and second best results are highlight in **bold** and underline.

Noise	Methods	Visible to Thermal					Thermal to Visible				
		Rank-1	Rank-10	Rank-20	mAP	mINP	Rank-1	Rank-10	Rank-20	mAP	mINP
0	AGW (TPAMI2021)	70.05	86.21	91.55	66.37	50.19	70.49	87.12	91.84	65.9	51.24
	DDAG (ECCV2020)	69.34	86.19	91.49	63.46	49.24	68.06	85.15	90.31	61.80	48.62
	LbA (ICCV2021)	74.17	–	–	67.64	–	72.43	–	–	65.46	–
	MPANet (CVPR2021)	83.70	–	–	**80.90**	–	82.8	–	–	**80.70**	–
	ADP (ICCV2021)	**85.03**	**95.49**	**97.54**	79.14	**65.33**	**84.75**	**95.33**	**97.51**	77.82	**61.56**
	DART (Ours)	83.60	93.74	96.22	75.67	60.60	81.97	92.98	95.53	73.78	56.70
20%	AGW (TPAMI2021)	47.77	74.33	81.80	31.35	12.43	47.18	73.73	81.42	30.86	11.85
	LbA (ICCV2021)	35.99	63.33	72.02	23.48	7.49	36.18	63.87	72.30	22.75	6.74
	DDAG (ECCV2020)	39.27	64.24	73.09	25.74	10.03	37.69	64.05	72.44	25.07	9.61
	MPANet (CVPR2021)	33.83	69.22	77.72	23.50	–	32.62	74.27	82.67	22.06	–
	ADP (ICCV2021)	50.71	78.97	85.33	35.92	14.12	49.98	77.46	84.35	34.75	12.62
	ADP-C (ICCV2021)	78.39	90.75	94.47	70.02	51.80	71.81	90.92	94.35	68.95	51.19
	DART (Ours)	**82.04**	**93.39**	**96.00**	**74.18**	**57.89**	**79.48**	**92.55**	**95.28**	**71.72**	**54.47**
50%	AGW (TPAMI2021)	21.87	48.16	58.14	13.40	3.93	20.98	47.70	58.76	12.95	3.70
	DDAG (ECCV2020)	24.03	50.00	60.73	14.44	4.25	21.46	44.90	54.81	13.38	4.28
	LbA (ICCV2021)	11.65	31.60	42.38	6.68	1.53	10.24	31.26	41.26	6.34	1.46
	MPANet (CVPR2021)	9.51	35.44	47.48	6.13	–	11.41	34.85	47.62	6.67	–
	ADP (ICCV2021)	17.04	41.17	51.46	11.25	3.55	20.28	47.30	58.43	12.31	3.24
	ADP-C (ICCV2021)	77.43	90.84	93.89	66.75	47.25	74.89	90.15	94.38	63.05	41.83
	DART (Ours)	**78.23**	**91.16**	**95.11**	**67.04**	**48.36**	**75.04**	**91.02**	**94.56**	**64.38**	**43.62**



(a) (b)

Figure 1. Performance comparisons between DART+ADP and ADP on SYSU-MM01 and RegDB with varying noise ratios.

industry. Despite the benefit, the potential negative impacts should be considered. Especially, the increasing accuracy of person ReID may raise the risk of privacy breaches and other security issues. Besides, DART may cause job loss since its tolerance on twin noisy labels, thus reducing the cost for annotations. Finally, It should be pointed that DART learns the recognization pattern from the training dataset which may have limited ability on adaptation to datasets of other domains.

References

[1] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel. Mixmatch: A holistic approach to semi-supervised learning. *arXiv:1905.02249*, 2019. 1

[2] Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv:2002.07394*, 2020. 1

[3] Mang Ye, Weijian Ruan, Bo Du, and Mike Zheng Shou. Channel augmented joint learning for visible-infrared recognition. In *ICCV*, pages 13567–13576, 2021. 1