Supplementary Materials for
Good is Bad: Causality Inspired Cloth-debiasing for Cloth-changing Person Re-identification

Abstract

This supplementary material accompanies our main manuscript: "Good is Bad: Causality Inspired Cloth-debiasing for Cloth-changing Person Re-identification". We present a further analysis of causal inference and provide a comprehensive study and more visualization results on PRCC [5] and LTCC [3] in this supplementary material.

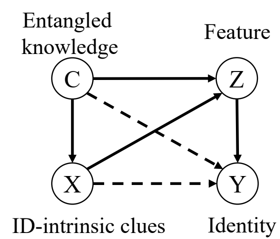


Figure 1. Causal graph for CC-ReID.

1. More Details on Causal Inference

1.1. Causal Graph

Here we will give a more detailed presentation of the proposed causal graph for Cloth-Changing person Re-IDtification (CC-ReID). In the causal graph of CC-ReID, $\boldsymbol{X}$ stands for the IDentity (ID)-intrinsic clues that only related to people; $\boldsymbol{Y}$ as the ID predicted by the ReID model, whose expectation value is equivalent to the ground truth; $\boldsymbol{Z}$ represents the feature produced by ReID model; $\boldsymbol{C}$ is term as the entangled knowledge that affect $\boldsymbol{X}$, which is also known as the confounder, which has two or more outer edges with $\boldsymbol{C}$ as the root node. Many potential elements are responsible for the entangled knowledge, where clothing, as part of a person's intuitive understanding, has a greater impact on ReID than others. The relationship between clothing and ID-intrinsic clues is so entangled that the two cannot be clearly separated, which is the primary focus of this paper.

$\boldsymbol{C} \rightarrow \boldsymbol{X}$ denotes the entangled representation contributing to the faulty association between clothing and ID. There are many potential elements responsible for the entangled knowledge, they may be from noisy backgrounds, properties bound to ID or from the training process (e.g. momentum and batch normalization), or both. Clothing knowledge is one of an attribute highly twisted by human beings. The entangled representation between clothing and ID-intrinsic clues easily allows the model to establish a false association between them, as reflected by the fact that the model's attention is mainly focused on clothing attributes rather than more ID-representative features.

$\boldsymbol{C} \rightarrow \boldsymbol{X} \rightarrow \boldsymbol{Z}$ denotes when an image is sent to the ReID model for representation learning, the entangled non-ID knowledge will be learned at the same time as the ID-intrinsic information, which interferes with the model's judgment of ID. Ideally, the model learns that $\boldsymbol{Z}$ is only related to $\boldsymbol{X}$, but clothing is entangled with ID, making it difficult for the model to separate these two types of features independently. Suppose there is a person with a striped shirt and blue pants, without a unique understanding of the clothing, when the same dressed person reappears, the model can easily determine that the two are the same ID based on the same striped T-shirt and blue pants. The key idea of counterfactual learning is to distill clothing bias and obtain the cloth-agnostic representation of present samples without affecting any other factors.

$(\boldsymbol{C}, \boldsymbol{X}) \rightarrow \boldsymbol{Z} \rightarrow \boldsymbol{Y}$ represents the complete CC-ReID prediction process, which can be divided into two links. The first is the ideal path $\boldsymbol{X} \rightarrow \boldsymbol{Z} \rightarrow \boldsymbol{Y}$, representing the ideal CC-ReID is performed through the feature of factual related to ID. $\boldsymbol{Z} \rightarrow \boldsymbol{Y}$ denotes that the extracted feature determines the final person recognition. Although $\boldsymbol{Z}$ appears as a mediator, it is inevitable for existing deep learning models. The second link is $\boldsymbol{C} \rightarrow \boldsymbol{X} \rightarrow \boldsymbol{Z} \rightarrow \boldsymbol{Y}$, which is the path affected by entangled knowledge.

2. More Details on Datasets

2.1. PRCC-ReID

The PRCC-ReID [5] is a recently proposed person ReID dataset. It consists of 221 IDs with three camera views. Each person has about 50 images in each camera view, including 33,698 images in total. Each person wears the same clothes

Table 1. Comparision on different evaluation strategies for cloth-changing settings on PRCC-ReID. **Bold** numbers are the best results under each strategy.

Method	Evaluation		PRCC-ReID			
	Single	Multi	R@1	R@5	R@10	mAP
IRANet [4]	○	●	54.9	-	-	53.0
CAL [1] (Baseline)	○	●	54.4	58.5	62.6	54.4
AIM (ours)	○	●	**57.9**	**62.1**	**63.9**	**58.3**
FSAM [2]	●	○	54.5	-	**86.4**	-
CAL [1] (Baseline)	●	○	54.6	70.8	82.6	62.7
AIM (ours)	●	○	**57.8**	**78.0**	85.5	**67.6**

in camera A and camera B, and different clothes in camera A and camera C. There are 150 IDs, 17,896 images used for training, and the remaining 71 IDs, 15,802 images used for testing.

2.2. LTCC-ReID

The LTCC-ReID [3] is a recently proposed long-term person ReID dataset that is captured over long periods of time, *e.g.*, days and months, with frequent changes of clothes. It contains 152 IDs with a total of 17,119 labeled images of people, each ID is captured by at least two cameras. The entire dataset is divided into two subsets: a cloth-changing dataset containing 91 IDs, 416 different outfits, and 14,783 images; a cloth-consistent subset containing the remaining 61 IDs, 2,236 images without a clothing change.

3. More Details on Performance

3.1. Evaluation Strategies

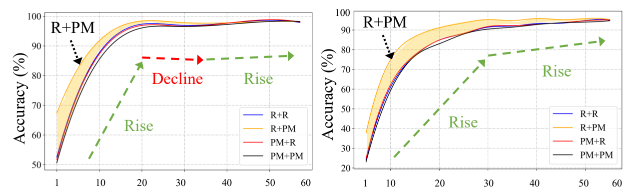
There exist two strategies in the evaluation of clothing-changing setting on PRCC-ReID: single-shot (SS) matching strategy and multi-shot (MS) matching strategy. The SS means randomly selecting an image of a different outfit for each ID as a gallery for testing. Some methods adopt SS several times and use the average result for evaluation, but MS is the widely used one. For a fair comparison, we adopt MS in the main manuscript. To better illustrate the performance of the proposed AIM, we also give a comparison by adopting SS metrics in Table. 1. The results demonstrate in both SS and MS settings, the proposed AIM is outperform the baseline method [1] and the second-best methods [2,4].

3.2. Analysis on complexity

We only use the simple ResNet50 of the ID branch for testing. The features obtained from the ID branch are constantly corrected in the training stage by causal auto-intervention, which does not require clothing bias estimation. We evaluate the number of parameters and running time of our model and other models in Table 2. Our method reduces a lot of parameters when compared to training with testing. In the

Table 2. Comparisons of network parameters (Params) and training and testing time. Note that all experiments are conducted fairly on two Tesla V100 GPUs and single-shot matching strategy is used for a fair comparison.

Method	Training		Testing		PRCC-ReID	
	Params	Time	Params	Time	R@1	R@10
RGA [6]	30.13M	0.8h	30.13M	40s	42.3	79.4
ISP [7]	31.68M	16.5h	31.68M	30s	36.6	66.5
FSAM [2]	164.27M	12h	23.82M	15s	54.5	86.4
CAL [1] (Baseline)	23.52M	2.2h	23.52M	58s	54.6	82.6
AIM (ours)	72.67M	4.1h	23.52M	58s	57.8	85.5



(a) Clothing Accuracy on PRCC-ReID (b) Clothing Accuracy on LTCC-ReID

Figure 2. Visualization of the training accuracy of the clothing branch. Different colors indicate different combinations.

testing stage, the proposed AIM achieves better performance and does not increase the parameters, compared with baseline and other methods. Our comparison with other methods is fair with regard to the parameters and the computational costs involved in testing.

3.3. More analysis of Pyramid Matching.

To verify the efficacy of PM in another dimension, we provide the training accuracy of the clothing branch. To a certain extent, this proves the advantages of adopting ResNet (R) as the backbone of ID branch and adopting the Pyramid Matching (PM) strategy for the clothing branch. Fig. 2 shows the changes in clothing accuracy through training. The orange curve represents the best performance by adding PM only for the clothing branch. Although the accuracy of each combination can gradually reach a high level, R + PM can maintain a higher accuracy throughout the training process, which is more conducive to the elimination of clothing bias.

4. More Visualization Results

As mentioned, the entangled representation of clothing and ID-intrinsic clues is an obstacle to the existing CC-ReID methods. The solution to this problem emphasizes eliminating the influence of clothing and focusing on the intrinsic properties of people. To notice the difference between the baseline [1] and ours and further demonstrate the ability of AIM to eliminate clothing bias, we give more results of visualization and retrieval ranks.

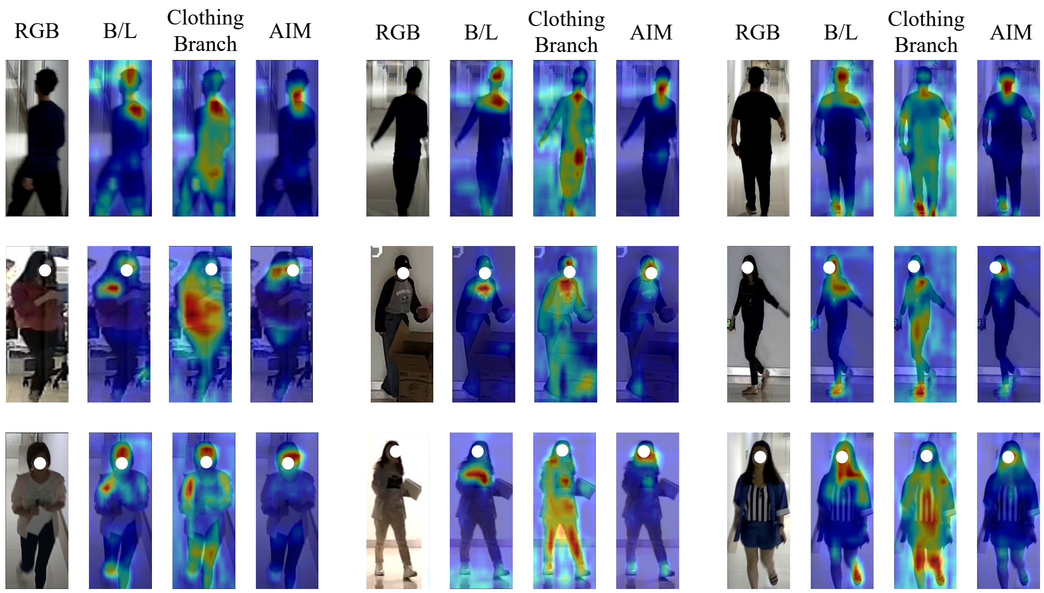
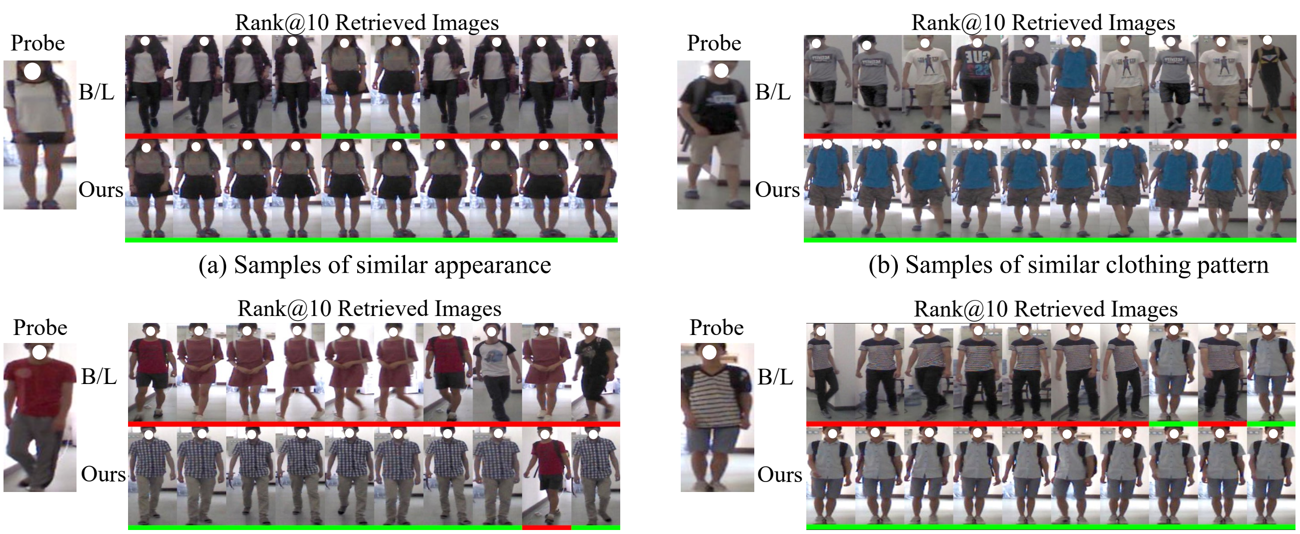


Figure 3. More visualization results. Each group consists of 4 images, which are RGB image, heatmap of features obtained from the baseline, heatmap of the feature obtained from the clothing branch of AIM, heatmap of the feature obtained from the ID branch of AIM, respectively.



(a) Samples of similar appearance

(b) Samples of similar clothing pattern

(c) Samples of similar clothing color

(d) Samples of similar clothing style

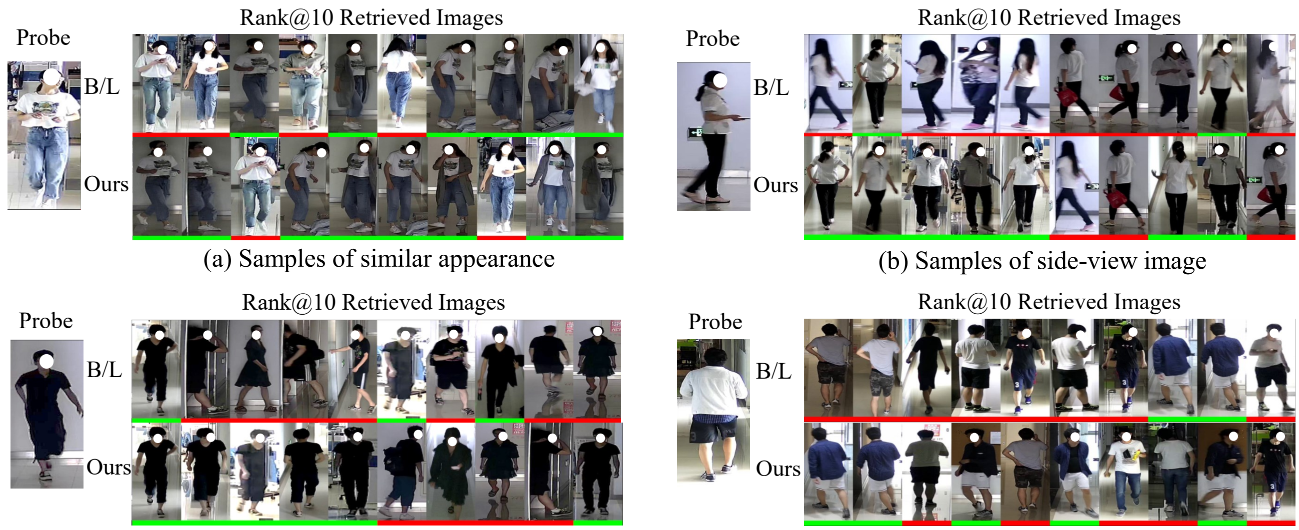
Figure 4. More retrieval results of samples in PRCC-ReID.

4.1. More Visualization of Heatmaps

Fig. 3 gives more visualization results of the feature heatmap, which shows features of interest. The comparison between AIM and baseline shows that AIM can better concentrate on ID-intrinsic clues without affecting by clothing bias.

4.2. More Visualization of Retrieval

Fig. 4 illustrates the results of the retrieval ranks on PRCC-ReID. 10 candidate gallery images with the highest similarity to the probe image are displayed. Specifically, we illustrate four scenarios: Fig. 4(a) shows samples with similar appearance and partially similar clothing, which is difficult to distinguish even for humans. Fig. 4(b) demonstrate the baseline method is misled by similar clothing pattern. Fig. 4(c) describes the clothing color as a misleading factor. Fig. 4(d) indicates that clothing style affects ID representation learning. Fig. 5 illustrates the results of the retrieval ranks on LTCC-ReID, including samples of (a) similar appearance and (c) clothing color, (b) side-view images, and (d) back-view images. The above results indicate that clothing bias, including pattern, color, and style, influences the ReID model to some extent.



(a) Samples of similar appearance

(b) Samples of side-view image

(c) Samples of similar clothing color

(d) Samples of back-view image

Figure 5. More retrieval results of samples in LTCC-ReID.

References

[1] Xinqian Gu, Hong Chang, Bingpeng Ma, Shutao Bai, Shiguang Shan, and Xilin Chen. Clothes-changing person re-identification with RGB modality only. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 1050–1059, 2022. 2

[2] Peixian Hong, Tao Wu, Ancong Wu, Xintong Han, and Wei-Shi Zheng. Fine-grained shape-appearance mutual learning for cloth-changing person re-identification. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 10513–10522, 2021. 2

[3] Xuelin Qian, Wenxuan Wang, Li Zhang, Fangrui Zhu, Yanwei Fu, Tao Xiang, Yu-Gang Jiang, and Xiangyang Xue. Long-term cloth-changing person re-identification. In *Proc. Asian Conf. Comput. Vis.*, pages 71–88, 2020. 1, 2

[4] Wei Shi, Hong Liu, and Mengyuan Liu. IRANet: identity-relevance aware representation for cloth-changing person re-identification. *Image Vis. Comput.*, 117:104335, 2022. 2

[5] Qize Yang, Ancong Wu, and Wei-Shi Zheng. Person re-identification by contour sketch under moderate clothing change. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(6):2029–2046, 2021. 1

[6] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, Xin Jin, and Zhibo Chen. Relation-aware global attention for person re-identification. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 3183–3192, 2020. 2

[7] Kuan Zhu, Haiyun Guo, Zhiwei Liu, Ming Tang, and Jinqiao Wang. Identity-guided human semantic parsing for person re-identification. In *Proc. Eur. Conf. Comput. Vis.*, pages 346–363, 2020. 2