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

Zhengwei Yang[1,2,3], Meng Lin[1,2,3], Xian Zhong[4], Yu Wu[1,2], Zheng Wang[1,2,3,*]

[1] National Engineering Research Center for Multimedia Software, Institute of Artificial Intelligence
[2] School of Computer Science, Wuhan University
[3] Hubei Key Laboratory of Multimedia and Network Communication Engineering
[4] School of Computer Science and Artificial Intelligence, Wuhan University of Technology

Abstract

Entangled representation of clothing and identity (ID)-intrinsic clues are potentially concomitant in conventional person Re-IDentification (ReID). Nevertheless, eliminating the negative impact of clothing on ID remains challenging due to the lack of theory and the difficulty of isolating the exact implications. In this paper, a causality-based Auto-Intervention Model, referred to as AIM[1], is first proposed to mitigate clothing bias for robust cloth-changing person ReID (CC-ReID). Specifically, we analyze the effect of clothing on the model inference and adopt a dual-branch model to simulate causal intervention. Progressively, clothing bias is eliminated automatically with model training. AIM is encouraged to learn more discriminative ID clues that are free from clothing bias. Extensive experiments on two standard CC-ReID datasets demonstrate the superiority of the proposed AIM over other state-of-the-art methods.

1. Introduction

Short-term person Re-IDentification (ReID) aims to match a person within limited time and space conditions, under the assumption that each individual maintains clothing consistency. Of both traditional and deep learning methods, the best way to deceive current ReID models is by having pedestrians alter their clothing. This highlights the inadequacy of existing short-term ReID methods [4, 42, 45]. To solve this issue, Cloth-Changing person ReID (CC-ReID) [2] has been recently explored, which is increasingly critical in public security systems for tracking down disguised criminal suspects. For example, witnesses typically provide descriptive details (*e.g.*, clothing, color, and stature) when describing suspects, but it is unlikely that criminals will wear the same clothes upon their reappearance. It follows that clothing information will disrupt the existing ReID system [40], which leads to a growing need felt by the research community to study CC-ReID task.

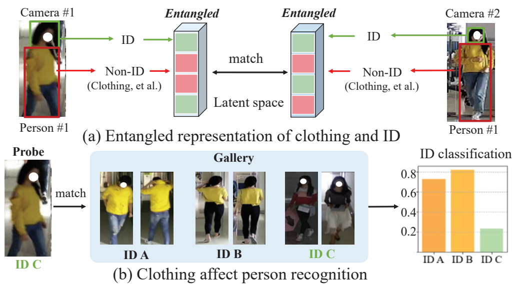


Figure 1. Illustration of the entangled representation of clothing and ID and how clothing bias affects model prediction.

As one of the accompanying objects of people, clothing is an essential factor in social life. There are two possible responses when people identify others: confusing the perception of identity (ID) or clearly distinguishing different IDs through immutable visual appearance (faces or soft-biometrics). The former manifests as the mix-up of IDs due to the similarity in flexible visual appearance (*e.g.*, clothing) of people, while the latter is caused by the high-level semantic (*e.g.*, ID-clues) perceived by humans, transcending the similarity that comes with clothing. The above reactions reflect that the relevance of clothing to ID is a double-edged sword. Traditionally, clothing is a helpful characteristic for ReID, where people wearing the same clothes are likely to have the same ID. However, entangled representation of clothing and ID leads the statistical-based neural networks to converge towards shallow clothing features that can be easily distinguished. This statistical association gives the ReID model a faulty perception that there is a strong relation between clothing and ID, which would undermine the ultimate prediction for seeking robust and sensible results.

Recent years have witnessed numerous deep learning attempts to model discriminative clues for person distinguishable learning. However, plenty of misleading information exists in these attempts, as some non-ID areas (*e.g.*, clothing

*Corresponding author: wangzwhu@whu.edu.cn. [1] Codes will publicly available at https://github.com/BoomShakaY/AIM-CCReID.

and background) may correlate highly with ID. As shown in Fig. 1(a), conventional ReID methods focus on image regions with distinct discrimination characteristics [33, 46], leading to the entanglement of clothing and ID-intrinsic clues. In CC-ReID, this phenomenon will lead to biased classification, misleading the ReID model to focus on the non-ID areas that appear to be ID-related. As shown in Fig. 1(b), clothing may mislead the classifier by giving high scores to person images with similar colors or patterns, but ignoring the faces and details that matter. To this end, if clothing bias can be captured and removed from the existing model, it will enhance the contribution of genuinely ID-relevant features to ID discrimination.

Lacking practical tools to alleviate clothing bias makes it challenging to correct the misleading attention of the current ReID model. Even knowing that clothing is a critical influencing factor, it is not apparent how to intervene in clothing directly in the representation space. Not to mention that rough negligence on clothing will damage the integrity of person representation, *e.g.*, the mask-based [12, 16, 41] and gait-based methods [6, 17] try to obtain cloth-agnostic representations by forcibly covering up or neglecting clothing information. While effective, these straightforward methods lose a plethora of semantic information and overlook the factual relation between clothing and real ID.

Causal inference is a recently emerging theory [19] widely used to extract causality [10] and explore the true association between two events [25]. Thanks to causal inference, we can revisit the issue of clothing bias from a causal perspective. The ladder of causality [7] divides cognitive abilities into three levels from low to high: association, intervention, and counterfactual analysis. Many aforementioned research works explore CC-ReID from the surface level of data association, while more advanced cognitive is not covered. Intervention allows us to incorporate clothing knowledge into the model prediction and eliminate the corresponding effects, in contrast to counterfactual data, which are difficult to obtain under strict variable control. Therefore, this paper attempts to start with intervention by examining the perturbation of clothing on the results and removing such perturbation from model predictions. Through the causal intervention, we attempt to remove the effect of clothing without destroying semantic integrity and further optimizing the learned discriminative features.

To bring the theoretical intervention into practice, we design a dual-branch model to capture clothing bias and ID clues separately and strip clothing inference from ID representation learning to simulate the entire intervention process. The clothing branch represents the model's perception of clothing, breaking the false association between clothing and ID brought by the entangled representation. Subsequently, while maintaining semantic integrity, this paper achieves bias elimination and improves the robustness of ID representation by constantly mitigating the influence of clothing on ID classification. Further, to improve the accuracy of clothing bias distillation, as clothing has top-middle-bottom characteristics, this paper adopts pyramid matching strategy [8] to enhance the partial feature representation of clothing. Additionally, we introduce two learning objectives explicitly designed to encourage clothing mitigation. A knowledge transfer objective is adopted to strengthen the perception of clothing bias entangled with ID-intrinsic representation. A bias elimination objective is utilized to cooperate with the causal auto-intervention for ID-intrinsic feature extraction.

Our contributions can be summarized threefold:

- We propose a novel causality-based Auto-Intervention Model (AIM) for Cloth-Changing person Re-IDentification (CC-ReID). The proposed AIM guarantees that the learned representation is unaffected by clothing bias. To the best of our knowledge, AIM is the first model to introduce causality into CC-ReID.
- A dual-branch model is proposed to simulate the causal intervention. Clothing bias is gradually stripped from the entangled ID-clothing representation without destroying semantic integrity, which optimizes the ID-intrinsic feature learning.
- We comprehensively demonstrate how clothing bias affects the current ReID model and highlight the significance of causal inference in CC-ReID. The experimental results on two CC-ReID datasets, PRCC-ReID [38] and LTCC-ReID [26], show that AIM outperforms state-of-the-art methods.

2. Related Work

2.1. Cloth-changing Person ReID

The quest for developing a ReID system that is simultaneously robust and discriminative has led to extensive research on ReID. The mainstream person ReID methods generally follow the paradigm that clothing is a stationary attribute [39], leading the statistic-based neural network to form an erroneous correlation between clothing and ID, known as clothing bias. Although clothing is helpful for traditional person ReID [46], it poses a significant obstacle in obtaining unbiased ID-intrinsic features for robust ReID.

To exclude the impact of clothing, many scholars [12, 29, 41] have attempted to use coercion by crude clothing masks to learn features beyond clothing. Others perform biometric learning by using shapes [3, 20], contour sketches [38], or gait [6, 17] to obtain cloth-agnostic representations. Specifically, Hong *et al*. [12] attempt to obtain the coarse ID mask with structure-related details, incorporating ID-relevant information for discriminative structural feature extraction. Jin *et al*. [17] use gait to capture biometric motions and concentrate on dynamic motion clues. Despite the remarkable progress these methods have made, they usually suffer from

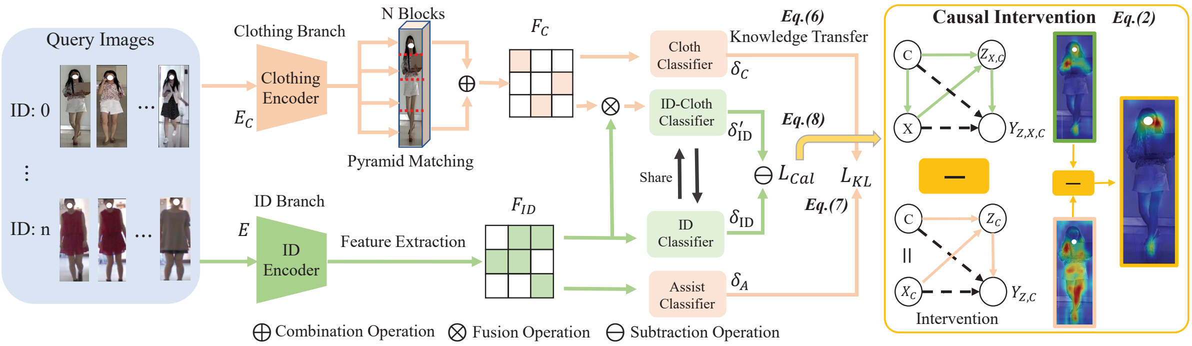


Figure 2. Structure of the proposed AIM. It consists of two branches, clothing branch (Orange) and ID branch (Green), which are training simultaneously. Such structure is to simulate causal intervention and distills clothing bias from the entangled representation automatically. The causal intervention is represented on the right. The X_C stands for the intervention on X, letting $X = c$. $Z_{X,C}$ and Z_C represent the feature of the entangled and the bias representation, respectively. $Y_{Z,X,C}$ and $Y_{Z,C}$ are the corresponding prediction.

estimation errors by overlooking the factual relationship between clothing and ID. In contrast, AIM utilizes causal auto-intervention to eliminate clothing bias and save the potential semantics.

2.2. Causality in Computer Vision

Causality helps to provide better learning and explainable models, within the broader context of computer vision, there is growing interest in causal discovery [32], causality distilling [14,50], incorporating causality within scene understanding downstream tasks [24, 27], stable learning [21, 43], disentanglement learning [36, 37], and debiasing [5, 22].

Causality models are designed to identify and analyze causal relationships among data, while conventional models focus on association. Theoretically, causal relationships are constructed on a deeper understanding of data, a higher dimensional abstraction of data relationships. Although randomized controlled experiments are a crucial criterion for establishing causal inference, controlling the specifics of feature extraction is challenging, making such experiments impractical. As a consequence, numerous approaches have surfaced to distill causality from the existing observational data. Recently, Tang *et al*. [31] alleviate context bias based on total direct effect (TDE) in causal inference for unbiased scene graph generation. CAL [27] leverages random attention to formulate counterfactual causality for visual categorization. CF-VQA [24] analyses the causal effect of questions on answers in visual question answering and mitigates language bias by subtracting the direct language effect from the total causal effect.

However, to our knowledge, no one has applied causality to CC-ReID. Enlightened by previous excellent works, we introduce causality into CC-ReID to separate the clothing bias and the ID-intrinsic clues. By utilizing causal auto-intervention, the clothing bias can be reasonably mitigated without compromising the integrity of the original semantics.

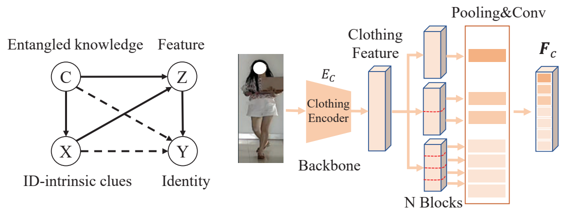
3. Proposed Method

3.1. CC-ReID form Causal View

The proposed AIM is trying to automatically eliminate clothing bias through causal intervention without destroying the semantic integrity within images. Fig. 2 shows the framework of AIM. Given a batch of person images, clothing knowledge is extracted with the assistance of pyramid matching strategy, while the ID branch is the ReID model for ID-intrinsic feature learning. With the knowledge transfer in training process, the clothing branch is gradually strengthened to perceive the clothing bias entangled with the ID representation. Subsequently, causal intervention is conducted by a simple subtraction operation between two ID classifiers to remove the effects of bias indirectly, which does not destroy semantic integrity. Finally, under the constraint of causality, the ID branch learns features unrelated to clothing as training continues, focusing the model's attention on ID-intrinsic features, without external cost.

Causal Analysis. Discovering causal relations from headless randomized controlled trials typically involves interventions [22] on data. Due to the severe entanglement of numerous influencing factors with the original data, it is either difficult or impossible to distill specific factors solely in the observational representation space. Causal theory sheds light on distilling specific effects without knowing all influencing factors, attributed to *do*-operation [7]. The $\mathrm{do}(X)$ denotes the *do*-operation on variable X, also known as making causal intervention upon X. Performing $\mathrm{do}(X)$ allows us to specify the exact value of X and isolate its cause.

Before further analysis, we first construct a CC-ReID based causal graph as the theoretical basis, as shown in Fig. 3(a). A causal graph is a directed acyclic graph showing how variables interact through causal paths, which provides a sketch of the causal relations behind the data. We begin with a brief description of the rationale behind the nodes and



(a) Causal graph for CC-ReID (b) Illustration of the pyramid matching

Figure 3. Causal graph for CC-ReID and the detailed illustration of the pyramid matching strategy.

paths to further elaborate on how the existing ReID models are misled by clothing bias.

Two nodes connected by a directed edge in a causal graph indicate that there are causalities between them, *e.g.*, ($\boldsymbol{X} \rightarrow \boldsymbol{Y}$) stands for $\boldsymbol{Y}$ is caused by $\boldsymbol{X}$. In the causal graph of CC-ReID, we denote $\boldsymbol{X}$ as the ID-intrinsic clues that only related to person; $\boldsymbol{Y}$ as the ID prediction 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 $\boldsymbol{X}$. 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 concomitant of clothing to people makes it closely entangled with intrinsic ID clues, which is the main concern of this paper.

$(\boldsymbol{C}, \boldsymbol{X}) \rightarrow \boldsymbol{Z} \rightarrow \boldsymbol{Y}$ shows the complete calculation process, which can be divided into two paths. The first is 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. $\boldsymbol{Z}$ appears as a mediator, which is inevitable for existing deep learning models. The second path is $\boldsymbol{C} \rightarrow \boldsymbol{X} \rightarrow \boldsymbol{Z} \rightarrow \boldsymbol{Y}$, where $\boldsymbol{C} \rightarrow \boldsymbol{X}$ is the entangled representation contributing to the faulty association between clothing and ID.

Additionally, the dashed arrow means there exists statistical dependence. As in $\boldsymbol{C} \rightarrow \boldsymbol{Y}$ and $\boldsymbol{X} \rightarrow \boldsymbol{Y}$, neither ID clues nor clothing bias of the person has a direct impact on the prediction results of the deep learning model.

Causal Intervention. Clothing bias stems from the entangled knowledge intertwined with ID-intrinsic clues, which can be challenging to discern and distinguish accurately. Causal intervention provides an opportunity to incorporate clothing knowledge into the model's prediction through backdoor adjustment without destroying semantic integrity:

$$P\left(Y|\mathrm{do}\left(X\right)\right) = \sum_{C} P\left(Y|X, C=c\right) P\left(C=c\right), \quad (1)$$

where do(X) is do-calculate, which is used for cutting the effect of $\boldsymbol{C} \rightarrow \boldsymbol{X}$. Specifically, through do(X), we can separate the effect of clothing bias and ID-intrinsic clues. However, $P(C=c)$ in the conventional backdoor adjustment needs to be aware of the specific impact of all clothing that appears, which remains difficult to implement.

Fortunately, TDE serves as a way to remove specific influences in causal inference [31], aligns with our goal, and provides guidance for the construction of AIM. For the model to separate the entangled representation and make predictions from the ID-intrinsic feature, an intuitive idea is to consistently eliminate the influence of clothing. To this end, given the observed outcome $Y_{x,c}(z)$ and the bias-specific prediction $Y_c(z)$, TDE can be formulated as:

$$\mathrm{TDE} = Y_{x,c}\left(z\right) - Y_c\left(z\right), \quad (2)$$

where the first term is from the regular prediction and the second is from the intervention by $P(Y|X=c)P(C=c)$.

Letting the clothing bias as the main effect, through a simple subtraction, we are able to remove the influence of the biased effect $\boldsymbol{C} \rightarrow \boldsymbol{Z}$ from the direct effect of $\boldsymbol{X} \rightarrow \boldsymbol{Z}$. Finally, the final prediction $\boldsymbol{Y}$ can be more robust by revealing the true relationship between unbiased feature $\boldsymbol{Z}$ and $\boldsymbol{X}$.

3.2. Model Construction

Following the TDE in causal theory, we build AIM, as shown in Fig. 2, which consists of two branches. The ID branch is to simulate $Y_{x,c}(z)$ to obtain the observed ID feature. The clothing branch is the realization of $Y_c(z)$. Further, the challenge lies in distilling clothing bias from the observation features. As preparation, we combine the samples' ID and the suit category as a separate clothing label for each suit, which denotes Y_C. The number of suits of all samples is summarised as N_C. Please note that this is a rough categorization strategy, and pedestrians are not necessarily sharing the same clothing label even if they wear similar clothes.

In the ID branch, the biased ID feature can be obtained through the ID encoder $\boldsymbol{F}_{\mathrm{ID}} = E(x^i)$ by minimizing the identification loss $\mathcal{L}_{\mathrm{ID}}$ as:

$$\mathcal{L}_{\mathrm{ID}} = -\sum_{i=1}^{N} y^i \log\left(\boldsymbol{p}_{\mathrm{ID}}\left(y^i|x^i\right)\right), \quad (3)$$

where $\boldsymbol{p}_{\mathrm{ID}}(y^i|x^i)$ is the probability of the i-th ID from the ID classifier δ_{ID} for image x^i.

In the clothing branch, two objectives are the basis of its design. Firstly, clothing information needs to be precisely extracted. Secondly, knowledge transfer learning from the ID branch is also required to understand how clothing bias affects the ID branch.

For the first purpose, we concentrate on the intrinsic top-middle-bottom characteristics of clothing (*i.e.*, shirts, pants, and shoes) and adopt pyramid matching strategy [8] to enhance the partial feature representation of clothing. Specifically, as shown in Fig. 3(b), pyramid features are extracted by dividing the deep features into different numbers of partial feature blocks. By incorporating feature blocks with diverse scales, the model can effectively capture both global and local information at varying spatial scales, resulting in more precise sensing of clothing details. The final clothing feature can be obtained from the clothing encoder with pyramid matching strategy as $\boldsymbol{F}_C = \mathrm{PM}(E_C(x^i))$.

Then, we adopt a cloth classifier δ_C trained by cloth classification loss $\mathcal{L}_C$ to leverage the ground-truth clothing labels while maintaining clothing information in feature space, which can be formulated as:

$$\mathcal{L}_C = -\sum_{i=1}^{N_C} y_C^i \log\left(\boldsymbol{p}_C\left(y_C^i|x^i\right)\right), \tag{4}$$

where $\boldsymbol{p}_C(y_C^i|x^i)$ is the probability of the i-th clothing, y_C^i is the corresponding clothing label.

For the second purpose, to enable $\boldsymbol{F}_C$ to perceive clothing bias in $\boldsymbol{F}_{\mathrm{ID}}$, we adopt Kullback Leibler (KL) Divergence as in mutual learning [44] to fit the distribution of clothing bias entangled with ID representation. To this end, we utilize an additional classifier δ_A trained by cloth classification loss in the ID branch. The clothing inference is distilled by δ_A and then transferred to the clothing branch for knowledge migration. The KL distance D_{KL} from $\boldsymbol{F}_C$ to $\boldsymbol{F}_{\mathrm{ID}}$ is computed as:

$$\hat{\boldsymbol{p}}_C = \exp\left(\delta_C\left(\boldsymbol{F}_C\right)\right), \quad \hat{\boldsymbol{p}}_{\mathrm{ID}} = \exp\left(\delta_A\left(\boldsymbol{F}_{\mathrm{ID}}\right)\right), \tag{5}$$

$$\mathcal{D}_{\mathrm{KL}}\left(\hat{\boldsymbol{p}}_C \| \hat{\boldsymbol{p}}_{\mathrm{ID}}\right) = \sum_{m=1}^{M} \hat{\boldsymbol{p}}_C^m \log \frac{\hat{\boldsymbol{p}}_C^m}{\hat{\boldsymbol{p}}_{\mathrm{ID}}^m}, \tag{6}$$

where M denotes the number of samples in a mini-batch. To be noticed, due to the asymmetry of Kullback–Leibler (KL) Divergence, we compute $D_{\mathrm{KL}}(\hat{\boldsymbol{p}}_{\mathrm{ID}} \| \hat{\boldsymbol{p}}_C)$ as well. The total KL Divergence can be formulated as:

$$\mathcal{L}_{\mathrm{KL}} = \mathcal{D}_{\mathrm{KL}}\left(\hat{\boldsymbol{p}}_C \| \hat{\boldsymbol{p}}_{\mathrm{ID}}\right) + \mathcal{D}_{\mathrm{KL}}\left(\hat{\boldsymbol{p}}_{\mathrm{ID}} \| \hat{\boldsymbol{p}}_C\right). \tag{7}$$

With the above preparations, to transfer TDE from theory to reality, (2) can be transferred into a causality loss:

$$\mathcal{L}_{\mathrm{CAL}} = -\sum_{i=1}^{N} y^i \log\left(\delta_{\mathrm{ID}}\left(\boldsymbol{F}_{\mathrm{ID}}\right) - \delta_{\mathrm{ID}}^{'}\left(\boldsymbol{F}_C \otimes \boldsymbol{F}_{\mathrm{ID}}\right)\right), \tag{8}$$

where $\delta_{\mathrm{ID}}^{'}$ is the ID-Cloth classifier sharing weight with δ_{ID}. $\boldsymbol{F}_C \otimes \boldsymbol{F}_{\mathrm{ID}}$ stands for the bilinear pooling to fuse both features. As bias and ID features are entangled, the model pays similar attention to both features, while clothing branch focuses only on the distillation of bias. Fusing both features can enhance the expression of same-located bias and suppress the attention on ID to achieve accurate intervention. It also promotes separating both features, whose effects on feature variation can be seen in ablation studies in Sec. 4.5. We perform $P(Y|X = c)P(C = c)$ as making clothing bias the source of ID prediction to remove it in the latent space without destroying semantic integrity. The right of Fig. 2 illustrates the intervention in graph view and the variation of the feature heatmap at each step. The green and orange arrows denote the normal and the biased routes, respectively. In summary, the clothing bias learned through model inference and knowledge transfer is automatically eliminated by causal constraints through training, resulting in unbiased ID representations for robust ReID.

3.3. Objective Function

In clothing branch, clothing representation is trained by minimizing $\mathcal{L}_C$. The ReID model of ID branch is trained by minimizing $\mathcal{L}_{\mathrm{ID}}$. $\mathcal{L}_{\mathrm{KL}}$ is minimized to perceive clothing bias in ReID training. The causal intervention of eliminating clothing bias is produced by minimizing $\mathcal{L}_{\mathrm{CAL}}$ The total objective function is a weighted sum of all the above losses:

$$\mathcal{L}_{\mathrm{Total}} = \mathcal{L}_C + \mathcal{L}_{\mathrm{ID}} + \lambda_{\mathrm{CAL}}\mathcal{L}_{\mathrm{CAL}} + \lambda_{\mathrm{KL}}\mathcal{L}_{\mathrm{KL}}, \tag{9}$$

where λ_{CAL} is the weight of causal auto-intervention loss to strengthen the effect of bias elimination, and λ_{KL} is used as the weight to control the transfer intensity of clothing knowledge. A detailed analysis of the hyper-parameters selection of λ can be referred to Sec. 4.3. The weights of other loss terms are basic terms [9] and set as 1. We fix the above weights during the training process in all experiments. Both branches are trained simultaneously, while only ID branch is employed for testing.

4. Experimental Results

4.1. Datasets and Metrics

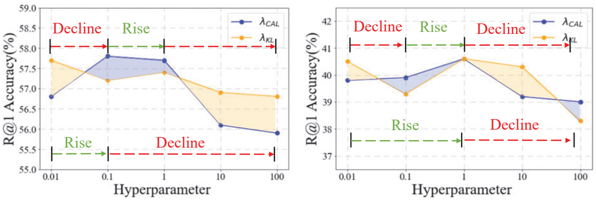
We evaluate AIM on two standard cloth-changing datasets, PRCC-ReID [38] and LTCC-ReID [26]. PRCC-ReID [38] consists of 221 IDs with three camera views, including 33,698 images. Each person wears the same clothes in camera A and camera B, and different clothes in camera A and camera C. LTCC-ReID [26] is a long-term person ReID dataset with frequent changes in clothing and multiple environmental changes. It is captured indoors with 12 camera views, containing 152 IDs and 478 outfits with 17,119 labeled images. Additionally, we follow the previous studies [9, 12] and leverage rank-K (R@K) and mean average precision (mAP) for evaluation.

4.2. Implementation Details

We conduct experiments on both cloth-changing and standard settings. The default settings of datasets [26, 38] with multi-shot matching strategy are employed for training and evaluation. In standard setting, images with the same ID and camera view in the testing set are discarded when evaluating. In contrast, in cloth-changing setting, images with the same ID, camera view, and clothing are discarded during testing to evaluate the model performance on unseen clothing.

Table 1. Comparison of R@K (%) and mAP (%) performance with the state-of-the-arts on PRCC-ReID and LTCC-ReID dataset. "†" denotes the methods that are designed for CC-ReID. "‡" indicates the reproduced results. "*" represent the single-shot result. **Bold** and <u>underline</u> numbers are the best and second-best results, same as the following. Type "RGB" means only RGB modality is utilized.

Method	Venue	Size	Type		PRCC-ReID				LTCC-ReID			
					Standard		Cloth-Changing		Standard		Cloth-Changing	
			RGB	Hybrid	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP
PCB [30]	ECCV 18	384×192	●	-	<u>99.8</u>	97.0	41.8	38.7	65.1	30.6	23.5	10.0
OSNet [48]	ICCV 19	384×192	●	-	-	-	-	-	67.9	32.1	23.9	10.8
HPM [8]	AAAI 19	384×128	●	-	99.4	96.9	40.4	37.2	-	-	-	-
IANet [13]	CVPR 19	384×192	●	-	99.4	98.3	46.3	46.9	63.7	31.0	25.0	12.6
ISP [49]	ECCV 20	256×128	●	-	92.8	-	36.6	-	66.3	29.6	27.8	11.9
3DSL [3] †	CVPR 21	256×128	-	+Shape	-	-	51.3	-	-	-	31.2	14.8
FSAM [12] †	CVPR 21	256×128	-	+Mask	98.8	-	54.5*	-	73.2	35.4	<u>38.5</u>	16.2
GI-ReID [17] †	CVPR 22	256×128	-	+Gait	86.0	-	33.3	-	63.2	29.4	23.7	10.4
UCAD [35] †	IJCAI 22	384×192	-	+Mask	96.5	-	45.3	-	74.4	34.8	32.5	15.1
ViT-VIBE [1] †	WACV 22	-	-	+Shape	99.7	-	47.0	-	71.4	35.8	-	-
IRANet [28] †	IVC 22	384×128	-	+Pose	99.7	97.8	<u>54.9</u>	53.0	-	-	-	-
AFD-Net [34] †	IJCAI 21	256×128	●	-	95.7	-	42.8	-	-	-	-	-
RCSANet [15] †	ICCV 21	336×336	●	-	99.6	96.6	48.6	50.2	-	-	-	-
CAL [9] (Baseline) ‡	CVPR 22	256×128	●	-	**100.0**	99.7	53.4	53.1	74.4	39.0	34.4	16.0
CAL [9] (Baseline) ‡	CVPR 22	384×192	●	-	**100.0**	<u>99.8</u>	54.4	54.4	73.4	<u>39.4</u>	38.0	<u>17.2</u>
AIM (Ours)		256×128	●	-	**100.0**	**99.9**	54.7	<u>55.0</u>	<u>76.1</u>	39.1	38.3	17.0
AIM (Ours)		384×192	●	-	**100.0**	**99.9**	**57.9**	**58.3**	**76.3**	**41.1**	**40.6**	**19.1**



(a) Performance on PRCC-ReID (b) Performance on LTCC-ReID

Figure 4. R@1 accuracy for different hyper-parameters on two datasets. Different colors indicate different loss terms.

Following CAL [9], ResNet50 [11] pre-trained on ImageNet is involved as the encoder of AIM, where the last two pooling layer and fully connected layer are removed. Following the top-middle-bottom characteristic of clothing, we adopt the maximum split number of $N = 4$ in the pyramid matching strategy. Following [26], in the training phase, the input images are regularized as 384×192 with random horizontal flipping, random cropping, and random erasing [47]. We adopt Adam [18] optimizer for the training of both branches. The learning rate is initialized to $3.5e{-4}$ and divided by 10 after every 20 epochs, totaling 80 epochs. $\mathcal{L}_{\text{CAL}}$ is used for training after the 25th epoch.

4.3. Parameter Analysis

The hyper-parameters of AIM including λ_{CAL} and λ_{KL} in (9). These two parameters control the intensity of causal auto-intervention and knowledge transferring. Inspired by the Gibbs sampling strategy, we first fix the initial value of λ_{CAL} to 1 and determine λ_{KL} by iterative experiments. Then, the λ_{KL} is fixed and reversely determines the value of the λ_{CAL} by the same iterative experiments. The results are presented in Fig. 4. The results on two datasets demonstrate that when λ_{CAL} is fixed, an increase of λ_{KL} does not improve accuracy. When λ_{KL} is fixed, an increase of λ_{CAL} leads to an increase and then a decrease in accuracy. Finally, we leverage the value of (0.1, 0.01) and (1, 1) for (λ_{CAL}, λ_{KL}) on PRCC-ReID and LTCC-ReID, respectively. The results show that λ_{CAL}, as a proxy for the ability of bias elimination, prefers a medium value, while λ_{KL} depends on different data environments and is therefore chosen accordingly.

4.4. Comparison with State-of-the-art Methods

We conduct comparative experiments with conventional short-term ReID and novel CC-ReID state-of-the-art methods on two standard CC-ReID datasets. Since CC-ReID is in its infancy, most methods are hybrids, consisting of multiple modalities, *e.g.*, sketch, keypoint, pose, mask, and gait. In contrast, only a few methods, including AIM, make their efforts to explore discriminative features from pure RGB modality. As shown in Table 1, the results of AIM show a significant improvement over hybrid and RGB state-of-the-art methods by causal auto-intervention.

Specifically, we compare AIM with 5 conventional methods in short-term ReID and 8 methods explicitly designed for CC-ReID. For PRCC-ReID, in the standard setting, AIM achieves the same metrics as the baseline [9] method and maintains a high level; In the cloth-changing setting, AIM surpasses the baseline with 3.5%/3.9% of R@1/mAP, and it also outperforms the second-best method IRANet [28], which leverages keypoint as external knowledge, by 3.0%/5.3% of R@1/mAP. The result demonstrates that AIM works well compared to existing CC-ReID meth-

Table 2. Ablation studies of each component of AIM in cloth-changing setting on PRCC-ReID and LTCC-ReID.

Basic		Causality		PRCC-ReID		LTCC-ReID	
B/L	Debias	⊗	KL	R@1	mAP	R@1	mAP
●	○	○	○	54.4	54.4	38.0	17.2
●	●	○	○	55.1	56.2	39.3	17.9
●	●	●	○	56.1	56.4	38.0	18.2
●	●	○	●	55.2	56.4	40.1	18.0
●	●	●	●	**57.9**	**58.3**	**40.6**	**19.1**

ods with unbiased ID-intrinsic feature learning. For LTCC-ReID, AIM improves by 2.9%/2.6% of R@1 and 1.7%/1.9% of mAP to the baseline method in two settings. The second-best method on LTCC-ReID is FSAM [12], which leverages fine-grained mask and multi-branch learning strategy for appearance and structure features. Compared to it, we attain 3.1%/2.1% of R@1 and 5.7%/2.9% of mAP in two settings. The results show that even under complex scenarios (resolution, illumination, viewpoint, *et al.*), AIM can still capture the existing discriminative features rather than being misled by the noticeable clothing bias. We also compare the conventional ReID methods (rows 1-5) and find that AIM significantly outperforms them. This is because capturing ID-intrinsic information becomes challenging under complex conditions, which increases the probability of the entangled representation misleading the conventional model.

4.5. Ablation Studies

Component Analysis. To verify the validity of the AIM model and each component, we perform ablation studies in Table 2. Specifically, there are two basic components, "B/L" is for baseline, and "Debias" is for the pure dual-branched causal intervention structure without other components. "⊗" stands for the fusion operation in (8). "KL" denotes the knowledge transfer process, specified as the KL Divergence loss in (7). The solid black dot indicates that the corresponding module is being used for training. By comparing rows 1-2, we observe that the causal auto-intervention has a certain effect but is still unsatisfactory. By comparing rows 2 with 3-4, the results indicate the effectiveness of feature fusion and clothing knowledge migration. Each component enhances the ability to distill clothing bias from the entangled representation. The last row demonstrates that AIM with all components gives the best performance. These results indicate that by auto-intervention and clothing bias distillation, the unbiased ID-intrinsic representation is exploited for better identification, and each component contributes equally to the final result.

Influence of Pyramid Matching. We evaluate the effect of adding pyramid matching (PM) strategy to the backbone of each branch. Table 3 shows the performance changes on two datasets. The results show that either adding PM to the

Table 3. Ablation studies of the backbone of each branch of AIM in cloth-changing setting on PRCC-ReID and LTCC-ReID.

Baseline		Debias		PRCC-ReID		LTCC-ReID	
Res	PM	Res	PM	R@1	mAP	R@1	mAP
●	-	-	-	54.4	54.4	38.0	17.2
●	○	●	○	55.8	56.9	39.3	18.1
○	●	○	●	57.1	57.6	37.0	16.8
○	●	●	○	56.2	56.6	35.7	16.7
●	○	○	●	**57.9**	**58.3**	**40.6**	**19.1**

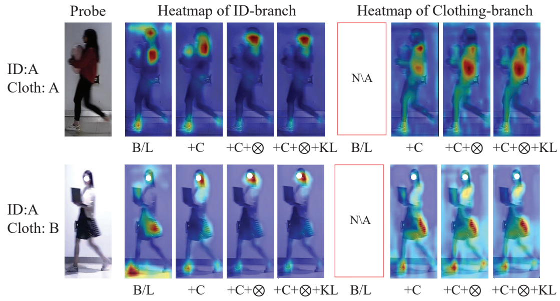


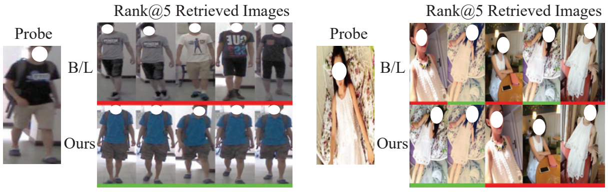
Figure 5. Visualization of the ID feature and the clothing feature.

ID branch (rows 3 & 4) or the clothing branch (rows 3 & 5) can improve the ReID accuracy. While, the combination of ResNet (R) + PM, which only adds PM for the clothing branch, gives the best results. A possible reason is that the model with PM has better clothing sensitivity through the top-middle-bottom structures, which increases the partial representation. For ID branch, although more detailed information is extracted, it somehow exacerbates the clothing bias, which violates the purpose to optimize ID-intrinsic representation for robust ReID.

4.6. Qualitative Results

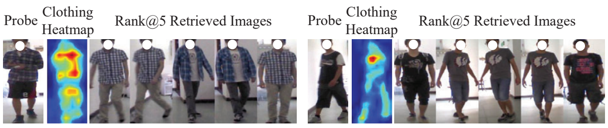
Analysis of Robust Representation. Fig. 5 shows how the heatmap of the ID and the clothing change as each component is added. The details of notations can be referred to Sec. 4.5. More components indicate more accurate clothing bias distillation and more robust ID-intrinsic representation learning. The comparison from left to right clearly shows the variation of feature attention areas. For the same person with cloth changing, the ID features of the baseline [9] method are misled by clothing bias, which has faulty attention on the wearings. While AIM successfully mitigates the impact of clothing bias, redirects the model's attention to the intrinsic areas (*e.g.*, head) and weakens for regions of the body that are not associated with the ID (*e.g.*, bag, shoes, and shoulder). Through continuous refinement of the clothing bias distillation, the clothing features become more concentrated on areas that affect the learning of ID-intrinsic information.

Clothing Bias Elimination. As mentioned, the entangled representation of clothing and ID-intrinsic clues is an obstacle to the existing CC-ReID methods. The solution to this



(a) Retrieval of dataset images (b) Retrieval of real world same-cloth images

Figure 6. Retrieval results of dataset images and real-world images. The green bar at the bottom represents the correct matching, while the red bar is the opposite.



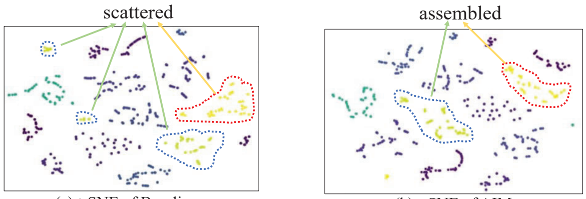
(a) Similar appearance (b) Similar pattern

Figure 7. Visualization of the clothing heatmap and top-5 retrieval results by feature of clothing branch in AIM.

problem emphasizes eliminating the influence of clothing and focusing on the intrinsic properties of people. To notice the difference between the baseline [9] and ours and further demonstrate the ability of AIM to eliminate clothing bias, Fig. 6 illustrates the results of the retrieval ranks at the laboratory and real-world level. Five candidate gallery images with the highest similarity to the probe image are displayed. The results in Fig. 6(a) demonstrate that similar clothing patterns or colors mislead the baseline method, while AIM captures ID information and yields the correct retrieval results, despite the dramatic shift in dressing style. As there are no samples for different people wearing the same cloth, we collect samples in DeepFashion [23] for evaluation, where all people wear the same kind of white dress. Fig. 6(b) demonstrates that AIM performs well even with interference of identical clothing. The above results reveal that the proposed AIM is more robust to clothing bias and accurately captures the correct candidates.

Clothing Knowledge Distillation. Additionally, to verify that the clothing branch indeed distills the knowledge of clothing and learns the representation of clothing bias, we also visualize the feature of the clothing branch and use such features as probes to produce retrieval in the database. Note that all clothing in the gallery is not identical to the probe. The heatmap of clothing and the top 5 images with the highest similarity scores are displayed. Fig. 7(a) shows the similarity of the striped plaid shirt, while Fig. 7(b) shows the similarity of the clothing with similar color style and logo on the chest. The results demonstrate that the clothing branch of AIM captures the clothing knowledge as expected.

Features Distribution Analysis. Moreover, to evaluate the methodology of AIM from another dimension, Fig. 8 visualizes the feature distributions of 10 randomly selected



(a) t-SNE of Baseline (b) t-SNE of AIM

Figure 8. Visualization of 10 classes randomly selected from PRCC-ReID. Our AIM separates different classes while effectively reducing the inter-class variance.

categories on PRCC-ReID, which compares the feature distributions in the latent space of the baseline [9] and AIM. The results show that AIM can reduce the intra-class distance and make the feature distribution more aggregated (as shown in the red dashed circle). The blue dashed circles in Fig. 8(a) enclose three groups of yellow points with very scattered distances, which are better clustered together in Fig. 8(b).

5. Conclusion

Clothing is an essential aspect of human identity and is crucial in general person ReID. However, it also poses a significant challenge in Cloth-Changing person ReID (CC-ReID) due to its confounding nature. In this paper, we analyze the effect of clothing on model prediction and adopt causal intervention to eliminate this effect automatically. Specifically, a causality-based Auto-Intervention Model (AIM) is first proposed to mitigate clothing bias for robust CC-ReID. A dual-branch structure of clothing and ID is utilized to simulate the causal intervention process and penalized by a causality loss. AIM is eventually encouraged to learn ID-intrinsic clues free from clothing bias. Furthermore, experimental results on two CC-ReID datasets demonstrate the effectiveness and the ability of clothing bias distillation of AIM, which achieves state-of-the-art performance.

Although causal theory provides sufficient guidance for AIM, limitations remain. In the entangled representation, clothing is merely one of several intuitive factors, and many other confounding factors are worth investigating. The design of the clothing branch requires the perception of clothing as a prerequisite, which is limited by clothing annotations and sample numbers. Besides, as bias is widely present in ReID tasks, a stronger backbone might lead to greater boosts.

Acknowledgments

This work was supported by National Key R&D Project (2021YFC3320301), National Natural Science Foundation of China (62171325, 62271361), Hubei Key R&D (2022BAA033) and CAAI-Huawei MindSpore Open Fund. The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

References

[1] Vaibhav Bansal, Gian Luca Foresti, and Niki Martinel. Cloth-changing person re-identification with self-attention. In *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops*, pages 602–610, 2022. 6

[2] Igor Barros Barbosa, Marco Cristani, Alessio Del Bue, Loris Bazzani, and Vittorio Murino. Re-identification with RGB-D sensors. In *Proc. Eur. Conf. Comput. Vis. Workshops*, pages 433–442, 2012. 1

[3] Jiaxing Chen, Xinyang Jiang, Fudong Wang, Jun Zhang, Feng Zheng, Xing Sun, and Wei-Shi Zheng. Learning 3d shape feature for texture-insensitive person re-identification. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 8146–8155, 2021. 2, 6

[4] Xiumei Chen, Xiangtao Zheng, and Xiaoqiang Lu. Bidirectional interaction network for person re-identification. *IEEE Trans. Image Process.*, 30:1935–1948, 2021. 1

[5] Saloni Dash, Vineeth N. Balasubramanian, and Amit Sharma. Evaluating and mitigating bias in image classifiers: A causal perspective using counterfactuals. In *Proc. IEEE/CVF Winter Conference on Appl. Comput. Vis.*, pages 3879–3888, 2022. 3

[6] Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He. GaitPart: temporal part-based model for gait recognition. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 14213–14221, 2020. 2

[7] Norman E. Fenton, Martin Neil, and Anthony C. Constantinou. The book of why: The new science of cause and effect, judea pearl, dana mackenzie. basic books (2018). *Artif. Intell.*, 284:103286, 2020. 2, 3

[8] Yang Fu, Yunchao Wei, Yuqian Zhou, Honghui Shi, Gao Huang, Xinchao Wang, Zhiqiang Yao, and Thomas S. Huang. Horizontal pyramid matching for person re-identification. In *Proc. AAAI Conf. Artif. Intell.*, pages 8295–8302, 2019. 2, 4, 6

[9] 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. 5, 6, 7, 8

[10] Ruocheng Guo, Lu Cheng, Jundong Li, P. Richard Hahn, and Huan Liu. A survey of learning causality with data: Problems and methods. *ACM Comput. Surv.*, 53(4):75:1–75:37, 2020. 2

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 770–778, 2016. 6

[12] 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, 5, 6, 7

[13] Ruibing Hou, Bingpeng Ma, Hong Chang, Xinqian Gu, Shiguang Shan, and Xilin Chen. Interaction-and-aggregation network for person re-identification. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 9317–9326, 2019. 6

[14] Xinting Hu, Kaihua Tang, Chunyan Miao, Xian-Sheng Hua, and Hanwang Zhang. Distilling causal effect of data in class-incremental learning. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 3957–3966, 2021. 3

[15] Yan Huang, Qiang Wu, Jingsong Xu, Yi Zhong, and Zhaoxiang Zhang. Clothing status awareness for long-term person re-identification. In *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pages 11875–11884, 2021. 6

[16] Xuemei Jia, Xian Zhong, Mang Ye, Wenxuan Liu, and Wenxin Huang. Complementary data augmentation for cloth-changing person re-identification. *IEEE Trans. Image Process.*, 2022. 2

[17] Xin Jin, Tianyu He, Kecheng Zheng, Zhiheng Yin, Xu Shen, Zhen Huang, Ruoyu Feng, Jianqiang Huang, Zhibo Chen, and Xian-Sheng Hua. Cloth-changing person re-identification from A single image with gait prediction and regularization. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 14258–14267, 2022. 2, 6

[18] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. Int. Conf. Learn. Represent.*, pages 1–15, 2015. 6

[19] Kun Kuang, Lian Li, Zhi Geng, Lei Xu, Kun Zhang, Beishui Liao, Huaxin Huang, Peng Ding, Wang Miao, and Zhichao Jiang. Causal inference. *Eng.*, 6(3):253–263, 2020. 2

[20] Yu-Jhe Li, Zhengyi Luo, Xinshuo Weng, and Kris M. Kitani. Learning shape representations for clothing variations in person re-identification. *arXiv:2003.07340*, 2020. 2

[21] Jiashuo Liu, Zheyan Shen, Peng Cui, Linjun Zhou, Kun Kuang, Bo Li, and Yishi Lin. Stable adversarial learning under distributional shifts. In *Proc. AAAI Conf. Artif. Intell.*, pages 8662–8670, 2021. 3

[22] Ruyang Liu, Hao Liu, Ge Li, Haodi Hou, Tinghao Yu, and Tao Yang. Contextual debiasing for visual recognition with causal mechanisms. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 12745–12755, 2022. 3

[23] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. DeepFashion: powering robust clothes recognition and retrieval with rich annotations. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 1096–1104, 2016. 8

[24] Yulei Niu, Kaihua Tang, Hanwang Zhang, Zhiwu Lu, Xian-Sheng Hua, and Ji-Rong Wen. Counterfactual VQA: A cause-effect look at language bias. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 12700–12710, 2021. 3

[25] Judea Pearl. Direct and indirect effects. *arXiv:1301.2300*, 2013. 2

[26] 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. 2, 5, 6

[27] Yongming Rao, Guangyi Chen, Jiwen Lu, and Jie Zhou. Counterfactual attention learning for fine-grained visual categorization and re-identification. In *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pages 1005–1014, 2021. 3

[28] Wei Shi, Hong Liu, and Mengyuan Liu. IRANet: identity-relevance aware representation for cloth-changing person re-identification. *Image Vis. Comput.*, 117:104335, 2022. 6

[29] Xiujun Shu, Ge Li, Xiao Wang, Weijian Ruan, and Qi Tian. Semantic-guided pixel sampling for cloth-changing person re-identification. *IEEE Signal Process. Lett.*, 28:1365–1369, 2021. 2

[30] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and A strong convolutional baseline). In *Proc. Eur. Conf. Comput. Vis.*, pages 501–518, 2018. 6

[31] Kaihua Tang, Yulei Niu, Jianqiang Huang, Jiaxin Shi, and Hanwang Zhang. Unbiased scene graph generation from biased training. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 3713–3722, 2020. 3, 4

[32] Lan Wang and Vishnu Naresh Boddeti. Do learned representations respect causal relationships? In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 264–274, 2022. 3

[33] Xiao Wang, Jun Chen, Zheng Wang, Wu Liu, Shin'ichi Satoh, Chao Liang, and Chia-Wen Lin. When pedestrian detection meets nighttime surveillance: A new benchmark. In *Proc. IJCAI Int. Joint Conf. Artif. Intell.*, pages 509–515, 2020. 2

[34] Wanlu Xu, Hong Liu, Wei Shi, Ziling Miao, Zhisheng Lu, and Feihu Chen. Adversarial feature disentanglement for long-term person re-identification. In *Proc. IJCAI Int. Joint Conf. Artif. Intell.*, pages 1201–1207, 2021. 6

[35] Yuming Yan, Huimin Yu, Shuzhao Li, Zhaohui Lu, Jianfeng He, Haozhuo Zhang, and Runfa Wang. Weakening the influence of clothing: Universal clothing attribute disentanglement for person re-identification. In *Proc. IJCAI Int. Joint Conf. Artif. Intell.*, pages 1523–1529, 2022. 6

[36] Kaiwen Yang and Xinmei Tian. Domain-class correlation decomposition for generalizable person re-identification. *IEEE Trans. Multimedia*, 2022. 3

[37] Mengyue Yang, Furui Liu, Zhitang Chen, Xinwei Shen, Jianye Hao, and Jun Wang. CausalVAE: disentangled representation learning via neural structural causal models. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 9593–9602, 2021. 3

[38] 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. 2, 5

[39] Zhengwei Yang, Xian Zhong, Hong Liu, Zhun Zhong, and Zheng Wang. Attentive decoupling network for cloth-changing re-identification. In *Proc. IEEE Int. Conf. Multimedia Expo*, pages 1–6, 2022. 2

[40] Zhengwei Yang, Xian Zhong, Zhun Zhong, Hong Liu, Zheng Wang, and Shin'ichi Satoh. Win-win by competition: Auxiliary-free cloth-changing person re-identification. *IEEE Trans. Image Process.*, 2023. 1

[41] Shijie Yu, Shihua Li, Dapeng Chen, Rui Zhao, Junjie Yan, and Yu Qiao. COCAS: A large-scale clothes changing person dataset for re-identification. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 3397–3406, 2020. 2

[42] Jianfu Zhang, Li Niu, and Liqing Zhang. Person re-identification with reinforced attribute attention selection. *IEEE Trans. Image Process.*, 30:603–616, 2021. 1

[43] Xingxuan Zhang, Peng Cui, Renzhe Xu, Linjun Zhou, Yue He, and Zheyan Shen. Deep stable learning for out-of-distribution generalization. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 5372–5382, 2021. 3

[44] Ying Zhang, Tao Xiang, Timothy M. Hospedales, and Huchuan Lu. Deep mutual learning. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 4320–4328, 2018. 5

[45] 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. 1

[46] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 2138–2147, 2019. 2

[47] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proc. AAAI Conf. Artif. Intell.*, pages 13001–13008, 2020. 6

[48] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. In *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pages 3701–3711, 2019. 6

[49] 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. 6

[50] Shengyu Zhu, Ignavier Ng, and Zhitang Chen. Causal discovery with reinforcement learning. In *Proc. Int. Conf. Learn. Represent.*, pages 1–16, 2020. 3