Coarse Attribute Prediction with Task Agnostic Distillation for Real World Clothes Changing ReID

Priyank Pathak
priyank@ucf.edu
Yogesh S Rawat
yogesh@ucf.edu

Center for Research in Computer Vision
University of Central Florida
Orlando, FL, USA

Abstract

This work focuses on Clothes Changing Re-IDentification (CC-ReID) for the real world. Existing works perform well with high-quality (HQ) images, but struggle with low-quality (LQ) where we can have artifacts like pixelation, out-of-focus blur, and motion blur. These artifacts introduce noise to not only external biometric attributes (e.g. pose, body shape, etc.) but also corrupt the model's internal feature representation. Models usually cluster LQ image features together, making it difficult to distinguish between them, leading to incorrect matches. We propose a novel framework **Robustness against Low-Quality (RLQ)** to improve CC-ReID model on real-world data. RLQ relies on Coarse Attributes Prediction (CAP) and Task Agnostic Distillation (TAD) operating in alternate steps in a novel training mechanism. CAP enriches the model with *external* fine-grained attributes via coarse *predictions*, thereby reducing the effect of noisy input. On the other hand, TAD enhances the model's *internal* feature representation by bridging the gap between HQ and LQ features, via an external dataset through task-agnostic self-supervision and distillation. Our RLQ is among the few works to successfully learn from external attributes on very low-quality on real-world datasets like LaST, and DeepChange. *Github:* `https://github.com/ppriyank/RLQ-CGAL-UBD`.

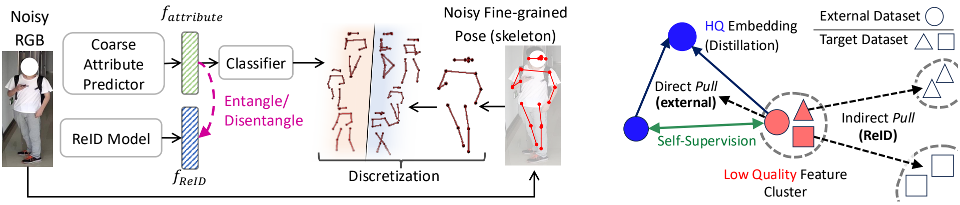
1 Introduction

Person Re-IDentification (ReID) involves identifying a person from a gallery of media, categorized as Short-Term and Long-Term ('Clothes Changing', 'CC') ReID. The former identifies people with similar backgrounds/clothes while the latter identifies across locations, time, and clothing. Short-term ReID is relatively easier, with non-biometric cues (background and timestamps) helping in identification [44]. We focus on image-based CC-ReID counteracting real-world artifacts, without any non-biometric shortcuts. Our work primarily focuses on **reducing the effect of low quality artifacts in real-world person ReID.**

Existing approaches often rely on external fine-grained attributes (*e.g*. silhouettes [21], body shapes [46], contours [57], *etc*.) as auxiliary inputs. These attributes, produced by high-quality (HQ) pre-trained models, have not seen much success on low-quality (LQ) datasets. When applied to real-world LQ images, which include *occlusions*, *pixelation*, *motion blur*,

(a) **Motion Blur** Wrong 3D Body & Pose (SMPL [31]) (b) **Motion Blur** Partial Contours (DexiNed [41]) (c) **Pixelation** Noisy Silhouette (SCHP [25]) (d) **Out-of-focus** Blurry Face (RetineNet [8]) (e) **Pixelation**: Partial Pose (AlphaPose [7])

Figure 1: **Noisy fine-grained attributes**: Samples from LaST, LTCC, DeepChange, PRCC.



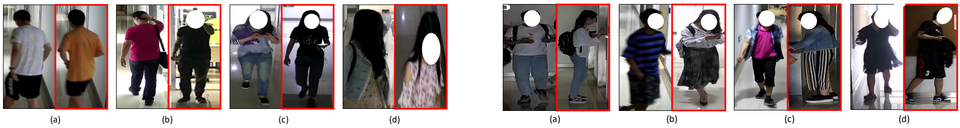
(a) **Coarse Attribute Prediction (CAP)** (b) **Task Agnostic Distillation (TAD)**

Figure 2: **Robustness against Low-Quality (RLQ)** (a) Fine-grained attribute (*e.g.* pose) is discretized/clustered into coarse classes which the model predicts using RGB images for learning attribute feature representation. Classification is an easy task that avoids input noise from attributes. (b) External (◯) HQ embeddings pull LQ ones towards it via self-supervision & distillation, dissolving the LQ cluster on the target dataset (△,□), especially helpful when the target dataset doesn't have HQ-LQ pairs.

and *out-of-focus (OOF)* blur, these attributes tend to be noisy (fig. 1). Artifacts not only negatively impact auxiliary models but also ReID models themselves. This results in a well-studied low-resolution issue, where models cluster LQ images regardless of identities, leading to indistinguishable erroneous matches (shown in *Ablation*). To address these issues we propose **Robustness against Low-Quality (RLQ)** framework which improves CC-ReID on real-world data. RLQ *alternates* between alleviating input noise from *external* fine-grained attributes via **Coarse Attributes Prediction (CAP)**, and reducing the negative impacts of LQ images on the model's *internal* features representation via **Task Agnostic Distillation (TAD)**. These two strategies when applied to our ReID model, enrich the model with fine-grained attributes (CAP) and distinguish LQ images (TAD).

CAP is proposed as an alternative to feeding additional input noise, via fine-grained attributes, while still learning their representation. Instead, features for these *external attributes* is learned by predicting attributes as coarse discrete labels, directly from input (fig. 2(a)) *i.e.* RGB-only inference. To show effectiveness of CAP, we focus on two attributes: one showing *entanglement* via gender awareness and other *disentanglement* via pose robustness. We classify images into pose clusters, and gender classes. Generative method [27] can learn attributes via predictions as well (*e.g.* generate poses) for learning pose, however, classification is an easier task than generating real-world data.

In addition to external attributes, ReID models themselves need robustness against real-world artifacts. Existing research [26, 34] uses pairs of high-quality (HQ) and low-quality (LQ) images to alleviate the clustering of LQ features. However, separating HQ images from those with artifacts *e.g.* pixelation in an upscaled image without HQ-LQ pairs is a challenging task. Our novel 'TAD' (fig. 2(b)) bridges the domain gap between the LQ and HQ image



(a) **Top-1 mismatch (shown in red) with similar Pose** (b) **Top-1 opposite gender mismatch (shown in red)**

Figure 3: **ReID against Pose and Gender:** LTCC mis-matches for CAL [11].

feature, using distillation and self-supervision, via synthetic HQ-LQ image pairs generated from an external dataset. This pulls synthetic LQ features towards their HQ counterpart, implicitly helping dissolve the LQ feature clusters on our target dataset, thus helping the model distinguish LQ images better.

In summary, (1) Our novel Robustness against Low-Quality (RLQ) framework for real-world CC-ReID alternatively applies Coarse Attribute Prediction (CAP) and Task Agnostic Distillation (TAD) on the base ReID model for robustness against low-quality artifacts. (2) CAP entangles gender and disentangles pose, via *external* fine-grained attribute without the additional input noise from these attributes. (3) TAD enhances the model's *internal* feature representation of low-quality images, via an external high-quality dataset reducing LQ image feature clusters. (4) Our RLQ is among the few works to successfully learn from external attributes on very low-quality on real-world datasets like LaST, and DeepChange. The extensive ablation analyses the performance benefits of CAP and TAD.

2 Related Work

Clothes Changing Person (CC) ReID Person ReID with same cloth [9, 38, 53] is a well-researched topic. Following the surge of CC-ReID [5, 39, 40], we primarily focus on image-based CC-ReID. Our design adopts a multi-branch design [13, 43, 57] where one branch is dedicated to clothing changes, while the other focuses on biometric features.

Attributes: Multi-modal methods entangle fine-grained attributes (silhouettes [21, 60], body shape [46], body parts [49, 55], contour [2], skeleton [13, 39], face [1, 47], and 3D shape [51]) *as input* via an external model to learn biometrics. These attributes are essential in solving ReID, *e.g.* pose robustness, but can also introduce additional input noise for real-world low-quality data. Generative models [22, 27, 52, 59] learn representations by generating attributes (*e.g.* generate pose to learn), similarly, our approach predicts attributes, for learning representations. Classifying coarse-grained attributes is an easier task than generating real-world data & fine-grained attributes, while still reducing the additional input noise.

Low Resolution: Prior research [42, 45] primarily addresses 'pixelation' in face recognition. However, the real world includes additional artifacts like blurring, which renders super-resolution [3, 58] and traditional low resolution ineffective, with unreliable HQ-LQ image pairs. Our solution is to create synthetic low-quality images [10, 19] with real-world artifacts using an external HQ dataset, bridging the gap between the two domains.

3 Method : Robustness against Low-Quality (RLQ)

Robustness against Low-Quality (RLQ) focuses on real-world CC-ReID by *alternating* Base model (baseline) between **Coarse Attributes Prediction (CAP)** to enrich the model

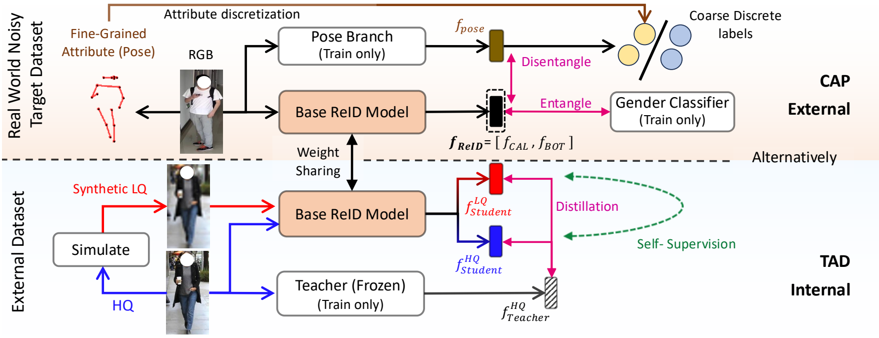


Figure 4: **Robustness against Low-Quality (RLQ)**: Base ReID model is our baseline with RGB-only input (orange). The base model is trained alternately between: *Upper*: Coarse Attributes Prediction (CAP) trained on target ReID dataset with train-only pose branch for pose disentanglement and gender classifier for gender entanglement. *Bottom*: Task Agnostic Distillation (TAD) trains a student base model using a teacher base model (pink, distillation) and self-supervision on classification logits (green). **Both CAP and TAD (white) are train-only, with base model f_{ReID} used during RGB-only inference.**

with *external* (pose, and gender) fine-grained attributes, and **Task Agnostic Distillation (TAD)** for enhancing the model's *internal* feature robustness on low-quality images (fig. 4). During testing, **RGB-only inference** on Base Model with no external attribute (CAP or TAD). **Note:** Our focus in this work is not on removing clothes bias; for that, we have the Base Model (CAL branch and cloth augmentations). Instead, our contribution is to induce robustness against low-quality artifacts as present in the real-world (inherently CC-ReID).

3.1 Baseline: Clothes Agnostic Base ReID Model (BM)

We adopt one of the most popular open-source CC-ReID model, CAL [11], for learning clothes invariant features. However, it lacks an understanding of pose and gender (fig. 3). Coupled with CAL's adversarial loss *sensitive* nature (table 5 & 6), its incompatible with both triplet loss and clothes augmentations, making it difficult to integrate additional losses in this architecture. Hence we adopt a two-branch architecture [4, 13], where Bag-of-tricks (BOT [33]) branch is added to CAL. BOT Branch can train with additional losses and transfer knowledge to the CAL branch via a shared backbone. The concatenation of features from both branches is our Base Model (used during inference) f_{ReID} (=[f_{BOT} , f_{CAL}]). Detailed description in *Supplmentary*, where results indicate two branch structures offer no performance boost, but simply stabilize CAL for additional training modules.

3.2 Coarse Attribute Prediction (CAP): External Attributes Features

The motivation behind Coarse Attribute Prediction (CAP) is to learn *essential external* fine-grained attributes without the additional input noise due to low-quality images. Thus, predicting these attributes is an obvious choice that CAP achieves by classifying attributes as coarse discrete labels from raw RGB images, as shown in fig. 4 (upper). This not only reduces the input noise but also removes the test time dependency on external attributes. CAP can disentangle attributes (remove pose bias) or entangle them (gender awareness). CAP applies a pose branch and a gender classifier on the BOT branch only during training.

Pose Branch: Pose embedding f_{POSE} is learned by classifying RGB images into pose clusters (similar to how ImageNet classification helps learn image class embeddings). Unlike pose *entanglement* [36], we believe pose is temporary and shouldn't play any role in identifying an individual; thus is disentangled. By maximizing f_{POSE} angular distance with f_{BOT} towards orthogonality *i.e.* minimize absolute cosine similarity, ($\mathcal{L}_{Cos}$), we disentangle pose bias from the features [24].

$$\mathcal{L}_{Cos} = \Big| \text{Cosine-Similarity}(f_{BOT}, f_{POSE}) \Big| \tag{1}$$

The pose branch shares the backbone with the base model, keeping features in the same feature space, easing disentanglement. Contrary to traditional methods [13, 37] of representing pose via location of **body joints**, we additionally include body line joining joints, consisting of **stretch** (length), and **bend** (angle) as shown in fig. 5 (left), adding fine-grained nuances to the pose vector. Pose vectors are then clustered using K-means, assigning each image a discrete pose cluster (clusters in *Supplementary*).

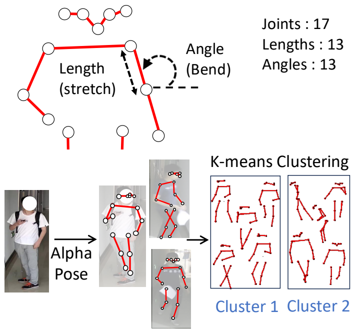


Figure 5: **Pose Clustering**: *Upper*: Pose Vector consists of the length and angle, used in L2 distance K-means clustering (*Lower*).

Gender Classifier: Gender often used for filtering out the wrong genders [8]. Simple classification of RGB images into gender labels can entangle gender-based cues (*e.g.* male/female body shape, pose) into ReID. Gender Classifier on f_{BOT} classifies RGB images into binary gender labels, *i.e.* male vs female, introducing gender awareness into BOT branch.

3.3 Task Agnostic Distillation (TAD): Model's Internal Features

Models cluster LQ image features as erroneous look-alike. Traditional solutions that low-quality features mimic high-quality ones are non-applicable without clean separable HQ-LQ pairs. Addressing these, TAD leverages external HQ datasets to bridge the gap between the two domains via synthetic HQ-LQ pairs. As shown in fig. 2(b), the external dataset HQ features pull LQ features via self-supervision and distillation. This implicitly helps ReID pull LQ features on the target dataset, dispersing low-quality clusters. This enhances the model's *internal* robustness against LQ artifacts (helping 'see better'). Unlike previous work [10, 18, 19], TAD is **1) trained parallel to the target ReID task**, in alternate epochs, **2) Any HQ dataset can help, as it plays no role in supervising ReID**, aka "task-agnostic".

Explanation: Matching HQ-LQ features risks pulling HQ embeddings towards LQ clusters [23], thus an *indirect* softer approach is adopted. Student HQ & LQ features are pulled toward the teacher's frozen HQ embeddings, reinforced by self-supervision on classifiers' logits *i.e.* HQ-LQ logit matching, instead of features. Teacher's distillation eliminates the need for labels from the external dataset, making any HQ dataset suitable regardless of the target ReID task (task-agnostic). Additionally, teacher fixed anchor embedding helps maintain the separability of features, serving two purposes: (1) **Prevents Collapsing**: Collapsing external dataset embeddings. (2) **No multi-tasking**: Since training on external dataset annotations is not required, a small subset can fix LQ clusters. All empirical evidence in ablation.

	Dataset	# of IDs		# of Imgs		# of Clothes		Px.	OOF	M. Blur	Evaluation Protocol
		Train	Test	Train	Test	Train	Test				
LQ	Celeb-ReID[14]†	632	420	20,208	13,978	20,208	-				-
LQ	LTCC[39]	77	75	9,576	7,543	256	221	✓	✓	✓	CC, General
LQ	PRCC[50]	150	71	17,896	10,800	300	142	✓	✓		CC, SC
VLQ	DeepChange[48]	450	521	75,083	80,483	1,219	1,336	✓	✓		General
VLQ	LaST[40]	5,000	5,807	71,248	135,529	16,035	-	✓	✓	✓	General

Table 1: **Dataset, artifact & Evaluation protocol** *Px.*: denotes pixelation, *OOF*: denotes out-of-focus blur, *M. Blur*: denotes motion blur, *CC*: means clothes changing, *SC*: means same clothes. †: **Celeb-reID** is chosen as the external dataset for TAD, (not for RLQ evaluation). PRCC and LTCC are real-world LQ datasets, while DeepChange and LaST are very LQ datasets (VLQ). All artifacts are empirically observed.

Implementation: External dataset x_{HQ} pre-trains the teacher base model and generates synthetic LQ image x_{LQ}. The frozen teacher only sees x_{HQ} producing normalized CAL and BOT features $[T^{HQ}_{CAL}, T^{HQ}_{BOT}]$, while the student sees both (x_{HQ}, x_{LQ}) generating normalized features as $[S^{HQ}_{CAL}, S^{HQ}_{BOT}]$ and $[S^{LQ}_{CAL}, S^{LQ}_{BOT}]$ (fig. 4 (bottom)). Classifiers produce ID logits $[Y^{\delta}_{CAL,ID}, Y^{\delta}_{BOT,ID}]$ and CAL's clothing logits $[Y^{\delta}_{CAL,CL}]$, where $\delta \in \{HQ, LQ\}$. $\mathcal{L}_D$ (pink) *distills* features, while KL divergence $\mathcal{L}_S$ (green) is used for soft *self-supervision*.

$$\mathcal{L}_D = \sum_{\delta}^{\{HQ,LQ\}} \|T^{HQ}_{CAL} - S^{\delta}_{CAL}\|^2_2 + \|T^{HQ}_{BOT} - S^{\delta}_{BOT}\|^2_2 \tag{2}$$

$$\mathcal{L}_S = KL\big(Y^{HQ}_{CAL,CL} \| Y^{LQ}_{CAL,CL}\big) + \sum_{\alpha}^{\{CAL,BOT\}} KL\big(Y^{HQ}_{\alpha,ID} \| Y^{LQ}_{\alpha,ID}\big) \tag{3}$$

3.4 Overall Learning Objective

Baseline inherits the traditional losses of the BOT and CAL branches. These include Label-smoothed cross-entropy loss (CE) for both branches identity classifiers, and clothes classifier for CAL, CAL adversarial loss for clothes invariance, and Triplet loss for f_{BOT}. We **introduce 1)** KL-divergence loss between the identity classifiers' logits of both branches (sync information across branches). *CAP* introduces **2)** CE loss for pose and **3)** CE for gender classification, and **4)** $\mathcal{L}_{Cos}$, (eq. (1)) for pose disentanglement. *TAD* losses are distillation **5)** $\mathcal{L}_D$, (eq. (2)), and **6)** self-supervision $\mathcal{L}_S$ (eq. (3)). **The weight for each loss (hyperparameter optimization *(left as future work)* is 1**.

4 Experiments and Results

Dataset & Evaluation Table 1 describes the four image CC-ReID datasets. Datasets are evaluated using Top-1 (*R-1*, CMC) and mAP evaluation. The evaluation protocol: (i) Cloth-Changing (CC): Query-Gallery wears different clothes. (ii) Same-Clothes (SC): Query-Gallery have matching clothes. (iii) General: All query and gallery samples.

Implementation Branches output is concat of max and avg pool [15]. Gender is manually annotated as 1 (male), and 0 (female). Pose vectors are generated using AlphaPose [7] clustered with an empirically found cluster size of 15 (0^{th} cluster for noisy poses). Celeb-ReID [14] is used as external HQ dataset. Synthetic LQ augmentation include **1) Pixelation**:

	Method	Additional Attributes	LTCC (%) CC R-1↑	LTCC (%) CC mAP↑	LTCC (%) General R-1↑	LTCC (%) General mAP↑	PRCC (%) CC R-1↑	PRCC (%) CC mAP↑	PRCC (%) SC R-1↑	PRCC (%) SC mAP↑
Prev. Work	BSGA (BMVC'22) [35]	S	-	-	-	-	<u>61.8</u>	58.7	99.6	97.3
	CCFA (CVPR'23) [12]	-	45.3	22.1	75.8	<u>42.5</u>	61.2	58.4	99.6	98.7
	CVSL (WACV'2024)	P	44.5	21.3	76.4	41.9	57.5	56.9	97.5	99.1
	CCPG (CVPR'24) [37]	P	<u>46.2</u>	**22.9**	**77.2**	**42.9**	<u>61.8</u>	58.3	**100**	99.6
	CLIP3D‡ (CVPR'24) [28]	B,Text	42.1	<u>21.7</u>	-	-	60.6	59.3	-	-
	Baseline: CAL (CVPR'22) [11]	-	40.1	18.0	74.2	40.8	55.2	55.8	**100**	<u>99.8</u>
	Baseline: Base Model	S	41.1	20.3	75.7	41.4	58.0	59.4	**100**	<u>99.8</u>
Our	CAP	S, P, G	$43.4^{+2.3}$	$20.7^{+0.4}$	$74.7^{-1.0}$	$41.9^{+0.5}$	$60.2^{+2.2}$	$60.3^{+0.9}$	**100**	**99.9**
	TAD	S	$43.4^{+2.3}$	$21.6^{+1.3}$	$76.1^{+0.4}$	$41.8^{+0.4}$	$61.2^{+3.2}$	<u>$61.1^{+1.7}$</u>	**100**	**99.9**
	RLQ (CAP + TAD)	S, P, G	**$46.4^{+5.3}$**	$21.5^{+1.2}$	<u>$76.9^{+1.2}$</u>	$41.8^{+0.4}$	**$64.0^{+6.0}$**	**$63.2^{+3.8}$**	**100**	<u>99.8</u>

Table 2: **Comparison on LTCC and PRCC (%)** Shorthand notations: "-": results were not reported, ↑ : higher the number better the performance, 'S': silhouettes/body parsing, 'P': Pose (2D/3D), 'B': Body Shape (2D/3D), and 'G': Gender. CLIP3D‡ [28] is short for 'CLIP3DReID'. '$^{+x}$' shows x% increment over baseline base model.

Method	Attributes	R-1↑	mAP ↑
M2Net-F (MM'22) [30]	S, C	<u>57.2</u>	19.6
IMS+GEP (TMM'23) [56]	-	55.1	18.3
MCSC-CAL (TIP'24) [17]	-	56.9	<u>21.5</u>
CLIP3D (CVPR'24) [28]	B,Text	56.7	20.8
Baseline: CAL (CVPR'22) [11]	-	54.0	19.0
Baseline: Base Model	S	57.7	21.4
RLQ (Our)	S, P, G	**58.8**	**22.1**

Table 3: **DeepChange General Results (%)**

Method	Attributes	R-1↑	mAP↑
LaST (TCSVT'22) [40])	-	71.0	28.0
MCL (MM'22) [20]	-	<u>75.0</u>	22.7
Lu and Jin (AIMS'23) [32]	-	68.9	24.1
IMS+GEP (TMM'23) [56]	-	73.2	<u>29.8</u>
Baseline: CAL (CVPR'22)[11]	-	73.7	28.8
Baseline: Base Model	S	75.7	31.7
RLQ (Our)	S, P, G	**77.9**	**35.3**

Table 4: **LaST General Results (%)**

downsample to resolution $\in [16 \times 16, 64 \times 64]$, and resizing back; **2) OOF blurring**: gaussian blur with random kernel $\in [5, 21]$; and **3) Motion blur**: random size kernel $\in [8, 20]$ of 1s rotated randomly $\in [0^\circ, 180^\circ]$. More details are provided in *Supplementary*.

4.1 Results

Ensuring fair comparison with all methods, accuracy-boosting techniques that can be universally applied to all models (including ours), *e.g*. Re-Ranking [54], Gallery enrichment [1], or Vivo learning [16] are not used in comparison. All results follow the color scheme: Our results, **SOTA**, and <u>second-best</u>. Results of the Celeb ReID dataset can be found under *Supplementary ('Limitations')*. Table 2 Formatting, applied to all rules (see caption).

- LTCC & PRCC (Real World LQ): Table 2 shows the comparison of our model with existing methods. Auxiliary attributes-based models in general outperform RGB-only models (CAL, CCFA). We surpass the previous best on the PRCC dataset with competitive results on the LTCC dataset, giving consistent improvement on CC protocol over the base model as <u>6.0% Top-1 and 3.8% mAP</u> on PRCC and <u>5.3% Top-1 and 1.2% mAP</u> on LTCC.

- DeepChange, and LaST (Very LQ): Table 3 & Table 4 shows results on the DeepChange and LaST. There are only a few fine-grained attributes work on these two datasets, partly because of their LQ artifacts. Attributes-related works include Contour (M2Net-F) and Text features w/ 3D body shape (CLIP3D), the rest are heuristic-based works. **Avoiding input noise & TAD** enables us to **use fine-grained attributes like pose, and gender for DeepChange and LaST (only work to do so)**. RLQ outperforms previous best by <u>1.6% and 2.9%</u> Top-1, and <u>0.6% and 5.5%</u> mAP on DeepChange and LaST, respectively.

Cloth Aug	Triplet Loss	CAL R-1	CAL mAP	BM R-1	BM mAP
		40.1	**18.0**	36.7	16.0
✓		36.9	18.0	40.1	17.9
	✓	34.7	16.6	38.8	18.3
✓	✓	35.5	17.0	**41.1**	**20.3**

Table 5: **CAL Fragile-A** Clothes augmentation & Triplet Loss

Branch	BM+G R-1	BM+G mAP	BM+P R-1	BM+P mAP
BOT	**43.5**	**21.7**	**42.5**	**20.2**
CAL	43.4	20.8	**42.4**	**20.2**
Both	41.9	20.4	41.5	19.5

Table 6: **CAL Fragile-B** Gender (G) entangle, Pose (P) disentangle, Both: both branches.

TAD HQ Datasets	PRCC R-1	PRCC mAP
BM	58.0	59.4
NTU	61.1	59.8
Celeb-ReID	**61.2**	**61.1**

Table 7: **TAD External dataset** NTU [29] instead of Celeb-ReID

5 Ablations and Analysis

Following has, *BM+G*: Base Model (BM) w/ only gender classifier, *BM+P*: w/ only pose branch, *BM+TAD*: w/ only TAD, and *BM+CAP*: RLQ w/o TAD *i.e.* BM+P+G. Ablations performed using the CC protocol for the LTCC (unless otherwise mentioned), due to its relatively smaller size compared to LaST and DeepChange. Design choices are highlighted.

- Base Model (Cal Sensitivity) Table 5 illustrates the *incompatibility of CAL's adversarial loss* with triplet loss and clothes augmentations. This is further evident from table 6, where pose disentanglement and gender entanglement perform optimally on the BOT-only branch. BM with extra BOT branch and foreground augmentation offers no performance boost (Base Model 41.1% vs vanilla RGB CAL 40.1%) but simply **fixes the CAL's sensitivity problem**.

- TAD (Task-agnostic) Comparable performance (table 7) of NTU (HQ dataset) [29] instead of Celeb-ReID in TAD, as a source of HQ-LQ image pairs indicates the **non-reliance/'task-agnostic' on the choice of HD dataset.** NTU limited viewpoints make Celeb-ReID preferable. Figure 7 (a) shows randomly sampling Celeb-ReID at ∼8% has consistent performance gain on TAD. Implying **small amount of HQ data & teacher's soft supervision is effective in learning the HQ-LQ relationship**, w/o the need for multi-tasking on external dataset.

- Each Component & Number of Params Table 8 shows the performance of each component on the Base model, w/ and w/o TAD. Gender has the largest impact on the LTCC, while pose performs well on the PRCC. TAD gives a performance boost to each component. The addition of pose with gender (CAP) for LTCC slightly reduces the performance, possibly because of **noisy pose clusters, which the model can fix (see better) upon application of TAD**. The final combination, RLQ, achieves the best performance. The gender classifier has minimal parameter overhead, while the pose branch results in a 50% parameter gain (ResNet branch). Also, 69M parameters are still relatively small compared to transformer-based approaches (LTCC sota) with ViT-Base/Swin-Base, with ∼87M parameters. Time analysis is tricky, as inference is only Base model with no TAD and CAP (Train-only).

- TAD (Teacher & External Dataset) Table 9 shows variations of TAD for improving robustness against LQ. Base Model with LQ augmentations (BM+LQ Aug.) doesn't impact the performance. Directly matching student's HQ-LQ features via MSE loss with teacher SS_{MSE}+TAD and without teacher SS_{MSE}, offers no performance gain, while later dropping performance. LTCC+SS_{MSE}, with noisy HQ-LQ pairs from LTCC instead of clean Celeb-ReID, has lower performance. This indicates the **need for clean HQ-LQ image pairs**, via an external dataset. Poor performance w/o teacher indicates possible feat collapse on the ext. dataset (distribution shift), justifying **need for feature separability via teacher**.

Aux.	#Par	BM				BM + TAD			
		LTCC		PRCC		LTCC		PRCC	
		R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP
BM	45.9	41.1	20.3	58.0	59.4	43.4	21.6	61.2	61.1
+G	45.9	43.5	21.7	58.8	59.8	44.8	**22.0**	62.4	62.6
+P	69.3	42.5	20.2	59.4	60.6	44.7	21.4	61.8	62.3
+CAP	69.3	43.4	20.7	60.2	60.3	**46.4**	21.5	**64.0**	**63.2**

Table 8: **Contribution of architectural component**. G.: Gender, P: Pose, CAP: Gender + Pose, #Par: parameters count in millions.

TAD Variants	T	ED	R-1	mAP
BM (Baseline)			41.1	20.3
BM + LQ Aug.			41.2	19.5
SS_{MSE} + TAD	✓	✓	43.3	21.2
SS_{MSE}		✓	32.4	16.3
LTCC+SS_{MSE}			38.3	18.5
TAD	✓	✓	**43.4**	**21.6**

Table 9: **TAD Variant** 'ED': external dataset, 'T': Teacher. SS_{MSE}: MSE on feats (Self-supervised).

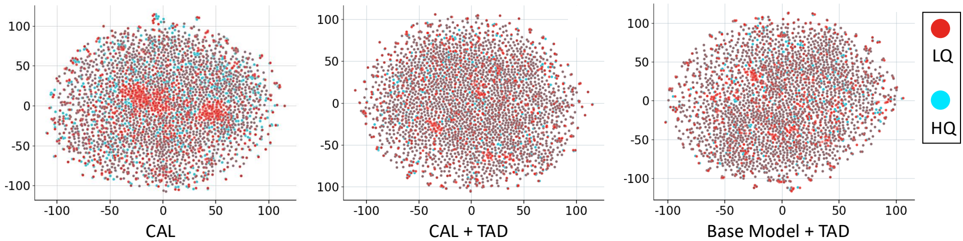
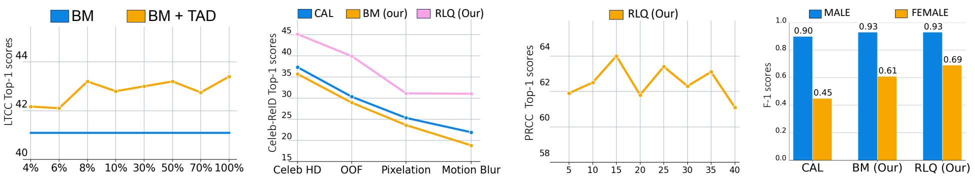


Figure 6: **TAD robustness against LQ clusters**: t-SNE of pairs of HQ (blue) & LQ (red) features for CelebReID dataset with synthetic artifacts. Lumps of Red dots (& not blue dots) indicate model cluster LQ features regardless of identities, with TAD reducing these clusters.



(a) Celeb-ReID Size (b) Impact of artifacts (c) K-means cluster Size (d) Gender classification

Figure 7: (a) Subset (8%) of external dataset improves performance. (b) Motion blur affects the performance most, and RLQ (w/ UBD) outperforms CAL and Base model. (c) PRCC pose vectors optimal cluster is 15. (d) RLQ is more gender aware than CAL and Base Model.

- TAD Visualization: Figure 6 shows t-SNE plots of the LTCC pre-trained model on the synthetic LQ image of the Celeb-ReID dataset (LTCC doesnt have HQ-LQ image pairs). Models cluster LQ image features, creating erroneous look-alike matches. **TAD tries to make LQ and HQ features indistinguishable, eliminating LQ clusters** (red dots), ideally overlapping with HQ features, thus inducing robustness against LQ artifacts. Figure 7 (b) shows RLQ, CAL, and the Base Model experience a significant drop in accuracy due to motion blur. Specifically addressing it is *left as future work*.

- Error Analysis Optimal cluster size (K-means) for pose is around 15 and 25 clusters as indicated Figure 7 (c). Figure 7 (d) shows a quantitative verification of gender awareness in RLQ compared to the Base Model and CAL. On their respective LTCC mispredictions, RLQ scores a 0.68 F-1 score on female classification (minority class), compared to CAL's 0.44 F-1 score and BM (0.61 F-1 score). Figure 8 instances where CAL misidentifies (a) gender,

(a) **Gender Robust** (b) **Pose Robust** (c) **LQ Robust** (d) Gender & Pose **Wrong**

Figure 8: **RLQ vs CAL Visual Error Analysis:** LTCC Top-1 match shown, with **correct** and **incorrect** matches. General improvements where CAL makes a mistake while RLQ doesn't are: (a) Gender recognition (b) Pose robustness (c) Low-quality robustness (LQ images match with other LQ images because of LQ feature clustering). When both models make mistake (d), RLQ still correctly identifies gender and avoids similar pose.

(b) pose (similar/flipped body), and (c) LQ images (motion blur matches motion blur), while RLQ correctly identifies. When both models make mistake (d), RLQ correctly identifies gender and selects different poses.

6 Limitation

We mainly focus on LQ images, deliberately disregarding fine-grained details that are available in HQ datasets. Coarse Attribute Prediction (CAP) tries to extract maximum information from artifact-ridden inputs by approximating fine-grained attributes as coarse labels. Coarse labels can fall short compared to their fine-grained counterparts. On evaluating CAP on Celeb-ReID (table 10), CAP offers marginal improvements over the Base model. We hypothesize the model has maximized its information gain from HQ RGB alone, leaving little scope for enhancement with coarse attributes.

Model	R-1	mAP
CAL	51.7	8.9
Base Model	58.8	14.4
Base Model + CAP	59.1	14.8

Table 10: **Celeb-ReID results**. TAD not applicable (already HQ).

7 Conclusion & Future Work

Our Robustness against Low-Quality (RLQ) focus on real world Clothes Changing Re-IDentification with real world low-quality (LQ) artifacts, namely pixelation, out-of-focus, and motion blur. Our goal is to incorporate traditional external attributes like Pose and gender, while avoiding the input noise they bring into ReID models via Coarse Attribute Prediction (CAP). Additionally, our TAD is designed to help see model better distinguish between low-quality images, in a task-agnostic setting, without the need for ReID. Adding more such attributes, modeling other noises (& improving motion blur), and a theoretical analysis on the benefit of TAD (and alternative training mechanisms) is *left as future work*. Additionally, hyper parameter optimization to re-weight (loses) CAP and TAD may provide a future direction of research. Our method "RLQ" is among the first works to successfully bring these attributes to very LQ datasets like LaST and DeepChange.

8 Acknowledgement

The authors gratefully acknowledge Steven Dick (UCF High-Performance Computing) for his invaluable assistance in maintaining the smooth-running cluster on which all the experiments were performed.

References

[1] Daniel Arkushin, Bar Cohen, Shmuel Peleg, and Ohad Fried. Geff: Improving any clothes-changing person reid model using gallery enrichment with face features. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024.

[2] Kunfeng Chen, Lei Shi, Zhisong Pan, Jiabao Wang, and Xi Zhan. Clothes-independent identity feature learning for long-term person re-identification. In *Image and Graphics: 11th International Conference, ICIG 2021, Haikou, China, August 6–8, 2021, Proceedings, Part II*, page 397–409, Berlin, Heidelberg, 2021. Springer-Verlag. ISBN 978-3-030-87357-8. URL https://doi.org/10.1007/978-3-030-87358-5_32.

[3] Yun-Chun Chen, Yu-Jhe Li, Xiaofei Du, and Yu-Chiang Frank Wang. Learning resolution-invariant deep representations for person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8215–8222, 2019.

[4] Zhenyu Cui, Jiahuan Zhou, Yuxin Peng, Shiliang Zhang, and Yaowei Wang. Dcr-reid: Deep component reconstruction for cloth-changing person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(8):4415–4428, 2023. doi: 10.1109/TCSVT.2023.3241988.

[5] Daniel Davila, Dawei Du, Bryon Lewis, Christopher Funk, Joseph Van Pelt, Roderic Collins, Kellie Corona, Matt Brown, Scott McCloskey, Anthony Hoogs, et al. Mevid: Multi-view extended videos with identities for video person re-identification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1634–1643, 2023.

[6] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *CVPR*, 2020.

[7] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu. Rmpe: Regional multi-person pose estimation. In *ICCV*, 2017.

[8] Hiren Galiyawala, Kenil Shah, Vandit Gajjar, and Mehul S Raval. Person retrieval in surveillance video using height, color and gender. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE, 2018.

[9] Jiyang Gao and Ram Nevatia. Revisiting temporal modeling for video-based person reid. *arXiv preprint arXiv:1805.02104*, 2018.

[10] Shiming Ge, Shengwei Zhao, Chenyu Li, Yu Zhang, and Jia Li. Efficient low-resolution face recognition via bridge distillation. *IEEE Transactions on Image Processing*, 29: 6898–6908, 2020. doi: 10.1109/TIP.2020.2995049.

[11] Xinqian Gu, Hong Chang, Bingpeng Ma, Shutao Bai, Shiguang Shan, and Xilin Chen. Clothes-changing person re-identification with rgb modality only. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1060–1069, 2022.

[12] Ke Han, Shaogang Gong, Yan Huang, Liang Wang, and Tieniu Tan. Clothing-change feature augmentation for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22066–22075, 2023.

[13] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10513–10522, June 2021.

[14] Yan Huang, Qiang Wu, Jingsong Xu, and Yi Zhong. Celebrities-reid: A benchmark for clothes variation in long-term person re-identification. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.

[15] Yan Huang, Qiang Wu, JingSong Xu, Yi Zhong, and ZhaoXiang Zhang. Clothing status awareness for long-term person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11895–11904, 2021.

[16] Yan Huang, Yan Huang, Zhang Zhang, Qiang Wu, Yi Zhong, and Liang Wang. Enhancing person re-identification performance through in vivo learning. *IEEE Transactions on Image Processing*, 33:639–654, 2024. doi: 10.1109/TIP.2023.3341762.

[17] Yan Huang, Qiang Wu, Zhang Zhang, Caifeng Shan, Yan Huang, Yi Zhong, and Liang Wang. Meta clothing status calibration for long-term person re-identification. *IEEE Transactions on Image Processing*, 33:2334–2346, 2024. doi: 10.1109/TIP.2024.3374634.

[18] Jiening Jiao, Wei-Shi Zheng, Ancong Wu, Xiatian Zhu, and Shaogang Gong. Deep low-resolution person re-identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), Apr. 2018. doi: 10.1609/aaai.v32i1.12284. URL https://ojs.aaai.org/index.php/AAAI/article/view/12284.

[19] Jiening Jiao, Wei-Shi Zheng, Ancong Wu, Xiatian Zhu, and Shaogang Gong. Deep low-resolution person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[20] Xin Jin, Tianyu He, Xu Shen, Tongliang Liu, Xinchao Wang, Jianqiang Huang, Zhibo Chen, and Xian-Sheng Hua. Meta clustering learning for large-scale unsupervised person re-identification. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, page 2163–2172, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392037. doi: 10.1145/3503161.3547900. URL https://doi.org/10.1145/3503161.3547900.

[21] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14278–14287, June 2022.

[22] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14278–14287, 2022.

[23] Syed Safwan Khalid, Muhammad Awais, Zhen-Hua Feng, Chi-Ho Chan, Ammarah Farooq, Ali Akbari, and Josef Kittler. Resolution invariant face recognition using a distillation approach. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(4):410–420, 2020. doi: 10.1109/TBIOM.2020.3007356.

[24] Minjung Kim, MyeongAh Cho, and Sangyoun Lee. Feature disentanglement learning with switching and aggregation for video-based person re-identification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1603–1612, 2023.

[25] Peike Li, Yunqiu Xu, Yunchao Wei, and Yi Yang. Self-correction for human parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. doi: 10.1109/TPAMI.2020.3048039.

[26] Yu-Jhe Li, Yun-Chun Chen, Yen-Yu Lin, Xiaofei Du, and Yu-Chiang Frank Wang. Recover and identify: A generative dual model for cross-resolution person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8090–8099, 2019.

[27] Feng Liu, Minchul Kim, ZiAng Gu, Anil Jain, and Xiaoming Liu. Learning clothing and pose invariant 3d shape representation for long-term person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19617–19626, 2023.

[28] Feng Liu, Minchul Kim, Zhiyuan Ren, and Xiaoming Liu. Distilling clip with dual guidance for learning discriminative human body shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 256–266, June 2024.

[29] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C Kot. Ntu rgb+d 120: A large-scale benchmark for 3d human activity understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10):2684–2701, 2020.

[30] Mengmeng Liu, Zhi Ma, Tao Li, Yanfeng Jiang, and Kai Wang. Long-term person re-identification with dramatic appearance change: Algorithm and benchmark. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6406–6415, 2022.

[31] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. *SMPL: A Skinned Multi-Person Linear Model*. Association for Computing Machinery, New York, NY, USA, 1 edition, 2023. ISBN 9798400708978. URL https://doi.org/10.1145/3596711.3596800.

[32] Yong Lu and Ming Zhe Jin. Dual-branch network fused with two-level attention mechanism for clothes-changing person re-identification. *Int. J. Web Serv. Res.*, 20 (1):1–14, apr 2023. ISSN 1545-7362. doi: 10.4018/IJWSR.322021. URL https://doi.org/10.4018/IJWSR.322021.

[33] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019.

[34] Fei Ma, Xiao-Yuan Jing, Yongfang Yao, Xiaoke Zhu, and Zhiping Peng. High-resolution and low-resolution video person re-identification: A benchmark. *IEEE Access*, 7:63426–63436, 2019. doi: 10.1109/ACCESS.2019.2912302.

[35] Jingyi Mu, Yong Li, Jun Li, and Jian Yang. Learning clothes-irrelevant cues for clothes-changing person re-identification. In *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*, page 337. BMVA Press, 2022. URL https://bmvc2022.mpi-inf.mpg.de/337/.

[36] Vuong D. Nguyen, Khadija Khaldi, Dung Nguyen, Pranav Mantini, and Shishir Shah. Contrastive viewpoint-aware shape learning for long-term person re-identification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1041–1049, January 2024.

[37] Vuong D. Nguyen, Pranav Mantini, and Shishir K. Shah. Contrastive clothing and pose generation for cloth-changing person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 7541–7549, June 2024.

[38] P. Pathak. Fine-grained re-identification. *ArXiv*, abs/2011.13475, 2020.

[39] 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. *arXiv preprint arXiv:2005.12633*, 2020.

[40] Xiujun Shu, Xiao Wang, Xianghao Zang, Shiliang Zhang, Yuanqi Chen, Ge Li, and Qi Tian. Large-scale spatio-temporal person re-identification: Algorithm and benchmark. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 4390–4403, 2022.

[41] X. Soria, E. Riba, and A. Sappa. Dense extreme inception network: Towards a robust cnn model for edge detection. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1912–1921, Los Alamitos, CA, USA, mar 2020. IEEE Computer Society. doi: 10.1109/WACV45572.2020.9093290. URL https://doi.ieeecomputersociety.org/10.1109/WACV45572.2020.9093290.

[42] Veeru Talreja, Fariborz Taherkhani, Matthew C Valenti, and Nasser M Nasrabadi. Attribute-guided coupled gan for cross-resolution face recognition. In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–10. IEEE, 2019.

[43] Dingyi Wang and Haishun Du. Double-stream network for clothes-changing person re-identification based on clothes related feature suppression and attention mechanism. In Wang Yongtian and Wu Lifang, editors, *Image and Graphics Technologies and Applications*, pages 208–219, Singapore, 2023. Springer Nature Singapore. ISBN 978-981-99-7549-5.

[44] Guangcong Wang, Jianhuang Lai, Peigen Huang, and Xiaohua Xie. Spatial-temporal person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8933–8940, 2019.

[45] Haihan Wang and Shangfei Wang. Low-resolution face recognition enhanced by high-resolution facial images. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–8, 2023. doi: 10.1109/FG57933.2023.10042552.

[46] Qizao Wang, Xuelin Qian, Yanwei Fu, and Xiangyang Xue. Co-attention aligned mutual cross-attention for cloth-changing person re-identification. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 2270–2288, December 2022.

[47] Jianbing Wu, Hong Liu, Wei Shi, Hao Tang, and Jingwen Guo. Identity-sensitive knowledge propagation for cloth-changing person re-identification. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 1016–1020, 2022. doi: 10.1109/ICIP46576.2022.9897243.

[48] Peng Xu and Xiatian Zhu. Deepchange: A large long-term person re-identification benchmark with clothes change. *arXiv preprint arXiv:2105.14685*, 2021.

[49] 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 Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 1523–1529. International Joint Conferences on Artificial Intelligence Organization, 7 2022. doi: 10.24963/ijcai.2022/212. URL https://doi.org/10.24963/ijcai.2022/212. Main Track.

[50] Qize Yang, Ancong Wu, and Wei-Shi Zheng. Person re-identification by contour sketch under moderate clothing change. *IEEE transactions on pattern analysis and machine intelligence*, 43(6):2029–2046, 2019.

[51] Hanlei Yu, Bin Liu, Yan Lu, Qi Chu, and Nenghai Yu. Multi-view geometry distillation for cloth-changing person reid. In *Pattern Recognition and Computer Vision: 5th Chinese Conference, PRCV 2022, Shenzhen, China, November 4–7, 2022, Proceedings, Part I*, page 29–41, Berlin, Heidelberg, 2022. Springer-Verlag. ISBN 978-3-031-18906-7. URL https://doi.org/10.1007/978-3-031-18907-4_3.

[52] Zhengxu Yu, Yilun Zhao, Bin Hong, Zhongming Jin, Jianqiang Huang, Deng Cai, and Xian-Sheng Hua. Apparel-invariant feature learning for person re-identification. *IEEE Transactions on Multimedia*, 24:4482–4492, 2022. doi: 10.1109/TMM.2021.3119133.

[53] Xianghao Zang, Ge Li, and Wei Gao. Multidirection and multiscale pyramid in transformer for video-based pedestrian retrieval. *IEEE Transactions on Industrial Informatics*, 18(12):8776–8785, 2022. doi: 10.1109/TII.2022.3151766.

[54] Renjie Zhang, Yu Fang, Huaxin Song, Fangbin Wan, Yanwei Fu, Hirokazu Kato, and Yang Wu. Specialized re-ranking: A novel retrieval-verification framework for cloth changing person re-identification. *Pattern Recognition*, 134:109070, 2023. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2022.109070. URL https://www.sciencedirect.com/science/article/pii/S0031320322005507.

[55] Zhiwei Zhao, Bin Liu, Qi Chu, Yan Lu, and Nenghai Yu. Joint color-irrelevant consistency learning and identity-aware modality adaptation for visible-infrared cross modality person re-identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(4):3520–3528, May 2021. doi: 10.1609/aaai.v35i4.16466. URL https://ojs.aaai.org/index.php/AAAI/article/view/16466.

[56] Zhiwei Zhao, Bin Liu, Yan Lu, Qi Chu, Nenghai Yu, and Chang Wen Chen. Joint identity-aware mixstyle and graph-enhanced prototype for clothes-changing person re-identification. *IEEE Transactions on Multimedia*, pages 1–12, 2023. doi: 10.1109/TMM.2023.3311143.

[57] Junhao Zheng, Xiaoman Hu, Tianyi Xiang, and Patrick P. K. chan. Dual-path model for person re-identification under cloth changing. In *2020 International Conference on Machine Learning and Cybernetics (ICMLC)*, pages 291–297, 2020. doi: 10.1109/ICMLC51923.2020.9469545.

[58] Wei-Shi Zheng, Jincheng Hong, Jiening Jiao, Ancong Wu, Xiatian Zhu, Shaogang Gong, Jiayin Qin, and Jianhuang Lai. Joint bilateral-resolution identity modeling for cross-resolution person re-identification. *International Journal of Computer Vision*, 130, 01 2022. doi: 10.1007/s11263-021-01518-z.

[59] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2138–2147, 2019.

[60] Haidong Zhu, Wanrong Zheng, Zhaoheng Zheng, and Ram Nevatia. Sharc: Shape and appearance recognition for person identification in-the-wild. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6290–6300, 2024.