CGCC: Towards Generalizable Clothes-Changing Person Re-Identification

Yizhi Wu[1], Fangyi Liu[2†], Wei Yu[1†], and Mang Ye[2]

[1] School of Artificial Intelligence, Wuhan University, Wuhan, China
[2] School of Computer Science, Wuhan University, Wuhan, China
{yizhiwu, fangyiliu, yuwei, yemang}@whu.edu.cn

Abstract. Clothes-Changing Person Re-Identification (CC-ReID) aims to match pedestrians across non-overlapping camera views despite variations in clothes. However, current research is severely constrained by two main shortcomings: existing datasets lack comprehensive diversity and semantic annotations, and current methods fail to globally model and eliminate identity-irrelevant interfering factors, severely limiting CC-ReID generalization in real-world scenarios. To address these limitations, we propose a unified framework for generalizable CC-ReID. First, we introduce CGCC, a Comprehensive dataset for Generalizable CC-ReID featuring extensive diversity and fine-grained semantic annotations, comprising 4,101 identities and 217,248 images. Additionally, we introduce Text-Guided Identity Refinement (TGIR), a novel framework that utilizes text descriptions to construct an identity-irrelevant subspace, and employs Singular Value Decomposition (SVD) to holistically extract and eliminate environmental and clothes interference. Together, CGCC and TGIR form a mutually reinforcing closed loop, where the diverse data and structured annotations of the former provide the essential foundation for the semantic disentanglement of the latter. Extensive experiments demonstrate that the comprehensive diversity of CGCC effectively improves model generalizability, and TGIR precisely extracts robust and generalizable identity representations. Dataset is available at https://github.com/zhi-time/CGCC.

Keywords: Clothes-Changing Person Re-identification · Generalization

1 Introduction

Person Re-Identification (ReID) [7, 9, 61, 68] aims to retrieve a specific person across non-overlapping camera networks. Traditional ReID methods [10, 15, 16, 23, 38, 55, 62, 66, 67, 69, 70] are often built on a strong assumption: pedestrians wear the same clothes within a short period. However, in practical long-term surveillance or cross-scene tracking, clothes-changing is a highly common and unavoidable phenomenon. Therefore, researchers have started to focus on Clothes-Changing Person Re-identification (CC-ReID) [11, 13, 28, 37]. The core difficulty

† Corresponding authors.

Table 1: Comparison of our CGCC dataset with existing CC-ReID datasets. Our dataset covers a greater diversity and has semantic annotations.

Datasets	# IDs	# Images	# Cams	Annotation	Diversity Variations
PRCC [59]	221	33,698	3	ID, Clothes	Clothes
VC-Clothes [45]	512	19,060	4	ID	Clothes
LTCC [39]	152	17,119	12	ID, Clothes	Clothes, Light
Celeb-reID [20]	1,052	34,186	-	ID	Clothes
LaST [42]	10,862	228,156	-	ID, Clothes	Clothes, Region
DeepChange [57]	1,121	178,407	17	ID	Clothes, Season, Weather
CGCC (Ours)	**4,101**	**217,248**	**15**	ID, Clothes, Semantics	Clothes, Light, Region, Culture, Season, Weather

of this task lies in the fact that clothes appearance features, which traditional models highly rely on, undergo changes or even become completely invalid after a person changes clothes. Consequently, learning more essential identity clues that remain stable before and after clothes changes has become a key challenge.

To address this challenge, numerous works in recent years have attempted to introduce various datasets [18, 39, 42, 45, 57, 59] containing clothes variations and proposed methods [2, 8, 19, 40, 54] leveraging contours [17, 59], poses [34, 65], 3D shapes [3, 30], gait [24] or attribute descriptions [37] to mitigate clothes interference. Nevertheless, existing research still suffers from two key shortcomings.

First, existing datasets [32, 39, 42, 45, 57, 59] suffer from insufficient diversity and a lack of semantic annotations, making it difficult to support models in learning clothes-changing representations with broad generalizability. Specifically, as shown in Table 1, some datasets only focus on the single factor of Clothes. Although a few datasets have started to introduce variations in factors like Light and Season, they completely ignore the profound impact of cultural differences on human dressing habits and fail to comprehensively incorporate multiple factors. Meanwhile, existing datasets primarily use discrete ID and clothes labels, lacking semantic annotations capable of accurately describing appearance changes. These issues make it difficult for existing datasets to support the model in learning identity representations with broad generalization capabilities.

Second, existing methods [28,37] based on semantic supervision often lack the global modeling and utilization of identity-irrelevant text descriptions. Specifically,as Figure 1(a), current works often treat various identity-irrelevant interfering factors in isolation, such as clothes and environments, failing to construct them into a unified "identity-irrelevant semantic space" for collaborative elimination. This semantic utilization strategy, which lacks a global perspective, makes it difficult for the model to characterize the distribution boundaries of identity-irrelevant features from a holistic viewpoint, thereby severely limiting the generalization capability of the learned identity representations in real-world clothes-changing scenarios.

To address the aforementioned shortcomings, this paper proposes a unified framework for clothes-changing person re-identification geared towards high generalizability. First, to address the limited diversity of existing datasets and the

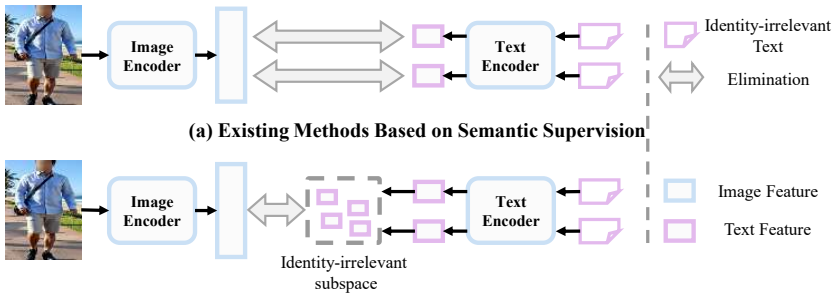

Fig. 1: Comparison of semantic supervision strategies. (a) Existing methods treat identity-irrelevant texts in isolation during elimination. (b) Our proposed TGIR constructs a unified identity-irrelevant subspace from texts to holistically extract and eliminate interference.

difficulty of collecting highly diverse real-world samples for the same identity, we construct a dataset, CGCC, which contains Comprehensive diversity and fine-grained semantic annotations for Generalizable CC-ReID. As shown in Table 1, relying on Multimodal Large Language Models (MLLMs) [6,27] and controllable generation mechanisms [49,53], we construct the CGCC dataset that significantly expands the training distribution of the same identity with multiple clothes and environments from the perspectives of season, region, culture, light, weather and so on. It also automatically generates structured semantic annotations for the samples, comprising a total of 4,101 IDs and 217,248 samples, providing a solid data foundation for model learning. Second, targeting the issue of lacking global semantic modeling in existing methods, we propose a novel Text-Guided Identity Refinement (TGIR) method. Motivated by the insight that textual semantics are naturally divisible into identity-relevant and identity-irrelevant descriptions [28], prompting the model to selectively absorb the identity features and avoid the irrelevant ones. TGIR approaches the problem from the innovative perspective of semantic subspace disentanglement. We decompose the descriptive information into identity-relevant biometric texts and identity-irrelevant clothes and environment texts. Based on this, TGIR, on the one hand, actively utilizes identity-relevant semantics in the representation space to guide the reinforcement of visual representations; on the other hand, as illustrated in Figure 1(b), it utilizes identity-irrelevant texts to construct an identity-irrelevant subspace. By conceptualizing the identity-irrelevant interference not as unstructured noise but as a specific semantic basis, it then applies Singular Value Decomposition (SVD) to holistically extract the distribution boundaries of identity-irrelevant features from a global perspective. This rigorous mathematical orthogonal projection is designed to reduce of environmental and clothes variations without compromising intrinsic biometric clues. Through the synergistic effect of data diversity expansion and semantic disentangled learning, our model can extract more stable, and generalizable identity representations.

Crucially, our dataset (CGCC) and method (TGIR) form a mutually reinforcing closed loop. CGCC provides the essential fine-grained annotations and diverse data for semantic disentanglement, while TGIR's explicit decoupling mechanism fully unlocks the supervisory potential of this data. This deep synergy ensures that the rich variations are effectively translated into identity representations with superior generalizability across diverse clothes-changing scenarios.

Our main contributions are as follows:

- **Proposed a unified framework for Generalizable CC-ReID:** We propose a unified framework that establishes a mutually reinforcing closed loop between data diversity expansion and semantic disentangled learning. This deep synergy ensures that the rich variations provided by our dataset are effectively translated into identity representations with superior generalizability across diverse clothes-changing scenarios.
- **Constructed a novel dataset, CGCC:** We propose a large-scale, highly diverse dataset for Generalizable CC-ReID, which significantly expands the sample distribution of identical identities across different clothes and environments, and provides fine-grained structured semantic annotations. This effectively alleviates the problems of limited scale, singular variation patterns, and lack of semantic supervision in existing datasets.
- **Proposed the TGIR framework:** We propose a novel Text-Guided Identity Refinement method. By disentangling text descriptions, it utilizes environmental and clothes texts to construct an identity-irrelevant subspace. Combined with SVD global boundary extraction, it effectively strips away environmental and clothes interference, achieving the precise learning of highly robust and generalizable identity representations.
- **Achieved superior performance:** We conducted extensive and in-depth experimental validation. The results not only prove that the CGCC dataset can effectively enhance the generalization capability of existing models but also fully verify the robustness and high generalizability of the proposed TGIR method in complex clothes-changing scenarios.

2 Related Works

2.1 Clothes-Changing Person Re-identification Datasets

Existing CC-ReID datasets are primarily categorized into real-world [20, 39, 42, 45, 57, 59, 63] and synthetic datasets [25, 45] based on their data sources.

Real-world Datasets. Early CC-ReID datasets, such as PRCC [59], LTCC [39], typically focused on clothes or light variations in restricted environments. While recent datasets like LaST [42] and DeepChange [57] expand the scale, they remain confined to specific scenes, ignoring the profound impact of multi-factor couplings, such as culture, season, and weather. More importantly, these datasets rely predominantly on discrete ID and clothes labels rather than fine-grained semantic annotations. Consequently, these issues make it difficult for

existing datasets to support models in learning identity representations with broad generalization capabilities.

Synthetic Datasets. To overcome high collection costs and privacy constraints, synthetic datasets like VC-Clothes [45] and SMPL-reID [25] utilize game engines or 3D rendering. However, these datasets remain bottlenecked by limited variation patterns. Crucially, like their real-world counterparts, they also suffer from a complete absence of fine-grained semantic supervision, similarly failing to support the learning of broadly generalizable representations.

Compared to all aforementioned datasets, our proposed CGCC dataset leverages the open-world knowledge of MLLMs and controllable generation mechanisms to break through these physical, privacy, and rendering constraints. It not only achieves unprecedented multi-factor diversity, seamlessly coupling region, culture, weather, and lighting, but also automatically provides fine-grained, structured semantic annotations. This synergistic combination of rich data distribution and dense semantic supervision successfully addresses the shortcomings of existing datasets, providing a solid data foundation for learning identity representations with broad generalization capabilities.

2.2 Clothes-Changing Person Re-identification Methods

Early CC-ReID methods [2, 4, 8, 14, 19, 21, 26, 36, 40, 41, 46–48, 50, 51, 54, 56, 58] extract clothes-invariant features using structural cues like contours [17, 59], human parsing [12], 3D shapes [30], or gait [24]. Others employ adversarial learning [28,34] or data augmentation [5,29] to mitigate clothes bias. However, strictly relying on structural modalities often discards fine-grained identity clues (e.g., age, gender, style), limiting generalization across unconstrained environments.

Recent works [28,37] leverage language or attributes as auxiliary supervision. Despite their progress, they typically treat interfering factors in isolation, failing to construct a unified identity-irrelevant semantic space for collaborative elimination. Lacking a global perspective, these methods struggle to characterize the distribution boundaries of interference, leading to suboptimal generalization.

Unlike existing methods, our TGIR framework employs semantic subspace disentanglement. By decomposing texts, TGIR constructs a unified identity-irrelevant subspace from clothes and environment descriptions and applies SVD for global boundary extraction. This orthogonal projection holistically strips away complex interference while preserving intrinsic biometric clues.

3 The CGCC dataset

The CGCC dataset is a large-scale synthetic dataset explicitly designed for Generalizable CC-ReID. To overcome real-world physical and privacy constraints, we construct this dataset using MLLMs and controllable generation mechanisms. CGCC breaks existing limitations by introducing unprecedented diversity that couples multiple factors such as clothes, season, region, culture, light,

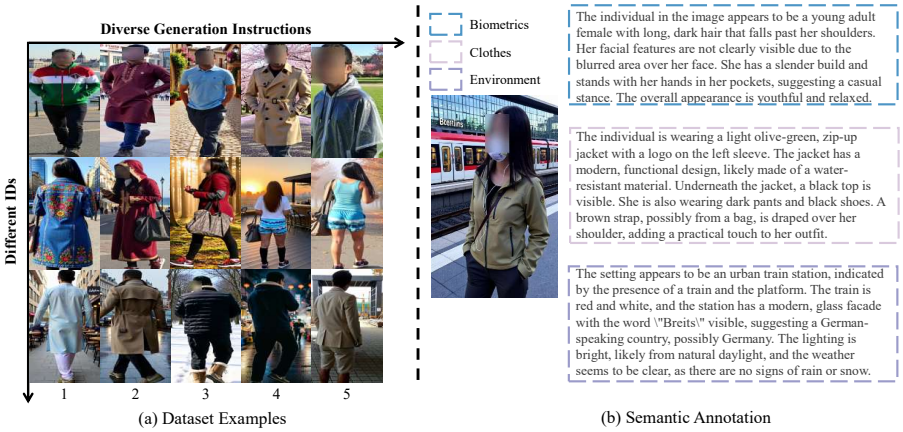

Fig. 2: Visual characteristics of the proposed CGCC datasets. (a) **Dataset Examples.** Dataset examples generated through diverse textual instructions. Each row represents a distinct identity, and the columns correspond to the five specific generation instructions coupling multiple factors such as culture, weather, season and so on. (b) **Semantic Annotation.** Fine-grained semantic annotations. The comprehensive descriptions are disentangled into biometrics, clothes, and environments to provide a solid foundation for robust representation learning.

and weather. Furthermore, by providing fine-grained semantic annotations instead of discrete labels, CGCC supplies a solid data foundation for learning highly generalizable identity representations. This section systematically details its construction and analysis.

3.1 Constructing CGCC via Data Generation

Collecting extreme diversity data for identical persons in the real world is physically impossible and restricted by privacy concerns. To break through the limitations of existing datasets, we utilize controllable text-to-image data generation techniques and by leveraging MLLMs to construct the CGCC dataset. The specific construction process is detailed as follows.

High-Quality Seed Data Screening. We selected MSMT17 [52] as our foundational source because it offers a massive scale of diverse identities, abundant samples per identity, and high overall image quality. To prevent the editing model from inheriting or amplifying real-world noise such as motion blur and severe occlusion, we established a strict screening mechanism. First, for Identity Integrity, we use SCHP [22] to filter out samples without a clear pedestrian torso. Second, for Perceptual Quality, we introduce a Portrait Quality Assessment model [44] to retain only the top half of the clearest images. Third, for Intra-class Diversity, we strictly select identities with more than 20 images to ensure a rich base of poses and viewpoints. These high-quality seeds provide a stable identity backbone for subsequent generative variations.

Diversity-Driven Image Editing Instruction Design. To overcome the limitations of manual editing prompts that often fail to cover the complex characteristics of diverse global scenarios, we leverage the powerful open-world knowledge of MLLMs to construct a comprehensive instruction pool encompassing two distinct dimensions. Specifically, utilizing the advanced Large Language Model ChatGPT5, we focus the first dimension on global-cultural and regional specificity. We generated descriptions of regional environments and their corresponding characteristic clothes, covering 6 continents and 66 countries. The second dimension focuses on spatio-temporal and environmental dynamics. We generated diverse descriptions of various seasons, weather conditions, and their corresponding functional clothes. These descriptions are then embedded into a standardized editing template, i.e., *"Change the background to {scene description}, and change the person's clothes to {clothes description}. Keep the person unchanged."*. Ultimately, this dual-path design fundamentally guarantees the comprehensive diversity of CGCC, enabling it to encompass the multifaceted variations required for generalizable CC-ReID in complex, real-world scenarios.

Instruction-Driven Image Editing. We employ the Qwen-Image-Edit model [53] to execute the synthesis process, leveraging its superior visual-semantic understanding to maintain identity structure while repainting complex attributes. For each high-quality seed image, we generate five augmented samples to ensure a balanced yet comprehensive data distribution. Specifically, two samples are derived from the global-cultural and regional pool, and another two are drawn from the spatially-temporal and environmental pool. To further enhance the model's robustness against rare or unforeseen scenarios, the fifth sample is generated using an Instruction Recombination Strategy. This strategy randomly pairs a geographic location from the first pool with a non-matching environmental or clothes description from the second. This strategy intentionally creates extreme visual scenarios and increases the diversity of CGCC.

Synthetic Image Filtration and Semantic Annotation. To ensure that the synthetic images do not suffer from identity information alteration or distortion, we deployed a post-processing module driven by Qwen3-VL. Since synthetic images are based on editing the clothes and background of real images, their ID labels should be consistent with the corresponding real image labels. Therefore, we first use Qwen3-VL [1] to check the edited images one by one, verifying whether the pedestrian appearance remains normal and whether there is a significant difference from the original identity, thereby filtering out samples with facial distortion or identity loss. Upon completing the filtration, we merge the high-quality synthetic images with the complete original MSMT17 dataset to constitute the final CGCC dataset. Finally, we utilize Qwen3-VL to generate detailed textual descriptions covering pedestrian biometrics, clothes, and environment. These automatically generated texts serve as explicit fine-grained semantic supervision signals.

3.2 Dataset Analysis

We analyze the proposed CGCC dataset from three primary dimensions: comprehensive clothes diversity, fine-grained semantic richness, and dataset scale and quality.

Comprehensive Clothes Diversity. As shown in Figure 2(a), CGCC contains comprehensive diversity through the seamless coupling of multiple factors. Specifically, to capture the profound impact of cultural differences on human dressing habits, the dataset encompasses diverse ethnic and daily clothes from multiple countries globally. Additionally, light variations explicitly cover transitions between day and night, while regional diversity spans worldwide urban settings and wilderness environments. Furthermore, it incorporates the four seasons along with various common and extreme weather conditions. By strictly integrating these multifaceted factors, CGCC successfully provides the rich diversity essential for generalizable CC-ReID.

Fine-Grained Semantic Richness. Unlike existing datasets using coarse discrete labels, CGCC provides open-vocabulary textual supervision. As illustrated in Figure 2(b), our annotations are meticulously structured into stable identity biometrics, variable clothes details, and environmental contexts. These explicit semantic signals more accurately annotate the identity-relevant invariant and the identity-irrelevant variations.

Scale, Quality, and Privacy. CGCC comprises 4,101 identities and 217,248 high-fidelity samples characterized by exceptional image quality. By utilizing MLLMs and controllable generation mechanisms, we create a large-scale benchmark that completely overcomes real-world privacy constraints while maintaining photorealistic visual details. This synergistic combination of rich data distribution and high image quality supports models in learning identity representations with broad generalization capabilities in complex clothes-changing scenarios.

4 Methodology

To address the limitation of existing methods that lack global modeling for identity-irrelevant interference, we propose the **Text-Guided Identity Refinement (TGIR)** framework, as illustrated in Figure 3. From the innovative perspective of semantic subspace disentanglement, TGIR decomposes descriptive information into identity-relevant biometric texts and identity-irrelevant clothes and environmental texts. It leverages these disentangled semantics to achieve precise feature refinement through two synergistic modules: **Biometric-Guided Prototype Alignment (BGPA)** actively reinforces intrinsic identity representations, while **Spectrum-Aware Orthogonal Projection (SAOP)** constructs a unified identity-irrelevant subspace and applies Singular Value Decomposition (SVD) to globally strip away interference. Note that all textual descriptions are employed solely as supervision signals during training, imposing no additional computational burden during inference.

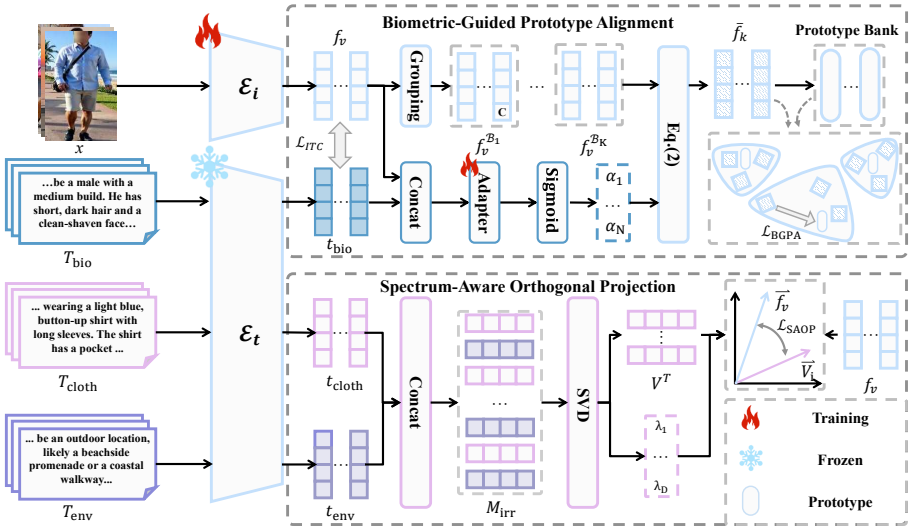

Fig. 3: Overview of the Text-Guided Identity Refinement (TGIR) framework. TGIR extracts robust, generalizable identity representations for CC-ReID via two core modules that fully align with our semantic subspace disentanglement paradigm: (1) **Biometric-Guided Prototype Alignment (BGPA)** leverages identity-relevant biometric text (T_{bio}) to calibrate stable identity prototypes, reinforcing intrinsic identity representations in the feature space. (2) **Spectrum-Aware Orthogonal Projection (SAOP)** constructs a unified identity-irrelevant semantic subspace from identity-irrelevant clothes (T_{cloth}) and environmental (T_{env}) texts, then holistically eliminates identity-irrelevant variations via orthogonal projection. Note that all textual inputs are used exclusively during training, with no additional computational burden during inference.

4.1 Biometric-Guided Prototype Alignment

In CC-ReID scenarios, extracting robust identity representations from drastically changing appearances is the core challenge. Ideally, diverse visual samples of the same pedestrian should be compactly distributed around an intrinsic identity center in the feature space [64]. Therefore, establishing a stable identity prototype by aggregating multiple visual features is a natural approach to align diverse instances. However, traditional unweighted updates [35] easily deviate from this ideal center. Since visual features cannot self-diagnose their identity purity, they are often overwhelmed by severe clothes and environmental biases, leading to semantic drift. To address this, we propose Biometric-Guided Prototype Alignment (BGPA), which leverages biometric text T_{bio} to adaptively evaluate feature reliability and accurately calibrate the identity prototype.

Specifically, given an input image x and its corresponding biometric description T_{bio}, we first employ an image encoder $\mathcal{E}_i$ and a text encoder $\mathcal{E}_t$ to extract the visual feature $f_v = \mathcal{E}_i(x)$ and the biometric text embedding $t_{bio} = \mathcal{E}_t(T_{bio})$,

respectively. For a training batch containing N samples, we first apply a grouping operation on the visual features based on their identity labels, yielding grouped features $f_v^{\mathcal{B}_k}$ for the k-th identity. Subsequently, we employ an MLP module as an Adapter to learn the semantic relationship between the visual and textual modalities and compute a Semantic Consistency Score α_i for each sample:

$$\alpha_i = \sigma(\text{Adapter}([f_v^i; t_{bio}^i])) \in (0, 1) \tag{1}$$

where $[\cdot;\cdot]$ denotes feature concatenation, and σ is the Sigmoid activation function. The score α_i reflects the degree of matching between the current visual feature and the biometric description, effectively indicating how "reliable" the model considers the image to be. Based on this score, we propose a Batch-based Adaptive Momentum Update strategy. Let $\mathcal{B}_k$ denote the set of indices belonging to identity k in the current batch. We utilize α_i to calculate the weighted feature center $\bar{f}_k$ for that identity:

$$\bar{f}_k = \frac{\sum_{i \in \mathcal{B}_k} \alpha_i \cdot f_v^i}{\sum_{i \in \mathcal{B}_k} \alpha_i} \tag{2}$$

Simultaneously, we calculate the average confidence of this batch of samples as $\bar{\alpha}_k = \frac{1}{|\mathcal{B}_k|} \sum_{i \in \mathcal{B}_k} \alpha_i$. To prevent low-quality samples or those dominated by strong regional styles from contaminating the global prototype, we dynamically adjust the step size of the momentum update. The update formula for the global identity prototype P_k is defined as:

$$\begin{aligned} \mu_k &= (1 - m) \cdot \bar{\alpha}_k \\ P_k^{(t)} &= (1 - \mu_k) \cdot P_k^{(t-1)} + \mu_k \cdot \bar{f}_k \end{aligned} \tag{3}$$

where $P_k^{(t)}$ and $P_k^{(t-1)}$ denote the identity prototype at the current and previous iteration steps, respectively, and m is the base momentum coefficient. This adaptive strategy dynamically calibrates the update weight μ_k based on the semantic reliability $\bar{\alpha}_k$. Specifically, for high-quality samples consistent with biometric descriptions ($\bar{\alpha}_k \to 1$), the update proceeds with the standard weight $(1-m)$; conversely, for noisy samples dominated by identity-irrelevant variations ($\bar{\alpha}_k \to 0$), μ_k vanishes to suppress prototype contamination.

Furthermore, to explicitly minimize the distance between visual features and the calibrated identity prototypes, we introduce a Prototype Contrastive Loss. This loss classifies the visual feature f_v to its corresponding prototype P_y:

$$\mathcal{L}_{BGPA} = -\log \frac{\exp(f_v \cdot P_y)}{\sum_{j=1}^{N_{id}} \exp(f_v \cdot P_j)} \tag{4}$$

where N_{id} is the total number of identities in a batch. By minimizing $\mathcal{L}_{BGPA}$, the model is forced to cluster samples tightly around the identity prototypes calibrated by biometric text in the feature space, thereby achieving intra-class compactness.

Finally, to establish a foundational alignment between the visual and textual modalities, we introduce an Image-Text Contrastive (ITC) loss. Before calculating the semantic consistency score, it is crucial to ensure that f_v and t_{bio} reside in a well-aligned joint embedding space. Therefore, we explicitly pull the visual feature and its corresponding biometric text embedding closer using the contrastive loss:

$$\mathcal{L}_{ITC} = -\log \frac{\exp(f_v \cdot t_{bio}^{+})}{\sum_{j=1}^{N} \exp(f_v \cdot t_{bio}^{j})} \tag{5}$$

where t_{bio}^{+} is the positive biometric text counterpart for f_v within the batch of size N.

4.2 Spectrum-Aware Orthogonal Projection

While BGPA ensures identity accuracy, existing methods often fail to generalize because they treat interfering factors in isolation, failing to construct a unified "identity-irrelevant semantic space" for collaborative elimination. This lack of a global perspective makes it difficult to characterize the distribution boundaries of non-identity features holistically.

To achieve true generalization, identity features must be disentangled at the subspace level. Instead of merely pushing features away from specific clothes instances, we conceptualize interference as specific semantic bases and force features to be orthogonal to the identity-irrelevant subspace.

To this end, we propose the Spectrum-Aware Orthogonal Projection (SAOP) module. Directly targeting this subspace-level decoupling, we leverage the clothes text T_{cloth} and environment text T_{env} provided in CGCC to model the comprehensive identity-irrelevant space. Specifically, we first extract their corresponding text embeddings $t_{cloth} = \mathcal{E}_t(T_{cloth})$ and $t_{env} = \mathcal{E}_t(T_{env})$. For a batch containing N samples, we collect these embeddings and concatenate them to form an identity-irrelevant matrix $M_{irr} = [t_{cloth}; t_{env}]$. Subsequently, we perform Singular Value Decomposition (SVD) on this matrix:

$$M_{irr} \xrightarrow{SVD} U \Sigma V^T \tag{6}$$

where the right singular vectors $V = \{v_1, v_2, \ldots, v_D\}$ span the identity-irrelevant Subspace, and D represents the dimensionality of the image and text embeddings. These vectors represent the principal directions of non-identity variations in the current data. To eliminate this non-identity information, we minimize the projection of the visual feature f_v onto this subspace. Considering that different basis vectors represent variations of different intensities measured by D singular values $\lambda_1, \ldots, \lambda_D$, we propose a Spectrum-Weighted orthogonal loss:

$$\mathcal{L}_{SAOP} = \sum_{j=1}^{D} \frac{\lambda_j}{\sum_{d=1}^{D} \lambda_d} \cdot \|f_v \cdot v_j\|^2 \tag{7}$$

where λ_j denotes the singular value corresponding to the j-th basis vector, quantifying the intensity of variation along that direction, and $\sum_{d=1}^{D} \lambda_d$ represents the sum of all D singular values, serving as a normalization factor to balance the orthogonality constraints across different basis vectors.

4.3 Optimization Objectives

The TGIR framework adopts an end-to-end joint training strategy. To extract robust discriminative identity features under clothes-changing scenarios, we define the overall optimization objective as:

$$\mathcal{L}_{total} = \mathcal{L}_{ID} + \mathcal{L}_{Triplet} + \mathcal{L}_{ITC} + \mathcal{L}_{BGPA} + \mathcal{L}_{SAOP} \tag{8}$$

where $\mathcal{L}_{ID}$ and $\mathcal{L}_{Triplet}$ denote the standard cross-entropy loss and triplet loss, respectively. Furthermore, $\mathcal{L}_{BGPA}$ represents the prototype contrastive loss formulated in the BGPA module, and $\mathcal{L}_{SAOP}$ designates the spectrum-weighted orthogonal loss applied in the SAOP module.

5 Experiment

5.1 Datasets and Evaluation Metrics

To comprehensively evaluate the proposed Text-Guided Identity Refinement (TGIR) framework and the constructed CGCC dataset, we conduct extensive experiments on four datasets: the small-scale LTCC [39] and PRCC [59], the spatiotemporally diverse LaST [42], the large-scale real-world DeepChange [57] (hereafter abbreviated as **DEEP**), and our proposed CGCC dataset. These experiments are designed to verify two core claims: first, that the multi-factor diversity of CGCC significantly enhances open-world generalization; and second, that TGIR effectively leverages disentangled textual supervision to extract robust, generalizable identity representations. Following standard person ReID protocols, we employ Rank-1, Rank-5 accuracy and mean Average Precision (mAP) as our primary evaluation metrics across all experiments.

5.2 Implementation Details

We adopt EVA02-CLIP-L [43] with a patch size of 14 as our visual backbone, and initialize the text encoder using the pre-trained weights of EVA02-CLIP-bigE-14 [43]. During the training phase, our models are optimized for a total of 60 epochs. Specifically, to accommodate the large scale of the CGCC dataset, the base learning rate is set to 2×10^{-5} with a batch size of 32. When training on the other comparative datasets, the learning rate is adjusted to 2×10^{-6} with a batch size of 8. For a fair comparison, all other comparative methods evaluated in our experiments strictly follow the hyperparameter settings and training strategies detailed in their respective original papers. All experiments are implemented using the PyTorch framework and accelerated by NVIDIA RTX 4090 GPUs.

Table 2: Cross-dataset generalization evaluation. "Source → Target" indicates training on the Source dataset and testing on the Target dataset.

Domains	TransReID [15]			CAL [11]			Eva02-CLIP [43]			MADE [37]		
	R1	R5	mAP	R1	R5	mAP	R1	R5	mAP	R1	R5	mAP
DEEP → LTCC	20.7	35.7	9.0	13.0	24.7	5.6	31.6	46.4	15.7	37.0	51.8	17.7
LaST → LTCC	12.2	25.3	7.1	8.4	18.4	5.2	26.8	47.7	13.0	35.5	51.5	17.2
CGCC → LTCC	**21.9**	**39.5**	**9.2**	**19.4**	**32.9**	**7.5**	**37.2**	**52.8**	**18.1**	**38.3**	**58.2**	**19.6**
DEEP → PRCC	33.1	38.7	31.0	28.8	34.3	25.7	37.2	43.6	36.7	43.0	50.7	41.8
LaST → PRCC	**37.1**	44.2	**32.9**	30.2	37.4	27.7	**55.5**	**66.4**	47.6	54.8	**67.2**	48.3
CGCC → PRCC	**37.1**	**44.3**	32.5	**35.2**	**42.9**	**31.0**	54.5	61.2	**50.6**	**60.3**	66.1	**55.9**
LTCC → DEEP	32.9	44.3	7.2	25.5	36.6	4.6	55.3	67.4	17.2	56.4	69.2	17.5
LaST → DEEP	42.3	53.8	10.8	28.9	38.3	6.2	42.6	57.0	10.7	51.4	64.8	14.7
CGCC → DEEP	**46.7**	**58.7**	**12.7**	**42.4**	**53.7**	**10.6**	**57.9**	**69.2**	**20.2**	**59.6**	**71.7**	**20.9**

Table 3: Cross-dataset generalization performance of different methods trained on our proposed CGCC dataset. "CGCC → Target" indicates models are trained on CGCC and evaluated on unseen Target datasets.

Method	CGCC → LTCC			CGCC → PRCC			CGCC → DEEP			Average		
	R1	R5	mAP	R1	R5	mAP	R1	R5	mAP	R1	R5	mAP
TransReID [15]	21.9	39.5	9.2	37.1	44.3	32.5	46.7	58.7	12.7	35.2	47.5	18.1
MADE [37]	38.3	**58.2**	**19.6**	60.3	66.1	55.9	59.6	71.7	20.9	52.7	65.3	32.1
Differ [28]	36.2	53.8	17.7	55.4	61.8	50.6	57.9	69.3	20.0	49.8	61.6	29.4
TGIR (Ours)	**40.3**	55.4	19.1	**62.0**	**69.5**	**56.9**	**62.1**	**72.7**	**22.4**	**54.8**	**65.9**	**32.8**

5.3 Cross-Dataset Generalization

To validate that CGCC's multi-factor diversity boosts open-world generalization, we conduct cross-dataset evaluations. Four representative architectures are trained on CGCC, LaST, and DEEP, and tested on three unseen datasets, LTCC, PRCC, DEEP(with LTCC replacing DEEP as the training source when evaluating on the DEEP target). As shown in Table 2, CGCC-trained models generally outperform those trained on existing datasets. For instance, MADE trained on CGCC achieves 59.6% Rank-1 on DEEP, outperforming the LaST-trained version by 8.2%. These results confirm that CGCC's rich variations effectively compel models to learn universally generalizable identity representations.

5.4 Comparison with State-of-the-Art Methods

To validate that our TGIR framework enhances representation generalizability while preserving competitive discriminability, we evaluate it in both cross-dataset and intra-dataset settings.

Superiority in Cross-Dataset Generalization. Evaluated across LTCC, PRCC, and DEEP when trained on CGCC (Table 3), TGIR achieves the best average performance across the three target datasets, with **54.8%** Rank-1, **65.9%** Rank-5, and **32.8%** mAP. Compared with the strongest competing results, TGIR improves the average Rank-1 and mAP by 2.1% and 0.7% over MADE,

Table 4: Comparison with State-of-the-Art Methods on LTCC. * denotes the reproduction results.

Method	Venue	LTCC			
		CC		General	
		R1	mAP	R1	mAP
TransReID [15]	CVPR'21	34.4	17.1	70.4	37.0
CAL [11]	CVPR'22	40.1	18.0	74.2	40.8
AIM [60]	CVPR'23	40.6	19.1	76.3	41.1
SCNet [12]	ACM'23	47.5	25.5	76.3	43.6
CVSL [33]	WACV'24	44.5	21.3	76.4	41.9
CLIP3DReID [31]	CVPR'24	42.1	21.7	-	-
CGPG [34]	CVPR'24	46.2	22.9	77.2	42.9
MADE [37]	TMM'24	47.4	24.4	84.2	48.2
Differ* [28]	CVPR'25	57.4	31.6	83.0	51.5
TGIR (Ours)	-	**57.9**	**32.7**	**84.5**	**53.6**

Table 5: Comparison with State-of-the-Art Methods on PRCC. * denotes the reproduction results.

Method	Venue	PRCC			
		CC		SC	
		R1	mAP	R1	mAP
TransReID [15]	CVPR'21	46.6	44.8	100	99.0
CAL [11]	CVPR'22	55.2	55.8	100	99.8
AIM [60]	CVPR'23	57.9	58.3	100	**99.9**
SCNet [12]	ACM'23	58.7	59.9	100	97.8
CVSL [33]	WACV'24	57.5	56.9	97.5	99.1
CLIP3DReID [31]	CVPR'24	60.6	59.3	-	-
CGPG [34]	CVPR'24	61.8	58.3	100	99.6
MADE [37]	TMM'24	64.3	59.1	100	98.6
Differ* [28]	CVPR'25	65.4	**62.6**	100	99.6
TGIR (Ours)	-	**66.3**	61.7	**100**	99.1

and by 5.0% and 3.4% over Differ, respectively. More importantly, TGIR consistently achieves the highest Rank-1 accuracy on all three unseen datasets, indicating that its advantage is not restricted to a specific target domain. These results demonstrate that, under the same CGCC training protocol, TGIR learns more generalizable identity representations than existing methods. This confirms that our unified subspace orthogonal projection effectively strips clothes and environmental biases, yielding highly generalizable representations.

Table 6: Comparison with State-of-the-Art Methods on our CGCC dataset.

Method	Venue	CGCC			
		CC		SC	
		R1	mAP	R1	mAP
TransReID [15]	ICCV'21	41.0	17.0	75.0	47.5
CAL [11]	CVPR'22	32.3	12.0	65.3	35.9
MADE [37]	CVPR'24	51.2	23.3	81.2	54.9
Differ [28]	CVPR'25	50.7	23.3	83.7	60.1
Eva02-CLIP [43]	ARXIV'24	54.7	26.3	85.0	62.0
TGIR (Ours)	-	**57.3**	**28.1**	**86.0**	**64.3**

Table 7: Ablation study of core components in TGIR.

Index	Components		CGCC → PRCC	
	BGPA	SAOP	R1	mAP
1 (Baseline)			54.5	50.6
2	✓		56.4	52.1
3		✓	59.0	55.3
4 (Full TGIR)	✓	✓	**62.0**	**56.9**

Superiority in Intra-Dataset Settings. TGIR also demonstrates competitive performance on standard intra-dataset evaluations across LTCC, PRCC, and CGCC (Tables 4, 5, 6). Notably, under the challenging CC setting on CGCC, TGIR reaches **57.3%** Rank-1 and **28.1%** mAP. This validates that our semantic disentanglement strategy prevents reliance on superficial visual cues, extracting intrinsic biometric features that remain stable.

5.5 Ablation Study

Table 7 details ablations on CGCC → PRCC. The baseline denotes an EVA02-CLIP model trained with only the standard ID loss and triplet loss, without using BGPA or SAOP, which achieves 54.5% Rank-1 and 50.6% mAP. Compared to the baseline, BGPA (Index 2, 56.4%) reduces source-domain overfitting by anchoring features to invariant biometric texts, while SAOP (Index 3, 59.0%) orthogonally projects out identity-irrelevant interference. Their integration achieves the optimal **62.0%** Rank-1 and **56.9%** mAP, confirming their synergy.

6 Limitations

Although CGCC and TGIR show strong generalization ability, several limitations remain. First, CGCC significantly enriches the diversity of clothes-changing scenarios, but it is still not intended to cover all real-world clothes-changing cases. Second, while the automatically generated textual descriptions provide fine-grained semantic supervision, they may still omit subtle visual details, especially for low-resolution or heavily occluded person images. Third, TGIR introduces additional training-stage complexity due to text encoding and SVD-based semantic subspace modeling, while no textual input or extra computation is required during inference.

7 Conclusion

In this paper, we address the lack of comprehensive data diversity and global semantic modeling in CC-ReID. To break these bottlenecks, we first introduce CGCC, a large-scale dataset featuring comprehensive diversity and fine-grained structured semantic annotations. Furthermore, we propose the TGIR framework, which utilizes the natural separability of textual semantics to achieve semantic subspace disentanglement; by applying SVD-based orthogonal projection, TGIR holistically strips away a unified identity-irrelevant subspace while preserving intrinsic biometric clues. By establishing a reinforcing closed loop where structured data and explicit decoupling supervision mutually empower each other, our approach ensures that rich variations are effectively translated into identity representations. Ultimately, our work provides a scalable and effective solution for practical, open-world clothes-changing person re-identification.

Acknowledgement. This work was supported by the National Natural Science Foundation of China under Grant Nos. 62376200 and T2541022, the Intelligent Computing Center of the National Cybersecurity Talent and Innovation Base, Wuhan, and the China Postdoctoral Science Foundation–Hubei Joint Support Program under Grant No. 2025T053HB.

References

1. Bai, S., Cai, Y., Chen, R., Chen, K., Chen, X., Cheng, Z., Deng, L., Ding, W., Gao, C., Ge, C., et al.: Qwen3-vl technical report. arXiv preprint arXiv:2511.21631 (2025)
2. Chan, P.P., Hu, X., Song, H., Peng, P., Chen, K.: Learning disentangled features for person re-identification under clothes changing. ACM Transactions on Multimedia Computing, Communications and Applications **19**(6), 1–21 (2023)
3. Chen, J., Jiang, X., Wang, F., Zhang, J., Zheng, F., Sun, X., Zheng, W.S.: Learning 3d shape feature for texture-insensitive person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8146–8155 (2021)
4. Chen, J., Zheng, W.S., Yang, Q., Meng, J., Hong, R., Tian, Q.: Deep shape-aware person re-identification for overcoming moderate clothing changes. IEEE Transactions on Multimedia **24**, 4285–4300 (2021)
5. Cui, Z., Zhou, J., Peng, Y., Zhang, S., Wang, Y.: 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)
6. Dang, Y., Huang, K., Huo, J., Yan, Y., Huang, S., Liu, D., Gao, M., Zhang, J., Qian, C., Wang, K., et al.: Explainable and interpretable multimodal large language models: A comprehensive survey. arXiv preprint arXiv:2412.02104 (2024)
7. Davila, D., Du, D., Lewis, B., Funk, C., Van Pelt, J., Collins, R., Corona, K., Brown, M., McCloskey, S., Hoogs, A., 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. pp. 1634–1643 (2023)
8. Gao, Z., Wei, H., Guan, W., Nie, J., Wang, M., Chen, S.: A semantic-aware attention and visual shielding network for cloth-changing person re-identification. IEEE Transactions on Neural Networks and Learning Systems **36**(1), 1243–1257 (2023)
9. Gong, S., Xiang, T.: Person re-identification. In: Visual Analysis of Behaviour: From Pixels to Semantics, pp. 301–313. Springer (2014)
10. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: European conference on computer vision. pp. 262–275. Springer (2008)
11. Gu, X., Chang, H., Ma, B., Bai, S., Shan, S., Chen, X.: Clothes-changing person re-identification with rgb modality only. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1060–1069 (2022)
12. Guo, P., Liu, H., Wu, J., Wang, G., Wang, T.: Semantic-aware consistency network for cloth-changing person re-identification. Proceedings of the 31st ACM International Conference on Multimedia pp. 8730–8739 (2023)
13. Han, K., Gong, S., Huang, Y., Wang, L., Tan, T.: Clothing-change feature augmentation for person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 22066–22075 (2023)
14. Han, X., Zhong, X., Huang, W., Jia, X., Yu, X., Kot, A.C.: See what you seek: Semantic contextual integration for cloth-changing person re-identification. arXiv preprint arXiv:2412.01345 (2024)
15. He, S., Luo, H., Wang, P., Wang, F., Li, H., Jiang, W.: Transreid: Transformer-based object re-identification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 15013–15022 (2021)
16. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017)

17. Hong, P., Wu, T., Wu, A., Han, X., Zheng, W.S.: 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. pp. 10513–10522 (2021)
18. Huang, Y., Wu, Q., Xu, J., Zhong, Y.: Celebrities-reid: A benchmark for clothes variation in long-term person re-identification. In: 2019 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2019)
19. Huang, Y., Wu, Q., Xu, J., Zhong, Y., Zhang, Z.: Clothing status awareness for long-term person re-identification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 11895–11904 (2021)
20. Huang, Y., Xu, J., Wu, Q., Zhong, Y., Zhang, P., Zhang, Z.: Beyond scalar neuron: Adopting vector-neuron capsules for long-term person re-identification. IEEE Transactions on Circuits and Systems for Video Technology **30**(10), 3459–3471 (2019)
21. Jin, X., He, T., Zheng, K., Yin, Z., Shen, X., Huang, Z., Feng, R., Huang, J., Chen, Z., Hua, X.S.: 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. pp. 14278–14287 (2022)
22. Li, P., Xu, Y., Wei, Y., Yang, Y.: Self-correction for human parsing. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(6), 3260–3271 (2020)
23. Li, S., Sun, L., Li, Q.: Clip-reid: exploiting vision-language model for image re-identification without concrete text labels. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 1405–1413 (2023)
24. Li, W., Hou, S., Zhang, C., Cao, C., Liu, X., Huang, Y., Zhao, Y.: An in-depth exploration of person re-identification and gait recognition in cloth-changing conditions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13824–13833 (2023)
25. Li, Y.J., Luo, Z., Weng, X., Kitani, K.M.: Learning shape representations for clothing variations in person re-identification. arXiv preprint arXiv:2003.07340 (2020)
26. Li, Y.J., Weng, X., Kitani, K.M.: Learning shape representations for person re-identification under clothing change. In: Proceedings of the IEEE/CVF Winter conference on applications of computer vision. pp. 2432–2441 (2021)
27. Liang, C.X., Tian, P., Yin, C.H., Yua, Y., An-Hou, W., Ming, L., Song, X., Wang, T., Bi, Z., Liu, M.: A comprehensive survey and guide to multimodal large language models in vision-language tasks. arXiv preprint arXiv:2411.06284 (2024)
28. Liang, X., Rawat, Y.S.: Differ: Disentangling identity features via semantic cues for clothes-changing person re-id. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 13980–13989 (2025)
29. Liu, F., Ye, M., Du, B.: Cloth-aware augmentation for cloth-generalized person re-identification. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 4053–4062 (2024)
30. Liu, F., Kim, M., Gu, Z., Jain, A., Liu, X.: 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. pp. 19617–19626 (2023)
31. Liu, F., Kim, M., Ren, Z., Liu, X.: 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. pp. 256–266 (2024)
32. Liu, M., Ma, Z., Li, T., Jiang, Y., Wang, K.: Long-term person re-identification with dramatic appearance change: Algorithm and benchmark. In: Proceedings of the 30th ACM international conference on multimedia. pp. 6406–6415 (2022)

33. Nguyen, V.D., Khaldi, K., Nguyen, D., Mantini, P., Shah, S.: Contrastive viewpoint-aware shape learning for long-term person re-identification. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) pp. 1030–1038 (2024)
34. Nguyen, V.D., Mantini, P., Shah, S.K.: Contrastive clothing and pose generation for cloth-changing person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7541–7549 (2024)
35. Ni, H., Li, Y., Gao, L., Shen, H.T., Song, J.: Part-aware transformer for generalizable person re-identification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 11280–11289 (2023)
36. Pan, W., Zhu, J., Cui, X., Zeng, H., Zhan, Y.: Focusing on pedestrians like human for clothes changing person re-identification. Neural Networks p. 107960 (2025)
37. Peng, C., Wang, B., Liu, D., Wang, N., Hu, R., Gao, X.: Masked attribute description embedding for cloth-changing person re-identification. IEEE Transactions on Multimedia (2024)
38. Prosser, B.J., Zheng, W.S., Gong, S., Xiang, T., Mary, Q., et al.: Person re-identification by support vector ranking. In: Bmvc. vol. 2, p. 6 (2010)
39. Qian, X., Wang, W., Zhang, L., Zhu, F., Fu, Y., Xiang, T., Jiang, Y.G., Xue, X.: Long-term cloth-changing person re-identification. In: Proceedings of the Asian conference on computer vision (2020)
40. Shi, W., Liu, H., Liu, M.: Iranet: Identity-relevance aware representation for cloth-changing person re-identification. Image and vision computing **117**, 104335 (2022)
41. Shu, X., Li, G., Wang, X., Ruan, W., Tian, Q.: Semantic-guided pixel sampling for cloth-changing person re-identification. IEEE Signal Processing Letters **28**, 1365–1369 (2021)
42. Shu, X., Wang, X., Zang, X., Zhang, S., Chen, Y., Li, G., Tian, Q.: Large-scale spatio-temporal person re-identification: Algorithms and benchmark. IEEE Transactions on Circuits and Systems for Video Technology **32**(7), 4390–4403 (2021)
43. Sun, Q., Fang, Y., Wu, L., Wang, X., Cao, Y.: Eva-clip: Improved training techniques for clip at scale. arXiv preprint arXiv:2303.15389 (2023)
44. Sun, W., Zhang, W., Jiang, Y., Wu, H., Zhang, Z., Jia, J., Zhou, Y., Ji, Z., Min, X., Lin, W., et al.: Dual-branch network for portrait image quality assessment. arXiv preprint arXiv:2405.08555 (2024)
45. Wan, F., Wu, Y., Qian, X., Chen, Y., Fu, Y.: When person re-identification meets changing clothes. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 830–831 (2020)
46. Wang, Q., Qian, X., Li, B., Chen, L., Fu, Y., Xue, X.: Content and salient semantics collaboration for cloth-changing person re-identification. In: ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2025)
47. Wang, Q., Qian, X., Li, B., Fu, Y., Xue, X.: Image-text-image knowledge transfer for lifelong person re-identification with hybrid clothing states. IEEE Transactions on Image Processing (2025)
48. Wang, Q., Qian, X., Li, B., Xue, X., Fu, Y.: Exploring fine-grained representation and recomposition for cloth-changing person re-identification. IEEE Transactions on Information Forensics and Security **19**, 6280–6292 (2024)
49. Wang, T., Ye, M.: Texfit: Text-driven fashion image editing with diffusion models. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 10198–10206 (2024)

50. Wang, Y., Yu, H., Yan, Y., Song, S., Liu, B., Lu, Y.: Exploring shape embedding for cloth-changing person re-identification via 2d-3d correspondences. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 7121–7130 (2023)
51. Wang, Z., Jiang, X., Xu, K., Sun, T.: A transformer-based cloth-irrelevant patches feature extracting method for long-term cloth-changing person re-identification. In: Computer Graphics International Conference. pp. 278–289. Springer (2022)
52. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer gan to bridge domain gap for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 79–88 (2018)
53. Wu, C., Li, J., Zhou, J., Lin, J., Gao, K., Yan, K., Yin, S.m., Bai, S., Xu, X., Chen, Y., et al.: Qwen-image technical report. arXiv preprint arXiv:2508.02324 (2025)
54. Wu, J., Liu, H., Shi, W., Tang, H., Guo, J.: Identity-sensitive knowledge propagation for cloth-changing person re-identification. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 1016–1020. IEEE (2022)
55. Xiao, T., Li, H., Ouyang, W., Wang, X.: Learning deep feature representations with domain guided dropout for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1249–1258 (2016)
56. Xiong, M., Yang, X., Chen, H., Aly, W.H.F., AlTameem, A., Saudagar, A.K.J., Mumtaz, S., Muhammad, K.: Cloth-changing person re-identification with invariant feature parsing for uavs applications. IEEE transactions on vehicular technology **73**(9), 12448–12457 (2024)
57. Xu, P., Zhu, X.: Deepchange: A long-term person re-identification benchmark with clothes change. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11196–11205 (2023)
58. Yan, Y., Yu, H., Li, S., Lu, Z., He, J., Zhang, H., Wang, R.: Weakening the influence of clothing: Universal clothing attribute disentanglement for person re-identification. In: IJCAI. pp. 1523–1529 (2022)
59. Yang, Q., Wu, A., Zheng, W.S.: Person re-identification by contour sketch under moderate clothing change. IEEE transactions on pattern analysis and machine intelligence **43**(6), 2029–2046 (2019)
60. Yang, Z., Lin, M., Zhong, X., Wu, Y., Wang, Z.: Good is bad: Causality inspired cloth-debiasing for cloth-changing person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1472–1481 (2023)
61. Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., Hoi, S.C.: Deep learning for person re-identification: A survey and outlook. IEEE transactions on pattern analysis and machine intelligence **44**(6), 2872–2893 (2021)
62. Yu, C., Liu, X., Zhu, J., Wang, Y., Zhang, P., Lu, H.: Climb-reid: A hybrid clip-mamba framework for person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 9589–9597 (2025)
63. Yu, S., Li, S., Chen, D., Zhao, R., Yan, J., Qiao, Y.: Cocas: A large-scale clothes changing person dataset for re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3400–3409 (2020)
64. Yuan, C., Zhang, G., Ma, C., Zhang, T., Niu, G.: From poses to identity: Training-free person re-identification via feature centralization. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 24409–24418 (2025)
65. Zhang, G., Liu, J., Chen, Y., Zheng, Y., Zhang, H.: Multi-biometric unified network for cloth-changing person re-identification. IEEE Transactions on Image Processing **32**, 4555–4566 (2023)

66. Zhang, L., Xiang, T., Gong, S.: Learning a discriminative null space for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1239–1248 (2016)
67. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3586–3593 (2013)
68. Zheng, L., Yang, Y., Hauptmann, A.G.: Person re-identification: Past, present and future. arXiv preprint arXiv:1610.02984 (2016)
69. Zheng, W.S., Gong, S., Xiang, T.: Reidentification by relative distance comparison. IEEE transactions on pattern analysis and machine intelligence **35**(3), 653–668 (2012)
70. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: Proceedings of the IEEE international conference on computer vision. pp. 3754–3762 (2017)