Clothing-Change Person Re-identification Based on Fusion of RGB Modality and Gait Features

Hongbin Tu
Chao Liu (1987456141@qq.com)
Yuanyuan Peng
Haibo Xiong
Haotian Wang

Research Article

Keywords: Clothing-Changing Peerson Re-identification, Gait Recognition, RGB Modality, Gait Prediction, Clothing Multi-Positive-Class Loss

Posted Date: October 17th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3440938/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Signal, Image and Video Processing on December 26th, 2023. See the published version at https://doi.org/10.1007/s11760-023-02913-4.

Clothing-Change Person Re-identification Based on Fusion of RGB Modality and Gait Features

Hongbin Tu [1], Chao Liu [1], Yuanyuan Peng [1], Haibo Xiong[1], Haotian Wang[1]

Abstract: Clothing-change person re-identification is an emerging research topic aimed at reidentifying individuals who have changed their clothing. This topic is highly challenging and has not received sufficient research attention to date. Most existing person re-identification methods primarily focus on clothing features, but in real-world scenarios where individuals change their attire, the identification accuracy of conventional person re-identification networks significantly decreases. The key challenge is how to effectively extract clothing-agnostic human features. This paper presents a method for clothing-change person re-identification that combines gait recognition and RGB modality. Initially, we use a ResNet50-based network for feature extraction from pedestrian images to obtain initial person features. We train the network using a clothing-aware loss function, which emphasizes features such as the face, body posture, or shoes that are less likely to change with different clothing. Additionally, we employ semantic segmentation to transform the original pedestrian images into gait energy maps. To address the limited gait information within a single image, we introduce a gait prediction module and use the GaitSet network to extract multi-frame pedestrian gait features. Finally, we fuse the extracted gait features with clothing-agnostic features to enhance the network's robustness. Experimental results demonstrate the effectiveness of this approach on a new dataset, showcasing improved robustness and higher accuracy compared to existing methods, thus confirming its superiority.

Keywords: Clothing-Changing Peerson Re-identification; Gait Recognition; RGB Modality; Gait Prediction; Clothing Multi-Positive-Class Loss

1. Introduction

When it comes to cross-camera person matching, challenges arise due to the widespread use of face recognition technology, which often struggles with profiles and

✉ Chao Liu

1987456141@qq.com

1. School of Electrical and Automation Engineering, East China Jiao Tong University, Nanchang City, Jiang xi Province, China

the back of the head. Additionally, limitations in camera resolution and shooting angles often result in lower-quality facial images, leading to ineffective matching outcomes. Consequently, cross-camera person matching frequently relies on Person Re-identification (Re-ID) technology [1]. Re-ID is the process of matching individuals across non-overlapping cameras, typically framed as an image ranking problem: given a query image of a person, it requires sorting all gallery images based on their similarity [2]. This technology finds extensive applications in intelligent video surveillance, security, and related fields [3].

With the rapid advancement of deep learning, the field of person re-identification has witnessed significant developments, with new research methods continuously emerging. Research in person re-identification primarily focuses on three main directions: first, through local feature extraction methods [4]; second, through distance learning methods [5,6]; and third, through deep learning methods [7-9]. In general, clothing information of the target person plays a significant role in existing person re-identification methods [10-12], dominating the local features. However, the performance of these algorithms significantly degrades when the target person changes clothing in query and gallery images [13,14], making it challenging to accurately match individuals who have changed their attire.

This highlights the heavy reliance of person re-identification methods on clothing information. In practical applications, the person in a query image may be wearing different clothing from those in gallery images. For instance, a suspect in a mall may change clothes to avoid being tracked after leaving the scene. Moreover, people change their attire every few days, so when query and gallery images are taken on different days, the clothing information of the target person may differ. This has prompted researchers to explore the domain of Clothing-Changing Person Re-identification, also known as long-term person re-identification.

Clothing-Changing Person Re-identification is considered a type of image classification task, where class distinctions are typically determined by the large-scale structure of images. Without intervention, networks tend to make predictions based on localized or smaller features, which may not be conducive to the classification task. In scenarios where people change clothing, unreliable external appearance features like clothing need to be disregarded, while features like gait remain consistent across attire changes. Features such as stride length, step frequency, and gait rhythm remain relatively stable and can be considered invariant features.

Based on these considerations, we propose a research method for Clothing-Changing Person Re-identification based on the RGB modality and gait features. This method utilizes two branches: one branch employs ResNet50 as the feature extraction network and modifies the loss function to encourage the network to focus on clothing-irrelevant and accurate person features, reducing the network's reliance on clothing features. The other branch extracts the person's gait features using the GaitSet network. To enrich single-image gait information, we introduce a gait sequence prediction module that takes gait energy maps as input and outputs images for the preceding and subsequent frames. Finally, the features obtained from both branches are fused to achieve effective Clothing-Changing Person Re-identification results.

2. Related Work

With the widespread adoption of surveillance systems in real-life scenarios, pedestrian re-identification (Re-ID) tasks have garnered increasing attention. As previously mentioned, nearly all existing Re-ID datasets are collected over short durations, resulting in relatively consistent clothing appearances for the same individual. In the era of deep learning, there has been significant progress in developing automatic pedestrian Re-ID methods through discriminative feature learning and distance metrics. These models exhibit robustness against variations induced by pose, lighting, and viewpoint changes, as these conditions are prevalent in these datasets. However, they are susceptible to changes in clothing since they heavily rely on the consistency of clothing appearances. Below, we'll discuss this field in the context of general pedestrian Re-ID and clothing-changing pedestrian Re-ID separately.

General Pedestrian Re-Identification:

General pedestrian Re-ID focuses on scenarios with relatively ideal conditions, considering minimal variations in lighting, pose, and viewpoint. It emphasizes feature representation and similarity measurement of individuals. A standard algorithmic workflow includes capturing raw video data from surveillance systems, extracting frames to obtain a series of scene images, using pedestrian detection algorithms to locate and crop pedestrians in the scenes, extracting robust feature representations of pedestrians through representation learning methods, computing similarity scores between pedestrians using metric learning methods, and sorting them in descending order. The target is to re-identify the desired pedestrian based on the ranking results. Furthermore, performance can be further optimized through re-ranking algorithms based on the initial sorting.

In [15], a deep model called AlignedReID++ was introduced to address the issue of global features being unable to handle misaligned images. It significantly improved global feature extraction and achieved human-level recognition rates. [16] designed a joint learning framework consisting of single-image representation and cross-image representation, trained with triplet loss. [17] proposed the Pose-Guided Feature Learning with Knowledge Distillation (PGFL-KD) network aimed at matching person images with occlusions. [18] addressed the occlusion problem with an Incremental Generation of Occlusion-Adversarial Suppression (IGOAS) network. [19] introduced the Texture Semantic Alignment (TSA) method, which attempted to solve occlusion issues while considering pose and viewpoint changes.

Clothing-Changing Pedestrian Re-Identification:

In contrast to general pedestrian Re-ID, clothing-changing scenarios introduce a new challenge, where traditional recognition methods struggle to perform well. The core problem in clothing-changing pedestrian Re-ID is extracting clothing-agnostic features. Gu et al. [20] introduced the Clothing Adversarial Loss (CAL) based on penalizing the model's ability to predict clothing, thereby extracting unrelated clothing features from raw RGB images. Wang et al. [21] designed a novel Shape Semantic Embedding (SSE) module to encode human body shape semantic information, which

is an important clue in recognizing pedestrian disguise. They proposed a Collaborative Attention Mutual Cross-Attention (CAMC) framework. Yu et al. [22] presented a semi-supervised approach called Attire Invariant Feature Learning (AIFL) to learn clothing-invariant pedestrian representations and an unsupervised Attire Simulation GAN (AS-GAN) to synthesize clothing variations. Chen et al. [13] introduced a new Re-ID learning framework called 3D Shape Learning (3DSL), which directly extracts texture-insensitive 3D shape embeddings from 2D images with the addition of 3D body reconstruction as an auxiliary task. Hong [14] proposed a Fine-grained Shape-Appearance Mutual Learning framework (FSAM), a dual-stream framework that learns fine-grained body shape knowledge from the shape stream and transfers it to the appearance stream to complement clothing-agnostic knowledge. Qian et al. [23] introduced a Shape Embedding Module and Clothing Elimination Shape Distillation Module to eliminate unreliable clothing appearance features and focus on body shape information. Gu et al. [13] presented Appearance-Preserving 3D Convolution (AP3D), comprising an Appearance-Preserving Module (APM) and 3D convolutional kernels, which align adjacent feature maps at the pixel level using APM and model temporal information while preserving appearance representation quality. Wu et al. [24] introduced an effective Identity-Sensitive Knowledge Propagation framework (DeSKPro) that incorporates clothing-agnostic spatial attention modules and leverages knowledge from a human parsing module to eliminate clothing appearance interference.

In summary, the aforementioned pedestrian Re-ID methods mostly rely on extracting appearance features. However, the performance of appearance-based person re-identification significantly drops when dealing with similar individuals or the same individual in different outfits, such as clothing and accessories.

3. Method

In the scenario of pedestrians changing clothes, the reliability of clothing and other appearance features becomes compromised, which hinders the implementation of pedestrian re-identification tasks. Therefore, our goal is to enable the network to learn to extract features unrelated to clothing or related to soft biometrics (such as body shape, gait), focusing on clothing-agnostic features and gait features in this work. Learning pedestrian identity features directly from images is a challenging task in itself, and considering the success of many recent papers in extracting and parsing human pose and gait features from raw RGB images, we propose a dual-branch network structure based on the fusion of RGB modality and gait features. The main branch network utilizes the original RGB images to extract identity recognition information while reducing the interference of irrelevant clothing information to enhance network robustness. Another branch employs semantic segmentation to output gait energy maps from RGB images and extract their features. Finally, the extracted gait features are fused as auxiliary features with clothing-agnostic features, and the network is trained with a loss function. The specific structure is shown in the diagram below (two branches).

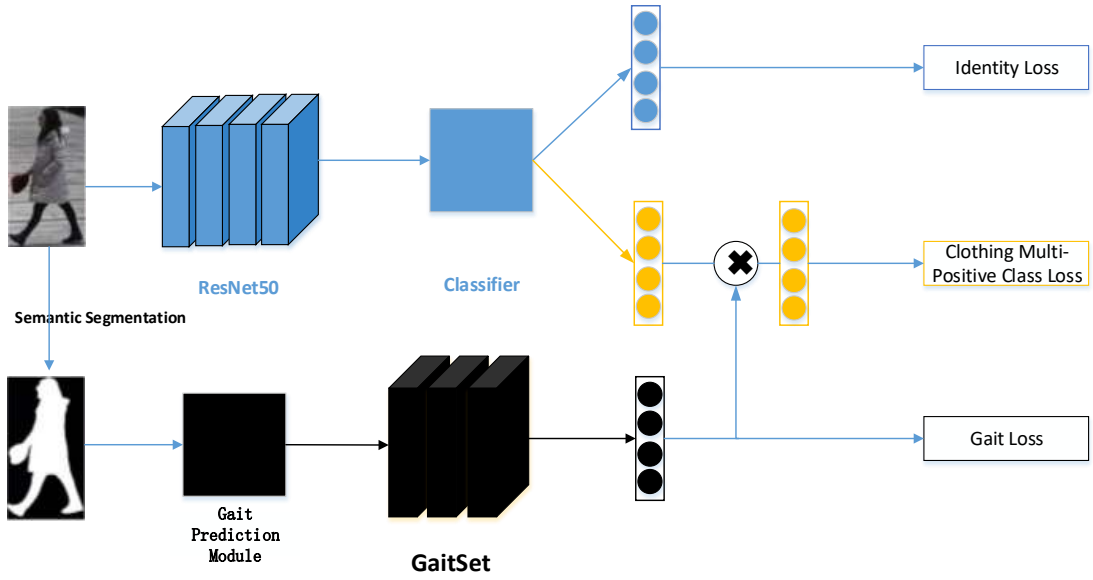

Figure 1: Network Architecture Diagram in this Paper

3.1 Clothing-agnostic Feature Extraction

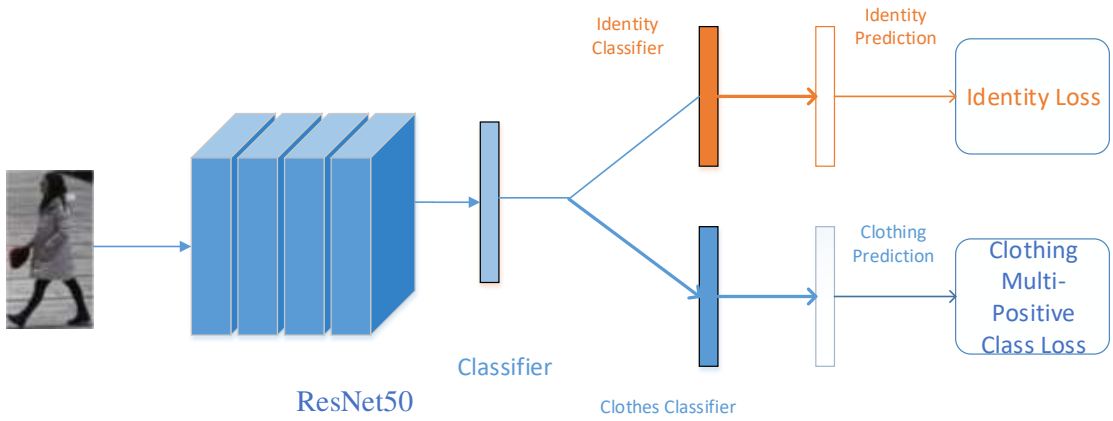

Figure 2: Clothing-Agnostic Feature Extraction Diagram

To fully leverage the pedestrian features carried by the original images, in this paper, we employ ResNet50 as the backbone network while removing the final downsampling layer to retain more complete and accurate feature information. Following this, we introduce a classifier to reduce the interference of certain local information in making predictions, aiming to better mimic the robustness exhibited by humans in the field of visual recognition. Specifically, as follows:

The N-tuple loss corresponds to a multi-class classification problem and optimizes the distance from instance to instance. In this context, N-1 instances act as reference nodes, and one instance acts as the query. By jointly optimizing multiple instances, it is possible to improve the relative order of distances between multiple instances, thus better matching the ReID (pedestrian re-identification) inference task. Therefore, the N-tuple loss will be employed to constrain the network in the clothing classifier.

To train the clothing classifier, we optimize it by minimizing the clothing

classification loss L_N (which is the N-tuple loss between the predicted clothing $F_\alpha^F\left(g_\theta(m_i)\right)$ and the clothing label x_i^C. This process can be represented as follows:

$$\min_{\alpha} L_N\left(F_\alpha^F\left(g_\theta(m_i)\right), x_i^F\right) \tag{1}$$

$$L_N = -log\frac{\exp\left(\frac{1}{\tau}S_a^+\right)}{\exp\left(\frac{1}{\tau}S_a^+ + \sum_{k=1}^{N-2}\exp\left(\frac{1}{\tau}S_{a,k}^-\right)\right)} \tag{2}$$

$$S_a^+ = S(x_a, x^+) \tag{3}$$

$$S_{a,k}^- = S(x_a, x_k^-) \tag{4}$$

Where N represents the number of clothing categories, including one anchor sample, one positive sample, and $N-2$ negative samples. Here, $x-k(k=1,\cdots,N-2)$corresponds to $N-2$ different classes of negative samples, i.e., different pedestrians, and x^+ corresponds to a positive sample, which belongs to the same class as the query sample.

Extracting clothing-agnostic features: We fix the parameters of the clothing classifier and enforce the backbone to learn features that are unrelated to clothing. To do this, we should penalize the re-identification model with respect to its predictive capability regarding clothing. The loss function for this is as follows:

$$L_D = \lambda_1 L_C + \lambda_2 L_{tri} + \lambda_3 L_{cos} \tag{5}$$

Note that L_D and L_C have some similarities in learning clothing-agnostic features. When we train using only L_D, the model tends to initially learn simple samples (same clothing) and then gradually learns to differentiate difficult samples (same identity, different clothing). This is consistent with curriculum learning. The goal of L_C is to bring features with the same identity closer together, which is similar to L_D. However, we do not discard L_D. The reason for this is that minimizing only L_C and forcing the model to differentiate hard samples in the early stages of optimization could potentially lead to local optima. Instead, in our experiments, we introduced L_C for training after the first learning rate reduction.

3.2 Gait Feature Extraction

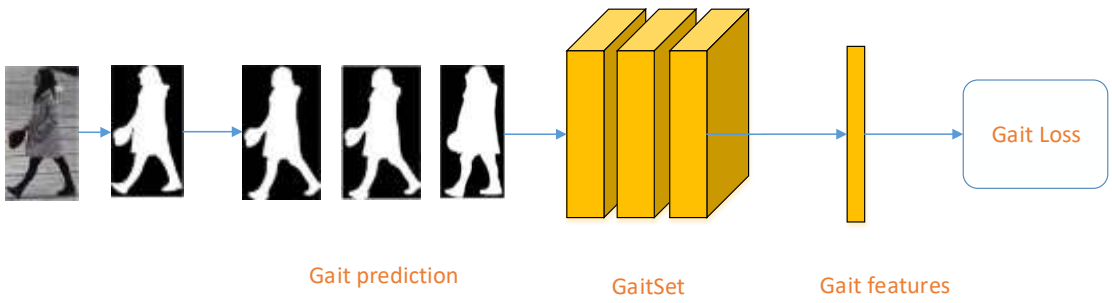

Figure 3: Gait Feature Extraction Diagram

When encountering a friend wearing new clothes on the street, people can easily recognize their identity because a person's biometric characteristics do not change with

clothing alterations. Alternatively, individuals may pause to observe their walking posture to determine if it matches their familiar friend's gait. This is one of the inspirations behind the idea in this paper. To enrich the gait information from a single image, a gait prediction module is employed, obtaining multiple frames of images, and the GaitSet network is utilized to extract gait features. This is explained in more detail in the following text.

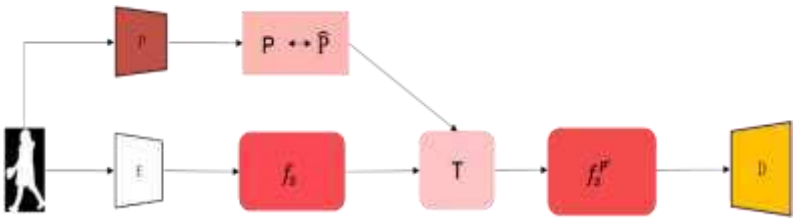

Figure 4: Gait Prediction Module

Encoder: Given a silhouette input S, the objective of the encoder E is to extract a compact feature with reduced dimensions.

$$f_s = E(s) \tag{6}$$

Position Embedder and Feature Aggregator: Taking into account the uncertainty in predictions, we introduce the mid-frame input principle, which assumes that the input silhouette always corresponds to the mid-frame of the predicted gait sequence. In GPM training, we take the gait frame at the middle position of the gait sequence as the input to GPM and use a one-dimensional vector $p_{mid} \epsilon \mathbb{R}$ to represent the position label for that frame.

$$\dot{p} = P(S) \epsilon \mathbb{R}^2 \tag{7}$$

$$L_p = \| \dot{p} - p_{mid} \|_2^2 \tag{8}$$

In this process, we compare the embedded position output ṗ with the ground truth to construct the position loss L_p. P represents a mapping between the input and the mid-position.

A feature aggregator A, implemented as a fully connected layer inserted between the encoder and the decoder, transforms the original encoded feature f_s into mid-position-aware feature $f_s^{\dot{p}}$. Considering the mid-position information $\dot{p}$ embedded in the next decoder, this explicitly informs the decoder that we need to predict the gait states before and after the current input's mid-gait state, thus reducing the prediction ambiguity in the results. This feature aggregation process is expressed as follows:

$$f_s^{\dot{p}} = A([f_s, \dot{p}]) \tag{9}$$

In gait analysis, the visual relationships between silhouette profiles in consecutive frames make it relatively easy to recognize the temporal patterns. Therefore, this paper no longer explicitly models the temporal relationships of gait silhouette profiles. Instead, it treats gait silhouette profiles as a collection of images without temporal ordering, allowing deep neural networks to optimize and utilize these relationships. This forms the basis of the GaitSet algorithm.

Using a set of pedestrian silhouette images as input implies that there's no requirement for aligning or ordering the input image sequence. In some traditional gait recognition algorithms, when comparing the similarity between two gait sequences, alignment operations are needed, such as adjusting both video sequences to start and end at the moment when the left foot is lifted and returned, completing one gait cycle. Using a set as input eliminates these preprocessing steps, making it more applicable to real-world scenarios.

4. Experiments

4.1Datasets

To validate the effectiveness of the proposed method in this paper, training and experiments were conducted on four different datasets: LTCC, PRCC, NKUP, and Celeb-ReID. Here is a description of these datasets:

LTCC Dataset: The LTCC dataset consists of 17,138 images, covering 152 unique identities and 478 different clothing variations, captured from 12 different camera viewpoints. This dataset is designed for person detection and long-term clothing-changing person re-identification. It has meticulous manual labeling of both individual identities and clothing identities.

Celeb-ReID Dataset: The Celeb-ReID dataset is divided into three subsets: training set, gallery set, and query set. This dataset exhibits significant clothing variations within the same person's identity. Specifically, over 70% of images for each person display different clothing. It comprises 1,052 unique person identities with a total of 34,186 images. The training set includes 632 person identities with 20,208 images, while the test set involves 420 person identities with 13,978 images. The test set further comprises 2,972 query images and 11,006 gallery images.

PRCC Dataset: The PRCC dataset contains 221 distinct pedestrian identities, encompassing a total of 33,698 pedestrian bounding box images. These images were captured from 3 different camera angles. Camera A and B have individuals with identical clothing, whereas Camera C features the same individuals wearing different clothing. The dataset is divided into training, testing, and query sets. The training set consists of 150 person identities with 17,896 images, the testing set includes 71 person identities with 3,384 images, and the query set has 71 person identities with 3,543 images for scenarios with the same clothing and 71 person identities with 3,873 images for clothing-changing scenarios.

These datasets were used to evaluate the performance of the proposed method in different re-identification scenarios.

4.2 Experimental Setup

The experiments in this paper were conducted in an Ubuntu environment with a

64-bit operating system. Python was the programming language used for implementing the algorithm models, with the primary deep learning framework being PyTorch. Additionally, CUDA 10.6 was employed for GPU acceleration. Due to the large dataset size and extended training times, two GPUs were utilized for image feature extraction. This approach not only improved computation speed but also resulted in higher output accuracy. The training process involved 100 epochs, with a total training time of h hours. The deep Re-ID model was implemented based on the PyTorch framework. A ResNet-50 architecture was used as the backbone, initialized with pre-trained parameters from ImageNet. To conserve GPU memory, the first two residual layers were fixed. Input image dimensions were resized to 384×192. Data augmentation techniques, including random flipping, random cropping, and random erasing, were employed during training, with a batch size set to 64. The base layers of ResNet-50 were trained with a learning rate of 3.5e-4 for the first 40 epochs, followed by a learning rate of 1e-4 for the next 60 epochs, with a learning rate reduction by a factor of 10. SGD optimization was used for training, and mini-batch sizes of 128 were applied for both the source and target images.

4.3Results and Analysis

The experiment validates our approach on three clothing-changing pedestrian re-identification datasets (Celeb-reID, LTCC, PRCC), demonstrating the effectiveness of the proposed method. Specifically, as shown in Table 1 below, our method achieves outstanding performance in terms of metrics such as mAP and top-1.

Table 1: Performance Data of Various Methods on Different Datasets

Method	Celeb-reID		LTCC		PRCC	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
PCB[12]	37.1	8.2	23.5	10.0	41.8	38.7
IANet[25]	36.2	9.6	25.0	12.6	46.3	45.9
CAL[20]	53.2	10.3	40.1	18.0	55.2	55.8
3DSL[13]	-	-	**51.3**	-	31.2	14.8
FSAM[14]	-	-	38.5	16.2	54.4	-
GI-ReID[26]	-	-	23.7	10.4	33.3	-
Ours	**55.6**	**13.4**	49.9	**19.3**	**58.6**	**57.5**

In the table above, we conducted experiments comparing two general pedestrian re-identification methods (PCB and IANet) and four clothing-changing pedestrian re-

identification methods across three different clothing-changing datasets. It is evident that clothing-changing methods outperform general pedestrian re-identification methods on clothing-changing datasets, while their recognition performance remains reasonably high. The reason for this is that not all images in the clothing-changing dataset depict individuals who have changed their clothing. Clothing-changing methods achieve better recognition results because the network architecture was originally designed to handle cases where the same person wears different clothing and different people wear similar clothing.

From the table, it can be observed that the fusion of RGB and gait features in the Celeb-reID dataset leads to an improvement in performance. Compared to the CAL method, there is a 2.4% improvement in Rank-1 and a 3.1% improvement in mAP. On the PRCC dataset, the Rank-1 accuracy reaches 58.6%, significantly higher than other methods, and there is also a substantial improvement in mAP. This improvement is attributed to the clothing multi-positive class loss proposed in this paper, which penalizes the network, enabling it to extract clothing-agnostic features. To further enhance the method's robustness, pedestrian gait features were introduced, contributing to improved recognition performance.

4.4Ablation Studies

The baseline used in this paper is a model that only extracts RGB images. To explore the robustness of our model, we conducted ablation experiments on the Celeb-reID, LTCC, and PRCC clothing-changing pedestrian re-identification datasets. Table 2 below presents the performance of the network structures with different modules added on different datasets.

Table 2: Performance of Network Structures with Different Modules Added on Different Datasets

Method	Celeb-reID		PRCC	
	Rank-1	mAP	Rank-1	mAP
Baseline	35.2	6.5	36.5	33.7
+CML	48.4	7.8	45.9	40.4
+GF	43.6	10.6	50.1	44.6
+GF+CML	52.7	11.2	56.9	45.6
+Ours	**55.6**	**13.4**	**58.6**	**57.5**

As shown in the table, we conducted experiments by adding the Clothing Multi-Positive Class Loss (CML), Gait Features (GF), and their joint combination to the

baseline model to demonstrate the advancements of the RGB and Gait dual-branch structure. The baseline structure includes only the ResNet50 backbone network trained with the common triplet loss function. When each module is added individually, it outperforms the methods in terms of Rank-1 and mAP metrics. Notably, the significant improvement in Rank-1 data from the addition of the Clothing Multi-Positive Class Loss indicates its strong model performance. Additionally, the addition of Gait Features alone results in a 4.1% higher mAP compared to the baseline, illustrating the value of pedestrian gait features, which remain consistent despite clothing changes. Our approach further enhances model recognition performance by incorporating the N-tuple loss on top of the Clothing Multi-Positive Class Loss and Gait Features.

Table 3: Performance of Different Methods on Datasets with Invisible Faces

Method	NKUP	
	Rank-1	mAP
Baseline	9.0	6.7
PCB[12]	16.9	12.4
MGN[27]	18.8	15.0
LSD[28]	19.7	15.6
Ours	**23.9**	**17.3**

To further validate the robustness of our model, we introduced an additional dataset (NKUP) where pedestrians' faces are not visible, which places higher demands on the network model. As shown in the table above, the performance of our model significantly improves when compared to two general pedestrian re-identification methods. When compared to the long-duration method LGD, our model also demonstrates advantages, achieving the best performance with a 4.2% improvement in Rank-1 and a 1.7% improvement in mAP. This indicates that even in scenarios with obscured faces, our model can still effectively perform re-identification tasks using the RGB and gait modalities.

4.5Visualization Analysis

Figure 5: Heatmaps Comparing the Baseline Model and Our Proposed Method

To visually demonstrate the effectiveness of our proposed gait-based comprehensive pedestrian feature acquisition method in clothing-changing scenarios, we conducted a visual comparison between the baseline method and our proposed method on the PRCC dataset. The baseline method utilized the ResNet-50 backbone network, inserted a BN layer between the final global pooling layer and the classification layer, and was trained on a pedestrian image dataset using cross-entropy loss and triplet loss, following the same training strategy as our proposed method. The visualization results are shown in the heatmap below, where darker colors indicate areas where the model pays more attention.

5. Conclusion

In scenarios involving clothing changes, the accuracy of general pedestrian re-identification methods significantly decreases. To extract clothing-agnostic features, this paper proposed leveraging the unique human gait as an auxiliary solution to address the clothing variation challenge in pedestrian re-identification. In summary, we introduced a novel dual-branch network framework that fuses gait with RGB modality, treating gait as an auxiliary pedestrian feature, and combined it with a network trained with clothing multi-positive class loss to accomplish re-identification tasks. Extensive experiments on multiple datasets demonstrated that our approach outperforms baseline models across multiple metrics and is more advanced than current methods.

Acknowledgements This work was supported by the Jiangxi Provincial Natural Science Foundation(20224BAB202024) and the Jiangxi Provincial Natural Science Foundation(20212BAB202007). The authors sincerely express thanks to the reviewers for taking the time to review the paper in a busy schedule.

Author contributions Tu and Peng prepared data of research, Liu wrote the main manuscript text and conducted experiments, and Xiong and Wang help revise the paper.

Data availability The data used to support the findings of this study are available from the corresponding author upon request.

Declarations

Conflict of interest The authors declare that there are no conflicts of interest regarding the publication of this paper.

[1] Luo, H., Jiang, W., Fan, X., & Zhang, S. P. (2019). Research Progress on Pedestrian Re-identification Based on Deep Learning. Acta Automatica Sinica, 45(11), 2032-2049.

[2] Zahra, A., Perwaiz, N., Shahzad, M., & Fraz, M. M. (2023). Person re-identification: A retrospective on domain specific open challenges and future trends. *Pattern Recognition*, 109669.

[3] Avola, D., Cascio, M., Cinque, L., Fagioli, A., & Petrioli, C. (2022). Person re-identification through Wi-Fi extracted radio biometric signatures. *IEEE Transactions on Information Forensics and Security*, *17*, 1145-1158.

[4] Wang, K., Dong, S., Liu, N., Yang, J., Li, T., & Hu, Q. (2021). PA-Net: Learning local features using by pose attention for short-term person re-identification. *Information Sciences*, *565*, 196-209.

[5] Yu, H. X., Wu, A., & Zheng, W. S. (2018). Unsupervised person re-identification by deep asymmetric metric embedding. *IEEE transactions on pattern analysis and machine intelligence*, *42*(4), 956-973.

[6] Sun, B., Ren, Y., & Lu, X. (2020). Semisupervised consistent projection metric learning for person reidentification. *IEEE Transactions on Cybernetics*, *52*(2), 738-747.

[7] Zheng, Z., Zheng, L., & Yang, Y. (2017). 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).

[8] Li, W., Zhao, R., Xiao, T., & Wang, X. (2014). Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 152-159).

[9] Wei, L., Zhang, S., Gao, W., & Tian, Q. (2018). 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).

[10] Gu, X., Chang, H., Ma, B., Zhang, H., & Chen, X. (2020). Appearance-preserving 3d convolution for video-based person re-identification. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16* (pp. 228-243). Springer International Publishing.

[11] Hou, R., Chang, H., Ma, B., Shan, S., & Chen, X. (2020). Temporal complementary learning for video person re-identification. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16* (pp. 388-405). Springer International Publishing.

[12] Sun, Y., Zheng, L., Yang, Y., Tian, Q., & Wang, S. (2018). Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *Proceedings of the European conference on computer vision (ECCV)* (pp. 480-496).

[13] Chen, J., Jiang, X., Wang, F., Zhang, J., Zheng, F., Sun, X., & Zheng, W. S. (2021). 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).

[14] Hong, P., Wu, T., Wu, A., Han, X., & Zheng, W. S. (2021). 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).

[15] Luo, H., Jiang, W., Zhang, X., Fan, X., Qian, J., & Zhang, C. (2019). Alignedreid++: Dynamically matching local information for person re-identification. *Pattern Recognition*, *94*, 53-61.

[16] Wang, F., Zuo, W., Lin, L., Zhang, D., & Zhang, L. (2016). Joint learning of single-image and cross-image representations for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1288-1296).

[17] Zheng, K., Lan, C., Zeng, W., Liu, J., Zhang, Z., & Zha, Z. J. (2021, October). Pose-guided feature learning with knowledge distillation for occluded person re-identification. In *Proceedings of the 29th ACM International Conference on Multimedia* (pp. 4537-4545).

[18] Zhao, C., Lv, X., Dou, S., Zhang, S., Wu, J., & Wang, L. (2021). Incremental generative occlusion adversarial suppression network for person ReID. *IEEE Transactions on Image Processing*, *30*, 4212-4224.

[19] Gao, L., Zhang, H., Gao, Z., Guan, W., Cheng, Z., & Wang, M. (2020, October). Texture semantically aligned with visibility-aware for partial person re-identification. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 3771-3779).

[20] Gu, X., Chang, H., Ma, B., Bai, S., Shan, S., & Chen, X. (2022). 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).

[21] Wang, Q., Qian, X., Fu, Y., & Xue, X. (2022). Co-attention Aligned Mutual Cross-Attention for Cloth-Changing Person Re-identification. In *Proceedings of the Asian Conference on Computer Vision* (pp. 2270-2288).

[22] Yu, Z., Zhao, Y., Hong, B., Jin, Z., Huang, J., Cai, D., & Hua, X. S. (2021). Apparel-invariant feature learning for person re-identification. *IEEE Transactions on Multimedia*, *24*, 4482-4492.

[23] Qian, X., Wang, W., Zhang, L., Zhu, F., Fu, Y., Xiang, T., ... & Xue, X. (2020). Long-term cloth-changing person re-identification. In *Proceedings of the Asian Conference on Computer Vision*.

[24] Wu, J., Liu, H., Shi, W., Tang, H., & Guo, J. (2022, October). Identity-Sensitive Knowledge Propagation for Cloth-Changing Person Re-Identification. In *2022 IEEE International Conference on Image Processing (ICIP)* (pp. 1016-1020). IEEE.

[25] Hou, R., Ma, B., Chang, H., Gu, X., Shan, S., & Chen, X. (2019). Interaction-and-aggregation network for person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9317-9326). (ianet)

[26] Jin, X., He, T., Zheng, K., Yin, Z., Shen, X., Huang, Z., ... & Hua, X. S. (2022). 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).

[27] Wang, G., Yuan, Y., Chen, X., Li, J., & Zhou, X. (2018, October). Learning discriminative features with multiple granularities for person re-identification. In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 274-282).

[28] Yaghoubi, E., Borza, D., Degardin, B., & Proença, H. (2021). You look so different! Haven't I seen you a long time ago?. *Image and Vision Computing*, *115*, 104288.