

OC⁴-ReID: Occluded Cloth-Changing Person Re-Identification

Zhihao Chen¹ Yiyuan Ge^{2*} Yanyan Lv³ Ziyang Wang^{4*} Mingya Zhang⁵

¹ Beijing University of Posts and Telecommunications ² South China University of Technology

³ Nanjing University of Information Science & Technology ⁴ University of Oxford ⁵ Nanjing University

zhihaochen666@bupt.edu.cn 202510182871@mail.scut.edu.cn yyly@nuist.edu.cn

ziyang.wang17@gmail.com dg20330034@smail.nju.edu.cn

Abstract—The study of Cloth-Changing Person Re-identification (CC-ReID) focuses on retrieving specific pedestrians when their clothing has changed, typically under the assumption that the entire pedestrian images are visible. Pedestrian images in real-world scenarios, however, are often partially obscured by obstacles, presenting a significant challenge to existing CC-ReID systems. In this paper, we introduce a more challenging task termed Occluded Cloth-Changing Person Re-identification (OC⁴-ReID), which simultaneously addresses two challenges of clothing changes and occlusions. Concretely, we construct two new datasets, Occ-LTCC and Occ-PRCC, based on original CC-ReID datasets to include random occlusions of key pedestrians components (e.g., head, torso). Moreover, a novel benchmark is proposed for OC⁴-ReID incorporating a Train-Test Micro Granularity Screening (T²MGS) module to mitigate the influence of occlusions and proposing a Part-Robust Triplet (PRT) loss for partial features learning. Comprehensive experiments on the proposed datasets, as well as on two CC-ReID benchmark datasets demonstrate the superior performance of proposed method against other state-of-the-art methods. The codes and datasets are available at: <https://github.com/1024AILab/OC4-ReID>.

Index Terms—Person Re-identification, Cloth-Changing, Occlusion.

I. INTRODUCTION

Cloth-Changing Person Re-identification (CC-ReID) aims to retrieve target pedestrians across instances where their clothing has changed [1, 2, 3]. To mitigate clothing bias, existing methods utilize deep neural networks to learn features that are less dependent on clothing, often assuming that the pedestrian images are minimally occluded [1, 4, 5, 6, 7]. In real-world application scenarios, pedestrian images captured by practical cameras, however, are often occluded. This occlusion reduces available feature information and can cause the model to make incorrect matches.

Fig. 1 illustrates the evaluation of classical CC-ReID architectures [8, 9] on both the standard CC-ReID dataset and our newly constructed OC⁴-ReID dataset, where parts of the pedestrian’s body are randomly occluded. Our experimental results indicate that occlusions into the CC-ReID dataset significantly impairs the model’s performance, with CAL [8] and AIM [9] reducing in Rank-1 by 9.87% and 10.80%, and in mAP by 12.62% and 13.32%, respectively.

Given the significant impact of occlusions on CC-ReID, we introduce a novel and more challenging task, termed **Occluded**

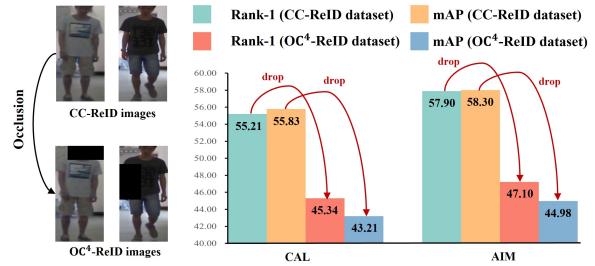


Fig. 1. We construct OC⁴-ReID dataset based on the CC-ReID dataset by introducing random occlusions on various human body parts (e.g., head, arms). Fair training and testing results indicate that traditional CC-ReID methods (e.g., CAL [8] and AIM [9]) perform worse on the OC⁴-ReID dataset.

Cloth-Changing Person Re-identification (OC⁴-ReID). This task aims to identify pedestrians while overcoming the dual challenges of occlusions and clothing changes. Particularly, OC⁴-ReID is critical for suspects tracking, as suspects often change clothes and utilize obstacles to hide their bodies during escape, which misleads conventional ReID tracking systems. Compared to CC-ReID, OC⁴-ReID poses two primary challenges: (i) occlusion leads to even less cloth-unrelated pedestrian discrimination information; (ii) blurred and low-quality occlusion regions further interfere with effective features extraction.

In this paper, we propose two specialized OC⁴-ReID datasets and a simple yet effective benchmark for OC⁴-ReID task. **Firstly**, We adopt a pre-trained human parser [10] to locate the body components of pedestrians in PRCC [11] and LTCC [12] datasets, then randomly select one component to occlude, constructing the OC⁴-ReID datasets. **Secondly**, to mitigate the impact of occlusion and mine discriminative local representations, we propose a Train-Test Micro Granularity Screening (T²MGS) module, in which a quality predictor is maintained to evaluate each micro granularity features. During testing, low-quality features are discarded to suppress occlusions. **Thirdly**, we also propose a **Part-Robust Triplet** loss (PRT) to supervise partial features, which aims to combine all the body parts and mine truly discriminative representations. In addition, PRT is robust to clothing features in pedestrians’ partial appearances and emphasises the extraction of cloth-

unrelated identity information. The main contributions of this paper are summarized as follows:

- To the best of our knowledge, the OC⁴-ReID task is introduced for the first time, and we construct two specialized datasets Occ-PRCC and Occ-LTCC.
- We propose a novel benchmark for OC⁴-ReID, introducing the Train-Test Micro Granularity Screening (T²MGS) module to against occlusion.
- We introduce a Part-Robust Triplet (PRT) loss to supervise the learning of body partial features in pedestrians, making model robust to clothing and occlusion regions.
- The experimental results in both cloth-changing and occluded cloth-changing scenarios demonstrate the effectiveness of the proposed method.

II. RELATED WORKS

A. Occluded Person Re-identification (OCC-ReID)

Person ReID matches the target pedestrian from the image gallery, crossing different scenarios and views [27, 28, 29, 30]. All of the above ReID methods assume that the entire body of the pedestrian is visible, ignoring the more challenging case of occlusion. Mainstream OCC-ReID methods utilize pose estimation [32] and body parsing [31, 33] to localize body parts, aiming to mitigate the effects of occlusion. Wang et al. [32] used labeled poses to learn visible local features as well as topological information during the training and testing phases, achieving high detection accuracy. Song et al. [31] utilized human parsing masks to generate attention maps, enabling the extraction of discriminative and robust features in occluded scenarios. However, these methods rely on maintaining an offline pose or segmentation encoder, which limits model flexibility and incurs significant computational overhead.

B. Cloth-changing Person Re-identification (CC-ReID)

CC-ReID considers a long-term retrieval situation where the pedestrian's clothing changes. To tackle this, the researchers employed different approaches including GAN [1], 3D information [34], human contour [3], and adversarial loss [8]. Xu et al. [1] enhanced the model's ability to learn cloth-unrelated features by reconstructing intra-class examples to minimize feature differences while generating adversarial dressing examples across classes. Chen et al. [34] introduced 3D Shape Learning (3DSL), leveraging 3D body reconstruction as an auxiliary task to extract texture-insensitive 3D shape features for CC-ReID. Cui et al. [3] achieved controlled reconstruction by introducing the cloth-unrelated, cloth-related, and pedestrian contour features. Gu et al. [8] introduced a cloth-based adversarial loss which first trains a cloth classifier to capture cloth-related features, then shifts focus to learning cloth-unrelated features in the second stage.

Although the above methods [1, 3, 8, 34] mitigate clothing variations to a certain extent, they are less effective when additional occlusion disturbances are introduced. Therefore, the aim of this paper is to challenge an even tougher scenario with integrated cloth changes and occlusions.

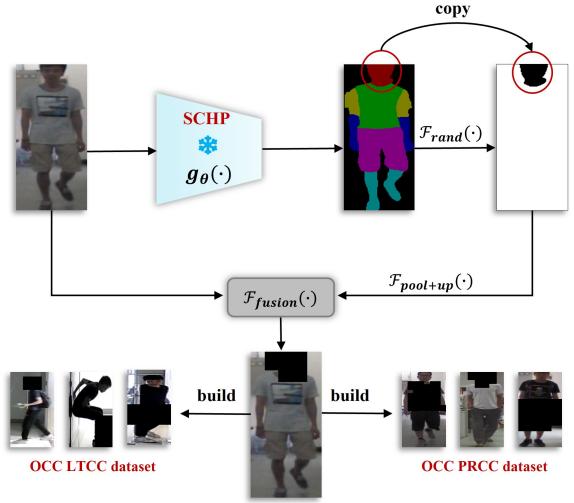


Fig. 2. Construction details of the occ-PRCC and occ-LTCC datasets.

III. OC⁴-REID DATASET

As shown in Fig. 2, to facilitate the study of OC⁴-ReID, we customise two datasets Occ-PRCC and Occ-LTCC based on existing cloth-changing datasets [11, 12]. Motivated by prior works [13, 14], we use black patches to represent occlusions. Specifically, a pre-trained human parser, SCHP [10], is employed to locate the key components of pedestrians (i.e., head, torso, upper or lower arms, upper or lower legs). We then randomly select a component to get the occlusion target f_{mask} , and this process is detailed as $f_{\text{mask}} = \mathcal{F}_{\text{rand}}(g_{\theta}(x))$, where x denotes the input image, $g_{\theta}(\cdot)$ is the processes of SCHP inference, and $\mathcal{F}_{\text{rand}}(\cdot)$ refers to the random selection process. Note that occlusions in real-world do not exactly match the segmentation regions, therefore we perform a pooling and an up-sampling operation on the segmentation results to obtain more arbitrary occlusion shapes. Finally, we combine the occluded and original image to get the processed image f_{occ} :

$$f_{\text{occ}} = \mathcal{F}_{\text{fusion}}(\mathcal{F}_{\text{up}}(\mathcal{F}_{\text{av-pool}}(f_{\text{mask}})), x), \quad (1)$$

where $\mathcal{F}_{\text{av-pool}}(\cdot)$ and $\mathcal{F}_{\text{up}}(\cdot)$ denotes the average pooling and up-sampling operations, and $\mathcal{F}_{\text{fusion}}(\cdot, \cdot)$ represents the images fusion process.

IV. METHOD

The overall architecture of our method is illustrated in Fig. 3. We first introduce the Train-Test Micro Granularity Screening (T²MGS) module in Section IV-A. Following that, we provide a detailed explanation of the Part-Robust Triplet (PRT) loss in Section IV-B.

A. Train-Test Micro Granularity Screening Module

a) Training stage: Let $F \in \mathbb{R}^{C \times H \times W}$ denotes the output feature of the backbone, where C , H , and W are the number of channels, height, and width. We divide F into k horizontal partitions denoted as $f_1, f_2, \dots, f_k \in \mathbb{R}^{C \times \frac{H}{k} \times W}$. For each partition, we maintain a quality predictor, which comprise

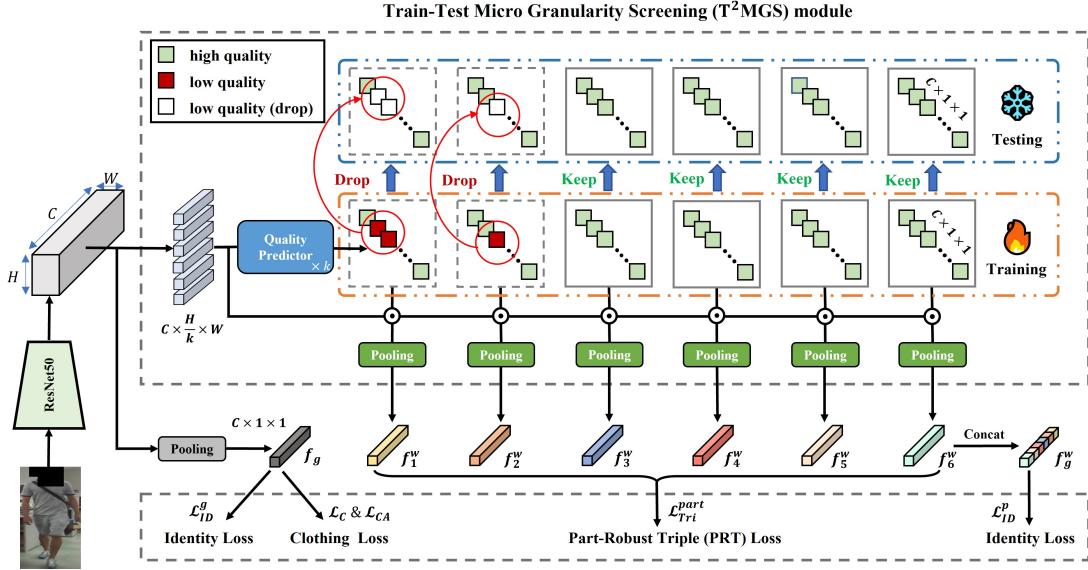


Fig. 3. We utilize the Train-Test Micro Granularity Screening ($T^2\text{MGS}$) module to adaptively discard low-quality features, and the model is supervised by identity loss [16], clothing loss [8], and the proposed Part-Robust Triple (PRT) loss for learning more discriminative representations.

two convolutional layers, a batch normalization layer, and a sigmoid activation layer. The above process can be formulated as:

$$\mathcal{F}_{QP}^i(\cdot) = \mathcal{F}_{\text{ad-pool}}(\sigma(\mathcal{F}_{\text{BN}}(\mathcal{F}_{1 \times 1}(\mathcal{F}_{1 \times 1}(\cdot, \theta), \theta)))) , \quad (2)$$

$$f_i^\phi = \mathcal{F}_{QP}^i(f_i), \quad i \in [1, k], \quad (3)$$

where $\mathcal{F}_{1 \times 1}(\cdot, \theta)$ denotes the convolution operation with a kernel size of 1×1 , and θ denotes the process parameters. $\mathcal{F}_{\text{BN}}(\cdot)$ and $\sigma(\cdot)$ represent the batch normalization and sigmoid function, respectively. $\mathcal{F}_{\text{ad-pool}}(\cdot)$ denotes the adaptive average pooling operation. $f_i^\phi \in \mathbb{R}^{C \times 1 \times 1}$ is the feature obtained by the quality predictor, and we dot product it on each partition and followed by a adaptive average pooling to obtain $f_i^w \in \mathbb{R}^{C \times 1 \times 1}$, which is detailed as $f_i^w = \mathcal{F}_{\text{ad-pool}}(f_i \cdot f_i^\phi)$. Subsequently, we concatenate the feature vector f_i^w ($i \in \{1, 2, \dots, k\}$) in the H -dimension and obtain the global feature f_g^w . In addition, we also maintain a pooled original feature f_g for the identity and clothing classifiers. The value of k is set to 6.

b) *Testing stage*: During the testing phase, for the quality prediction result f_i^ϕ ($i \in \{1, 2, \dots, k\}$) of k partitions, we conduct a micro granularity screening in the channel dimension. Specifically, channels with quality scores lower than the threshold λ (set to 0.35 in this paper) are discarded. Through this, our model can adaptively discard occluded features while focusing on discriminative representations. The detailed procedure is outlined in Algorithm 1.

B. Part-Robust Triplet Loss

In part-based ReID methods [17, 18, 19, 20], identity-related triplet loss [15] is directly used to supervise partial features. However, these approaches encounter limitations in the OC⁴-ReID scenario: (i) Generally, pedestrian clothing is

Algorithm 1 Testing Stage

```

1: if not training then
2:   for each tensor in  $f_i^\phi$  do
3:     for each channel in tensor do
4:       if channel <  $\lambda$  then
5:         Set channel to 0
6:       end if
7:     end for
8:   end for
9:   for i = 0 to k do
10:     $f_i^\phi = \mathcal{F}_{\text{ad-pool}}(f_i^\phi \cdot f_i)$ 
11:  end for
12:   $f_{tmp} \leftarrow$  Concatenate tensors in  $f_i^\phi$  along dimension H
13:   $f_g^w = \mathcal{F}_{\text{ad-pool}}(f_{tmp})$ 
14: end if
15: return  $f_g^w$ 

```

more likely to act as a discriminative feature and promote the ReID task. However, clothing features are interference information in OC⁴-ReID, supervising the clothing region with identity-related loss impairs model performance. (ii) Occluded regions mainly contain blurred and low-quality features with little identity information. Using identity-related loss for supervision introduces misleading information. To address these issues, we propose the Part-Robust Triplet (PRT) Loss, which considers the average distances of all partial features and focuses on mining the overall discriminative representations:

$$\mathcal{D}(x, y) = \frac{1}{k} \sum_{i=1}^k \mathcal{F}_{\text{dis}}^{\text{eucl}}(f_i^x, f_i^y), \quad (4)$$

where $\mathcal{F}_{\text{dis}}^{\text{eucl}}(\cdot)$ denotes the Euclidean distance. For the anchor sample a , we compute the hardest positive distance $\mathcal{D}_{(a,p)}$ and the hardest negative distance $\mathcal{D}_{(a,n)}$, similar to [15]. Finally,

TABLE I
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE STANDARD CC-ReID DATASETS AND OC⁴-ReID DATASETS.

Method	Occ-PRCC		Occ-LTCC		PRCC		LTCC	
	Rank@1	mAP	Rank@1	mAP	Rank@1	mAP	Rank@1	mAP
HACNN [22]	10.12	7.71	12.73	5.92	21.86	17.61	21.61	9.32
PCB [23]	30.64	28.73	14.94	7.41	41.87	38.71	23.52	10.12
IANet [24]	35.51	35.51	16.54	8.64	46.32	45.92	25.21	12.61
RCSANet [25]	41.95	36.93	22.56	9.04	50.24	48.67	29.61	13.87
CAL [8]	45.34	43.21	29.94	9.23	55.21	55.83	34.41	16.12
AIM [9]	47.10	44.98	31.51	11.43	57.90	58.30	40.60	19.10
Ours*	58.81	56.52	40.73	19.89	62.28	62.12	43.84	22.43

the part-robust triplet loss can be expressed as follows:

$$\mathcal{L}_{\text{Tri}}^{\text{part}} = \sum \mathcal{F}_{\max} (\mathcal{D}_{(a,p)} - \mathcal{D}_{(a,n)} + M, 0), \quad (5)$$

where $\mathcal{F}_{\max}(\cdot, \cdot)$ returns the maximum value, and M is the margin in triplet loss, which is set to 0.3 in this paper. For global features f_g and f_g^w , we utilize cross-entropy loss [16] as the identity loss to supervise, which are denoted as \mathcal{L}_{ID}^g and \mathcal{L}_{ID}^p . Additionally, similar to [8], we introduce a clothing-based adversarial loss \mathcal{L}_{CA} that collaborates with the clothing classification loss \mathcal{L}_C [16] to learn the cloth-unrelated features. The final total loss is calculated as follows:

$$\mathcal{L} = \mathcal{L}_{\text{Tri}}^{\text{part}} + \mathcal{L}_{ID}^p + \mathcal{L}_{ID}^g + \mathcal{L}_C + \mathcal{L}_{CA}. \quad (6)$$

V. EXPERIMENTS

A. Datasets and Evaluation Metrics

We evaluate the performance of our method on two CC-ReID benchmark datasets PRCC [11] and LTCC [12], as well as the proposed Occ-PRCC and Occ-LTCC datasets, under the cloth-changing configuration. PRCC [11] is a large CC-ReID dataset consists of 33,698 pedestrian images from 221 identities, and each person wears the same clothes in cameras 1 and 2, and different clothes in cameras 1 and 3. LTCC [12] contains 17,119 images from 152 identities and 478 outfits, which is captured with 12 different views. Following the previous studies [8, 9], this paper use Rank@1 and mean Average Precision (mAP) for evaluation.

B. Implementation Details

We implement our method with PyTorch and train the model on two NVIDIA RTX 4090 with 24 Gigabyte memory. ResNet-50 [10] pre-trained on ImageNet is selected as the backbone. All the input images are resized to 384 × 192. Random horizontal flipping, cropping and erasing are utilized for data augmentation. The batch size is set to 64, and initial learning rate is set to 3.5×10^{-4} , which decays at 30th and 50th epochs with a decay factor of 0.1 (120 epochs in total).

C. Compare with State-of-the-art Methods

We compare the proposed model with three traditional ReID architectures (i.e. HACNN [22], PCB [23], and IANet [24]) and three classical open-source CC-ReID methods (i.e. RCSANet [25], CAL [8], AIM [9]) in Table I.

Result on Occ-PRCC and Occ-LTCC: As shown in Table I, our model outperforms existing state-of-the-art methods on

TABLE II
ABLATION EXPERIMENTS OF OUR METHOD ON OCC-PRCC DATASET.

Index	T ² MGS	PRT loss	Triplet loss	Occ-PRCC	
				Rank@1	mAP
0(baseline)				45.34	43.21
1	✓			54.73	54.65
2	✓		✓	52.12	52.11
3(Ours*)	✓	✓		58.81	56.52

the OC⁴-ReID datasets. Taking the Occ-PRCC dataset as an example, our model achieves 58.81% Rank-1 and 56.52% mAP, which is 11.71% and 11.54% higher than the second-best method. Similarly, we also achieve advanced performance on the Occ-LTCC dataset, with Rank-1 and mAP improving by 9.22% and 8.46% compared to the second-best method. The experimental results demonstrate the superior performance of our method in handling the OC⁴-ReID task, because the following two reasons: (i) By conducting screening mechanisms on local and channel features, T²MGS effectively suppresses low-quality features while further mining fine-grained discriminative parts. (ii) Using PRT loss to supervise partial features, this key design motivates the model to focus on truly discriminative features at each training step, which in turn mitigates the effect of occlusions and clothing features.

Result on PRCC and LTCC: In addition, we compare our method with current state-of-the-art methods on standard CC-ReID datasets: PRCC [11] and LTCC [12]. On the LTCC dataset, the Rank-1 and mAP of our method under the cloth-changing scenario are 43.84% and 22.43%, outperforming the second-best method by 3.24% and 3.33%, respectively. Similarly, our method also achieve the optimal performance on PRCC dataset. The experimental results demonstrate that our method can also achieve excellent performance in a separate cloth-changing scenario.

D. Ablation Study of The Proposed Method

Table II reports the ablation experimental results of our method in Occ-PRCC. Our baseline achieves 45.34% Rank@1 and 43.21% mAP accuracy. When T²MGS is introduced for occlusion mitigation (Index 1), the Rank@1 and mAP are improved by 9.39% and 11.44%, respectively. Then, we use classic triplet loss [15] for supervising partial features (Index 2), which reduce the Rank@1 and mAP by 2.61% and 2.54%. The results show that acting triplet loss directly on partial

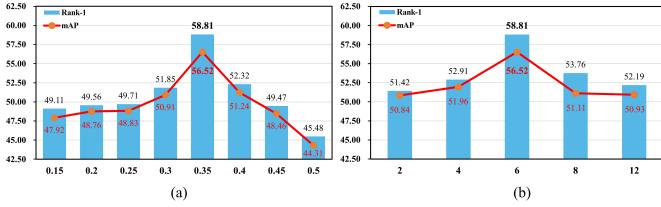


Fig. 4. Parameter-sensitivity experiments. (a) and (b) evaluate our method with different values of λ and k on Occ-PRCC dataset.

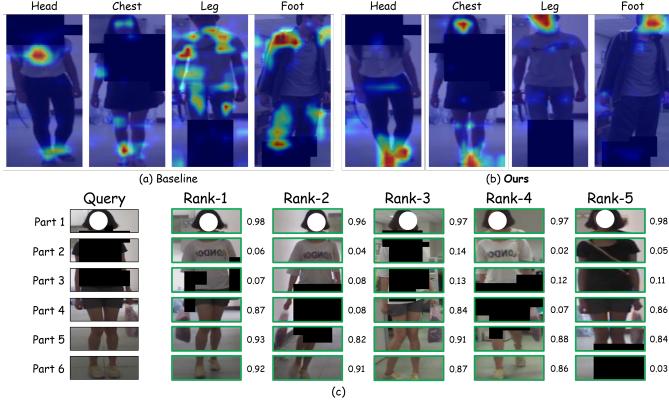


Fig. 5. (a) and (b) are the Grad-CAM visualization of the baseline and our method on PRCC datasets. (c) visualizes retrieval results and part-based similarities/distances.

features inevitably damages the feature extraction process and introduces interference information. Finally, when replacing triplet loss with the proposed PRT loss (Index 3), our method achieves optimal performance with Rank@1 and mAP of 58.81% and 56.52%.

E. Sensitivity Analysis of Hyper-Parameter Setting

To explore the optimal value of threshold λ and partition number k , we conduct parameter-sensitivity experiments on the Occ-PRCC dataset. We first set λ and k to 0.15 and 2, then increase them by 0.05 and 2, respectively. As shown in Fig. 4 (a), the model's performance reduces when the threshold λ once exceeds 0.35, which indicates that visible features are removed along with occluded features. In addition, as illustrated in Fig. 4 (b), the model achieves best performance when the number of partitions k is set to 6.

F. Visualization

We conduct visualization experiments to compare OC⁴-ReID with the baseline. In Fig.5 (a) and (b), we present the Grad-CAM [26] visualization results, showing that our method focuses more effectively on discriminative pedestrian appearance, whereas the baseline is frequently influenced by identity-independent fragmented information. Additionally, Fig.5 (c) visualizes part-based similarities. The results demonstrate that the model effectively assigns low scores to occluded and cloth-related areas while allocating high scores to cloth-unrelated discriminative body parts, highlighting the model's ability to

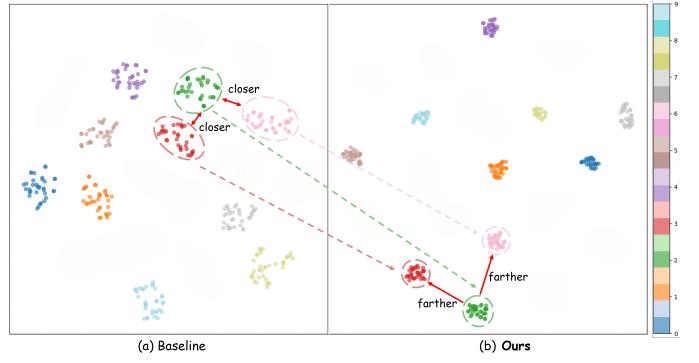


Fig. 6. t-SNE feature visualization of baseline and our method on PRCC dataset. Different colors represent different identities.

handle occlusion and cloth-changing. As demonstrated in Fig. 6, we utilize t-SNE [27] to visualize the feature distribution of random ten identities of baseline and our method. It can be observed that our method encourages more compact intra-class representations while increasing inter-class discrimination, which further validates the effectiveness of our proposed approach.

VI. CONCLUSION

In this paper, we propose the task of cloth-changing person re-identification under occlusion conditions, termed OC⁴-ReID, for the first time. To support this new task, we construct two specialized datasets, Occ-PRCC and Occ-LTCC. To address the challenges presented in this task, we design a new benchmark, which includes a Train-Test Micro Granularity Screening (T²MGS) module. This module effectively mitigates occlusions while facilitating the extraction of fine-grained cloth-unrelated features. Furthermore, we propose a Part-Robust Triplet (PRT) loss, which supervises the partial features learning properly. Experimental results on both CC-ReID and OC⁴-ReID datasets demonstrate the effectiveness of our proposed method.

REFERENCES

- [1] Xu, W., Liu, H., Shi, W., Miao, Z., Lu, Z., & Chen, F. Adversarial Feature Disentanglement for Long-Term Person Re-identification. In IJCAI (pp. 1201-1207).
- [2] Xu, P., & Zhu, X. (2023). 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).
- [3] Cui, Z., Zhou, J., Peng, Y., Zhang, S., & Wang, Y. (2023). Dcr-reid: Deep component reconstruction for cloth-changing person re-identification. IEEE Transactions on Circuits and Systems for Video Technology, 33(8), 4415-4428.
- [4] 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.

- [5] Ge, Y., Yu, M., Chen, Z., Lu, W., Dai, Y., & Shi, H. (2024). Attention-Enhanced Controllable Disentanglement for Cloth-Changing Person Re-identification.
- [6] Chen, Z., Ge, Y., & Yue, Q. (2024, August). Features Reconstruction Disentanglement Cloth-Changing Person Re-identification. In International Conference on Intelligent Computing (pp. 390-403). Singapore: Springer Nature Singapore.
- [7] Tu, M., Zhu, K., Guo, H., Miao, Q., Zhao, C., Zhu, G., ... & Wang, J. (2022). Multi-granularity mutual learning network for object re-identification. *IEEE Transactions on Intelligent Transportation Systems*, 23(9), 15178-15189.
- [8] Gu, X., Chang, H., Ma, B., Bai, S., Shan, S., & Chen, X. (2022). Cloth-changing person re-identification with rgb modality only. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1060-1069).
- [9] Yang, Z., Lin, M., Zhong, X., Wu, Y., & Wang, Z. (2023). 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).
- [10] Li, P., Xu, Y., Wei, Y., & Yang, Y. (2020). Self-correction for human parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6), 3260-3271.
- [11] Yang, Q., Wu, A., & Zheng, W. S. (2019). Person re-identification by contour sketch under moderate clothing change. *IEEE transactions on pattern analysis and machine intelligence*, 43(6), 2029-2046.
- [12] 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.
- [13] Shen D, Zhao S, Hu J, et al. Es-net: Erasing salient parts to learn more in re-identification[J]. *IEEE Transactions on Image Processing*, 2020, 30: 1676-1686.
- [14] Huang H, Li D, Zhang Z, et al. Adversarially occluded samples for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5098-5107.
- [15] Hermans, Alexander, Lucas Beyer, and Bastian Leibe. "In defense of the triplet loss for person re-identification." arXiv preprint arXiv:1703.07737 (2017).
- [16] Zhang, Zhilu, and Mert Sabuncu. "Generalized cross entropy loss for training deep neural networks with noisy labels." *Advances in neural information processing systems* 31 (2018).
- [17] Zhao, L., Li, X., Zhuang, Y., & Wang, J. (2017). Deeply-learned part-aligned representations for person re-identification. In Proceedings of the IEEE international conference on computer vision (pp. 3219-3228).
- [18] Suh, Y., Wang, J., Tang, S., Mei, T., & Lee, K. M. (2018). Part-aligned bilinear representations for person re-identification. *ECCV* (pp. 402-419).
- [19] Guo, J., Yuan, Y., Huang, L., Zhang, C., Yao, J. G., & Han, K. (2019). Beyond human parts: Dual part-aligned representations for person re-identification. *ICCV* (pp. 3642-3651).
- [20] Li, Y., He, J., Zhang, T., Liu, X., Zhang, Y., & Wu, F. (2021). Diverse part discovery: Occluded person re-identification with part-aware transformer. *CVPR* (pp. 2898-2907).
- [21] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *CVPR* (pp. 770-778).
- [22] Li W, Zhu X, Gong S. Harmonious attention network for person re-identification[C]//CVPR 2018: 2285-2294.
- [23] 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). *ECCV*(pp. 480-496).
- [24] Hou R, Ma B, Chang H, et al. Interaction-and-aggregation network for person re-identification[C] *CVPR* 2019: 9317-9326.
- [25] Huang, Y., Wu, Q., Xu, J. (2021). Clothing status awareness for long-term person re-identification. *ICCV*(pp. 11895-11904).
- [26] Selvaraju R R, Cogswell M, Das A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//CVPR 2017: 618-626.
- [27] Van der Maaten L, Hinton G. Visualizing data using t-SNE[J]. *Journal of machine learning research*, 2008, 9(11).
- [28] Xu, Wanlu, et al. "Adversarial Feature Disentanglement for Long-Term Person Re-identification." *IJCAI*. 2021.
- [29] Liu H, Feng J, Qi M, et al. End-to-end comparative attention networks for person re-identification[J]. *IEEE transactions on image processing*, 2017, 26(7): 3492-3506.
- [30] Zheng Z, Zheng L, Yang Y. A discriminatively learned cnn embedding for person reidentification[J]. *ACM transactions on multimedia computing, communications, and applications (TOMM)*, 2017, 14(1): 1-20.
- [31] Song C, Huang Y, Ouyang W, et al. Mask-guided contrastive attention model for person re-identification[C] *CVPR*. 2018: 1179-1188.
- [32] Wang, Guan'an, et al. "High-order information matters: Learning relation and topology for occluded person re-identification." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
- [33] He, Lingxiao, and Wu Liu. "Guided saliency feature learning for person re-identification in crowded scenes." *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII* 16. Springer International Publishing, 2020.
- [34] Chen J, Jiang X, Wang F, et al. Learning 3D shape feature for texture-insensitive person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 8146-8155.