2024 IEEE International Conference on Multimedia and Expo (ICME) | 979-8-3503-9015-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICME57554.2024.10687575

Clothing Sampling Based on Active Learning For Cloth-Changing Person Re-identification

Yujie Liu
China University of Petroleum
Qingdao, China
liuyujie@upc.edu.cn

Mingyue li✉
China University of Petroleum
Qingdao, China
S22070005@s.upc.edu.cn

Jiansen Jing
China University of Petroleum
Qingdao, China
S22070034@s.upc.edu.cn

Yante Li
University of Oulu
Oulu, Finland
yante.li@oulu.fi

Guoying Zhao
University of Oulu
Oulu, Finland
guoying.zhao@oulu.fi

***Abstract*—Cloth-Changing Person Re-Identification (CC-ReID) aims to match the same person with clothing changes. The challenges mainly include two types: same person wearing different clothing and different person wearing similar clothing. The current methods are usually limited by the number and variation of clothing in training data, making it difficult to cope with the latter. To address this issue, this article proposes a clothing sampler (CS) based on active learning. The main idea is actively selecting valuable clothing images, which can ensure both "clothing diversity with the same identity" and "identity diversity with the similar clothing" in batches, forcing the model to learn features that are independent of clothing. In addition, a multi-clothing loss (MC) is also designed to guide the network to learn clothing-independent features. Experiment results on two cloth-changing datasets show the effectiveness of our proposed CS.**

***Index Terms*—Person re-identification, Cloth-changing, Active learning**

I. Introduction

Person Re-Identification (ReID) [1] aims to match images of the same person in different scenes and is widely used in urban surveillance systems. In early ReID methods [2]–[4], it was assumed that people would not change clothing in a short period, making clothing features helpful for recognition. However, when a person changes clothing, if the model continues to emphasize clothing features, it becomes challenging to match the same person. In practical scenarios, this research has gained increasing interest, especially in the context of tracking criminal suspects. During pursuits, criminal suspects often attempt to evade the police by frequently changing their clothing.

The current methods for Cloth-Changing Person Re-Identification (CC-ReID) can be broadly classified into two categories: multi-modal methods and single-modal methods.

✉: Corresponding author. The work was partly supported by the National Key Research and Development Program of China (Grant no. 2019YFF0301800), the China National Scholarship Fund, the Research Council of Finland (former Academy of Finland) Academy Professor project EmotionAI (grants 336116, 345122), the University of Oulu & Research Council of Finland Profi 7 (grant 352788).

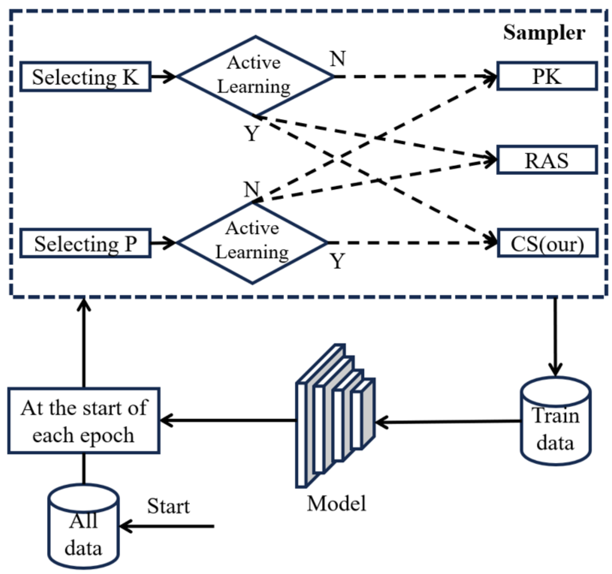

Fig. 1. Comparison of our proposed sampler with other samplers. The random sampler (PK) randomly selects P persons in a batch, with each person randomly choosing K images. The random appearance sampler (RAS) selects K valuable images. Our proposed clothing sampler (CS) not only selects K valuable images, but also selects P valuable persons.

Multi-modal methods utilize auxiliary models to extract features such as silhouette sketches [5], [6], human body parsing maps [7], [8], human poses [9], [10], and gait [11]. These features assist the model in making more accurate judgments in the presence of clothing changes. However, these methods do not fully exploit the inherent identity information present in the original RGB images, which is independent of clothing. Moreover, they may be susceptible to the quality of auxiliary modality estimation. Single-modal methods can extract more clothing-independent features from RGB images by leveraging clothing labels through adversarial learning [12] or causal distillation [13]. But the clothing label has the wrong association with identity. The reason is that in daily life, everyone's

979-8-3503-9015-5/24/$31.00 ©2024 IEEE

clothing combination is often as unique as their personality: a person can wear many sets of clothing, but few people wear the same clothing. Therefore, even if different people wear the similar clothing, they will be marked with different labels.

Previous methods have primarily focused on model design but are constrained by the impact of auxiliary models or auxiliary labels. In this regard, we abandoned the supervision information of incorrect clothing labels. Based on rich life experience [14], we believe that the correlation between clothing and identity diminishes significantly when persons wear uniform clothing or diverse sets of clothing. What we need to do is provide these samples in small batches. The most commonly used random sampler (PK) [15] randomly selects P persons in a batch, and each person randomly chooses K images, which fails to achieve this, as shown in Fig. 1. Even though random appearance sampler (RAS) [6] began actively selecting K valuable images for everyone, it still did not achieve the results we wanted. Therefore, this paper proposes a clothing sampler (CS) based on active learning, which can not only find different identities with similar clothing, but also different clothing with same identity. In a data-driven manner, it compels the model to abandon learning clothing features and shift towards learning robust identity features.

From the study of face recognition loss functions [16], [17], it is evident that emphasizing difficult samples contributes to enhancing the discriminative capability of the model. By employing the CS method, we ensure the diversity of clothing within the batch. Consequently, we design a multi-clothing loss (MC), which constrains the distribution of positive samples with different clothing through the incorporation of negative samples with the same clothing. This approach further guides the model to learn clothing-independent features.

In summary, the contributions of this paper include:

- We propose a clothing sampler (CS) based on active learning, which has been proven to perform better than a random sampler (PK) in CC-ReID.
- We design a multi-clothing Loss (MC) to assist the model in extracting more clothing-independent features to adapt to clothing changes.
- Experiment results on the cloth-changing datasets PRCC [5] and NKUP+ [6] indicate a substantial improvement over the baseline using our method, showcasing competitive performance compared to state-of-the-art methods.

II. Related Work

A. Cloth-Changing Person Re-Identification

Cloth-Changing Person Re-Identification (CC-ReID) is an inevitable scenario in practical situations. The key to solve this problem is to suppress clothing interference and extract more features unrelated to clothing. The existing methods are mainly divided into two categories: multi-modal methods and single-modal methods.

Multi-modal methods utilize the features extracted by auxiliary models to enrich identity information such as silhouette sketches [5], [6], human body parsing maps [7], [8], human poses [9], [10], and gait [11], which help the model make more accurate judgments in situations of clothing changes. Yang et al. [5] used the outline sketch information of character images to extract and aggregate human body shape features from multiple angles, mitigating the interference of clothing color on the features. Qian et al. [9] proposed a disentanglement learning method using human body key points to extract identity-related features and distill clothing features. Jin et al. [11] tried to obtain gait features from a single image to supplement the image features. However, these multimodal methods do not fully leverage the underlying identity information in the original RGB images that is unrelated to clothing. Moreover, they may also be influenced by the quality of auxiliary modality estimation.

Single-modal methods can extract clothing-independent features from RGB images by leveraging clothing labels through adversarial learning [12] or causal distillation [13]. Gu et al. [12] proposed to mine clothing-independent features from original RGB images by penalizing the predictive ability of the model to identify clothing. Yang et al. [13] used a dual-branch model to simulate causal intervention to gradually eliminate clothing bias. However, as a result of the incorrect association between clothing labels and identity, these methods also introduce new biases related to clothing while using clothing labels as auxiliary.

To sum up, the above methods are easily influenced by auxiliary models or auxiliary labels. Our method only mines more clothing-independent features on RGB images without supervised information from incorrect auxiliary labels.

B. Active Learning

The primary objective of active learning is to select the most valuable data for annotation from a large pool of unlabeled data, thus achieving higher model performance with reduced cost. This concept has found broad applicability across various domains [6], [15]. The method proposed by Liu et al. [6], is called random appearance sampler (RAS), which ensures the diversity of clothing with the same identity, effectively suppressing the interference of clothing information and enhancing model accuracy during training. However, in the case of an uneven distribution of different clothing types in the dataset, oversampling or undersampling may be necessary. Additionally, this method overlooks the impact of similar clothing images with different identities on the model. Our proposed method (CS) can effectively balance the diversity of clothing within the same identity and the diversity of identity within the same clothing in batches, achieving better results.

III. Method

In the standard person re-identification, the well-known PK sampler is commonly employed in metric learning to select batch samples. However, due to its randomness, it may not be able to provide valuable and challenging examples in small batches, let alone sufficient information for discriminative learning [15]. Therefore, we propose a new sampler based on active learning called clothing sampler (CS), ensuring "clothing diversity with the same identity" and "identity diversity

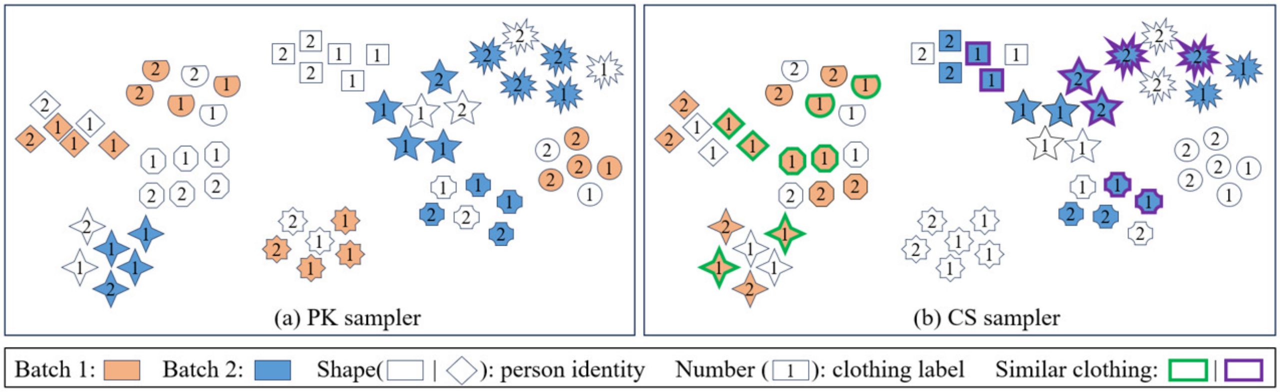

Fig. 2. Two different samplers: (a) PK sampler and (b) the proposed CS sampler. Different colors indicate different batches. Different shapes indicate different persons. Different numbers indicate different clothing. PK sampler randomly selects P persons, each person randomly selects K images. CS sampler actively selects P persons with similar clothing, and each person uniformly selects K images with different clothing.

with the same clothing" within the small batches. This forces the model to disentangle clothing features and identity features in a data-driven manner, thereby obtaining robust clothing-independent features.

A. CS

At the beginning of each epoch, a feature similarity matrix is constructed for all identity classes, and the nearest neighbor relationships are filtered and retained. CS sampler selects P persons with similar clothing in each batch, and each person selects multiple sets of clothing images. The specific implementation of the algorithm is divided into two stages.

The first stage is to construct a feature similarity matrix at the beginning of each epoch:

(1) randomly select one image from each person to form a clothing gallery.

(2) extract features from each image in the gallery using the trained model at the current stage.

(3) calculate the similarity between features to construct a similarity matrix and sort each row in descending order.

The second stage is to select P persons for each batch and K (A × N) images for each person:

(4) using each person as an anchor, find similar top P-1 persons and form an id set with the anchor.

(5) each person in the id set selects A types of clothing, each clothing selects N images.

Finally, we obtained the same number of batches as our identity. In Fig. 2, we only show two batches, where P is 4, A is 2, and N is 2.

B. Loss Function

Following [5], in order to equip the model with the basic ability to extract unique person features, we use cross-entropy loss and triplet loss to train the model as follows.

$$L_{ID} = L_{Ce} + L_{Tri} \tag{1}$$

where L_{Ce} and L_{Tri} are respectively the cross-entropy and triplet loss. It is worth noting that since we discovered

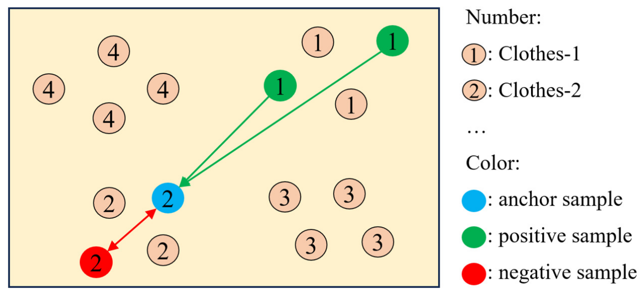

Fig. 3. Visualization of the instance consistency loss. Blue, green, and red represent anchor samples, positive samples with different clothing, and negative samples with the same clothing, respectively.

different persons wearing similar clothing during the sampling process, adding online hard sample mining to expand inter class distance and reduce intra class distance can effectively extract features unrelated to clothing.

To mitigate significant differences in feature changes after a person changes clothing, we designed an instance consistency loss to further suppress clothing feature learning.

Following [6], for anchor, the multi-appearance loss (MA) selected a furthest negative sample with the same clothing and identity, and a furthest positive sample with the same identity but different clothing, which can be formalized as follows:

$$\begin{aligned} L_{MA} = \sum_{p=1}^{P}\sum_{a=1}^{A}\sum_{n=1}^{N}\sum_{b=1,b\neq a}^{A'} \\ [\beta + \max_{n'=1,\ldots,N'} Dist(f\left(x_a^n\right), f(x_b^{n'})) \\ - \max_{n''=1,\ldots,N,n''\neq n} Dist(f(x_a^n), f(x_a^{n''}))]_+ \end{aligned} \tag{2}$$

where $f(x_a^n)$ represents the feature of n-th image x_a^n $(1 \leq a \leq A, 1 \leq n \leq N)$ of the a-th clothing of the person in the batch, $f(x_a^{n''})$ represents other images of the same person and clothing, $f(x_b^{n'})$ represents other clothing images of the same person, $Dist$ represents the distance function, and β is a hyper-parameter set to 0. By limiting the distance from the anchor to these positive samples be less than the distance to negative samples, the network can learn robust features of cross clothing images.

However, if only the furthest different clothing samples are used as positive samples, the model focuses more on feature differences caused by posture or angle rather than clothing. To alleviate this issue, we designed a multi-clothing loss (MC), which can be formalized as follows:

$$L_{MC} = \sum_{p=1}^{P} \sum_{a=1}^{A} \sum_{n=1}^{N} \sum_{b=1, b \neq a}^{A'} \\ [\beta + \min_{n'=1,\dots,N'} Dist(f(x_a^n), f(x_b^{n'})) \\ - \max_{n''=1,\dots,N, n'' \neq n} Dist(f(x_a^n), f(x_a^{n''}))]_+ \tag{3}$$

In MC loss, we use the nearest different clothing samples as positive samples, where feature differences are more attributed to changes in clothing, effectively solving the lack of response to clothing changes in MA loss. Please note that b in Eq. 3 and b in Eq. 2 represent the same clothing.

We use both MA Loss and MC Loss to effectively bring the features of different clothing images with the same identity closer, as shown in Fig. 3. The instance consistency loss is represented as follows:

$$L_{IC} = \lambda_1 L_{MA} + \lambda_2 L_{MC} \tag{4}$$

where λ_1 and λ_2 are used as the weights of MA loss and MC loss. Typically, they are the same, with a default value of 1.

Finally, the total loss function is calculated as follows:

$$L_{Total} = L_{ID} + L_{IC} \tag{5}$$

IV. EXPERIMENT

A. Datasets and Evaluation Protocol

We mainly conducted experiments and evaluated our methods on publicly available PRCC [5] and NKUP+ [6] datasets.

PRCC is a widely used dataset of moderate clothing changes, which consists of 33698 images of 221 identities with 3 camera views. The person images in cameras A and B have the same clothing, while the person images in camera C have changed clothing.

NKUP+ is one of the latest CC-ReID datasets, which consists of 40217 images of 361 identities with 29 camera views. Compared to other datasets, NKUP+ provides real scene information such as multiple scenes, perspectives, and costumes, with a shooting cycle of up to 10 months, including seasonal changes in clothing.

Evaluation Protocol. Following the previous method, we evaluated our method using rank-1 (R1) and mAP. To ensure fair comparison with existing methods, we will continue to use the previous two evaluation modes: same clothing retrive (SC) and cross clothing retrive (CC). On PRCC, SC and CC represent retrieving all images in camera A from images in cameras B and C. On NKUP+, SC and CC represent retrieving all images in query from the same clothing gallery and the cross clothing gallery.

B. Implementation Details

We use ResNet-50 as the backbone of the ReID model and use the pre-trained model on ImageNet. Following [12], we use random horizontal flipping, random cropping, and random erasing for data augmentation and delete the last downsampling of ResNet-50 to enrich fine-grained feature. We use global average pooling and global max pooling to extract features, then concatenate them in the channel dimension. All input images have been adjusted to 384 × 192. This model was trained by Adam for 90 epochs. The learning rate is initialized to 3.5e-4 and divided by 10 after every 30 epochs.

C. Comparison with State-of-the-Art Methods

In our research, we compared our method with four standard person re-identification methods and seven cloth-changing person re-identification methods on PRCC and NKUP+ in Table I. In previous studies, the methods that incorporate multiple modalities have achieved significant results. However, the use of multiple modalities can lead to increased computational costs, especially for FSAM [22] and M2Net [6], which utilize three modalities. In contrast, the approach of employing only RGB images for disentanglement learning yields better results without relying on auxiliary model. But these single-modal methods generate new associations between clothing and identity due to incorrect clothing labels.

Our method (CS) outperforms all comparative methods on the PRCC and NKUP+, achieving state-of-the-art results. We use the RAS sampler as the baseline [6], supervised by ID Loss and MA Loss. For PRCC dataset, CS outperforms the baseline with 7.8% and 9.2% on R1 and mAP, and it is also better than the R1 and mAP of the suboptimal method CCFA [14] by 3.6% and 3.1% in CC mode. For NKUP+ dataset, CS outperforms the baseline with 2.2% and 2.4% on R1 and mAP, and it is also better than the R1 and mAP of the suboptimal method CAL [12] by 5.2% and 0.1% in CC mode.

D. Ablation Study

In this section, we conduct ablation studies to validate the effectiveness of the key components of our model. And more comprehensive ablation experiments are provided in supplemental materials.

Comparison with other sampler. In conjunction with the joint guidance of the triplet loss and cross-entropy loss, we replicated the actual performance of three samplers on PRCC dataset, as presented in Table II. In CC mode, the CS method proposed in this paper outperforms other commonly used sampling techniques. It achieves a 4.5% improvement in R1 and a 2.5% improvement in mAP compared to PK sampling. These results indicate that the CS method aggregates samples with similar clothing but different identities and samples with same identity but different clothing into batch, which can help the model learn features unrelated to clothing.

Loss Functions. Table III presents ablation studies for each loss function on the PRCC dataset. 1) Adding MA Loss separately to the model significantly improves performance, indicating that MA Loss can help the model effectively resist

TABLE I
COMPARISON OF CS AND OTHER SOTA REID NETWORKS ON PRCC AND NKUP+ DATASETS. BOLD INDICATES BEST EFFECT, UNDERLINE INDICATES SUBOPTIMAL EFFECT.'SKE', 'SIL', 'POSE', 'PAR' AND 'GAIT' REPRESENT THE CONTOUR SKETCHES, SILHOUETTES, HUMAN POSES, HUMAN PARSING AND HUMAN GAIT, RESPECTIVELY.

Method	Venue	Size	Type		PRCC				NKUP+			
					SC		CC		SC		CC	
			RGB	Other	R1↑	mAP↑	R1↑	mAP↑	R1↑	mAP↑	R1↑	mAP↑
PCB [18]	ECCV 18	384×192	●		99.8	97.0	41.8	38.7	76.0	60.9	9.3	5.1
MGN [19]	MM 18	-	●		99.5	98.4	33.8	35.9	81.6	69.3	15.4	9.0
IANet [20]	CVPR 19	384×192	●	-	99.4	98.3	46.3	46.9	-	-	-	-
TransReID [21]	ICCV 21	256×128	●	-	99.6	98.7	44.6	46.5	81.1	69.6	19.7	10.9
SPT [5]	TPAMI 18	256×128	-	ske	64.2	-	34.4	-	-	-	-	-
FSAM [22]	CVPR 21	256×128	●	sil+pose	98.8	-	54.5	-	-	-	-	-
GI-ReID [11]	CVPR 22	256×128	●	gait	86.0	-	33.3	-	-	-	-	-
M2Net-F [6]	MM 22	256×128	●	ske+par	99.5	99.1	59.3	57.7	86.2	72.3	24.0	11.0
CAL [12]	CVPR 22	384×192	●	-	**100**	99.8	55.2	55.8	99.7	**92.4**	24.8	13.2
AIM [13]	CVPR 23	384×192	●	-	**100**	**99.9**	57.9	58.3	-	-	-	-
CCFA [14]	CVPR 23	384×128	●	-	99.6	98.7	61.2	58.4	-	-	-	-
M2Net [6](baseline)		384×192	●	-	99.5	97.0	57.0	52.3	99.7	80.8	26.8	10.9
CS(ours)		384×192	●	-	99.9	99.2	**64.8**	**61.5**	**99.8**	88.6	**29.0**	**13.3**

TABLE II
ABLATION EXPERIMENTS OF DIFFERENT SAMPLER ON PRCC.

Method	Sampler			PRCC	
	PK	RAS	CS	R1	mAP
1	✓	×	×	49.7	52.0
2	×	✓	×	53.8	53.1
3	×	×	✓	54.2	54.5

TABLE III
ABLATION EXPERIMENTS OF LOSS FUNCTION ON PRCC.

Method	Loss			SC		CC	
	ID	MA	MC	R1	mAP	R1	mAP
1	✓	×	×	100	99.7	54.2	54.5
2	✓	✓	×	99.8	99.0	58.7	57.3
3	✓	×	✓	99.9	99.5	55.5	55.0
4	✓	✓	✓	99.9	99.2	64.8	61.5

TABLE IV
EXPERIMENTS OF RAS AND CS STRATEGIES ON PRCC.

Loss	Sampling Strategy	SC		CC	
		R1	mAP	R1	mAP
MA	RAS	99.5	97.0	57.0	52.3
	CS	99.8	99.0	58.7	57.3
MA+MC	RAS	99.8	99.0	59.8	58.5
	CS	99.9	99.2	64.8	61.5

pose angle changes and suppress the influence of clothing information. 2) Adding MC Loss alone will also improve model performance, proving that it does have a suppressive effect on clothing information. 3) When MA Loss and MC Loss jointly guide the model, significant performance improvements can be observed, verifying that our proposed consistency loss can better guide model training.

Comparison of RAS and CS Strategies. In addition to identifying different persons wearing similar clothing, CS also enhances the effectiveness of the Triplet loss. When randomly selecting different identities in a small batch, it is impossible to provide diversity in clothing for the same identity and diversity in identity for similar clothing, and the performance loss of TriHard will decrease. The previous RAS sampling method can ensure the diversity of clothing with same identity to improve the performance of TriHard loss. On the basis of this method, CS further provides the diversity of identity with similar clothing to improve model performance. Therefore, we tested two sampling strategies based on MA Loss and MA+MC Loss on PRCC, and the results are shown in Table IV. Compared with RAS, we can observe that in CC mode, CS improves the rank 1 accuracy and mAP of MA by 1.7%/5.0% respectively, and improves the rank 1 accuracy and mAP of MA+MC by 5.0%/3.0% respectively.

V. VISUALIZATION

Visualization of feature distribution. To provide a more intuitive evaluation of our CS method, we used the t-SNE [23] method to visualize the high-dimensional feature vectors output by the network, as shown in Fig. 4. Compared to PK, RAS has compact intra class features that can effectively respond to clothing changes of the same person. Unfortunately, RAS has also compact inter class characteristics, making it difficult to cope with the impact of similar clothing. In comparison, CS can simultaneously handle situations where the same person is wearing different clothing and different people are wearing similar clothing.

Visualization of the activation feature maps .As shown in Fig 5, we visualized activation feature maps based on PK sampling method, baseline, and our CS on the PRCC and

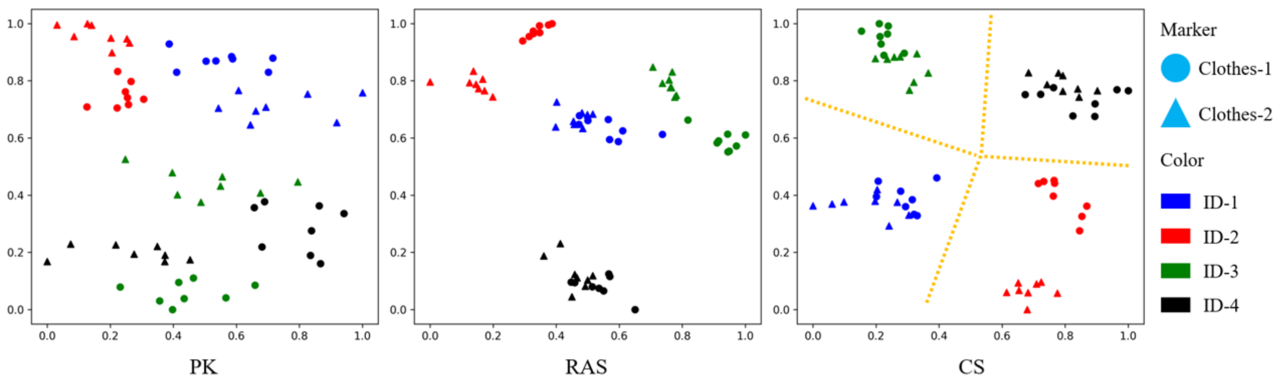

Fig. 4. On the PRCC dataset, t-SNE [23] visualization of image features. Different colors represent different identities, while circles and triangles respectively represent the first and second sets of clothing for a person. We have chosen four persons with similar clothing here. It is evident that RAS encounters difficulty in distinguishing between individuals wearing similar clothing. In contrast, our method (CS) effectively discerns between different identities even when they have similar clothing.

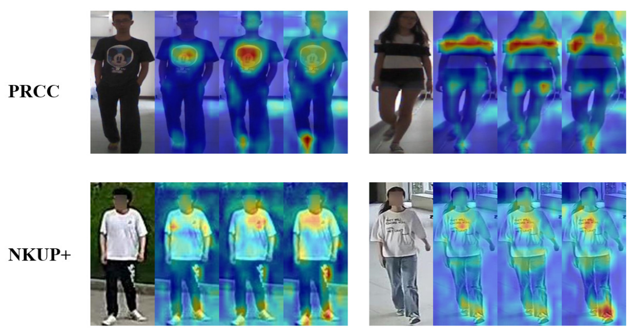

Fig. 5. Visualization of the activation feature maps. In each group, the second, third, and fourth columns display activation maps based on PK sampling method, baseline, and our CS, respectively.

NKUP+datasets, respectively. It can be intuitively observed that our method focuses more on identity related areas such as the head, feet, and joints rather than clothing related areas.

VI. Conclusion

In the absence of accurate clothing labels, this paper proposes a clothing sampler (CS) based on active learning methods. Existing methods may have new clothing impacts while using clothing labels. The CS achieves higher performance while only utilizing partial correct labels. A multi-clothing loss (MC) is proposed to better guide the model in extracting inherent character features. Experimental results show that our approach significantly improves the robustness of person re-identification models when confronted with clothing changes.

References

[1] X. Gu, B. Ma, H. Chang, S. Shan, and X. Chen, "Temporal knowledge propagation for image-to-video person re-identification," in *ICCV*, 2019.

[2] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in *CVPR*, 2018.

[3] Z. Zhang, C. Lan, W. Zeng, X. Jin, and Z. Chen, "Relation-aware global attention for person re-identification," in *CVPR*, 2020.

[4] J. Zhang, L. Niu, and L. Zhang, "Person re-identification with reinforced attribute attention selection," *TIP*, 2020.

[5] Q. Yang, A. Wu, and W.-S. Zheng, "Person re-identification by contour sketch under moderate clothing change," *TPAMI*, 2019.

[6] M. Liu, Z. Ma, T. Li, Y. Jiang, and K. Wang, "Long-term person re-identification with dramatic appearance change: Algorithm and benchmark," in *ACM MM*, 2022.

[7] X. Shu, G. Li, X. Wang, W. Ruan, and Q. Tian, "Semantic-guided pixel sampling for cloth-changing person re-identification," *SPL*, 2021.

[8] H. Chao, Y. He, J. Zhang, and J. Feng, "Gaitset: Regarding gait as a set for cross-view gait recognition," in *AAAI*, 2019.

[9] X. Qian, W. Wang, L. Zhang, F. Zhu, Y. Fu, T. Xiang, Y.-G. Jiang, and X. Xue, "Long-term cloth-changing person re-identification," in *ACCV*, 2020.

[10] Y.-J. Li, Z. Luo, X. Weng, and K. M. Kitani, "Learning shape representations for clothing variations in person re-identification," *arXiv*, 2020.

[11] X. Jin, T. He, K. Zheng, Z. Yin, X. Shen, Z. Huang, R. Feng, J. Huang, Z. Chen, and X.-S. Hua, "Cloth-changing person re-identification from a single image with gait prediction and regularization," in *CVPR*, 2022.

[12] X. Gu, H. Chang, B. Ma, S. Bai, S. Shan, and X. Chen, "Clothes-changing person re-identification with rgb modality only," in *CVPR*, 2022.

[13] Z. Yang, M. Lin, X. Zhong, Y. Wu, and Z. Wang, "Good is bad: Causality inspired cloth-debiasing for cloth-changing person re-identification," in *CVPR*, 2023.

[14] K. Han, S. Gong, Y. Huang, L. Wang, and T. Tan, "Clothing-change feature augmentation for person re-identification," in *CVPR*, 2023.

[15] S. Liao and L. Shao, "Graph sampling based deep metric learning for generalizable person re-identification," in *CVPR*, 2022.

[16] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *CVPR*, 2019.

[17] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *ECCV*, 2016.

[18] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in *ECCV*, 2018.

[19] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *ACM MM*, 2018.

[20] R. Hou, B. Ma, H. Chang, X. Gu, S. Shan, and X. Chen, "Interaction-and-aggregation network for person re-identification," in *CVPR*, 2019.

[21] S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, "Transreid: Transformer-based object re-identification," in *ICCV*, 2021.

[22] P. Hong, T. Wu, A. Wu, X. Han, and W.-S. Zheng, "Fine-grained shape-appearance mutual learning for cloth-changing person re-identification," in *CVPR*, 2021.

[23] G. Hinton and L. van der Maaten, "Visualizing data using t-sne journal of machine learning research," 2008.