Hierarchical Proxy Learning for Cloth-Changing Person Re-Identification

Chenyang Yu*, Xuehu Liu†, Ju Dai‡, Pingping Zhang§ and Huchuan Lu*§

*School of Information and Communication Engineering, Dalian University of Technology, Dalian, China

†School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan, China.

‡Peng Cheng Laboratory, Shenzhen, China

§School of Future Technology, School of Artifcial Intelligence, Dalian University of Technology, Dalian, China

***Abstract*—Cloth-Changing person Re-Identification (CC-ReID) depends significantly on learning discriminative features under the cloth-changing scenario. It is quite challenging due to the large intra-person variance and small inter-person variance caused by clothes changing. To address these issues, in this work we propose a Hierarchical Proxy Learning (HPL) framework to extract clothes-irrelevant and person-invariant features. Specifically, we employ person labels as the main proxy. Instead of leveraging clothing labels as sub proxy, we further propose a clustering-based automatic sub-proxy mining scheme. More specifically, we first construct a person-aware Main Proxy Learning (MPL) to improve the separability of different persons. Then, a Sub Proxy Learning (SPL) is constructed to enhance the intra-person compactness. Finally, a Sub-to-Main Proxy Learning (S2MPL) is proposed to promote the cooperation between the main proxies and sub proxies. In addition, to weed out the negative effect of clothes, we propose a Sample Balance and Diversity (SBD) module, which balances the number of sub proxies in a mini-batch and utilizes semantic guidance to enrich the diversity of clothes, simultaneously. Extensive experiments on two public CC-ReID datasets demonstrate the superiority of our proposed method over most state-of-the-art methods.**

***Index Terms*—cloth-changing person re-identification, hierarchical proxy learning, sample balance, joint training.**

I. Introduction

Cloth-Changing person Re-Identification (CC-ReID) is a long-term retrieval task, which aims at re-identifying target persons across non-overlapping cameras. Compared with traditional ReID [1]–[5], CC-ReID [6], [7] is encountering more realistic challenges. Despite being quite challenging, CC-ReID is receiving more and more interest from researchers due to its crucial role in more realistic scenario applications.

To address CC-ReID, previous methods [8]–[10] aim to eliminate the impact of clothes, and extract the inherent characteristics of pedestrians, such as 3D human shape, gait information, contour sketches, etc. However, these inherent characteristics are not as effective as appearance features, leading to some performance deteriorations on the same-clothing ReID. Furthermore, various approaches rooted in metric learning [11], [12] and data augmentation techniques [13], [14] have been introduced to tackle the CC-ReID problem.

In fact, a critical step in ReID is to design a good distance metric [15]. As shown in Fig. 1 (a) and (b), due to the large intra-person variance and small inter-person variance caused by changing clothes, the instance-level triplet loss [16] and contrastive loss [17], [18] cannot achieve satisfactory performance. Recently, some works [19]–[21] perform ReID by proxy-based metric learning. For example, Wang *et al.* [22] propose intra-camera and inter-camera proxy contrastive learning. For CC-ReID, Gu *et al.* [23] design a clothes-based adversarial loss to further pull the features with the same identity closer. Unfortunately, both of them focus on instance-to-proxy interactions, and neglect inter-proxy relations. Different from previous methods, as shown in Fig. 1 (c), we propose a Hierarchical Proxy Learning (HPL) framework, which consists of a Main Proxy Learning (MPL), a Sub Proxy Learning (SPL) and a Sub-to-Main Proxy Learning (S2MPL). In MPL, we first create main proxies for individuals, then bring the proxies of the same person closer while distancing those of different people, enhancing inter-person separability. In SPL, we create sub proxies for each person and group instances with different sub proxies to improve intra-person compactness. Unlike [23] which directly using clothing labels as sub proxies, we propose a clustering-based automatic sub-proxy mining scheme. In S2MPL, each sub-proxy acts as an anchor, being pulled toward its corresponding main proxy and pushed away from others, promoting inter-person diversity and intra-person compactness. Such a hierarchical structure contributes to extracting person-invariant and clothes-irrelevant features.

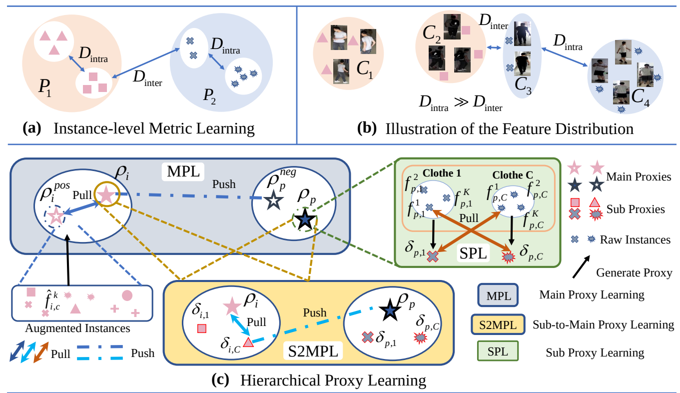

Fig. 1. Our motivations. (a) Geometry interpretation of instance-level metric learning. (b) Illustration of the feature distribution of randomly selected persons from CC-ReID datasets. (c) Geometry interpretation of the proposed Hierarchical Proxy Learning (HPL). Different colored dots and shapes represent different persons and sub proxies identities, respectively.

As shown in Fig. 1 (c), when constructing a hierarchical

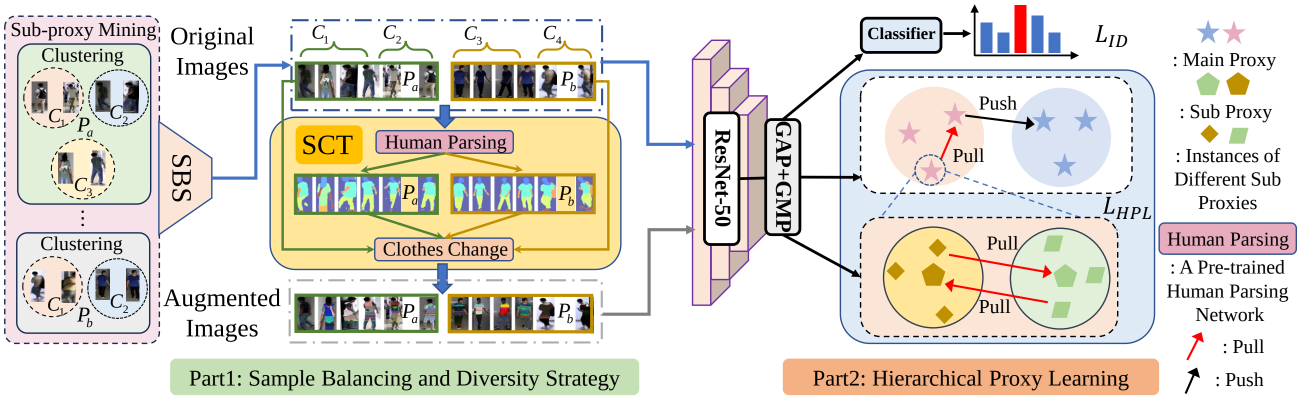

Fig. 2. Illustration of the proposed framework.

structure, if there is no assistance, the main proxy P_1 in a mini-batch will have no corresponding positive samples. What's more, if there is no constraint, the distribution of sub proxies will be random. Meanwhile, due to the annotation limitation in current CC-ReID datasets, it is highly possible for a person who wears one clothes all the time. Considering the above issues, we further propose a Sample Balance and Diversity (SBD) module, which balances the number of sub proxies in a mini-batch and utilizes semantic guidance to enrich the diversity of clothes, simultaneously. Specifically, we first explore a Sub-proxy Balanced Sampling (SBS) strategy taking the balance and diversity of sub proxies into consideration, which is more suitable for CC-ReID. Then, a Semantic-guided Clothes Transfer (SCT) is proposed to enrich the diversity of clothes, which utilizes a pre-trained human parsing network [24] to guide clothing changing. Thanks to SCT, we can get the main proxy positive samples corresponding to pedestrians based on the augmented samples. Experimental results demonstrate that our method significantly outperforms most state-of-the-art works on two public CC-ReID datasets.

The contributions of our work can be summarized as: (1) We propose an effective data processing module named SBD for CC-ReID. (2) We propose a novel proxy-level metric learning method with a hierarchical structure to extracting person-invariant and clothes-irrelevant features. (3) Extensive experiments demonstrate that our proposed method outperforms most state-of-the-art cloth-changing methods on two widely-used CC-ReID datasets, *i.e.*, PRCC and VC-Clothes.

II. Methods

As illustrated in Fig. 2, our proposed framework mainly includes two components: Sample Balance and Diversity (SBD) module and Hierarchical Proxy Learning (HPL). Detailed descriptions are presented in the following sections.

A. Sample Balance and Diversity

Recent methods [23], [25], [26] utilize clothes labels to improve the performance of CC-ReID. However, obtaining the clothes labels requires a certain price. Inspired by the recent success of unsupervised person ReID methods, we propose a clustering-based automatic sub-proxy mining scheme. Specifically, before each round of network training, we cluster all the feature representations $\{f_n^p\}_{n=1}^{N_p}$ for each person p into C_p clusters whose pseudo-labels are used as the sub-proxies. In practice, we adopt the DBSCAN [27] method for clustering.

Sub-proxy Balanced Sampling. The sampling strategy [16] in traditional ReID mainly considers the balance of different persons but ignores the balance of different clothes. Intuitively, it is useful to choose balanced sub-proxy in each batch for CC-ReID. Therefore, we propose a Sub-proxy Balanced Sampling (SBS) strategy. We choose P persons in each mini-batch, where C sub proxies per person and K images per sub proxy. Our SBS strategy performs a balanced optimization of persons and sub-proxy, thereby promoting the learning efficacy.

Semantic-guided Clothes Transfer. In CC-ReID datasets, some persons may wear only one clothes all the time. Data augmentation is an effective strategy to enrich the diversity of training samples in CC-ReID. We propose a Semantic-guided Clothes Transfer (SCT) to change clothes among different persons. Specifically, given one image $x_i \in \{x_{p,c}^k\}_{p=1,c=1,k=1}^{P,C,K}$, we first randomly select another image x_j with different person and sub proxy in a mini-batch. Then, a pre-trained human parsing network [24] is employed to obtain semantic masks of x_i and x_j. Considering that the most common dressing parts for persons are upper-clothes and pants, we perform SCT based on the masks of upper-clothes and pants, respectively. Given the upper-clothes masks m_i and m_j of two pedestrians, we can transfer the upper-clothes of x_j to x_i,

$$\hat{x}_i = x_i \odot (1 - m_i) + Reshape(Mean(x_j \odot m_j)) \odot m_i, \quad (1)$$

where $\odot$ means the matrix multiplication. $Mean(\cdot)$ calculates the average pixel value of the upper-clothes to address the variability in the clothing area of different persons. $Reshape(\cdot)$ duplicates the pixel value to the same shape of the target image x_i. Similarly, we can change pants from x_j to x_i.

In one mini-batch, we get the corresponding augmented image $\hat{x}$ of each image $x \in \{x_{p,c}^k\}_{p=1,c=1,k=1}^{P,C,K}$ through SCT. Meanwhile, its person label remains unchanged while the clothes label has been changed. Thus, our SCT can generate more training samples for one person dressing in different clothes, which enriches the diversity of samples for CC-ReID.

B. Hierarchical Proxy Learning

For feature extraction, we feed the original images $x_{p,c}^k$ and augmented images $\hat{x}_{p,c}^k$ into the ResNet-50 following a

GAP and a GMP to obtain the feature vectors $\left(f_{p,c}^{k}, \hat{f}_{p,c}^{k}\right)$. In our baseline method, we adopt the cross-entropy loss for person classification. However, merely person classification is hard to reduce the intra-person variance and increase the inter-person variance. As shown in Fig. 1 (c), we propose a novel Hierarchical Proxy Learning (HPL) framework, including a Main Proxy Learning (MPL), a Sub Proxy Learning (SPL) and a Sub-to-Main Proxy Learning (S2MPL).

Main Proxy Learning. As stated in [23], the instance-level metric learning may lead to a sub-optimization for CC-ReID, because it only mine the hard cases in a mini-batch and is sensitive to noisy positives and negatives. Thus, to alleviate this problem, we propose a Main Proxy Learning (MPL). The illustration of our MPL is shown in the blue part of Fig. 1 (c). Specifically, based on the features of persons, the main proxy ρ_i can be constructed by:

$$\rho_i = \frac{1}{CK} \sum_{c=1,k=1}^{C,K} f_{i,c}^{k}, i \in [1, P]. \tag{2}$$

The main proxy ρ_i can be seen as an anchor. Then, we can obtain the corresponding positive main proxy ρ_i^{pos} from augmented samples by:

$$\rho_i^{pos} = \frac{1}{CK} \sum_{c=1,k=1}^{C,K} \hat{f}_{i,c}^{k}, i \in [1, P]. \tag{3}$$

Afterward, the negative main proxy ρ_i^{neg} which has a different person label with the anchor can be defined as:

$$\rho_p^{neg} = \frac{1}{CK} \sum_{c=1,k=1}^{C,K} f_{p,c}^{k}, p \neq i, p \in [1, P]. \tag{4}$$

For one anchor, we have $J = (P-1) \times 2$ negative main proxies. Thus, the loss of MPL can be defined as:

$$L_{MPL} = \frac{1}{P} \sum_{i=1}^{P} \left[\alpha + D(\rho_i, \rho_i^{pos}) - \min D(\rho_i, \rho_p^{neg})\right]_+, \tag{5}$$

where $D(,)$ is the Euclidean distance, $\min$ represents the minimized distances among negative pairs for obtaining the hardest negative main proxies in the mini-batch. α is a margin hyper-parameter and $[\cdot]_+$ represents the hinge loss. Different from previous methods, our proposed MPL can suppress the influence of noisy samples in feature optimization.

Sub Proxy Learning. Our MPL does not take the intra-person compactness into account. Thus, as shown in the green part of Fig. 1 (c), we further propose the Sub Proxy Learning (SPL) to resolve this problem. Specifically, thanks to SBS, we can sample C sub proxies for the p-th person in a mini-batch and construct sub proxy by:

$$\delta_{p,c} = \frac{1}{K} \sum_{k=1}^{K} f_{p,c}^{k}, c \in [1, C], p \in [1, P]. \tag{6}$$

TABLE I
COMPARISON WITH OTHER METHODS ON PRCC.

Method	Clothes label	PRCC			
		CC		SC	
		Rank-1	mAP	Rank-1	mAP
PCB [28]	×	41.8	38.7	99.8	97.0
IANet [29]	×	46.3	45.9	99.4	98.3
FSAM [30]	×	54.5	-	98.8	-
Pixel [31]	×	65.8	61.2	99.5	96.7
RCSANet [21]	✓	50.2	48.6	**100**	97.2
BSGA [26]	✓	61.8	58.7	99.6	97.3
CAL [23]	✓	55.2	55.8	**100**	**99.8**
HPL(Ours)	×	**74.3**	**69.0**	99.8	98.5

In SPL, we constrain the feature learning to pull the different sub proxies of the same person closer. Thus, the loss of our SPL can be expressed as:

$$L_{SPL} = \sum_{p=1}^{P} \sum_{c=1}^{C} \sum_{k=1}^{K} \sum_{i=1, i \neq c}^{C} D(\delta_{p,c}, f_{p,i}^{k}), \tag{7}$$

where the first three summation items represent traversing all samples obtained by the SBS strategy, and each sample is regarded as an anchor. The last summation item aims to calculate the distance between each sample and its proxy of instance samples with the same person but different sub proxies. Considering that the mined sub proxies act as substitutes for clothing labels. Under the constraint of SPL, the intra-person variance caused by clothes transformation will be reduced.

Sub-to-Main Proxy Learning. To achieve the collaboration between the main proxies and sub-proxies, we further propose a novel component called Sub-to-Main Proxy Learning (S2MPL) to effectively address the inter-person diversity and intra-person compactness. Specifically, as shown in the yellow part of Fig. 1 (c), each sub proxy $\delta_{i,c}$ is treated as an anchor. S2MPL pulls it towards the corresponding main proxy ρ_i, and pushes it away from the others. The S2MPL is formulated as:

$$L_{S2MPL} = -\frac{1}{PC} \sum_{i=1}^{P} \sum_{c=1}^{C} \log \frac{\exp(d(\delta_{i,c}, \rho_i)/\tau)}{\sum_{p=1}^{P} \exp(d(\delta_{i,c}, \rho_p)/\tau)}, \tag{8}$$

where $d(,)$ is the cosine distance and τ is a temperature factor. It should be emphasized that, to ensure a reasonable tolerance level of intra-person variances, we do not impose strict constraints on the similarity between instances in the sub-proxy. Similarly, the augmented sample $\hat{x}_{i,c}^{k}$ generated from the original sample $x_{i,c}^{k}$ is also not strictly enforced to be the same as $x_{i,c}^{k}$. The overall HPL loss is given by the combination of all the above losses:

$$L_{HPL} = L_{SPL} + L_{MPL} + L_{S2MPL}. \tag{9}$$

C. Joint Training

In this paper, we also apply the cross-entropy loss L_{ID} on the feature vectors to train the baseline network. Thus, the total loss function L_{total} can be formulated as:

$$L_{total} = L_{ID} + \lambda L_{HPL}, \tag{10}$$

where λ is a balanced parameter.

TABLE II
COMPARISON WITH OTHER METHODS ON VC-CLOTHES.

Method	Clothes label	VC-Clothes			
		CC		SC	
		Rank-1	mAP	Rank-1	mAP
PCB [28]	×	62.0	62.2	94.7	94.3
ISP [34]	×	72.0	72.1	94.5	94.7
FSAM [30]	×	78.6	78.9	94.7	94.8
3DSL [25]	✓	79.9	81.2	-	-
BSGA [26]	✓	84.5	84.3	94.9	94.4
CAL [23]	✓	81.4	81.7	95.1	**95.3**
HPL(Ours)	×	**84.9**	**84.7**	**95.3**	**95.3**

TABLE III
THE ABLATION STUDIES OF DIFFERENT COMPONENTS.

Components	PRCC				VC-Clothes			
	CC		SC		CC		SC	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
Baseline	57.0	57.2	100	99.6	73.9	74.0	94.7	94.8
+ SBS	59.3	58.5	100	99.6	75.7	75.8	94.9	94.7
+ SCT	68.6	63.9	100	98.6	77.9	77.6	94.7	94.7
+ HPL	74.3	69.0	99.8	98.5	84.9	84.7	95.3	95.3

III. EXPERIMENTS

A. Datasets and Evaluation Protocols

We evaluate our approach on two benchmark datasets, *i.e.*, PRCC [32] and VC-Clothes [33]. There are two test settings in CC-ReID, namely cloth-changing setting (CC), and same-clothes setting (SC). To measure the performance, we adopt the Cumulative Matching Characteristic (CMC) table and mean Average Precision (mAP).

B. Implementation Details

Our model is trained with two NVIDIA Tesla A30 GPUs (24G memory) and implemented with the PyTorch toolbox. We use the ResNet-50 [35] pre-trained on ImageNet [36] as the feature encoder. For automatic sub-proxy mining, we use DBSCAN [27] for clustering, and set the minimum number of samples for each cluster to 4 and the scanning radius to 0.8. During training, we adopt random flipping, random cropping and random erasing [37] for data augmentation. The input images are resized to 384×192. For the SBS strategy, we sample $P = 8$ identities, $C = 2$ different sub proxies for each identity, and each sub proxy with $K = 4$ images. We train the framework for 60 epochs in total by the Adam [38] optimizer. The initial learning rate is 0.00035 and is divided by 10 at every 20 epochs. As for the hyper-parameter, α is set to be 0.3, τ is set to be 1/16 and λ is set to be 0.1. We use the original features during testing, and the cosine similarity is employed as the distance metric for ranking.

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

We compared our method with various advanced CC-ReID methods on PRCC and VC-Clothes. Experimental results are reported in Tab. I and II. On the PRCC and VC-Clothes datasets, our method achieves the best results of 74.3% and 84.9% in Rank-1 accuracy under the CC setting, respectively. RCSANet [21] utilizes a clothes-proxy loss that encourages more consistency when a person wears the same clothes, and pulls away different clothes. In contrast, the purpose of our HPL loss is to pull the samples of different clothes with the same identity together, which is similar to CAL [23]. Moreover, unlike CAL [23], BSGA [26] and 3DSL [25], the proposed HPL don't use any clothes information and achieves better performance with more limited information. And all the above methods ignore the unbalanced distribution of training samples in a mini-batch. As a result, our method achieves 84.9% Rank-1 accuracy on VC-Clothes under the CC setting, which surpasses CAL by 3.5%. These comparisons fully demonstrate the effectiveness and superiority of our method.

TABLE IV
THE ABLATION STUDIES OF HPL ON PRCC.

Method	Components			PRCC			
	SPL	MPL	S2MPL	CC		SC	
				Rank-1	mAP	Rank-1	mAP
1	×	×	×	68.6	63.9	100	98.6
2	✓	×	×	72.4	67.7	99.5	97.2
3	×	✓	×	71.3	67.5	99.8	98.2
4	×	×	✓	71.5	67.4	99.9	98.3
5	✓	×	✓	73.5	68.4	99.5	97.6
6	✓	✓	×	73.9	68.6	99.6	97.4
7	×	✓	✓	72.7	67.9	99.9	98.6
8	✓	✓	✓	74.3	69.0	99.8	98.5

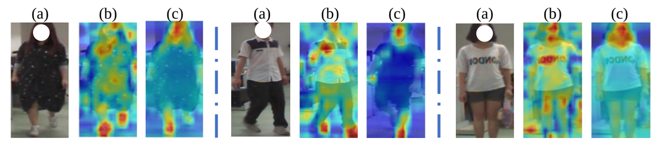

Fig. 3. The visualization of feature maps on PRCC. (a) original image; the feature map of the baseline method (b) and our proposed method (c).

D. Ablation Study

To verify the impact of each component, we conduct several experiments, and show the results in Tab. III. It can be observed that the proposed SBS strategy is superior to the baseline. And models leveraging SCT also achieved remarkable performance improvements. The reason can be attributed to the enrichment of training examples with clothes-changing. In addition, the proposed HPL further improves the performance of CC-ReID. A reasonable explanation is that a hierarchical structure is contributed to extracting person-invariant and clothes-irrelevant features. What's more, as shown in Tab. IV, we further verify the effectiveness of each component in HPL on PRCC. Fig. 3 also clearly shows that, with the help of HPL, the proposed method will pay more attention to non-clothing regions, *e.g.*, head, face and arms.

IV. CONCLUSION

In this paper, we propose a novel Hierarchical Proxy Learning (HPL) framework for CC-ReID. To get rid of the dependence on clothing labels, we propose a clustering-based automatic sub-proxy mining scheme. We further propose a Sample Balance and Diversity (SBD) module, which can generate auxiliary person examples for better ReID performance. Meanwhile, we propose a Hierarchical Proxy Learning (HPL) to extract clothes-irrelevant and person-invariant features. Extensive experiments demonstrate the superiority of our method.

References

[1] S. Gao, C. Yu, P. Zhang, and H. Lu, "Part representation learning with teacher-student decoder for occluded person re-identification," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2024, pp. 2660–2664.

[2] C. Yu, X. Liu, Y. Wang, P. Zhang, and H. Lu, "Tf-clip: Learning text-free clip for video-based person re-identification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 7, 2024, pp. 6764–6772.

[3] Z. Ran, X. Lu, and W. Liu, "Anomaly-aware semantic self-alignment framework for video-based person re-identification," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2024, pp. 2820–2824.

[4] X. Liu, P. Zhang, C. Yu, H. Lu, and X. Yang, "Watching you: Global-guided reciprocal learning for video-based person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13 334–13 343.

[5] X. Liu, C. Yu, P. Zhang, and H. Lu, "Deeply coupled convolution–transformer with spatial–temporal complementary learning for video-based person re-identification," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–11, 2023.

[6] G. Zhang, J. Liu, Y. Chen, Y. Zheng, and H. Zhang, "Multi-biometric unified network for cloth-changing person re-identification," *IEEE Transactions on Image Processing*, vol. 32, pp. 4555–4566, 2023.

[7] P. Zhang, J. Xu, Q. Wu, Y. Huang, and X. Ben, "Learning spatial-temporal representations over walking tracklet for long-term person re-identification in the wild," *IEEE Transactions on Multimedia*, vol. 23, pp. 3562–3576, 2020.

[8] J. Wu, H. Liu, W. Shi, H. Tang, and J. Guo, "Identity-sensitive knowledge propagation for cloth-changing person re-identification," in *Proceedings of the IEEE International Conference on Image Processing*, 2022, pp. 1016–1020.

[9] Y. Yan, H. Yu, S. Li, Z. Lu, J. He, H. Zhang, and R. Wang, "Weakening the influence of clothing: Universal clothing attribute disentanglement for person re-identification," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2022, pp. 1523–1529.

[10] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 14 278–14 287.

[11] X. Li, S. Wei, J. Wang, Y. Du, and M. Ge, "Adaptive multi-proxy for remote sensing image retrieval," *Remote Sensing*, vol. 14, no. 21, pp. 5615–5621, 2022.

[12] Z. Yang, M. Bastan, X. Zhu, D. Gray, and D. Samaras, "Hierarchical proxy-based loss for deep metric learning," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1859–1868.

[13] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv:1708.04552*, 2017.

[14] M. Ye, J. Shen, X. Zhang, P. C. Yuen, and S.-F. Chang, "Augmentation invariant and instance spreading feature for softmax embedding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 2, pp. 924–939, 2020.

[15] Z. Zhang, C. Lan, W. Zeng, Z. Chen, and S.-F. Chang, "Rethinking classification loss designs for person re-identification with a unified view," *arXiv:2006.04991*, 2020.

[16] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," *arXiv:1703.07737*, 2017.

[17] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 18 661–18 673, 2020.

[18] H. Chen, B. Lagadec, and F. Bremond, "Ice: Inter-instance contrastive encoding for unsupervised person re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 14 960–14 969.

[19] X. Li, B. Liu, Y. Lu, Q. Chu, and N. Yu, "Cloth-aware center cluster loss for cloth-changing person re-identification," in *Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision*. Springer, 2022, pp. 527–539.

[20] Y. Ge, F. Zhu, D. Chen, R. Zhao *et al.*, "Self-paced contrastive learning with hybrid memory for domain adaptive object re-id," *Advances in Neural Information Processing Systems*, vol. 33, pp. 11 309–11 321, 2020.

[21] Y. Huang, Q. Wu, J. Xu, Y. Zhong, and Z. Zhang, "Clothing status awareness for long-term person re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 11 895–11 904.

[22] M. Wang, B. Lai, J. Huang, X. Gong, and X.-S. Hua, "Camera-aware proxies for unsupervised person re-identification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 4, 2021, pp. 2764–2772.

[23] X. Gu, H. Chang, B. Ma, S. Bai, S. Shan, and X. Chen, "Clothes-changing person re-identification with rgb modality only," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1060–1069.

[24] P. Li, Y. Xu, Y. Wei, and Y. Yang, "Self-correction for human parsing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 3260–3271, 2022.

[25] J. Chen, X. Jiang, F. Wang, J. Zhang, F. Zheng, X. Sun, and W.-S. Zheng, "Learning 3d shape feature for texture-insensitive person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8146–8155.

[26] J. Mu, Y. Li, J. Li, and J. Yang, "Learning clothes-irrelevant cues for clothes-changing person re-identification," in *Proceedings of the British Machine Vision Conference*, 2022, pp. 1–15.

[27] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226–231.

[28] 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 *Proceedings of the European Conference on Computer Vision*, 2018, pp. 480–496.

[29] R. Hou, B. Ma, H. Chang, X. Gu, S. Shan, and X. Chen, "Interaction-and-aggregation network for person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9317–9326.

[30] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 513–10 522.

[31] X. Shu, G. Li, X. Wang, W. Ruan, and Q. Tian, "Semantic-guided sampling for cloth-changing person re-identification," *IEEE Signal Processing Letters*, vol. 28, pp. 1365–1369, 2021.

[32] Q. Yang, A. Wu, and W.-S. Zheng, "Person re-identification by contour sketch under moderate clothing change," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 2029–2046, 2019.

[33] F. Wan, Y. Wu, X. Qian, Y. Chen, and Y. Fu, "When person re-identification meets changing clothes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 830–831.

[34] K. Zhu, H. Guo, Z. Liu, M. Tang, and J. Wang, "Identity-guided human semantic parsing for person re-identification," in *Proceedings of the European Conference on Computer Vision*. Springer, 2020, pp. 346–363.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[36] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[37] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 13 001–13 008.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.