Long-Term Cloth-Changing Person Re-identification (Supplementary Material)

Xuelin Qian[1], Wenxuan Wang[1], Li Zhang[2], Fangrui Zhu[2], Yanwei Fu[2], Tao Xiang[3], Yu-Gang Jiang[1], and Xiangyang Xue[1,2]

[1] School of Computer Science, Shanghai Key Lab of Intelligent Information Processing, Fudan University
{xlqian15,wxwang19,ygj,xyxue}@fudan.edu.cn
[2] School of Data Science, and MOE Frontiers Center for Brain Science, Shanghai Key Lab of Intelligent Information Processing, Fudan University
{lizhangfd,18210980021,yanweifu}@fudan.edu.cn
[3] University of Surrey
t.xiang@surrey.ac.uk

1 More details on BIWI dataset

Due to space limitation, in the main paper, we only reported results on BIWI dataset. Here, we discuss more about this existing and smaller cloth-changing re-ID dataset and the implementation details.

BIWI dataset. BIWI [1] is a small-scale person re-identification dataset with cloth changes. It contains 50 identities, 28 of where appeared in two outfits. It is collected from two cameras and people are asked to perform a certain routine of motions in front of the cameras such as, a rotation around the vertical axis, several head movements (*i.e.*, the action of 'still') or walking towards the camera (*i.e.*, the action of 'walk'). The dataset consists of three subsets: 'Train' subset with images of all 50 identities from one camera, 'Still' and 'Walking' subsets of both with images of 28 cloth-changing people with above two actions from another camera. Some examples are shown in Fig. 5(d).

Implementation details. We randomly select 14 cloth-changing identities as training and the rest of 14 for testing. Considering the small-scale nature of BIWI, we pretrain all models on the LTCC dataset, and then fine-tune them on the training set of BIWI for 80 epochs with initial learning rate of 0.001 and decay rate of 0.1 for every 30 epochs. At the testing stage, we set the test images from 'Train' subset as query images, and those from 'Still' or 'Walking' subset as gallery images. Therefore, our experiment includes two test settings: Still and Walking setting. The evaluation metrics used in [2, 3] are adopt to calculate the scores of Rank-k under the multi-shot setting.

2 Results on Celeb-reID dataset

We further carry out new experiments on Celeb-reID [4, 5], and compare with several strong competitors, including MGN [6], ReIDCaps [5] and HACNN [7].

Celeb-reID dataset. Celebrities-reID dataset [4, 5] contains 1052 IDs with 34, 186 images. The images are collected by the street snap-shots of celebrities on Internet. Considering the same person usually does not wear the same clothing twice in snap-shots, this dataset is also suitable for the clothes variation person re-ID study. Specifically, more than 70% of the images of each person show different clothes on average. Some examples are shown in Fig. 5(e).

Implementation details. Following the standard training and testing setting in [5], we use 632 identities as training set and 420 identities for testing. Besides, we set a random unique number for each training image as the cloth label since it doesn't contain any clothes annotation, and replace ResNet-50 backbone with DenseNet121 for a fair comparison.

Results. As shown in Table 1, our model surprisingly gets Rank1/mAP 50.9%/9.8% with the dummy cloth label, while PCB 37.1%/8.2%, HACNN 47.6%/9.5% and ReIDCaps 51.2%/9.8%. It clearly suggests the effectiveness of our model.

Table 1. Results on Celeb-reID dataset. '$*$' denotes the results are reported under the setting of 'without human body parts partition'.

Methods	Rank-1	Rank-5	mAP
IDE+ (DenseNet121) [8]	42.9	56.4	5.9
PCB [9]	37.1	57.0	8.2
HACNN [7]	47.6	63.3	9.5
ResNet-Mid [10]	43.3	54.6	5.8
Two-Stream [11]	36.3	54.5	7.8
MLFN [12]	41.4	54.7	6.0
MGN [6]	49.0	64.9	**10.8**
ReIDCaps* [5]	**51.2**	65.4	9.8
Ours	50.9	**66.3**	9.8

3 Evaluations on model generalization

Intuitively, the body shape information should be more robust to the domain gap than other appearance information, *e.g.*, clothing, since the distribution of appearance can be easily changed by illumination or camera viewing conditions. Here, we conduct experiments by directly applying the best model on LTCC dataset to Market-1501 and DukeMTMC-reID without any fine-tuning.

The results are listed in Tab. 2, and our model significantly outperforms other approaches. For baseline 'ResNet-50' and 'PCB', 'Ours' outperforms them with a large margin of more than 6% accuracy improvement of Rank-1 in both settings. To be notice, our model beats all of other methods, considering MuDeep and OSNet both can achieve good performance under cross-domain task, which further indicates that our method has strong generalization ability.

Table 2. Results of the models under cross-domain Re-ID, that is, trained on LTCC dataset and evaluated directly on Market-1501 and DukeMTMC-reID. '†' denotes that only the images with cloth changes are used for training.

Methods	LTCC → Market-1501 [13]				LTCC → DukeMTMC-reID [14]			
	Standard Setting		Standard Setting†		Standard Setting		Standard Setting†	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
ResNet-50 [15]	24.70	9.57	22.06	7.89	16.87	6.26	13.64	4.90
PCB [9]	31.15	13.35	28.47	11.32	18.22	7.91	14.72	6.03
HACNN [7]	26.90	10.35	23.18	8.29	17.35	6.82	13.94	5.76
MuDeep [16]	29.36	11.22	22.27	8.21	18.53	7.65	14.27	5.89
OSNet [17]	34.33	15.59	32.77	14.01	13.15	8.92	21.49	7.87
Ours	**37.38**	**16.97**	**34.44**	**14.09**	**24.15**	**9.80**	**23.47**	**9.47**

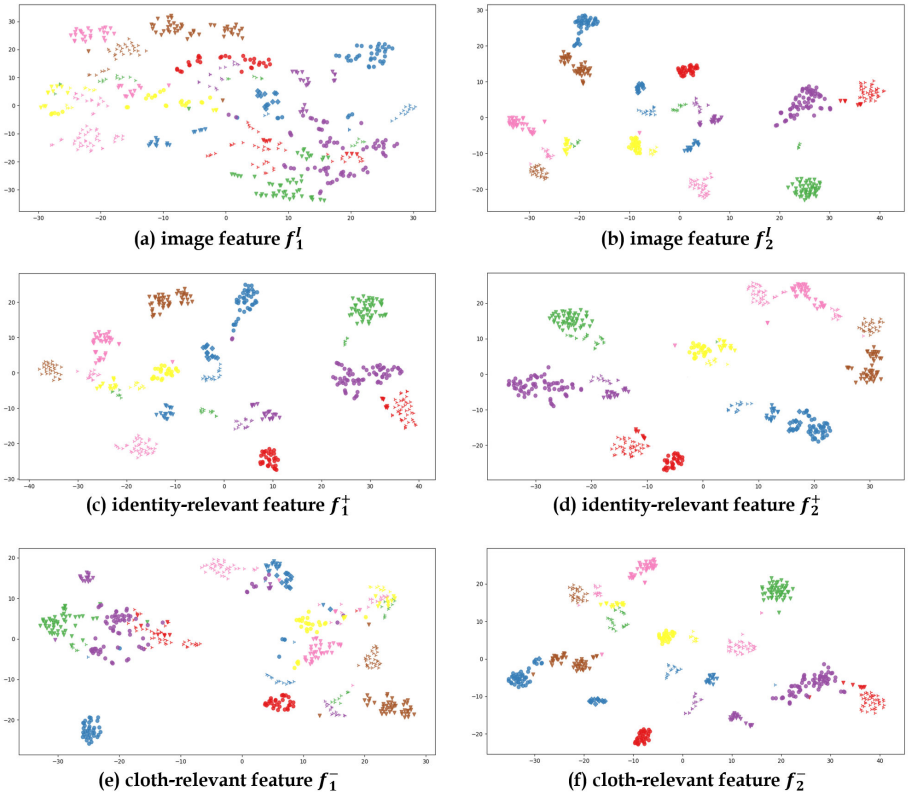

Fig. 1. Visualization of six different features extracted from our model using t-SNE [18]. The f_1^I, f_1^+ and f_1^- come from the first CESD module, and the f_2^I, f_2^+ and f_2^- are generated from the second CESD module. Each color represents one identity which is randomly selected from the testing set, and each symbol (circle, rhombus, triangle, *etc.*) with various color indicates different clothes. Best viewed in color and zoom in.

4 More visualizations and discussions

Visualization of features learned from CESD. Due to space limitation, we only show the distribution of two disentangled features, the identity-relevant feature f^+ and the cloth-relevant feature f^- from the last CESD module in the main paper. Here, we further visualize all the features associated with two CESD modules in our model to comprehensively analyze the importance and effectiveness of our proposed CESD module. Specially, we denote the input feature, two outputs of the identity-relevant feature and the cloth-relevant feature in the first CESD module (followed by *res3* block) as f_1^I, f_1^+ and f_1^-, respectively. Similarly, the associated features in the second CESD module (followed by *res4* block) are defined as f_2^I, f_2^+ and f_2^-. As shown in Fig. 1, we add more identities in the testing set for visualization, each color represents one identity, and each symbol (*e.g.*, circle, rhombus, triangle, cross) with various color indicates different clothes. From Fig. 1, we can make the following observations:

First, the distribution of features in (a) is relatively more chaotic. After going through the first CESD module, it becomes ordered and have the rudiment of clustering, as shown in (c) and (b). The final output features in (d) are clearly aggregated according to the identity information. It strongly indicates the effectiveness of our proposed CESD module in tackling the problem of LTCC Re-ID.

Secondly, from top to bottom (*i.e.*, (a)-(c)-(e) and (b)-(d)-(f)), we observe that our proposed CESD module is able to disentangle the cloth-relevant feature from the input image feature. In the space of the identity-relevant feature f_i^+, different symbols with the same color are grouped together based on the information of identities. On the contrary, the images with similar clothes are clustered regardless of identities in the space of the cloth-relevant feature f_i^-.

Lastly, from left to right (*i.e.*, (a)-(b), (c)-(d) and (e)-(f)), it is shown that applying two CESD modules can better remove the cloth-sensitive features and distill more discriminative features for person Re-ID. It is not only because the deeper features are more relevant to the specific task, but also because our proposed CESD module plays a key role of cloth-elimination and shape-distillation.

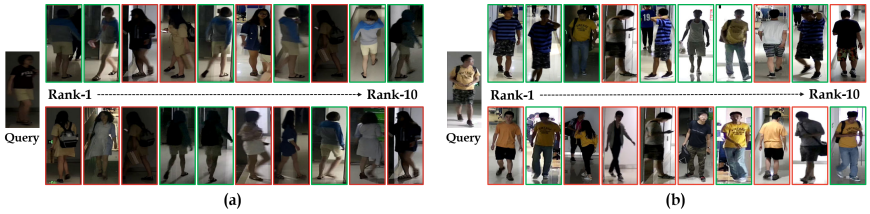

Fig. 2. Visualization of retrieval results generated by 'Ours' (upper) and 'ResNet-50' (lower) under the cloth-changing setting. The images with green borders are the correct results, those with red border are wrong. Best viewed in color and zoomed in.

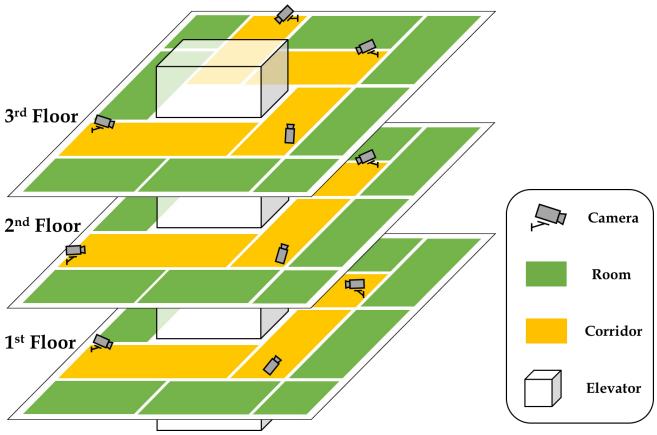

Fig. 3. The plan of the camera layout for collecting LTCC data.

Visualization of retrieval samples. To intuitively demonstrate the ability of our proposed framework in addressing the task of LTCC Re-ID, given same query images, we visualize the top 10 ranked retrieval results of our full model and ResNet-50 under the cloth-changing setting in Fig. 2. We can discover that our proposed model can better match the required images with different clothes. For example, the top-2 images in Fig. 2 (b), which are correctly retrieved by our model, are dressed in different clothes. On the contrary, the ResNet-50 model pays more attention to the similar color and type of yellow shorts. Interestingly, we notice from Fig. 2 (a) that on account of task-driven training, the ResNet-50 model also can learn some cloth-insensitive features. However, comparing the top-10 results, it cannot tackle the problem of LTCC Re-ID problem well. As a result, these clearly demonstrate that our method devotes more to the shape information rather than the appearance, so it can solve the dramatic changes of appearance caused by cloth changes to some extent.

5 More details on LTCC dataset

More details on data collection. To collect uniform and diverse data for long-term cloth-changing Re-ID, we utilize an existing CCTV system to record videos in 24 hours a day over two months. This CCTV system contains twelve cameras installed on three floors in an office building, shown in Fig. 3. After carefully annotation work, we release the first version of LTCC dataset, which contains 17,138 person images of 152 identities with 478 outfits. More identities will be followed in the future.

Comparison with Previous Datasets. As shown in Fig. 4, our proposed LTCC dataset contains cloth-changing identities with huge occlusion, illumination, camera view, carrying and pose variations, which is more realistic for the study of long-term person re-identification. We also compare our LTCC

Fig. 4. Samples from LTCC dataset. Given the queries, we show the corresponding gallery images with the same clothing, and five different outfits under other variations.

Fig. 5. Samples from Market-1501, DukeMTMC-reID, PRCC, BIWI and Celeb-reID datasets.

dataset with several widely-used STCC Re-ID datasets (Market-1501 [13] and DukeMTMC-reID [14]) and two related cloth-changing Re-ID datasets (BIWI [1], Celeb-reID [4, 5] and PRCC [19]) in Fig. 5.

Specifically, (1) comparing with Market-1501 and DukeMTMC-reID, these two general STCC Re-ID datasets are limited in clothing, carrying, illumination and other variations for each person. (2) The BIWI dataset, which only contains 28 people with two different clothes, is collected under several strict constraints, making it nearly have no variations of view-angle, occlusion, carrying or illumination. (3) The Celeb-reID dataset contains the images from the street snap-shots of celebrities crawled on Internet. And considering that people are usually in the front view from the snap-shots, there only a few back view images are included in the collection. (4) Comparing with PRCC dataset[4], it contains less drastic clothing changes and bare hairstyle changes. Furthermore, with only 3 cameras instead of 12 in our LTCC dataset, it is limited in view-angle, carrying and illumination changes.

References

1. Munaro, M., Fossati, A., Basso, A., Menegatti, E., Van Gool, L.: One-shot person re-identification with a consumer depth camera. In: Person Re-Identification. (2014)
2. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition. (2014)
3. Li, W., Zhao, R., X.Wang: Human re-identification with transferred metric learning. In: Asian Conference on Computer Vision. (2012)
4. Huang, Y., Wu, Q., Xu, J., Zhong, Y.: Celebrities-reid: A benchmark for clothes variation in long-term person re-identification. In: International Joint Conference on Neural Networks. (2019)
5. Huang, Y., Xu, J., Wu, Q., Zhong, Y., Zhang, P., Zhang, Z.: Beyond scalar neuron: Adopting vector-neuron capsules for long-term person re-identification. IEEE Transactions on Circuits and Systems for Video Technology (2019)
6. Wang, G., Yuan, Y., Chen, X., Li, J., Zhou, X.: Learning discriminative features with multiple granularities for person re-identification. In: ACM International Conference on Multimedia. (2018)
7. Li, W., Zhu, X., Gong, S.: Harmonious attention network for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition. (2018)
8. Zheng, L., Zhang, H., Sun, S., Chandraker, M., Tian, Q.: Person re-identification in the wild. arXiv preprint arXiv:1604.02531 (2016)
9. Sun, Y., Zheng, L., Yang, Y., Tian, Q., Wang, S.: Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In: European Conference on Computer Vision. (2018)
10. Yu, Q., Chang, X., Song, Y.Z., Xiang, T., Hospedales, T.M.: The devil is in the middle: Exploiting mid-level representations for cross-domain instance matching. arXiv preprint arXiv:1711.08106 (2017)

[4] PRCC dataset is not yet available at the time of submission, so samples in Fig. 5(c) are directly obtained from [19] rather than from the actual dataset.

11. Zheng, Z., Zheng, L., Yang, Y.: A discriminatively learned cnn embedding for person reidentification. ACM Transactions on Multimedia Computing, Communications, and Applications (2017)
12. Chang, X., Hospedales, T.M., Xiang, T.: Multi-level factorisation net for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition. (2018)
13. Zheng, L., Shen, L., Tian, L., S.Wang, J.Wang, Tian, Q.: Scalable person re-identification: A benchmark. In: IEEE International Conference on Computer Vision. (2015)
14. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: IEEE International Conference on Computer Vision. (2017)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. (2015)
16. Qian, X., Fu, Y., Xiang, T., Jiang, Y.G., Xue, X.: Leader-based multi-scale attention deep architecture for person re-identification. IEEE Transactions on Pattern Analysis and Machine Intelligence (2019)
17. Zhou, K., Yang, Y., Cavallaro, A., Xiang, T.: Omni-scale feature learning for person re-identification. In: IEEE International Conference on Computer Vision. (2019)
18. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research (2008)
19. Yang, Q., Wu, A., Zheng, W.S.: Person re-identification by contour sketch under moderate clothing change. IEEE Transactions on Pattern Analysis and Machine Intelligence (2019)