# Clothes-Changing Person Re-identification with RGB Modality Only
## *Supplementary Materials*

Xinqian Gu[1,2], Hong Chang[1,2], Bingpeng Ma[2], Shutao Bai[1,2], Shiguang Shan[1,2], Xilin Chen[1,2]
[1]Institute of Computing Technology, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences

## A. Convergence Analysis

To facilitate the convergence analysis, we define the classification probability of true positive clothes class (ground truth clothes class of $f_i$), pseudo positive clothes class (with the same identity but different clothes from $f_i$), and negative clothes class (with different identities from $f_i$) as $p_{Pos}$, $p_{PPos}$, and $p_{Neg}$. In the first step of optimization, minimizing $\mathcal{L}_C$ will maximize $p_{Pos}$ and minimize $p_{PPos}$ and $p_{Neg}$. In the second step, minimizing $\mathcal{L}_{CA}$ will maximize $p_{Pos}$ and $p_{PPos}$, and minimize $p_{Neg}$. So the probability distribution of the convergence state should be $p_{Pos} \gg p_{PPos} \gg p_{Neg}$. To verify this, we count the average probability for these three types of clothes classes of all samples on LTCC at the last training epoch. As shown in Fig. 1, $p_{Pos}$ of most samples converges to 0.6∼1; $p_{PPos}$ converges to 1e-4∼0.1; and $p_{Neg}$ converges to 1e-5∼1e-2. The convergence results are consistent with our analysis.

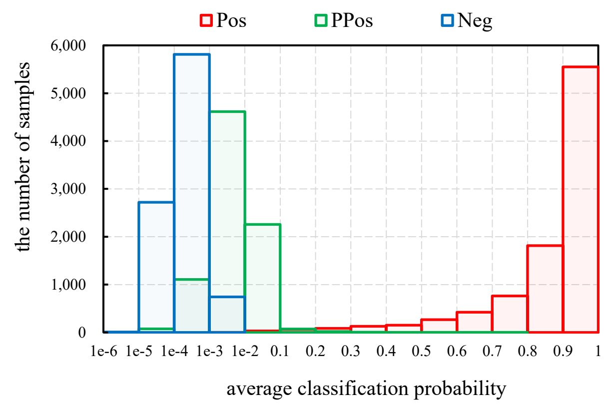


Figure 1. The statistics of the average probability for three types of clothes classes on LTCC at the last training epoch. Best viewed in color.

## B. Experiments on VC-Clothes

VC-Clothes [10] is a virtual dataset synthesized by GTA5. It contains 19,060 images from 512 identities and 4 cameras (scenes). Each identity has 1∼3 suits of clothes and all samples of each identity captured by camera 2&3 must wear the same clothes. Hence, most current works [3, 10] report the results on the subset from camera 2&3 as the accuracy in the same-clothes (SC) setting. Besides, they report the results on the subset from camera 3&4 as the accuracy in clothes-changing (CC) setting. To make a fair comparison, we follow the settings in these works.

We compare the proposed CAL with two single-modality-based (RGB) re-id methods (*i.e.* MDLA [6] and PCB [9]) and three multi-modality-based re-id methods (*i.e.* Part-aligned [8], FSAM [3], and 3DSL [1]) on VC-Clothes in Tab. 1. It can be seen that using RGB images only, the proposed CAL outperforms the baseline and all these state-of-the-art methods in general, the same-clothes, and clothes-changing settings. This comparison can demonstrate the effectiveness of CAL. Since these state-of-the-art methods do not report the accuracy in the same-clothes and clothes-changing settings on the whole dataset from all cameras, we only compare our method with the baseline in these settings in Tab. 2. The experimental results show CAL outperforms the baseline, especially in clothes-changing setting.

Table 1. Comparison with state-of-the-arts on VC-Clothes.

| method | general (all cams) | | SC (cam2&cam3) | | CC (cam3&cam4) | |
|---|---|---|---|---|---|---|
| | top-1 | mAP | top-1 | mAP | top-1 | mAP |
| MDLA [6] | 88.9 | 76.8 | 94.3 | 93.9 | 59.2 | 60.8 |
| PCB [9] | 87.7 | 74.6 | 94.7 | 94.3 | 62.0 | 62.2 |
| Part-aligned [8] | 90.5 | 79.7 | 93.9 | 93.4 | 69.4 | 67.3 |
| FSAM [3] | - | - | 94.7 | 94.8 | 78.6 | 78.9 |
| 3DSL [1] | - | - | - | - | 79.9 | 81.2 |
| baseline | 88.3 | 79.2 | 94.1 | 94.3 | 67.3 | 67.9 |
| CAL | **92.9** | **87.2** | **95.1** | **95.3** | **81.4** | **81.7** |

Table 2. The results in the same-clothes and clothes-changing settings for the data from all cameras on VC-Clothes.

| method | SC (all cams) | | CC (all cams) | |
|---|---|---|---|---|
| | top-1 | mAP | top-1 | mAP |
| baseline | 94.5 | 93.9 | 74.2 | 66.5 |
| CAL | **96.0** | **95.7** | **85.8** | **79.8** |

Table 3. Comparison with state-of-the-art methods on LaST and DeepChange.

| method | LaST | | DeepChange | |
|---|---|---|---|---|
| | top-1 | mAP | top-1 | mAP |
| OSNet [12] | 63.8 | 20.9 | 39.7 | 10.3 |
| ReIDCaps [4] | - | - | 39.5 | 11.3 |
| BoT [5] | 68.3 | 25.3 | 47.5 | 13.0 |
| mAPLoss [7] | 69.9 | 27.6 | - | - |
| baseline | 69.4 | 25.6 | 50.6 | 15.9 |
| CAL | **73.7** | **28.8** | **54.0** | **19.0** |

## C. Experiments on LaST and DeepChange

LaST [7] and DeepChange [11] are two large-scale long-term person re-id datasets. LaST provides clothes labels of the training set, while DeepChange does not provide any clothes labels. We notice that the collection date of each image on DeepChange is given. Since the samples of the same person captured on different days have a high probability of wearing different clothes, we attempt to use the collection date as pseudo clothes labels to train CAL. To make a fair comparison, we set batch size as 64 following [7] and each batch contains 16 persons and 4 images for each person. On LaST, triplet loss [2] is used as the metric learning loss for both baseline and CAL (removing it will cause severe performance degradation). Finally, we report the performance in general setting to compare with state-of-the-art methods. Especially, on DeepChange, we allow the true matches coming from the same camera but different tracklets as query following [11].

The comparisons with the baseline and state-of-the-art methods on LaST and DeepChange are shown in Tab. 3. It can be seen that the proposed CAL outperforms the baseline and these state-of-the-art methods significantly. Especially, on DeepChange, when the collection date is used as pseudo clothes label, CAL still works well. It demonstrates that the proposed method does not rely on accurate clothes annotations.

## References


[1] Jiaxing Chen, Xinyang Jiang, Fudong Wang, Jun Zhang, Feng Zheng, Xing Sun, and Wei-Shi Zheng. Learning 3d shape feature for texture-insensitive person re-identification. In *CVPR*, 2021. 1

[2] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 2

[3] Peixian Hong, Tao Wu, Ancong Wu, Xintong Han, and Wei-Shi Zheng. Fine-grained shape-appearance mutual learning for cloth-changing person re-identification. In *CVPR*, 2021. 1

[4] Yan Huang, Jingsong Xu, Qiang Wu, Yi Zhong, Peng Zhang, and Zhaoxiang Zhang. Beyond scalar neuron: Adopting vector-neuron capsules for long-term person re-identification. *TCSVT*, 30(10):3459–3471, 2020. 2

[5] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *CVPR Workshops*, 2019. 2

[6] Xuelin Qian, Yanwei Fu, Yu-Gang Jiang, Tao Xiang, and Xiangyang Xue. Multi-scale deep learning architectures for person re-identification. In *ICCV*, 2017. 1

[7] Xiujun Shu, Xiao Wang, Shiliang Zhang, Xianghao Zhang, Yuanqi Chen, Ge Li, and Qi Tian. Large-scale spatio-temporal person re-identification: Algorithm and benchmark. *arXiv preprint arXiv: 2105.15076*, 2021. 2

[8] Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, and Kyoung Mu Lee. Part-aligned bilinear representations for person re-identification. In *ECCV*, 2018. 1

[9] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *ECCV*, 2018. 1

[10] Fangbin Wan, Yang Wu, Xuelin Qian, and Yanwei Fu. When person re-identification meets changing clothes. In *CVPR Workshop*, 2020. 1

[11] Peng Xu and Xiatian Zhu. Deepchange: A large long-term person re-identification benchmark with clothes change. *arXiv preprint arXiv:2105.14685*, 2021. 2

[12] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. In *ICCV*, 2019. 2