DeepChange:  
A Long-Term Person Re-Identification Benchmark

XU Peng  
peng.xu@aliyun.com
ZHU Xiatian  
eddy.zhuxt@gmail.com

Abstract
Existing person re-identification (Re-ID) works mostly consider a short-term search problem assuming unchanged clothes and personal appearance. However, in real-world we often dress ourselves differently across locations, time, dates, seasons, weather, and events. As a result, the existing methods are unsuitable for long-term person Re-ID with clothes change involved. Whilst there are several recent long-term Re-ID attempts, a large realistic dataset with clothes change is lacking and indispensable for enabling extensive study as already experienced in short-term Re-ID setting. In this work, we contribute a large, realistic long-term person re-identification benchmark. It consists of $178K$ bounding boxes from $1.1K$ person identities, collected and constructed over 12 months. Unique characteristics of this dataset include: (1) Natural/native personal appearance (e.g., clothes and hair style) variations: The clothes-change and dressing styles are all highly diverse, with the reappearing gap in time ranging from minutes, hours, and days to weeks, months, seasons, and years. (2) Diverse walks of life: Persons across a wide range of ages and professions appear in different weather conditions (e.g., sunny, cloudy, windy, rainy, snowy, extremely cold) and events (e.g., working, leisure, daily activities). (3) Rich camera setups: The raw videos were recorded by 17 outdoor security cameras with various resolutions operating in a real-world surveillance system for a wide and dense block. (4) Largest scale: It covers the largest number of (17) cameras, (1, 121) identities, and (178, 407) bounding boxes, as compared to alternative datasets. Further, we conduct a large spectrum of experiments with existing deep networks and state-of-the-art person Re-ID models on this newly introduced benchmark and provide insightful observations and analysis with more realistic performance metrics. Our dataset and benchmark codes are available on 
https://github.com/PengBoXiangShang/deepchange

1 Introduction
Person re-identification (Re-ID) aims to match the person identity classes of bounding box images extracted from non-overlapping camera views [13, 54, 47]. Extensive Re-ID models were developed in the past decade [33, 55, 14, 17, 30, 3, 4, 27, 26, 59, 58, 49, 50, 23, 52, 45, 56, 38, 57, 48, 6, 51, 2]. The majority of these methods consider the short-term search scenario with a strong assumption that the appearance (e.g., clothes, hair style) of each person is stationary. We call this conventional setting as short-term Re-ID in the follows. Unfortunately, this assumption would be easily broken once the search time span is enlarged to long periods (such as days, weeks, or even months) as an average person often changes the outfit during different day time and across different weathers, daily activities and social events. As shown in Figure 1, a specific person was dressed the same only in a short time (e.g., minutes or hours) but with different clothes/hairs and associations over a long time and across seasons and weathers. Relying heavily on the clothes appearance, the previous short-term Re-ID models are unsuitable and ineffective in dealing with unconstrained clothes changes over time.
Figure 1: **Motivation:** An average person often changes the appearance (e.g., clothes, hair) over time and location, as shown by three random persons (color coded) here. Conventional Re-ID settings usually assume stationary appearance and are hence valid only for short-term person search. For the long-term counterpart, we must consider the unique challenges imposed by unconstrained appearance changes. Note, for clarity only part of true (red arrow) and false (blue arrow) matches are plotted. Best viewed in color.

Recently, there have been a few studies for tackling the long-term Re-ID situations focusing on clothes change [46, 20, 34, 31, 40, 42, 43]. Since there are no appropriate datasets publicly available, to enable research these works introduced several small-scale long-term Re-ID datasets by using web celebrity images [21], synthesizing pseudo person identities [40, 31], collecting person images under simulated surveillance settings [46, 34, 40, 42]. While these dataset construction efforts are useful in initiating the research, it is obvious that a real, large-scale long-term Re-ID benchmark is missing and highly demanded. The impact and significance of such a benchmark in driving the research progress has been repeatedly demonstrated in the short-term Re-ID case [53, 29, 44, 43]. However, it is much more challenging to establish a large Re-ID dataset with clothes change as compared to those short-term counterparts. This is due to a couple of reasons: (1) More video data need to be collected and processed over long time periods; (2) Labeling person identity becomes much more difficult when a person is dressed up with different outfit from time to time. Regardless, we believe that it is worthwhile to overcome all these challenges.

To facilitate the research towards the applications of long-term person search in reality, we contribute the first large-scale person Re-ID dataset with native/natural appearance (mainly clothes) changes, termed as DeepChange. Different from existing datasets, DeepChange has several unique characteristics: (1) The raw videos are collected in a real-world surveillance camera network deployed at a residential block where rich scenes and realistic background are presented over time and space. (2) The videos cover a period of 12 months with a wide variety of different weathers and contexts. To the best of our knowledge, this is the longest time duration among all Re-ID datasets, presenting natural/native personal appearance (e.g., clothes and hair style) variations with people from all walks of life. (3) Compared to existing long-term Re-ID datasets, it contains the largest number of (17)
cameras, (1121) identities, and (178K) bounding boxes. Overall, DeepChange is the only realistic, largest long-term person Re-ID dataset, created using the real-world surveillance videos.

We make the following contributions: (1) A large scale long-term person Re-ID dataset, called DeepChange, is introduced. Compared with existing alternative datasets, this dataset defines more realistic and more challenging person Re-ID tasks over long time, characterized by native appearance changes. With much larger quantity of person images for model training, validation, and testing, this new benchmark provides a more reliable and indicative test bed for future research. (2) We conduct extensive experiments for performance benchmark on the DeepChange dataset, including widely used CNNs [16, 19, 36, 39], state-of-the-art long-term Re-ID models [21], and multi-modality models (e.g., gray images, edge maps [60], key points [1]). We provide in-depth model performance analysis. Besides, it is possible that negative social impact can be imposed if a Re-ID system trained with the DeepChange dataset is used improperly or immorally; However, preventing this from being happened involves complex and systematic regularization. On our side, what we should be cautious is in disseminating the DeepChange dataset and we will do our best in this regard.

2 Benchmark creation

Venue and climate Our raw videos were collected from a real-world surveillance system for a wide (14 hectares) and dense block, with the middle temperate continental monsoon climate. This venue has various scenes, including crowded streets, shops, restaurants, construction sites, residential buildings, physical exercise areas, car parking, sparsely populated corner, etc. Thus it has two major advantages: (i) Identity Diversity: Persons cross a wide range of ages and professions, e.g., lactating baby, very older person, office lady, elementary student, high school student, worker, deliveryman, religious. Some identity examples are illustrated in Figure 2. (ii) Weather Diversity: During our collecting, we have observed a temperature variation ranging from $-30^\circ C$ (in winter) to $35^\circ C$ (in summer). Therefore, we have collected persons appearing in various weathers, e.g., sunny, cloudy, windy, rainy, snowy, extremely cold. Some image samples with snow are shown in Figure 3, where the snow can cause noticeable appearance changes on both clothes or background. Altogether, the identity and weather diversities will be embodied in drastic appearance changes, enabling realistic long-term Re-ID benchmarking with our DeepChange dataset. Figure 4 demonstrates some bounding boxes of an identity randomly selected from our dataset, where we can observe dramatic appearance changes across weathers, seasons, etc.

Security camera Our raw videos were recorded by 17 security cameras with a speed of 25 FPS (frames per second) and different resolutions ($1920 \times 1080$ spear camera $\times 14$, $1280 \times 960$ spherical camera $\times 3$). These 17 cameras are part of a large-scale video surveillance system. In particular, these cameras are mounted on the exterior walls of buildings or on lamp posts, and their height is approximately 3 to 6 meters. These cameras are monitoring various views including crowded streets, construction sites, physical exercise areas, car parking, sparsely populated corner, etc. Therefore, these cameras provide diverse scenes, identities, behaviors, events, etc.

Video collecting Our collecting is a very long course over 12 months across two different calendar years. Every month, we collected raw videos on 7 to 10 randomly selected days to cover as much weather as possible. On each selected day, we collected video from dawn to night to record comprehensive light variations. A huge amount of videos can provide highly diverse clothes changes and dressing styles, with the reappearing gap in time ranging from minutes, hours, and days to weeks, months, seasons, and years. The permission of using these videos was granted by the owner/authority for research purposes only.

Labeling and pedestrian detection It is much more difficult to recognize persons who have changed clothes or hair styles in the surveillance videos, even by human eyeballing. To minimize mistakes and errors, before labeling, we need to pay vast efforts to watch thousands of hours of videos to familiarize the persons recorded in the videos. Further, it is extremely challenging when persons wear masks, hats, or heavy winter clothes. During labeling, we inspected the labeled persons frequently to avoid duplicate identities assigned to the same person. We used Faster RCNN [35] to detect bounding boxes. For each selected bounding box, we annotated person ID, camera ID, tracklet ID, and time stamp.
Figure 2: Image samples of random identities in DeepChange. Identities from top left to bottom right: an aunt (bbox#1-#19), an office lady (bbox#20-#27), a pupil (bbox#28-#36), a newspaper delivery (bbox#37-#41), an older aunt (bbox#42-#49), a worker (bbox#50-#51), a nun (bbox#52-#53), a Muslim man (bbox#54-#56), a chef (bbox#57-#58), a disabled person (bbox#59-#60), a dustman (bbox#61-#70). Best viewed in color.
Figure 3: Images collected in snow (top two lines, bbox#1-#20) and rain (bottom line, bbox#21-#30) weather from DeepChange. Best viewed in color.

Statistics All identities were captured by at least two cameras, and most identities were captured by 2 ∼ 6 cameras (as shown in Figure 5(c)). Figure 5(c) indicates that the labeled bounding boxes are distributed from 6 am to 9 pm. As illustrated in Figure 5(f) the bounding box ratios of persons wearing spring&autumn, summer, and winter clothes are 59.97%, 32.99%, and 7.03%, respectively. More detailed statistics can be found in Figure 5.

Comparison with other datasets We compared the DeepChange dataset with existing Re-ID datasets with and without clothes changes in Table 1 and Table 2 respectively. Note that here we only discuss the publicly available and non-synthetic datasets. As summarized in Table 1, compared with existing long-term datasets, ours has the largest amounts of identities, cameras, and bounding boxes, providing the longest time range. DeepChange is the only one to offer four seasonal dressing styles. Moreover, when compared with previous traditional datasets, ours still has the largest amounts of cameras, and bounding boxes (see Table 2).

Splitting We shuffled all our collected identities, and then orderly picked 450, 150, and 521 IDs for training, validation, and test, respectively. In validation and test sets, given a tracklet, we randomly chose ∼ 5 bounding boxes as queries/probes, and the remaining boxes were split into the gallery. Details were summarized in Table 3.

Diversity and challenge As aforementioned, this wide (14 hectares) and dense block provides various identities (as shown in Figure 5), and middle temperate continental monsoon climate causes diverse clothes changes (as demonstrated in Figure 4 and Figure 2). Our long-term video collection makes full use of these characteristics of this venue and the climate. We observed that obvious hair-style changes often happened in long-term surveillance videos. In Figure 6, we present some random cases with simultaneous clothes and hair style changes. It is interesting to see that hair style changes should also be considered in long-term Re-ID, as this might lead to non-neglectable appearance alternation.

3 Experiments

In this section, we benchmarked the performance of commonly adopted deep CNN models and state-of-the-art methods on our DeepChange dataset.
Protocols and metrics  In the traditional short-term person Re-ID, it is assumed that the appearance of a specific person does not change across time and cameras. However, this is not true for long-term setting. Hence, in our benchmark we allow the true matches coming from the same camera as the probe/query image but from a different time and trajectory. Following \[53, 37, 44, 10, 8, 11\], we used both Cumulated Matching Characteristics (CMC) and mean average precision (mAP) as retrieval accuracy metrics.

Baselines  We evaluated several commonly used CNN in Re-ID models, including ResNet \[16\], Inceptionv3 \[39\], DenseNet \[19\], MobileNetv2 \[36\], ReIDCaps \[21\] is a capsule network designed
Table 1: Comparison with existing long-term image-based Re-ID datasets with clothes change. (‘Fas.’: Faster RCNN [35], ‘Mas.’: Mask RCNN [15], ‘ind.’: indoor, ‘out.’: outdoor, ‘CD’: cross-day, ‘CM’: cross-month, ‘CS’: cross-season, ‘CY’: cross-year, ‘spr.’: spring, ‘sum.’: summer, ‘aut.’: autumn, ‘win.’: winter, ‘sim.’: simulated, ‘sur.’: surveillance, ‘-’: unknown)

| dataset      | # person | # bbox | # cam | source | detector | scene | time range | cloth style | course | CD | CM | CS | CY | spr. | sum. | aut. | win. |
|--------------|----------|--------|-------|--------|----------|-------|------------|-------------|--------|----|----|----|----|------|------|------|------|
| Real28 [40]  | 28       | 4.3K   | 4 sim.| Mas.   | out., ind.| 3 days| ✓          | ✓           | ✓      | ✓  | ✓  | ✓  | ✓  | ✓    | ✓    | ✓    | ✓    |
| NKUP [42]    | 107      | 9.7K   | 15 sim.| -      | out., ind.| 4 months| ✓          | ✓          | ✓      | ✓  | ✓  | ✓  | ✓  | ✓    | ✓    | ✓    | ✓    |
| LTCC [34]    | 152      | 17K    | 12 sim.| Mas.   | ind.     | 2 months| ✓          | ✓          | ✓      | ✓  | ✓  | ✓  | ✓  | ✓    | ✓    | ✓    | ✓    |
| PRCC [46]    | 221      | 33K    | 3 sim.| -      | ind.     | ✓      | ✓          | ✓          | ✓      | ✓  | ✓  | ✓  | ✓  | ✓    | ✓    | ✓    | ✓    |
| DeepChange   | 1,121    | 178K   | 17 sur.| Fas.   | out.     | 12 months| ✓          | ✓          | ✓      | ✓  | ✓  | ✓  | ✓  | ✓    | ✓    | ✓    | ✓    |

Table 2: Comparison with conventional short-term image-based Re-ID datasets without clothes change. (‘Fas.’: Faster RCNN [35], ‘DPM’: Deformable Part Model [9], ‘ind.’: indoor, ‘out.’: outdoor)

| dataset      | DeepChange | MSMT17 [44] | Duke [50] | Market [53] | CUHK03 [29] | CUHK01 [28] | VIPeR [14] | PRID [18] | CA VIAR [5] |
|--------------|------------|-------------|-----------|-------------|-------------|-------------|-----------|-----------|-------------|
| # person     | 1,121      | 4,101       | 1,812     | 1,501       | 1,467       | 971         | 632       | 934       | 72          |
| # bbox       | 178K       | 126K        | 36K       | 32K         | 26K         | 3.3K        | 1.2K      | 1.1K      | 0.6K        |
| # camera     | 17         | 15          | 8         | 6           | 2           | 10          | 2         | 2         | 2           |
| detector     | Fas.       | Fas.        | hand      | DPM         | DPM, hand   | hand        | hand      | hand      | hand        |
| scene        | out.       | out. & ind. | out.      | out.        | ind.        | out.        | out.      | ind.      |             |

Table 3: Dataset splitting of DeepChange.

| train set | validation set | test set |
|-----------|----------------|----------|
| # person  | 450            | 521      |
| # bbox    | 75,083          | 150 probe: 4,976 | gallery: 17,865 |

for long-term person Re-ID. Moreover, we designed some multi-branch network baselines, taking multi-modal inputs for exploiting the complementary effect.
Implementation details  For a fair comparison, all experiments were implemented in PyTorch [32] with Adam [24] optimizer and run on a single Nvidia TITAN Xp GPU card. During training, only random horizontal flipping was used for data augmentation. For reproducibility, we conducted a unified early stop strategy for all experiments. We empirically adopted mAP on evaluation set as early stop metric and set patience as 30 epochs, thus the checkpoints with the highest validation performance were chosen to report test performance. As all the backbones were initialized by the weights pretrained on ImageNet [7], we empirically used an initial learning rate of $1 \times 10^{-4}$, multiplied by a decay factor 0.1 every 20 epochs. ReIDCaps [21] uses a customized loss function, while the other models were trained by minimizing the softmax based cross-entropy loss. During evaluating and testing, we extracted the features of bounding boxes from the penultimate layer to conduct Re-ID matching. Edge Box [60] and Open Pose [11] toolboxes were used to extract edge maps and detect body keypoints (15 keypoints, 54D vector), respectively. In our multi-branch baselines, body keypoint vectors were passed into a simple module of two fully-connected layers with Batch Normalization [22] and ReLU [12] activations, while CNN backbones were used to encode pixel space inputs, i.e., RGB bounding boxes, grayscale bounding boxes, edge maps.

Results and discussion  In Table 4, we make the following observations: (i) For the results with RGB input, deeper models often work better than shallower models as expected. For example, we can see a clear upward trend from #1 ResNet18 (rank@1: 34.45%, mAP: 08.44%) to #9 ResNet152 (rank@1: 39.84%, mAP: 11.49%) with ResNet architectures. Meanwhile, it is observed that ResNet and DenseNet outperform MobileNetv2 and Inceptionv3, by a clear margin. As a dedicated long-term Re-ID model, ReIDCaps [21] obtains the best mAP performance with a strong backbone network DenseNet121 (#12). However, the benefit of using capsule layers is considerable (#12 vs. #16). (ii) For the results with different modalities: #5 ResNet50+RGB, #6 ResNet50+grayscale, and #7 ResNet50+edge map, it is observed that RGB outperforms grayscale and edge map. Whilst grayscale images may be more tolerant with clothes change, they also deliver less information due to lacking
Table 4: Person Re-ID performance comparisons on the test set of DeepChange. Both rank accuracy (%) and mAP (%) are reported. ('R': RGB, 'G': grayscale, 'E': edge map, 'K': body key point, '2br': two branches, '3br': three branches)

| network/model | input modalities | dimensions | batch size | rank @1  | rank @5  | rank @10 | rank @20 | mAP |
|---------------|-----------------|------------|------------|---------|---------|---------|---------|-----|
| #1 ResNet18 [16] | R | 256×128 | 256 | 34.45 | 46.01 | 51.72 | 58.26 | 08.44 |
| #2 ResNet18 [16] | G | 256×128 | 256 | 26.61 | 39.02 | 45.45 | 53.06 | 05.49 |
| #3 ResNet34 [16] | R | 256×128 | 256 | 35.21 | 47.37 | 53.61 | 60.03 | 09.49 |
| #4 ResNet34 [16] | G | 256×128 | 256 | 28.60 | 41.53 | 47.98 | 54.87 | 06.39 |
| #5 ResNet50 [16] | R | 256×128 | 192 | 36.62 | 49.88 | 55.46 | 61.92 | 09.62 |
| #6 ResNet50 [16] | G | 256×128 | 192 | 30.04 | 43.12 | 49.82 | 57.04 | 06.96 |
| #7 ResNet50 [16] | E | 256×128 | 192 | 16.05 | 28.51 | 35.59 | 43.28 | 03.17 |
| #8 ResNet101 [16] | R | 256×128 | 128 | 39.31 | 51.65 | 57.36 | 63.72 | 11.00 |
| #9 ResNet152 [16] | R | 256×128 | 96 | 39.84 | 52.51 | 58.35 | 64.75 | 11.49 |
| #10 MobileNetv2 [36] | R | 256×128 | 24 | 44.29 | 56.44 | 62.01 | 68.01 | 13.25 |
| #11 Inceptionv3 [39] | R | 299×299 | 96 | 35.02 | 47.71 | 53.91 | 60.64 | 08.85 |
| #12 DenseNet121 [19] | R | 256×128 | 128 | 38.26 | 50.27 | 55.91 | 62.40 | 09.12 |
| #13 DenseNet161 [19] | R | 256×128 | 64 | 45.92 | 56.72 | 61.79 | 67.41 | 12.30 |
| #14 DenseNet169 [19] | R | 256×128 | 64 | 43.40 | 54.80 | 60.11 | 65.90 | 11.25 |
| #15 DenseNet201 [19] | R | 256×128 | 64 | 44.98 | 56.13 | 61.32 | 66.98 | 11.71 |
| #16 ReIDCaps [21] | R | 224×224 | 24 | 44.29 | 56.44 | 62.01 | 68.01 | 13.25 |
| #17 2br ResNet50 | R, K | 256×128 | 192 | 36.53 | 48.87 | 54.86 | 61.47 | 09.54 |
| #18 2br ResNet50 | R, E | 256×128 | 96 | 40.26 | 52.91 | 59.11 | 65.47 | 09.43 |
| #19 2br ResNet50 | R, G | 256×128 | 96 | 40.52 | 53.65 | 59.61 | 65.60 | 10.22 |
| #20 3br ResNet50 | R, G, E | 256×128 | 64 | 41.67 | 54.28 | 60.04 | 66.37 | 11.03 |
| #21 2br DenseNet121 | R, E | 256×128 | 64 | 44.55 | 56.40 | 62.03 | 67.85 | 11.21 |
| #22 2br DenseNet121 | R, G | 256×128 | 64 | 44.80 | 56.79 | 62.48 | 68.06 | 11.36 |
| #23 3br DenseNet121 | R, G, E | 256×128 | 32 | 45.36 | 57.36 | 62.91 | 69.29 | 11.73 |

color. Hence it is not necessary a stronger modality than RGB. Besides, the CNN model is pretrained on ImageNet data with color, which is a disadvantage for grayscale images. On 13.45% (23K out of 178K) bounding boxes, keypoint detection fails. This is not surprising as detecting body keypoint on low-resolution surveillance images itself is a challenging problem. Thus, we did not test keypoint modality in isolation. (iii) For the results with multi-modal inputs, it can be observed that edge map and grayscale inputs can bring performance gains for RGB (compare #5, #18, #19, and #20). This suggests that multi-modal fusion is promising for long-term Re-ID, as each individual modality focuses on some distinct aspects of personal appearance. It is expected that higher-quality edge maps would contribute more to the performance gain. Also, it is observed that using deeper networks (e.g., #21, #22, #23 by DenseNet121) can further improve the accuracy results along with multi-modal inputs. This means that multi-modal fusion and network architecture are two compatible aspects for searching improvement.

4 Conclusions

In this paper, we introduced the only realistic, large-scale long-term person Re-ID dataset, called DeepChange, aiming for facilitating the research towards more realistic person search applications without short-time constraints. Compared to existing alternative datasets, it is established uniquely using real-world surveillance video sources without any artificial simulation. Constructed with a huge amount of annotation efforts, DeepChange contains the largest number of cameras, identities, and bounding boxes, with the longest time period covered and native appearance (e.g., clothes, hair style) change involved. Further, we have conducted an extensive set of experiments using a variety of deep CNN models and state-of-the-art methods for providing performance evaluation on this new dataset, offering rich baselines for future research works.
5 Future work

In the future efforts, we would try to extend this work mainly in following aspects: (i) continually collecting video to provide longer term test bed, e.g., three years, (ii) annotating more person identities to enlarge training/validation/testing subsets and provide more complex appearance changes, (iii) accommodating more cameras to provide more view variety, (iv) creating a video-based long-term Re-ID dataset.

References

[1] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: realtime multi-person 2d pose estimation using part affinity fields. TPAMI, 2019.

[2] Xiaobin Chang, Timothy M Hospedales, and Tao Xiang. Multi-level factorisation net for person re-identification. In CVPR, 2018.

[3] Yanbei Chen, Xiatian Zhu, and Shaogang Gong. Person re-identification by deep learning multi-scale representations. In ICCVW, 2017.

[4] Ying-Cong Chen, Xiatian Zhu, Wei-Shi Zheng, and Jian-Huang Lai. Person re-identification by camera correlation aware feature augmentation. TPAMI, 2017.

[5] Dong Seon Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino. Custom pictorial structures for re-identification. In BMVC, 2011.

[6] Zhiyi Cheng, Qi Dong, Shaogang Gong, and Xiatian Zhu. Inter-task association critic for cross-resolution person re-identification. In CVPR, 2020.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In CVPR, 2009.

[8] Chanho Eom and Bumsub Ham. Learning disentangled representation for robust person re-identification. In NeurIPS, 2019.

[9] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2009.

[10] Yixiao Ge, Zhuowan Li, Haiyu Zhao, Guojun Yin, Shuai Yi, Xiaogang Wang, and Hongsheng Li. Fd-gan: Pose-guided feature distilling gan for robust person re-identification. In NeurIPS, 2018.

[11] Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, and Hongsheng Li. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In NeurIPS, 2020.

[12] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In AISTATS, 2011.

[13] Shaogang Gong, Marco Cristani, Shuicheng Yan, and Chen Change Loy. Person re-identification. Springer, January 2014.

[14] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In ECCV, 2008.

[15] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In ICCV, 2017.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016.

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

[18] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof. Person re-identification by descriptive and discriminative classification. In SCIA, 2011.

[19] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In CVPR, 2017.

[20] Yan Huang, Qiang Wu, Jingsong Xu, and Yi Zhong. Celebrities-reid: A benchmark for clothes variation in long-term person re-identification. In IJCNN, 2019.
[21] 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*, 2019.

[22] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015.

[23] Jiening Jiao, Wei-Shi Zheng, Ancong Wu, Xiatian Zhu, and Shaogang Gong. Deep low-resolution person re-identification. In *AAAI*, 2018.

[24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2014.

[25] Xu Lan, Hanxiao Wang, Shaogang Gong, and Xiatian Zhu. Deep reinforcement learning attention selection for person re-identification. In *BMVC*, 2017.

[26] Minxian Li, Xiatian Zhu, and Shaogang Gong. Unsupervised person re-identification by deep learning tracklet association. In *ECCV*, 2018.

[27] Minxian Li, Xiatian Zhu, and Shaogang Gong. Unsupervised tracklet person re-identification. *TPAMI*, 2019.

[28] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *ACCV*, 2012.

[29] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014.

[30] Wei Li, Xiatian Zhu, and Shaogang Gong. Harmonious attention network for person re-identification. In *CVPR*, 2018.

[31] Yu-Jhe Li, Zhengyi Luo, Xinshuo Weng, and Kris M Kitani. Learning shape representations for clothing variations in person re-identification. *arXiv preprint arXiv:2003.07340*, 2020.

[32] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019.

[33] Bryan James Prosser, Wei-Shi Zheng, Shaogang Gong, Tao Xiang, Q Mary, et al. Person re-identification by support vector ranking. In *BMVC*, 2010.

[34] Xuelin Qian, Wenxuan Wang, Li Zhang, Fangrui Zhu, Yanwei Fu, Tao Xiang, Yu-Gang Jiang, and Xiangyang Xue. Long-term cloth-changing person re-identification. In *ACCV*, 2020.

[35] Shaqoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NeurIPS*, 2015.

[36] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *CVPR*, 2018.

[37] Arulkumar Subramaniam, Moitreya Chatterjee, and Anurag Mittal. Deep neural networks with inexact matching for person re-identification. In *NeurIPS*, 2016.

[38] 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.

[39] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *CVPR*, 2016.

[40] Fangbin Wan, Yang Wu, Xuelin Qian, Yixiong Chen, and Yanwei Fu. When person re-identification meets changing clothes. In *CVPRW*, 2020.

[41] Jingya Wang, Xiatian Zhu, Shaogang Gong, and Wei Li. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In *CVPR*, 2018.

[42] Kai Wang, Zhi Ma, Shiyian Chen, Jinni Yang, Keke Zhou, and Tao Li. A benchmark for clothes variation in person re-identification. *IIS*, 2020.

[43] Taqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. Person re-identification by video ranking. In *ECCV*, 2014.
Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In CVPR, 2018.

Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. Learning deep feature representations with domain guided dropout for person re-identification. In CVPR, 2016.

Qize Yang, Ancong Wu, and Wei-Shi Zheng. Person re-identification by contour sketch under moderate clothing change. TPAMI, 2019.

Mang Ye, Jianbing Shen, Gaofei Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. Deep learning for person re-identification: A survey and outlook. TPAMI, 2021.

Jiahang Yin, Ancong Wu, and Wei-Shi Zheng. Fine-grained person re-identification. IJCV, 2020.

Hong-Xing Yu, Ancong Wu, and Wei-Shi Zheng. Cross-view asymmetric metric learning for unsupervised person re-identification. In ICCV, 2017.

Hong-Xing Yu, Wei-Shi Zheng, Ancong Wu, Xiaowei Guo, Shaogang Gong, and Jian-Huang Lai. Unsupervised person re-identification by soft multilabel learning. In CVPR, 2019.

Li Zhang, Tao Xiang, and Shaogang Gong. Learning a discriminative null space for person re-identification. In CVPR, 2016.

Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Unsupervised salience learning for person re-identification. In CVPR, 2013.

Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In ICCV, 2015.

Liang Zheng, Yi Yang, and Alexander G Hauptmann. Person re-identification: Past, present and future. arXiv preprint arXiv:1610.02984, 2016.

Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Reidentification by relative distance comparison. TPAMI, 2012.

Zhendong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In ICCV, 2017.

Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. Learning to adapt invariance in memory for person re-identification. TPAMI, 2020.

Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Learning generalisable omni-scale representations for person re-identification. TPAMI, 2021.

Xiangping Zhu, Xiaotian Zhu, Minxian Li, Pietro Morerio, Vittorio Murino, and Shaogang Gong. Intra-camera supervised person re-identification. IJCV, 2021.

C. Lawrence Zitnick and Piotr Dollár. Edge boxes: Locating object proposals from edges. In ECCV, 2014.