Supplementary Materials for Rethinking the Distribution Gap of Person Re-identification with Camera-based Batch Normalization

Zijie Zhuang[1], Longhui Wei[2,4], Lingxi Xie[2], Tianyu Zhang[2], Hengheng Zhang[3], Haozhe Wu[1], Haizhou Ai[1], and Qi Tian[2]

[1]Tsinghua University [2]Huawei Inc.
[3]Hefei University of Technology [4]University of Science and Technology of China
{jayzhuang42,weilh2568,198808xc,tianyu1949,imhmhm}@gmail.com
wuhz1997@163.com, ahz@tsinghua.edu.cn, tian.qi1@huawei.com

1 Camera-based Testing Scheme in Section 3.3

In this section, we introduce the testing scheme of our camera-based formulation. Unlike the conventional BN [9], which only calculates the statistics in the training stage and directly uses the recorded value for testing, our camera-based formulation with CBN utilizes a symmetrical approach, *i.e.*, estimating the camera-related statistics in both training and testing stages.

Algorithm 1 Inference with CBN layers

Input: a trained feature extractor $\mathbf{f}(\cdot)$, images from the testing camera set $\mathcal{C}$.
Initialize: grouping testing images according to their camera ID and randomly samples N mini-batches from each group, denoted as $\{\mathbf{I}_i\}^{(c)}$
for all $c \leftarrow 1$ to $|\mathcal{C}|$ **do**
  Forward all images from $\{\mathbf{I}_i\}^{(c)}$ in N mini-batches
  for all CBN layers in $\mathbf{f}(\cdot)$ **do**
    Collect the corresponding mini-batch mean μ_n and variance σ_n^2
    $\hat{\mu}_{(c)} = \text{accumulate}\{\mu_1, \mu_2, ..., \mu_N\}$
    $\hat{\sigma}^2_{(c)} = \text{accumulate}\left\{\sigma_1^2, \sigma_2^2, ..., \sigma_N^2\right\}$
    Inject $\hat{\mu}_{(c)}$ and $\hat{\sigma}^2_{(c)}$ into the corresponding CBN layer
  end for
  for all images $\mathbf{I}^{(c)}$ from camera c **do**
    Compute final features $\mathbf{f}\left(\mathbf{I}^{(c)}\right)$
  end for
end for

The method used in the training stage is introduced in Section 3.3. In the testing stage, before generating the final features for each testing image, we first cluster these images according to their camera labels. For each of these camera-related clusters, we randomly collect several unlabeled images. Then, we group these images into mini-batches and forward them across the ReID network. In this stage, the standardization procedure in every CBN uses mini-batch statistics, *i.e.*, the same procedure in training. For each mini-batch, we collect the mini-batch mean and variance of every CBN layer. After forwarding all related mini-batches, we approximate the overall mean and variance of each CBN layer with these mini-batch statistics using the same way in the conventional BN. Finally, we inject our estimated results into each CBN layer and generate the final features of all images from this specific camera. The above procedure ends when images from all testing cameras are processed. The detailed algorithm is presented in Algorithm. 1.

2 The Warm-Up Strategy in Section 4.1

In this section, we describe the warm-up strategy for initializing fully-connected classifiers in incremental learning tasks. Given a model that has already been trained on one or multiple ReID datasets, when fine-tuning it on a new training set, a new fully-connected classifier for classifying images from this specific dataset is required. Since this classifier is randomly initialized, if we directly fine-tune the entire model in an end-to-end manner, this classifier will introduce lots of noises to the feature extractor and heavily damage the previously learned knowledge. To alleviate the knowledge forgetting in the early stage of training, we warm-up the newest classifier before the formal training. Note that in the Replay incremental learning, there could be classifiers and images that correspond to multiple training sets (the exemplar memory and the current training set). However, in the warm-up stage of all incremental learning tasks, we only consider the latest training set and the corresponding new classifier. The details of this warm-up strategy are presented in Algorithm 2. In short, we freeze all previously learned layers and only iteratively fine-tune the new classifier on the latest training set until the loss becomes stable. After the warm-up stage, we start to train the entire network in a conventional end-to-end manner.

Our experiments show that this warm-up stage is essential for preserving the previously-learned knowledge. Without the warm-up, the conventional BN-based method loses another 5.3% Rank-1 accuracy on average, while our formulation loses 3.7% on average.

3 Exemplar Memory in Section 4.2

The exemplar memory is built for the Replay incremental learning task. Its goal is to reinforce the discriminative knowledge of the previous training sets with the least amount of old images. In this paper, we design a straightforward approach to achieve this goal. For each old training set, we propose a greedy algorithm that saves one image for each identity and tries to keep an equal number of images for each old camera. The details are presented in Algorithm 3. With this approach, the size of the exemplar memory for Market [51], Duke [53],

Algorithm 2 Warm-up the latest classifier

Input: a trained ReID model with the feature extractor $\mathbf{f}(\cdot)$, image $\mathbf{I}_i$ and the corresponding ID $\mathbf{y}_i$ from the latest training set $\mathcal{D}$
Initialize: freeze all trainable parameters in $\mathbf{f}(\cdot)$, randomly initialize a new classifier $\mathbf{g}(\cdot)$ for $\mathcal{D}$, set counter $n = 0$, set an empty list $\mathfrak{L} = []$
repeat
  Randomly sample a mini-batch $\{\mathbf{I}_i\}$ and the corresponding $\{\mathbf{y}_i\}$ from $\mathcal{D}$
  $\mathcal{L} = \text{get\_loss}\left(\mathbf{g}\left(\mathbf{f}\left(\{\mathbf{I}_i\}\right)\right), \{\mathbf{y}_i\}\right)$
  Backward $\mathcal{L}$ and only update $\mathbf{g}(\cdot)$
  Append $\mathcal{L}$ to $\mathfrak{L}$
  Truncate $\mathfrak{L}$ and only preserve the latest 50 items
  if ($\mathfrak{L}$ has 50 items) & ($|\mathcal{L} - \text{mean}(\mathfrak{L})| <= 0.1$) **then**
    $n = n + 1$
  else
    $n = 0$
  end if
until $n = 5$

Algorithm 3 Build the exemplar memory

Input: a ReID set $\mathcal{D}$ with an identity set $\mathcal{K}$ and a camera set $\mathcal{C}$
Output: the exemplar memory $\mathcal{M}$ in which each identity from $\mathcal{D}$ has exactly one image
Initialize: create a dict Ω that records the number of already picked images from each camera.
for all identity k in $\mathcal{K}$ **do**
  Collect all images that belong to the identity k
  Collect the camera ID of the above images as $\{c\}$
  Query Ω with $\{c\}$ and find the camera c that has the least picked images
  Randomly pick an image that simultaneously belongs to camera c and identity k, and add it to $\mathcal{M}$
  $\Omega[c] = \Omega[c] + 1$
end for

and MSMT17 [42] is only 5.5%, 4.2%, and 3.4% of their original training set, respectively.

Another thing worth noting is the way of utilizing these exemplars together with the data from the latest training set. On the one hand, in the exemplar memory, there are only very few samples that describe the previous cameras, and each old identity only has one image. On the other hand, as described in Section 3.3 and Section 4.1, for the latest training set, each identity has multiple images in the mini-batch, so does each camera. To make sure that our method can accurately approximate the CBN statistics of all previous and current cameras, we design a mixed sampling strategy. As shown in Fig. 3, when handling images from the latest training set, we follow the pipeline presented in Section 3.3. When sampling identities from the exemplar memory, we cluster images from the exemplar memory and make sure that each group has four successive old images

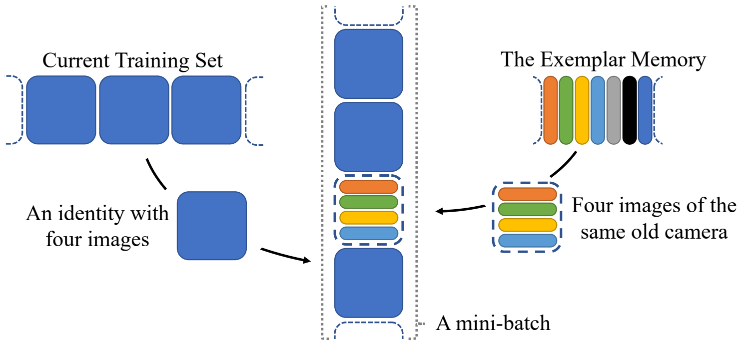

Fig. 3. The demonstration of a mini-batch. (1) A blue rectangle denotes four images of the same identity. (2) The rectangles in other colors represent the images from the exemplar memory. Each rectangle corresponds to one image of an old identity. We group these exemplars according to their camera ID, and randomly fuse these groups with the data sampled from the current training set.

that correspond to the same old camera. Then, these groups are randomly fused with the images sampled from the latest training set.

4 Experiments on partially Replacing BN with CBN

These are supplementary experiments for demonstrating the necessity of replacing **all** BN layers with CBN layers, rather than only part of them. We go back to our baseline and divide the BN layers into six parts: the BN that appears before all residual blocks, the BN within each of the four residual stages, and the BN that appears after all blocks. The following table summarizes the *direct transfer* performance when the model trained on Duke is tested on Market. Since the vanilla BN is below satisfaction in the *direct transfer* experiments, we utilize AdaBN for adapting testing set statistics.

Table 1. The direct transfer performance from Duke to Market. ✓ marks the component in which all its BN layers are replaced with CBN layers.

First BN	Block 1	Block 2	Block 3	Block 4	Last BN	Rank-1	mAP
						55.8	28.1
✓						60.6	31.6
✓	✓					61.9	32.9
✓	✓	✓				65.0	35.3
✓	✓	✓	✓			65.7	35.7
✓	✓	✓	✓	✓		67.3	37.0
✓	✓	✓	✓	✓	✓	**72.7**	**43.0**

These results indicate that replacing all BN layers with CBN layers obtains the best results in the *direct transfer*. More importantly, we emphasize that only replacing part of BN layers contradicts the fundamental idea of this paper, because we believe that distribution statistics should only be collected within a camera, and all camera-related distributions should be aligned explicitly.

References

1. Almazan, J., Gajic, B., Murray, N., Larlus, D.: Re-id done right: towards good practices for person re-identification. arXiv preprint arXiv:1801.05339 (2018)
2. Chang, X., Hospedales, T.M., Xiang, T.: Multi-level factorisation net for person re-identification. In: CVPR. IEEE (2018)
3. Deng, W., Zheng, L., Ye, Q., Kang, G., Yang, Y., Jiao, J.: Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: CVPR. IEEE (2018)
4. Fan, H., Zheng, L., Yan, C., Yang, Y.: Unsupervised person re-identification: Clustering and fine-tuning. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) **14**(4), 83 (2018)
5. Fan, X., Luo, H., Zhang, X., He, L., Zhang, C., Jiang, W.: Scpnet: Spatial-channel parallelism network for joint holistic and partial person re-identification. In: ACCV. Springer (2018)
6. Fu, Y., Wei, Y., Wang, G., Zhou, Y., Shi, H., Huang, T.S.: Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In: ICCV. IEEE (2019)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. IEEE (2016)
8. Huang, H., Yang, W., Chen, X., Zhao, X., Huang, K., Lin, J., Huang, G., Du, D.: Eanet: Enhancing alignment for cross-domain person re-identification. arXiv preprint arXiv:1812.11369 (2018)
9. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
10. Jiao, J., Zheng, W.S., Wu, A., Zhu, X., Gong, S.: Deep low-resolution person re-identification. In: AAAI (2018)
11. Kalayeh, M.M., Basaran, E., Gökmen, M., Kamasak, M.E., Shah, M.: Human semantic parsing for person re-identification. In: CVPR. IEEE (2018)
12. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al.: Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences **114**(13), 3521–3526 (2017)
13. Li, W., Zhu, X., Gong, S.: Harmonious attention network for person re-identification. In: CVPR. IEEE (2018)
14. Li, W., Zhu, X., Gong, S.: Harmonious attention network for person re-identification. In: CVPR. IEEE (2018)
15. Li, Y., Wang, N., Shi, J., Liu, J., Hou, X.: Revisiting batch normalization for practical domain adaptation. arXiv preprint arXiv:1603.04779 (2016)
16. Li, Z., Hoiem, D.: Learning without forgetting. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(12), 2935–2947 (2018)
17. Lin, S., Li, H., Li, C.T., Kot, A.C.: Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification. In: BMVC (2018)

18. Lin, Y., Dong, X., Zheng, L., Yan, Y., Yang, Y.: A bottom-up clustering approach to unsupervised person re-identification. In: AAAI (2019)
19. Lin, Y., Xie, L., Wu, Y., Yan, C., Tian, Q.: Unsupervised person re-identification via softened similarity learning. In: CVPR. IEEE (2020)
20. Liu, H., Feng, J., Qi, M., Jiang, J., Yan, S.: End-to-end comparative attention networks for person re-identification. IEEE Transactions on Image Processing **26**(7), 3492–3506 (2017)
21. Liu, J., Ni, B., Yan, Y., Zhou, P., Cheng, S., Hu, J.: Pose transferrable person re-identification. In: CVPR. IEEE (2018)
22. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: CVPRW. IEEE (2019)
23. Ma, N., Zhang, X., Zheng, H.T., Sun, J.: Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: ECCV. Springer (2018)
24. Mao, S., Zhang, S., Yang, M.: Resolution-invariant person re-identification. arXiv preprint arXiv:1906.09748 (2019)
25. Miao, J., Wu, Y., Liu, P., Ding, Y., Yang, Y.: Pose-guided feature alignment for occluded person re-identification. In: ICCV. IEEE (2019)
26. Pan, X., Luo, P., Shi, J., Tang, X.: Two at once: Enhancing learning and generalization capacities via ibn-net. In: ECCV. Springer (2018)
27. Peng, P., Xiang, T., Wang, Y., Pontil, M., Gong, S., Huang, T., Tian, Y.: Unsupervised cross-dataset transfer learning for person re-identification. In: CVPR. IEEE (2016)
28. Qian, N.: On the momentum term in gradient descent learning algorithms. Neural networks **12**(1), 145–151 (1999)
29. Rannen, A., Aljundi, R., Blaschko, M.B., Tuytelaars, T.: Encoder based lifelong learning. In: ICCV. IEEE (2017)
30. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: CVPR. IEEE (2018)
31. Shen, Y., Li, H., Yi, S., Chen, D., Wang, X.: Person re-identification with deep similarity-guided graph neural network. In: ECCV. Springer (2018)
32. Song, C., Huang, Y., Ouyang, W., Wang, L.: Mask-guided contrastive attention model for person re-identification. In: CVPR. IEEE (2018)
33. Song, J., Yang, Y., Song, Y.Z., Xiang, T., Hospedales, T.M.: Generalizable person re-identification by domain-invariant mapping network. In: CVPR. IEEE (2019)
34. Song, L., Wang, C., Zhang, L., Du, B., Zhang, Q., Huang, C., Wang, X.: Unsupervised domain adaptive re-identification: Theory and practice. Pattern Recognition (2020)
35. Suh, Y., Wang, J., Tang, S., Mei, T., Mu Lee, K.: Part-aligned bilinear representations for person re-identification. In: ECCV. Springer (2018)
36. Sun, H., Chen, Z., Yan, S., Xu, L.: Mvp matching: A maximum-value perfect matching for mining hard samples, with application to person re-identification. In: ICCV. IEEE (2019)
37. Sun, Y., Zheng, L., Deng, W., Wang, S.: Svdnet for pedestrian retrieval. In: ICCV. IEEE (2017)
38. 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: ECCV. Springer (2018)
39. Tian, M., Yi, S., Li, H., Li, S., Zhang, X., Shi, J., Yan, J., Wang, X.: Eliminating background-bias for robust person re-identification. In: CVPR. IEEE (2018)
40. Van Der Maaten, L.: Accelerating t-sne using tree-based algorithms. JMLR **15**(1), 3221–3245 (2014)

41. Wang, J., Zhu, X., Gong, S., Li, W.: Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: CVPR. IEEE (2018)
42. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer gan to bridge domain gap for person re-identification. In: CVPR. IEEE (2018)
43. Wei, L., Zhang, S., Yao, H., Gao, W., Tian, Q.: Glad: Global-local-alignment descriptor for pedestrian retrieval. In: ACMMM. ACM (2017)
44. Wu, A., Zheng, W.S., Guo, X., Lai, J.H.: Distilled person re-identification: Towards a more scalable system. In: CVPR. IEEE (2019)
45. Wu, A., Zheng, W.S., Lai, J.H.: Unsupervised person re-identification by camera-aware similarity consistency learning. In: ICCV. IEEE (2019)
46. Yu, H.X., Wu, A., Zheng, W.S.: Cross-view asymmetric metric learning for unsupervised person re-identification. In: ICCV. IEEE (2017)
47. Yu, H.X., Wu, A., Zheng, W.S.: Unsupervised person re-identification by deep asymmetric metric embedding. TPAMI (2018)
48. Yu, H.X., Zheng, W.S., Wu, A., Guo, X., Gong, S., Lai, J.H.: Unsupervised person re-identification by soft multilabel learning. In: CVPR (2019)
49. Zhang, T., Xie, L., Wei, L., Zhang, Y., Li, B., Tian, Q.: Single camera training for person re-identification. AAAI (2020)
50. Zhang, X., Luo, H., Fan, X., Xiang, W., Sun, Y., Xiao, Q., Jiang, W., Zhang, C., Sun, J.: Alignedreid: Surpassing human-level performance in person re-identification. arXiv preprint arXiv:1711.08184 (2017)
51. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: ICCV. IEEE (2015)
52. Zheng, Z., Zheng, L., Yang, Y.: A discriminatively learned cnn embedding for person reidentification. ACM Transactions on Multimedia Computing, Communications, and Applications **14**(1), 13 (2017)
53. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: ICCV. IEEE (2017)
54. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: CVPR. IEEE (2017)
55. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In: AAAI (2020)
56. Zhong, Z., Zheng, L., Li, S., Yang, Y.: Generalizing a person retrieval model hetero- and homogeneously. In: ECCV. Springer (2018)
57. Zhong, Z., Zheng, L., Luo, Z., Li, S., Yang, Y.: Invariance matters: Exemplar memory for domain adaptive person re-identification. In: CVPR. IEEE (2019)
58. Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: Camera style adaptation for person re-identification. In: CVPR. IEEE (2018)
59. Zhou, J., Yu, P., Tang, W., Wu, Y.: Efficient online local metric adaptation via negative samples for person reidentification. In: ICCV. IEEE (2017)
60. Zhou, K., Yang, Y., Cavallaro, A., Xiang, T.: Omni-scale feature learning for person re-identification. In: ICCV. IEEE (2019)
61. Zhu, X., Zhu, X., Li, M., Murino, V., Gong, S.: Intra-camera supervised person re-identification: A new benchmark. In: ICCVW. IEEE (2019)
62. Zhu, Z., Jiang, X., Zheng, F., Guo, X., Huang, F., Sun, X., Zheng, W.: Viewpoint-aware loss with angular regularization for person re-identification. In: AAAI (2020)