

# CAMERA BIAS REGULARIZATION FOR PERSON RE-IDENTIFICATION

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Person re-identification (Re-ID) is to match persons captured by non-overlapping cameras. Due to the discrepancies between cameras caused by illumination, background, or viewpoint, the underlying difficulty for Re-ID is the camera bias problem, which leads to the large gap of within-identity features from different cameras. With limited cross-camera annotated data, Re-ID models tend to learn camera-related features, instead of identity-related features. Consequently, Re-ID models suffer from poor transfer ability from seen domains to unseen domains. In this paper, we investigate the camera bias problem in supervised learning, unsupervised learning, and their variants. In particular, we propose a novel Camera Bias Regularization (CBR) term to reduce the feature distribution gap between cameras by enlarging the intra-camera distance and reducing the inter-camera distance simultaneously. Extensive experiments on person Re-ID tasks validate the effectiveness and universality of the proposed CBR.

## 1 INTRODUCTION

Person re-identification (Re-ID) is to match persons across non-overlapping cameras. It has been an active research field due to its potential in the application of smart cities. Re-ID is performed in different settings, including supervised learning (Zhong et al. (2018); Sun et al. (2018); Wang et al. (2018); Luo et al. (2019); Zhou et al. (2019)) and unsupervised learning (Fu et al. (2019); Wang & Zhang (2020); Ge et al. (2020a;c); Zhai et al. (2020)), which have witnessed significant progress in recent years. However, one of the main obstacles that limit the performance of Re-ID models and prevent them from wide applications is the *camera bias problem* (Yu et al. (2017); Zhong et al. (2018); Wu et al. (2019); Zhuang et al. (2020); Yang et al. (2021)). It comes from the large discrepancies between different cameras. As the illumination, background and viewpoint vary from camera to camera, it easily leads to the mismatch of feature distributions between cameras. As a result, it inevitably forces Re-ID models to learn camera-related features rather than identity-related features. In this way, many Re-ID tasks suffer from the problem, as the similarity measurement between persons may be inaccurate.

To alleviate the camera bias problem, it requires a large amount of annotated cross-camera data for supervised learning. However, the huge amount of annotated data does not lead to a model that generalizes well to an unseen domain. For example, a sharp decline of rank-1 accuracy from 94.1% to 26.7% is observed in Luo et al. (2019) when applying a well-trained model on Market-1501 (Zheng et al. (2015)) to DukeMTMC (Ristani et al. (2016)). The problem will become worse when there is no cross-camera data. Zhang et al. (2020) has shown that when each identity only appears in one camera, a serious camera bias problem will occur. Without constraints, the learned features are strongly related to cameras instead of person identities. Consequently, it shows poor performance even in the same domain. For unsupervised learning, the similarity measurement between persons is crucial. However, due to the camera bias problem, the measurement is usually inaccurate. Figure 1a shows the feature visualization of a subset of MSMT17 (Wei et al. (2018)) using T-SNE (van der Maaten & Hinton (2008)). It can be seen that intra-camera instances tend to have closer distances than inter-camera instances. In the iterative clustering-and-finetuning state-of-the-art unsupervised learning methods (Ge et al. (2020a;c)), errors of the clustering results based on the inaccurate similarity measurement will be accumulated, leading to limited performances (Xuan & Zhang (2021)).

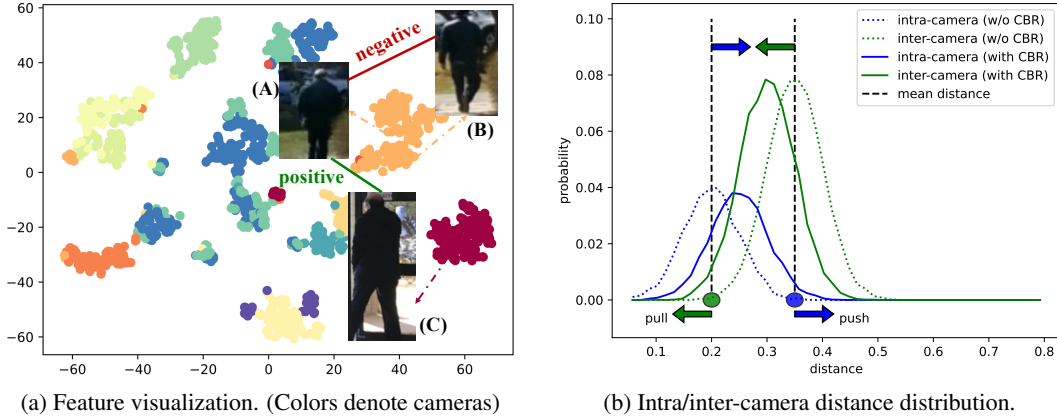


Figure 1: Illustration of the camera bias problem and our proposed camera bias regularization. **(a):** The positive pair (A) and (C) are far away from each other, while the negative pair (A) and (B) are close as they share the same background. **(b):** The gap between intra-camera and inter-camera distance distribution is reduced by the proposed CBR.

In this paper, we propose a *Camera Bias Regularization (CBR)* term to close the gap between intra-camera and inter-camera distance distribution. It regularizes the conventional training of Re-ID models by properly enlarging the intra-camera distance and reducing the inter-camera distance simultaneously. By default, the distance distribution discussed in this paper is from *within-identity* (for supervised learning) or *within-cluster* (for unsupervised learning). So, for simplicity, we will omit *within-identity* and *within-cluster* in rest of the paper. It should be noted that the proposed CBR works within each identity, while the conventional Re-ID loss, such as identity loss and triplet loss, mainly works between different identities to distinguish one from others. From the perspective of the conventional Re-ID loss, both the intra-camera and inter-camera distance will be smaller during training, but the gap between them is ignored. Figure 1b shows an example of the distance distribution of intra-camera and inter-camera pairs. By enlarging the farthest intra-camera distance (blue circle) and reducing the nearest inter-camera distance (green circle), the two distributions will get closer to each other. Consequently, the camera bias problem will be alleviated. For supervised learning, the proposed CBR will enhance the generalization ability to unseen instances or domains. The unsupervised learning can also benefit from more accurate similarity measurement.

To conclude, our contributions can be summarized as following:

- We investigate the camera bias problem in person Re-ID tasks: by aligning the distribution between cameras, the generalization and transfer ability of Re-ID models can be strengthened, thus the accuracy of similarity measurement is improved.
- We propose a Camera Bias Regularization (CBR) term to reduce the gap between intra-camera and inter-camera distance distribution. Additionally, a Cross-Camera (CC) clustering algorithm is also designed to cooperate with CBR for unsupervised learning tasks.
- Extensive experiments on person Re-ID tasks, including fully supervised learning, direct transfer, purely unsupervised learning and unsupervised domain adaptation, validate the effectiveness and universality of our method.

## 2 RELATED WORK

In this section, we mainly discuss existing works that aim to solve the camera bias problem. For a comprehensive Re-ID survey, please refer to Ye et al. (2021). Yu et al. (2017) projects data from different cameras to a unified feature space, but it is necessary to learn an individual projection function for each camera. Zhong et al. (2018) proposes to augment data in each camera with style-transferred data from other cameras in supervised learning. Qi et al. (2019) adopts a generative adversarial network to align feature distributions from different cameras in unsupervised learning. Wu et al. (2019) globally aligns the intra-camera and inter-camera similarity distribution by penalizing the difference

of mean and variance of the two distributions. In addition, it also requires a strong source-domain model to preserve the intra-camera similarity. Zhuang et al. (2020) also designs a camera-specific batch normalization layer to normalize data from each camera to a unified space, but it requires test data to estimate the statistics of the layer. Yang et al. (2021) proposes to use meta-learning (Finn et al. (2017)) to learn camera-invariant features for fast adaptation to unseen domains. For this purpose, the training data is split according to their camera views to form the meta-train and meta-test. Xuan & Zhang (2021) decomposes the similarity measurement into intra-camera and inter-camera instances, and trains the model in a two-step manner. Differently, our proposed method regularizes the conventional training of Re-ID models by enlarging the intra-camera distance and reducing the inter-camera distance for each identity simultaneously. It is concise, effective and universal for Re-ID tasks. Without modifications to the original training methods or backbone models, the proposed CBR can be easily integrated into state-of-the-arts.

### 3 PROPOSED APPROACH

In this section, we first review the background of person Re-ID problem (section 3.1). Then we introduce our proposed *Camera Bias Regularization (CBR)* term (section 3.2) which aims to alleviate the camera bias problem for person Re-ID tasks. Finally, we illustrate the application of CBR in supervised and unsupervised learning tasks.

#### 3.1 PRELIMINARY

For a person dataset  $D = \{(x_i, y_i) |_{i=1}^N\}$ , where  $x_i$  and  $y_i$  denote the  $i$ -th training samples and its associated identity label,  $N$  is the number of training samples. A Re-ID model  $\mathcal{M}$  is usually composed of a feature extraction module  $\mathcal{F}$  and an identity classifier  $\mathcal{C}$ . For each sample  $x_i$ , the feature extraction module outputs its feature  $f_i$ , which will be further used in both training and inference, and the identity classifier outputs its predicted identity probability  $\hat{y}_i$ . Conventional Re-ID methods usually apply identity loss and metric loss to learn person features:

$$\mathcal{L}_{reid} = \mathcal{L}_{id}(y_i, \hat{y}_i) + \mathcal{L}_{tri}(x_i, x_i^n, x_i^p) \quad (1)$$

The identity loss is the cross-entropy loss. The metric loss is usually the hard mining triplet loss (Hermans et al. (2017)), where  $x_i^p$  and  $x_i^n$  are the hardest positive and negative samples for  $x_i$ .

#### 3.2 CAMERA BIAS REGULARIZATION FOR PERSON RE-ID

In this work, we propose a camera bias regularization term to alleviate the camera bias problem for person Re-ID tasks. First, let  $x_i^{p,c}$  denotes the  $i$ -th instance of the person with identity label  $p$  and camera label  $c$ . Then we define the **nearest** instance of  $x_i^{p,c}$  as  $x_n^{p,c'}$  that has the **same** identity label  $p$  but a **different** camera label  $c'$  ( $c' \neq c$ ). Similarly, we also define the **farthest** instance of  $x_i^{p,c}$  as  $x_j^{p,c}$  ( $j \neq i$ ) that has the **same** identity label and **same** camera label  $c$ .

$$\begin{aligned} x_n^{p,c'} &= \arg \min_{n=1 \dots K, c'=1 \dots C, n \neq i, c' \neq c} d(\mathcal{F}(x_i^{p,c}), \mathcal{F}(x_n^{p,c'})) \\ x_j^{p,c} &= \arg \max_{j=1 \dots K, j \neq i} d(\mathcal{F}(x_i^{p,c}), \mathcal{F}(x_j^{p,c})) \end{aligned} \quad (2)$$

where  $K$  is the number of instances of person  $p$ ,  $C$  is the number of cameras. It should be noted that  $x_i^{p,c}$ ,  $x_j^{p,c}$  and  $x_n^{p,c'}$  share the same identity label  $p$ . Then we define the *Camera Bias Regularization (CBR)* term as following:

$$\mathcal{R}_{cbr} = \frac{1}{N} \sum_{i=1}^N \log(1 + \exp(d(\mathcal{F}(x_i^{p,c}), \mathcal{F}(x_n^{p,c'})) - d(\mathcal{F}(x_i^{p,c}), \mathcal{F}(x_j^{p,c})))) \quad (3)$$

With CBR, the total loss function for person Re-ID tasks will be:

$$\mathcal{L}_{reid}^{new} = \mathcal{L}_{reid} + \lambda * \mathcal{R}_{cbr} \quad (4)$$

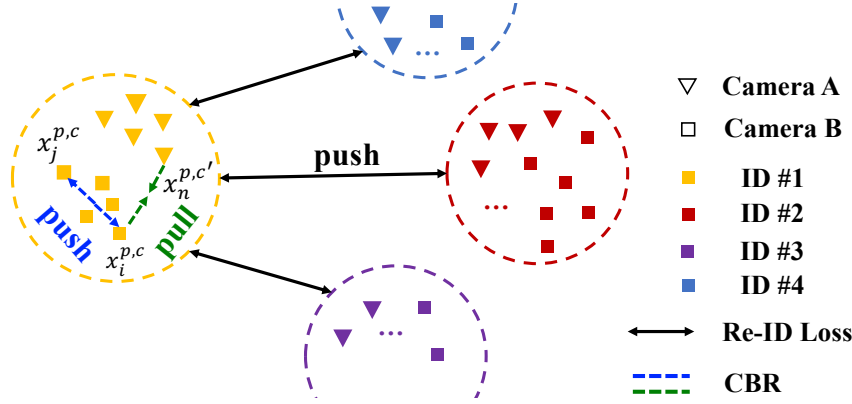


Figure 2: The difference between conventional Re-ID loss and camera bias regularization.

Table 1: Comparison between different unsupervised Re-ID tasks.

Unsupervised Re-ID Tasks	Training Data	
	Labeled Source	Unlabeled Target
<b>Direct Transfer (DT)</b>	✓	✗
<b>Purely Unsupervised Learning (PUL)</b>	✗	✓
<b>Unsupervised Domain Adaptation (UDA)</b>	✓	✓

where  $\lambda$  is a hyper-parameter to trade off between conventional Re-ID loss  $\mathcal{L}_{reid}$  and the camera bias regularization  $\mathcal{R}_{cbr}$ .

To further explain how the proposed CBR works, Figure 2 shows the difference between conventional Re-ID loss and CBR. The Re-ID loss mainly works between different identities by pushing them away from each other, but it neglects the within-identity distribution, which causes the camera bias problem. The proposed CBR works in each identity to align the distribution of intra-camera and inter-camera distance.

Notably, different from loss functions, the proposed CBR works as a regularization term. Without considering the Re-ID loss, CBR can be infinitely small by pushing the intra-camera distance  $d(\mathcal{F}(x_i^{p,c}), \mathcal{F}(x_j^{p,c}))$  far away. But in this way, the distance distribution of positive pairs (person images of the same identity) and negative pairs (person images of different identities) will severely overlap, which further leads to the increase of the Re-ID loss function  $\mathcal{L}_{reid}$ . So, similar to  $L_1$  and  $L_2$  regularization, the proposed CBR is beneficial to Re-ID tasks to some extent, but too much emphasis on this term may be harmful, so the tradeoff between  $\mathcal{L}_{reid}$  and  $\mathcal{R}_{cbr}$  is required.

### 3.3 CBR FOR FULLY SUPERVISED LEARNING

As the groundtruth identity annotation is provided in *Fully Supervised Learning (FSL)*, the overall loss function for supervised Re-ID is the same as Equation 4.

### 3.4 CBR FOR UNSUPERVISED LEARNING

The unsupervised Re-ID tasks can be roughly divided into *Direct Transfer (DT)*, *Unsupervised Domain Adaptation (UDA)* and *Purely Unsupervised Learning (PUL)*. The main difference between them is the training data. The comparison are shown in Table 1. As there is only labeled source data in DT, the loss function is also the same as that in FSL. In UDA and PUL, as the unlabeled target data is available, currently the effective approach is the iterative clustering-and-finetuning pipeline, which is adopted in most state-of-the-arts (Fu et al. (2019); Lin et al. (2019; 2020); Wang & Zhang (2020); Zeng et al. (2020); Ge et al. (2020c;a)). The pseudo labels can be used as groundtruth labels, so the loss function is also the same as that in FSL. Nevertheless, as pseudo labels may not be perfect, the effect of CBR is a little different from that in FSL. Besides bridging the gap

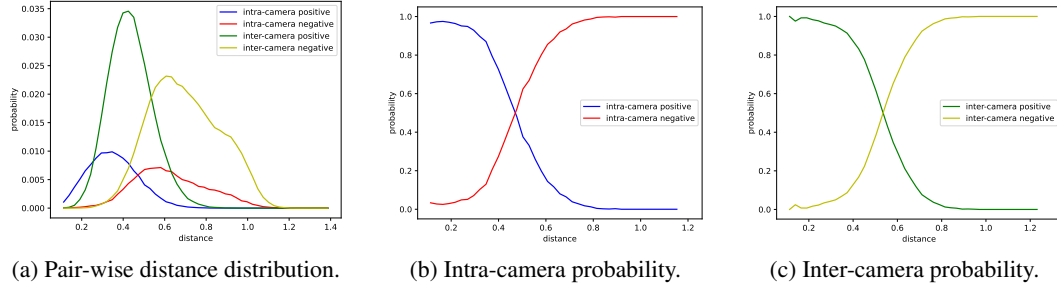


Figure 3: Pair-wise distance distribution from a clustering-and-finetuning based unsupervised Re-ID learning method. The experiment is conducted on Market-1501 (Zheng et al. (2015)) dataset. The method is based on the baseline proposed in (Ge et al. (2020a)).

between intra-camera and inter-camera distance, it also has the ability to resist noises of pseudo labels. **Within each cluster (or pseudo label)**, four types of pairs can be obtained by examining their groundtruth identity labels and camera labels: *intra-camera positive*, *intra-camera negative*, *inter-camera positive* and *inter-camera negative*. Figure 3a shows the distribution of them from a clustering-and-finetuning baseline method. It can be found that: First, the positive and negative distribution (including both intra-camera and inter-camera) overlap with each other. Second, the distribution of intra-camera positive and inter-camera positive are not aligned. The mean distance of intra-camera positive pairs is smaller than inter-camera positive pairs.

We also compare the distance distribution of positive and negative pairs of intra-camera and inter-camera separately in Figure 3b and Figure 3c. For intra-camera, with the growth of distance, the probability of negative pairs increases rapidly. For example, when the distance is larger than 0.8, nearly all the intra-camera pairs are negative. According to Equation 3, the farthest intra-camera distance  $d(\mathcal{F}(x_i^{p,c}), \mathcal{F}(x_j^{p,c}))$  will be pushed away. For all the intra-camera pairs in a cluster, if the farthest distance is large enough, it is more likely to be a negative pair (as shown in Figure 3b). So it is beneficial to push the pair away from each other. In the next clustering, they may not be grouped together. In this way, the CBR resists the noises of pseudo labels. If the farthest distance is small, it is more likely to be a positive pair (as shown in Figure 3b). The camera bias problem occurs under this circumstance, so the CBR will work the same as in FSL to align the intra-/inter-camera distribution. In a word, the proposed CBR can work properly whether the farthest intra-camera pair is positive or negative. Similarly, as shown in Figure 3c, the nearest inter-camera pair is more likely to be a positive pair, so it is beneficial to pull the distance  $d(\mathcal{F}(x_i^{p,c}), \mathcal{F}(x_n^{p,c}'))$  in Equation 3.

**Cross-Camera Clustering** An important prerequisite of CBR is that person images of the same label or pseudo label are required to come from at least two camera views. Otherwise, the inter-camera instance  $x_n^{p,c'}$  in Equation 3 does not exist. For FSL, as each identity contains images captured from different cameras, it naturally meets the condition. However, for unsupervised learning, the clustering algorithm (such as DBSCAN (Ester et al. (1996))) cannot guarantee it. It is more likely to cluster images from the same camera because of the large biases of features from different cameras (as shown in Figure 1a). To further alleviate the camera bias problem and ensure that instances in each cluster are at least from two cameras. We propose a DBSCAN-based **Cross-Camera (CC)** clustering algorithm. It puts more emphasis on inter-camera pairs than intra-camera pairs. The proposed CC is implemented by modifying  $\epsilon$  parameter in DBSCAN. Originally,  $\epsilon$  is the maximum distance of two instances to be considered as the neighborhood. It is a pre-defined hyper-parameter in DBSCAN, but it is dynamically adjusted in CC according to their camera views as:

$$\epsilon_{i,j} = \begin{cases} \epsilon, & C(x_i) \neq C(x_j) \\ 0, & C(x_i) = C(x_j) \end{cases} \quad (5)$$

where  $C(x)$  returns camera for  $x$ . The intra-camera pairs will no longer be directly reachable, because the minimum distance for two different instances is greater than  $\epsilon_{i,j} = 0$ . Nevertheless, they may be indirectly linked by the instance in another camera and finally be clustered into a group.

Table 2: Comparison with state-of-the-arts on fully supervised setting.

Methods	Market		Duke		MSMT	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
CamStyle (Zhong et al. (2018))	68.72	88.12	53.48	75.27	-	-
PCB (Sun et al. (2018))	77.40	92.30	66.10	81.70	-	-
PCB+RPP (Sun et al. (2018))	81.60	93.80	69.20	83.30	-	-
MGN (Wang et al. (2018))	<u>86.90</u>	<b>95.70</b>	<b>78.40</b>	<b>88.70</b>	52.10	76.90
OSNet (Zhou et al. (2019))	84.90	94.80	73.50	<u>88.60</u>	<u>52.90</u>	<b>78.70</b>
CBN (Zhuang et al. (2020))	77.30	91.30	67.30	82.50	42.90	72.80
BoT (Luo et al. (2019))	85.70	94.10	75.90	86.20	-	-
BoT (fast-reid)	86.30	94.27	76.52	86.27	51.15	74.70
BoT + CBR (ours)	<b>87.44</b> (+1.14)	<u>94.83</u> (+0.56)	<u>77.89</u> (+1.37)	88.24 (+1.97)	<b>53.43</b> (+2.28)	<u>78.42</u> (+3.72)

## 4 EXPERIMENTS

### 4.1 SETTINGS

**Datasets** The proposed CBR is evaluated on three popular benchmark datasets: Market-1501 (Zheng et al. (2015)), DukeMTMC-reID (Ristani et al. (2016)) and MSMT17 (Wei et al. (2018)). The Market-1501 dataset consists of 32,668 images of 1,501 identities captured from 6 cameras, where 12,936 images of 751 identities are used for training and the remaining 19,732 images of 750 identities are for testing. DukeMTMC-reID contains 16,522 images of 702 identities for training, and the remaining images of another 702 identities are for testing. All images are collected from 8 cameras. MSMT17 is currently the largest Re-ID dataset, which contains 32,621 images of 4,101 identities for training and about 100,000 images for testing. For simplicity, we will use Market, Duke and MSMT to refer to the three datasets in the rest of the paper.

**Implementation Details** For fully supervised learning and direct transfer, we implement CBR based on *fast-reid* (He et al. (2020)). The training details are exactly the same as the commonly used strong baseline BoT (Luo et al. (2019)). For unsupervised learning, including unsupervised domain adaptation and purely unsupervised learning, the implementation is based on *OpenUnReID* (Ge et al. (2020a;c)). ResNet-50 (He et al. (2016)) is adopted as the backbone for all tasks. For all the compared methods, we do not make modifications to the original implementations. Please refer to the two repositories and corresponding papers for more details.

**Evaluation Protocol** Mean average precision (mAP) and rank-1 accuracy are adopted to evaluate the performance of the proposed CBR. No post-processing technique is adopted, such as re-ranking (Zhong et al. (2017)) or query expansion (Chum et al. (2007)).

### 4.2 EVALUATION ON RE-ID TASKS

The proposed camera bias regularization is a general approach to alleviate the camera bias problem for person Re-ID tasks. In this section, we conduct extensive experiments to validate CBR in *Fully Supervised Learning (FSL)*, *Direct Transfer (DT)*, *Purely Unsupervised Learning (PUL)* and *Unsupervised Domain Adaptation (UDA)* settings.

#### 4.2.1 FULLY SUPERVISED LEARNING

The baseline method for fully supervised learning adopts the strong baseline (BoT) proposed in Luo et al. (2019). The original loss function of BoT is cross-entropy loss and hard-mined triplet loss. The loss function of BoT + CBR is the same as Equation 4.

**Comparison with State-of-the-arts** As shown in Table 2, the performance of BoT implementation based on *fast-reid* is slightly better than the reported results in the paper. With the proposed CBR, BoT+CBR can boost the performance on all three datasets, especially on the challenging dataset MSMT: 2.28% mAP improvement and 3.72% rank-1 improvement compared with BoT (fast-reid).

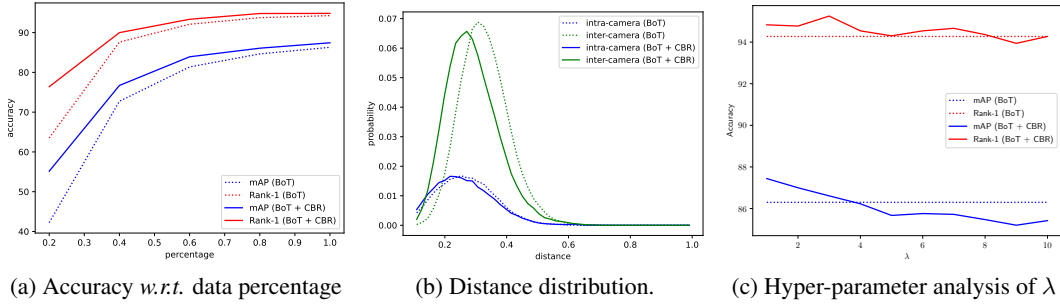


Figure 4: Experimental results of fully supervised learning on Market.

Table 3: Experimental results on the direct transfer setting. (The numbers in parentheses denote the improvement over BoT(fast-reid). Random erasing is turned off due to its negative effect.)

Methods	Duke-Market		Market-Duke		Market-MSMT		Duke-MSMT	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
BoT(Paper)	<b>25.50</b>	<b>54.30</b>	25.70	41.40	-	-	-	-
BoT(fast-reid)	24.43	52.46	24.40	41.29	4.90	14.31	6.03	18.84
BoT + CBR (ours)	24.87 (+0.44)	52.58 (+0.12)	<b>26.11</b> (+1.71)	<b>43.36</b> (+2.07)	<b>5.26</b> (+0.36)	<b>15.02</b> (+0.71)	<b>6.63</b> (+0.60)	<b>19.89</b> (+1.05)

**Impact of data size** Figure 4a shows the accuracy *w.r.t* data percentage. CBR shows significant improvement over BoT when the amount of data is small. It also indicates that the conventional training requires more cross-camera labeled data to alleviate the camera bias problem, while ours needs less. Similar results can also be obtained on Duke and MSMT datasets in Appendix A.1.

**Analysis of hyper-parameter  $\lambda$**  Figure 4c shows the mAP and rank-1 on Market dataset by varying  $\lambda$  from 1.0 to 10.0. The performance first improves with increasing  $\lambda$  (rank-1 improves but mAP drops), then drops when  $\lambda$  is too large. Similar results can also be obtained on Duke and MSMT datasets in Appendix A.2. In practice, when the dataset is complicated, especially in the variations of cameras,  $\lambda$  can be set to a large value. Nevertheless, it is generally a good choice to set  $\lambda = 1.0$ .

**Analysis of pair-wise distance distribution** Figure 4b shows the comparison between pair-wise distance distribution of BoT and BoT+CBR. The inter-camera distance distribution moves left when applying CBR to BoT, which shrinks the distance between intra-camera and inter-camera pairs. The mean distance between intra-camera and inter-camera pairs is reduced by CBR from 0.0998 to 0.0701. It verifies the claim that the proposed CBR can alleviate the camera bias problem by closing the gap between intra-camera and inter-camera distance.

#### 4.2.2 DIRECT TRANSFER

When the target domain data is unavailable, DT applies the model trained on source domain to target domain. It tests the model’s generalization ability. As the camera bias is one of the reasons to hinder the generalization ability (Zhuang et al. (2020)), it will be an effective approach to improve the target domain performance by alleviating the camera bias. As shown in Table 3, compared with BoT, the proposed CBR can boost the performance on all datasets.

#### 4.2.3 PURELY UNSUPERVISED LEARNING

PUL is to train the model only from target domain. In practice, the model is initialized by the ImageNet (Deng et al. (2009)) pre-trained weights. Currently, the clustering-and-finetuning-based method is an effective approach for both PUL and UDA person Re-ID. In this section, we first adopt a simple clustering-and-finetuning baseline to evaluate the effectiveness of the proposed CBR and then apply CBR to state-of-the-art methods to further boost their performances.

**Ablation Study** Table 4 shows results of baseline and its combination with CC and CBR. First, CC and CBR alone can already improve the mAP and rank-1. Specifically, the improvement of Base-line+CBR over Baseline is higher than Baseline+CC. Second, when combining CC and CBR, Base-

Table 4: Ablation study of cross-camera DBSCAN (CC) and camera bias regularization (CBR). (The numbers in parentheses denote the improvement over Baseline.)

Methods	Market		Duke		MSMT	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Baseline	67.0	84.9	53.4	69.1	13.5	31.8
Baseline + CC	69.8 (+2.8)	85.3 (+0.4)	55.6 (+2.2)	71.2 (+2.1)	13.7 (+0.2)	33.5 (+1.7)
Baseline + CBR	71.7 (+4.7)	87.6 (+2.7)	57.5 (+4.1)	75.6 (+6.5)	<b>23.0</b> (+9.5)	53.0 (+21.2)
Baseline + CC + CBR	<b>73.2</b> (+6.2)	<b>88.7</b> (+3.8)	<b>61.5</b> (+8.1)	<b>77.4</b> (+8.3)	22.8 (+9.3)	<b>54.5</b> (+22.7)

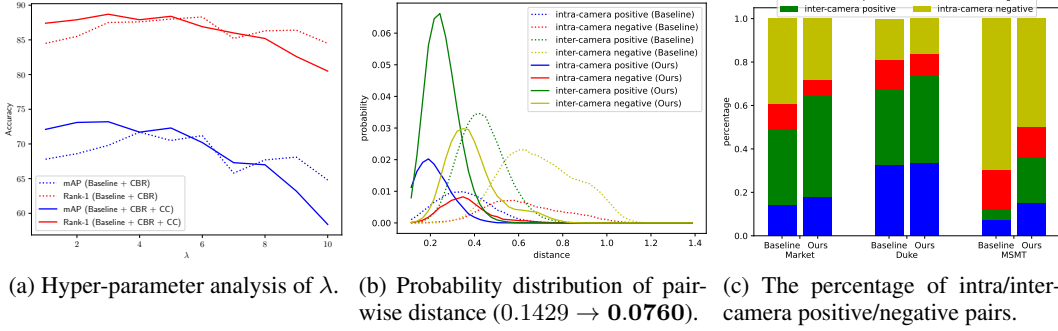


Figure 5: Experimental results of purely unsupervised learning on Market.

line+CC+CBR shows significant improvement over Baseline on all three datasets. It indicates that the proposed CBR can work better with the help of CC in unsupervised learning. We also note that, on MSMT, the improvement of CC alone is very limited. The reason may be the inaccurate similarity measurement, as the rank-1 of Baseline is only 31.8%. For the same reason, Baseline+CC+CBR shows little improvement in rank-1 over Baseline+CBR. Particularly, Baseline+CC+CBR is 0.2% lower than Baseline+CBR in mAP. We will see in Table 5 that as the similarity measurement becomes more accurate, CC will boost the performance again.

**Analysis of hyper-parameter  $\lambda$**  Figure 5a shows the performance on Market by varying  $\lambda$  from 1.0 to 10.0. The performance first improves with increasing  $\lambda$ , then drops when  $\lambda$  is too large. Similar results can also be obtained on Duke and MSMT dataset in Appendix B.1. Differently,  $\lambda$  that corresponds to the best performance (3.0 to 5.0) is larger than that in fully supervised learning (1.0 to 3.0). Also similarly, a more complicated dataset requires a larger  $\lambda$ .

**Analysis of pair-wise distance distribution** Figure 5b shows the comparison between the within-cluster pair-wise distance distribution of Baseline and Ours (Baseline+CBR). CBR can close the distance between intra-camera and inter-camera pairs. Although all four types of distributions move to the left, the mean distance between intra-camera and inter-camera positive pairs is reduced by CBR from 0.1429 to 0.0760. As stated in Section 3.4, another important effect of CBR in unsupervised learning is to resist noises of pseudo labels. In order to validate it, Figure 5c shows the percentage of each type of pair for Baseline and Ours. For all three datasets, the percentages of noisy pairs, including intra-camera negative (yellow bar) and inter-camera negative pairs (red bar), are significantly reduced.

**Comparison with State-of-the-arts** We integrate our method into the recent state-of-the-art MMT (Ge et al. (2020a)), and compare MMT+CC, MMT+CBR and MMT+CC+CBR with others in Table 5. First, our three methods outperform the original MMT by a large margin. Second, except for CAP, our full method, MMT+CC+CBR, also outperforms all other baselines that adopt the general clustering and training method. Third, for CAP, a specialized camera-based clustering and training method is designed, and it also adopts a stronger base method SpCL. Ours still shows competitive performance with CAP. Moreover, ours can work with a variety of methods on Re-ID tasks, which can not be easily done for CAP.



Table 5: Comparison with state-of-the-arts on purely unsupervised setting.

Methods	Market		Duke		MSMT	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
BUC (Lin et al. (2019))	38.3	66.2	27.5	47.4	-	-
SSL (Lin et al. (2020))	37.8	71.7	28.6	52.5	-	-
MMCL (Wang & Zhang (2020))	45.5	80.3	40.2	65.2	11.2	35.4
HCT (Zeng et al. (2020))	56.4	80.0	50.7	69.6	-	-
SpCL (Ge et al. (2020c))	72.6	87.7	65.3	<u>81.2</u>	19.1	42.3
MetaCam (Yang et al. (2021))	61.7	83.9	53.8	73.8	15.5	35.2
IICS (Xuan & Zhang (2021))	72.9	89.5	64.4	80.0	26.9	56.4
CAP (Wang et al. (2021))	<b>79.2</b>	<u>91.4</u>	<b>67.3</b>	81.1	<b>36.9</b>	<b>67.4</b>
MMT (Ge et al. (2020b))	69.9	85.1	59.7	73.5	18.3	39.5
MMT + CC	75.0	88.5	62.5	76.0	27.7	53.3
MMT + CBR	76.9	90.5	61.8	76.1	23.4	48.9
MMT + CC + CBR	<u>78.2</u>	<b>92.0</b>	<u>67.0</u>	<b>81.4</b>	<u>31.2</u>	<u>60.2</u>

Table 6: Comparison with state-of-the-arts on unsupervised domain adaptation setting.

Methods	Duke-Market		Market-Duke		Duke-MSMT		Market-MSMT	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
SPGAN (Deng et al. (2018))	22.8	51.5	22.3	41.4	-	-	-	-
UCDA-CCE (Qi et al. (2019))	34.5	64.3	36.7	55.4	-	-	-	-
CASCL (Wu et al. (2019))	35.6	64.7	30.5	51.5	-	-	-	-
SSG (Fu et al. (2019))	58.3	80.0	53.4	73.0	13.3	32.2	13.2	31.6
AD-Cluster (Zhai et al. (2020))	68.3	86.7	54.1	72.6	-	-	-	-
MMT (Ge et al. (2020a))	73.8	89.5	62.3	76.3	25.1	52.9	24.0	50.1
SpCL (Ge et al. (2020c))	76.7	90.3	68.8	<b>82.9</b>	26.5	53.1	25.4	51.6
ECN++ (Zhong et al. (2021))	63.8	84.1	54.4	74.0	16.0	42.5	15.2	40.4
HGA (Zhang et al. (2021))	70.3	89.5	67.1	80.4	26.8	58.6	25.5	55.1
UNRN (Zheng et al. (2021))	78.1	<b>91.9</b>	<u>69.1</u>	<u>82.0</u>	26.2	54.9	25.3	52.4
MMT <sup>+</sup> (Ge et al. (2020b))	76.7	89.7	65.4	78.2	23.4	47.6	22.2	46.3
MMT <sup>+</sup> + CC	<u>79.7</u>	<b>91.9</b>	67.7	80.2	<u>32.3</u>	<u>59.0</u>	<u>31.8</u>	<u>58.7</u>
MMT <sup>+</sup> + CBR	79.4	90.8	65.8	79.0	26.2	52.6	26.7	53.0
MMT <sup>+</sup> + CC + CBR	<b>81.0</b>	<u>91.7</u>	<b>69.2</b>	81.4	<b>35.0</b>	<b>63.7</b>	<b>33.1</b>	<b>61.9</b>

#### 4.2.4 UNSUPERVISED DOMAIN ADAPTATION

This setting aims to transfer the knowledge on source domain to target domain. MMT+ (Ge et al. (2020b)) is used as our base method, which improves original MMT (Ge et al. (2020a)) by jointly training on source domain labeled data and target domain pseudo-labeled data. The comparisons of its variants and state-of-the-arts are listed in Table 6. First, ours, including MMT<sup>+</sup>+CC, MMT<sup>+</sup>+CBR and MMT<sup>+</sup>+CC+CBR, improves over the original MMT<sup>+</sup> by a large margin. Second, compared with the best baseline, ours show competitive results on Duke-Market and Market-Duke. Third, on the more challenging Duke-MSMT and Market-MSMT, ours outperform the best baseline by 8.2% and 7.6% in terms of mAP, which is a new state-of-the-art on UDA person Re-ID.

## 5 CONCLUSIONS AND FUTURE WORK

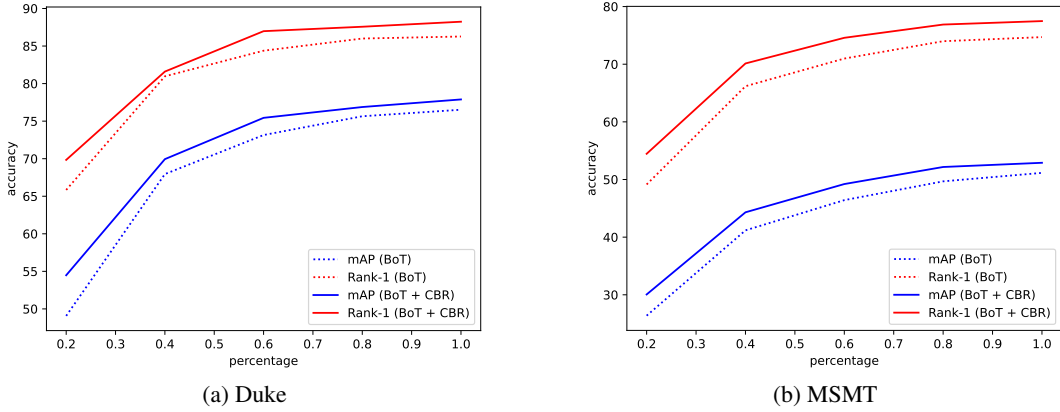
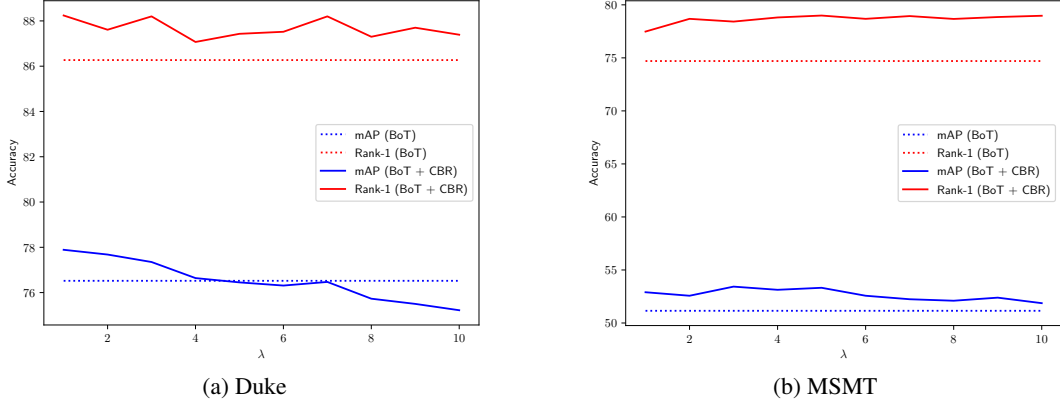
In this paper, we present a camera bias regularization (CBR) solution to alleviate the camera bias problem for person Re-ID. It significantly boosts the performance of several common Re-ID tasks, including fully supervised learning, direct transfer, purely unsupervised learning, and unsupervised domain adaptation. As the proposed CBR is concise, effective and universal, we believe it will be commonly adopted as identity and triplet loss in Re-ID. In the future, we will evaluate CBR on more vehicle Re-ID datasets and also generalize it to more side information other than camera views.

## REFERENCES

- Ondrej Chum, James Philbin, Josef Sivic, Michael Isard, and Andrew Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8. IEEE, 2007.
- Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.
- Weijian Deng, Liang Zheng, Guoliang Kang, Yezhou Yang, Qixiang Ye, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 994–1003, 2018.
- Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pp. 226–231, 1996.
- Chelsea Finn, P. Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, 2017.
- Yang Fu, Yunchao Wei, Guanshuo Wang, Xi Zhou, Humphrey Shi, and Thomas S. Huang. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6111–6120, 2019.
- Yixiao Ge, Dapeng Chen, and Hongsheng Li. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In *International Conference on Learning Representations*, 2020a. URL <https://openreview.net/forum?id=rJlnOhVYPS>.
- Yixiao Ge, Shijie Yu, and Dapeng Chen. Improved mutual mean-teaching for unsupervised domain adaptive re-id. *ArXiv*, abs/2008.10313, 2020b.
- 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 *Advances in Neural Information Processing Systems*, 2020c.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- Lingxiao He, Xingyu Liao, Wu Liu, Xinchun Liu, Peng Cheng, and Tao Mei. Fastreid: A pytorch toolbox for general instance re-identification. *arXiv preprint arXiv:2006.02631*, 2020.
- Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.
- Yutian Lin, Xuanyi Dong, Liang Zheng, Yan Yan, and Yi Yang. A bottom-up clustering approach to unsupervised person re-identification. In *AAAI*, 2019.
- Yutian Lin, Lingxi Xie, Yu Wu, Chenggang Clarence Yan, and Qi Tian. Unsupervised person re-identification via softened similarity learning. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3387–3396, 2020.
- Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1487–1495, 2019.
- Lei Qi, L. Wang, Jing Huo, Luping Zhou, Yinghuan Shi, and Yang Gao. A novel unsupervised camera-aware domain adaptation framework for person re-identification. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8079–8088, 2019.
- Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking*, 2016.

- Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling. In *ECCV*, 2018.
- Laurens van der Maaten and Geoffrey E. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 2008.
- Dongkai Wang and Shiliang Zhang. Unsupervised person re-identification via multi-label classification. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10978–10987, 2020.
- Guanshuo Wang, Yufeng Yuan, Xiong Chen, Jiwei Li, and Xi Zhou. Learning discriminative features with multiple granularities for person re-identification. *Proceedings of the 26th ACM international conference on Multimedia*, 2018.
- Menglin Wang, Baisheng Lai, Jianqiang Huang, Xiaojin Gong, and Xiansheng Hua. Camera-aware proxies for unsupervised person re-identification. In *AAAI*, 2021.
- Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 79–88, 2018.
- Ancong Wu, Wei-Shi Zheng, and Jianhuang Lai. Unsupervised person re-identification by camera-aware similarity consistency learning. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6921–6930, 2019.
- Shiyu Xuan and Shiliang Zhang. Intra-inter camera similarity for unsupervised person re-identification. In *CVPR*, 2021.
- Fengxiang Yang, Zhun Zhong, Zhiming Luo, Yuanzheng Cai, Yaojin Lin, Shaozi Li, and N. Sebe. Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification. In *CVPR*, 2021.
- Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven C. H. Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE transactions on pattern analysis and machine intelligence*, PP, 2021.
- Hong-Xing Yu, Ancong Wu, and Wei-Shi Zheng. Cross-view asymmetric metric learning for unsupervised person re-identification. *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 994–1002, 2017.
- Kaiwei Zeng, Munan Ning, Yaohua Wang, and Yang Guo. Hierarchical clustering with hard-batch triplet loss for person re-identification. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13654–13662, 2020.
- Yunpeng Zhai, Shijian Lu, Qixiang Ye, Xuebo Shan, Jie Chen, Rongrong Ji, and Yonghong Tian. Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9018–9027, 2020.
- Minying Zhang, Kai Liu, Yidong Li, Shihui Guo, Hongtao Duan, Yimin Long, and Yi Jin. Unsupervised domain adaptation for person re-identification via heterogeneous graph alignment. In *AAAI*, 2021.
- Tianyu Zhang, Lingxi Xie, Longhui Wei, Yongfei Zhang, Bo Li, and Qi Tian. Single camera training for person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 12878–12885, 2020.
- Kecheng Zheng, Cuiling Lan, Wenjun Zeng, Zhizheng Zhang, and Zhengjun Zha. Exploiting sample uncertainty for domain adaptive person re-identification. In *AAAI*, 2021.
- Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, pp. 1116–1124, 2015.

- Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1318–1327, 2017.
- Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. Camera style adaptation for person re-identification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5157–5166, 2018.
- Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yezhou Yang. Learning to adapt invariance in memory for person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43:2723–2738, 2021.
- Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3701–3711, 2019.
- Zijie Zhuang, Longhui Wei, Lingxi Xie, Tian-Yu Zhang, Hengheng Zhang, Haozhe Wu, Haizhou Ai, and Qi Tian. Rethinking the distribution gap of person re-identification with camera-based batch normalization. In *ECCV*, 2020.

Figure 6: Accuracy *w.r.t* data percentage for fully supervised learning.Figure 7: Analysis of hyper-parameters  $\lambda$  for fully supervised learning.

## A MORE EXPERIMENTS ON FULLY SUPERVISED LEARNING

### A.1 IMPACT OF DATA SIZE

Figure 6 shows the accuracy *w.r.t* data percentage. CBR shows significant improvement over BoT when the amount of data is small. It also indicates that conventional training requires more cross-camera labeled data to alleviate the camera bias problem. For example, ours can achieve similar results with 60% data (Rank-1 74.58) to BoT with all data (Rank-1 74.7) on MSMT.

### A.2 ANALYSIS OF HYPER-PARAMETER $\lambda$

Figure 7 shows the mAP and rank-1 on Duke and MSMT datasets by varying  $\lambda$  from 1.0 to 10.0. The performance first improves with increasing  $\lambda$ , then drops when  $\lambda$  is too large. In practice, when the dataset is complicated, especially in the variations of cameras,  $\lambda$  can be set to a large value. Nevertheless, it is generally a good choice to set  $\lambda = 1.0$ .

### A.3 ANALYSIS OF PAIR-WISE DISTANCE DISTRIBUTION

Figure 8 shows the comparison between pair-wise distance distribution of BoT and BoT+CBR. The inter-camera distance distribution moves left when applying CBR to BoT. It makes the intra-camera and inter-camera distance closer. The mean distance between intra-camera and inter-camera pairs is reduced by CBR from 0.0695 to 0.0219 on Duke (from 0.1681 to 0.1136 on MSMT). It verifies the claim that the proposed CBR can alleviate the camera bias problem by closing the gap between intra-camera and inter-camera distance.

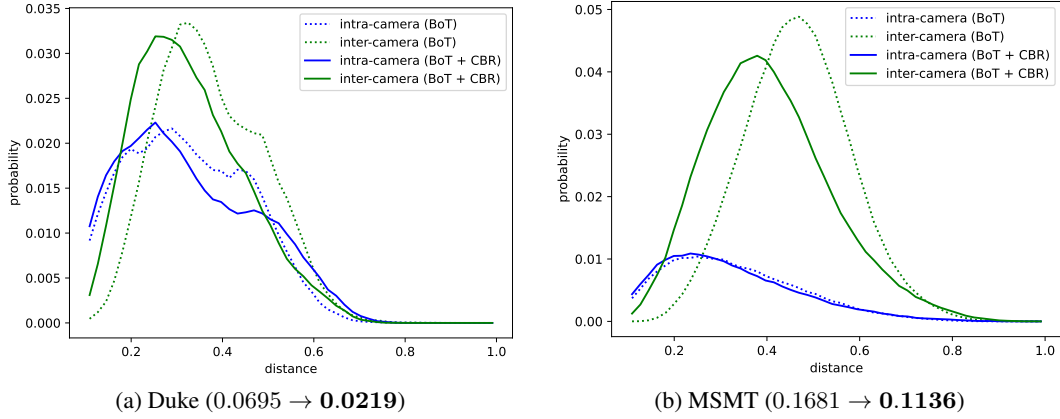


Figure 8: Comparison of distance distribution between BoT and BoT + CBR. The mean distances between intra-camera and inter-camera positive pairs of BoT and BoT + CBR are on the left and right side of  $\rightarrow$ , respectively.

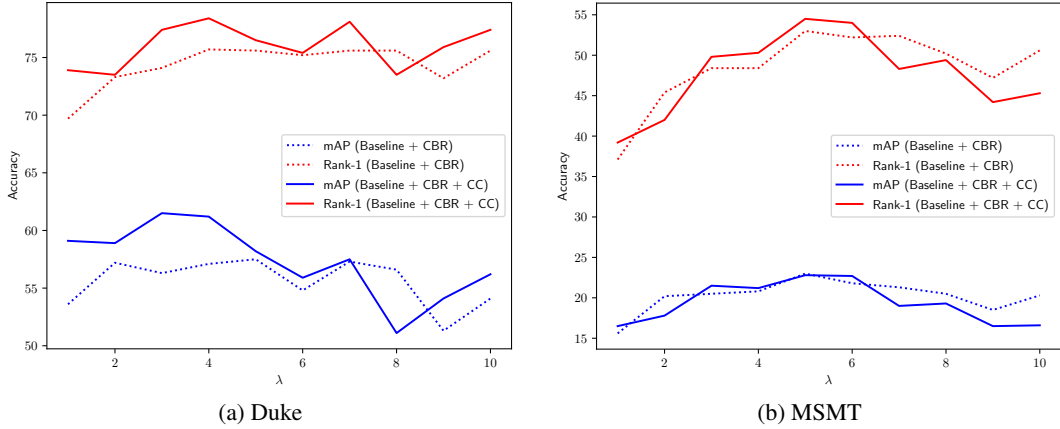


Figure 9: Analysis of hyper-parameters  $\lambda$  for purely unsupervised learning.

## B MORE EXPERIMENTS ON PURELY UNSUPERVISED LEARNING

### B.1 ANALYSIS OF HYPER-PARAMETER $\lambda$

Figure 9 shows the performance on Duke and MSMT by varying  $\lambda$  from 1.0 to 10.0. The performance first improves with increasing  $\lambda$ , then drops when  $\lambda$  is too large. Differently,  $\lambda$  that corresponds to the best performance (3.0 to 5.0) is larger than that in fully supervised learning (1.0 to 3.0). Also similarly, a more complicated dataset requires a larger  $\lambda$ .

### B.2 ANALYSIS OF PAIR-WISE DISTANCE DISTRIBUTION

Figure 10 shows the comparison between pair-wise distance distribution of Baseline and Ours (Baseline+CBR). CBR can close the distance between intra-camera and inter-camera pairs. Although all four types of distributions move to the left, the mean distance between intra-camera and inter-camera positive pairs is reduced by CBR from 0.1614 to 0.0843 on Duke (from 0.2026 to 0.1544 on MSMT).

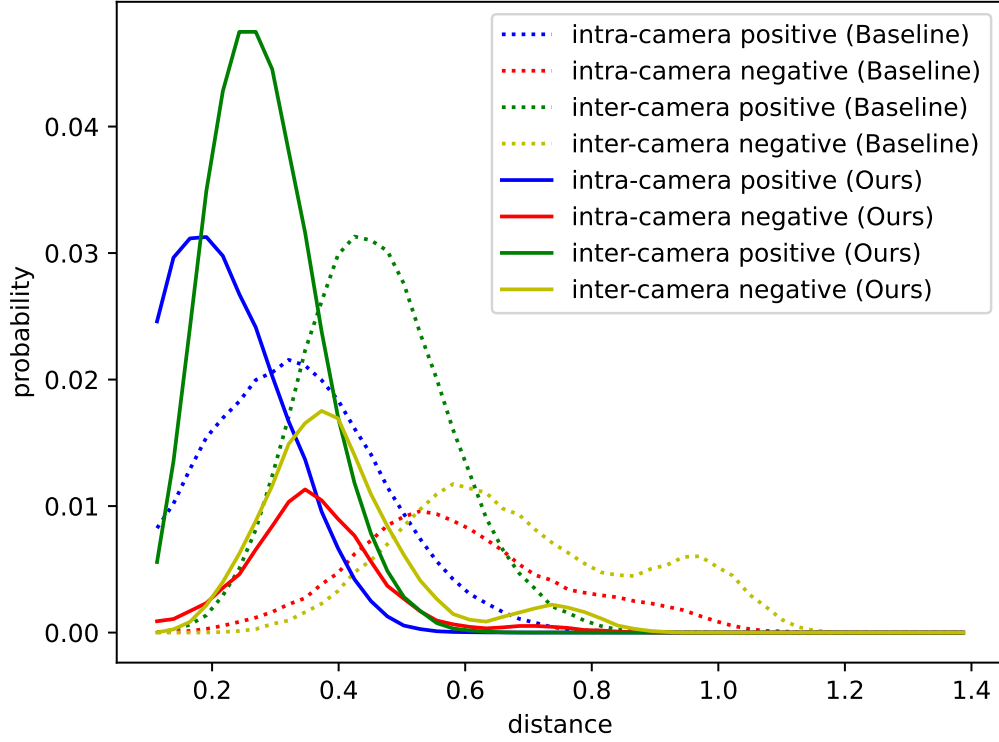
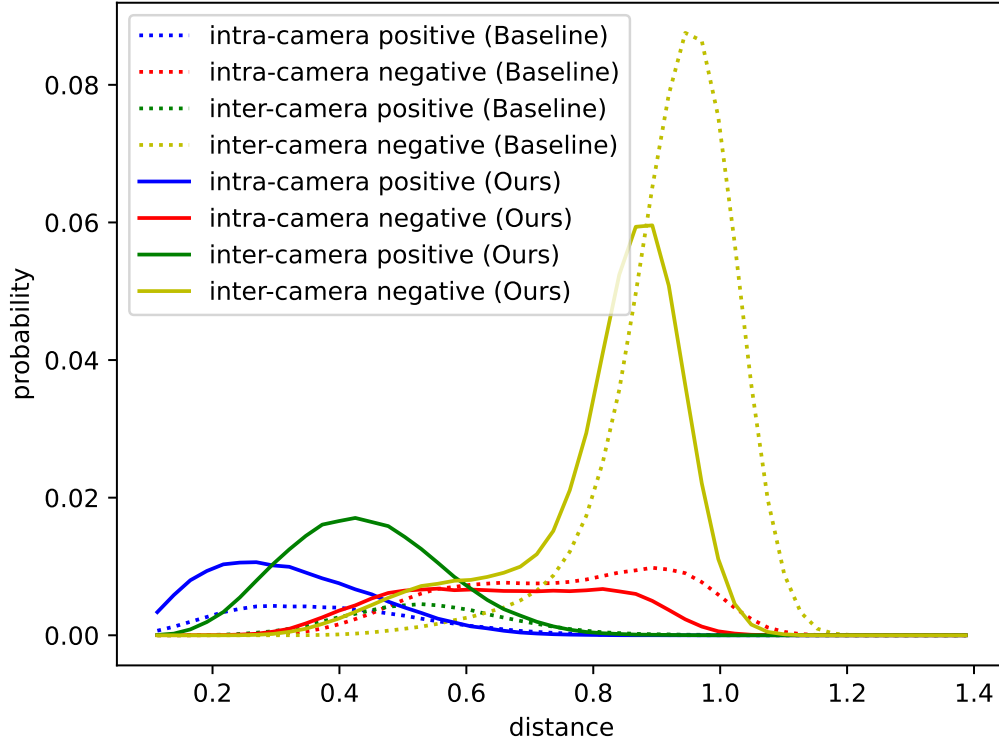
(a) Duke (0.1614  $\rightarrow$  **0.0843**)(b) MSMT (0.2026  $\rightarrow$  **0.1544**)

Figure 10: Comparison of distance distribution between baseline and ours (baseline + CBR). The mean distances between intra-camera positive pairs and inter-camera positive pairs of baseline and ours are on the left and right side of  $\rightarrow$ , respectively.