Unsupervised Object Re-identification via Irregular Sampling

Qing Tang, Ge Cao, Kang-Hyun Jo
Department of Electrical, Electronic and Computer Engineering
University of Ulsan
Ulsan, Korea
zucchini.tang@hotmail.com; caoge9706@gmail.com; acejo@ulsan.ac.kr

2022 International Workshop on Intelligent Systems (IWIS) | 979-8-3503-9893-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/IWIS56333.2022.9920902

***Abstract*—Object re-identification (Re-ID), is a fundamental task in intelligent systems, that aims to find the same object, i.e., person or vehicle under different camera views or scenes. This paper studies the fully unsupervised object re-ID problem which can learn re-ID without any human-annotated labeled data. Recent works show that self-supervised momentum contrastive learning is an effective method for unsupervised object re-ID, but they neglect to optimize one important component - sampling strategy. Here we investigate and analyze the performances of the current sampling strategy in different numbers of positive samples in a mini-batch under the same learning framework and loss function, then we proposed a more effective and robust sampling strategy - Irregular Sampling (IS). Experimental results show that sampling strategy is also an important factor in model performance, and the proposed sampling strategy IS can effectively boost the model performance. Extensive experiments are performed on one vehicle re-ID dataset and two mainstream person re-ID datasets.**

***Index Terms*—Person re-identification, fully unsupervised learning, vehicle re-identification**

I. Introduction

The discrepancy in learning strategy causes the advantage of self-supervised contrastive learning mechanism to be not fully utilized in object re-ID tasks. This paper focuses on one aspects - sampling strategy. The purpose of the sampling strategy is to split a whole dataset into mini-batches for training. Instance discrimination tasks [1]–[3] used random sampling, which treats each instance as a single class and samples $P = 1$ numbers of instance in each mini-batch. Random sampling results in a bad performance in re-ID task because of lack of intra-/inter-class learning [4], [5]. State-of-the-art re-ID methods [4], [6] performed triplet sampling to perform intra-/inter-class learning and achieved impressive performance. Triplet sampling [5], [7] samples a fixed number $P > 1$ of same identity instances in each mini-batch. Therefore, small clusters need to be sampled repeatedly to ensure P instances in each mini-batch. Here we found out and demonstrated that triplet sampling harms the network performance because 1) same patterns in a mini-batch leads model over-fitting and 2) selecting repeat samples introduces an imbalanced problem between small and large clusters. Therefore, we introduce a more effective and robust sampling strategy, called Irregular Sampling (IS) in this paper. The MoCo-based self-supervised contrastive learning framework [1] is adopted as the baseline framework in this work. The illustration of our proposed framework is shown in Fig. 1.

In summary, the contributions of this work are two-fold.

- This work analyze the shortcomings of the common and wide-used random and triplet sampling strategies.
- This work proposed a simple but effective sampling strategy Irregular Sampling (IS) to overcome two problems (1) information leakage within a batch and (2) training imbalance between small and large clusters. Extensive experiments demonstrate that the proposed sampling strategy IS shows exceptionally strong performances in three mainstream object re-ID datasets.

The remainder of this paper is organized as follows. Section II summarizes the related work. Section III describes the proposed method. In section IV, the implementation details and extensive experimental results are reported and analyzed. Finally, Section V concludes this paper.

II. Related Works

A. Fully Unsupervised object re-ID

Common unsupervised object re-ID problems include person re-ID [4], [9] and vehicle re-ID [10]. Lin et al. [11] first proposed a fully unsupervised method for re-ID, called Bottom-Up Clustering (BUC), which merges a fixed number of clusters to fine-tune the model step by step. Wang et al. [12] formulated re-ID as a multi-label classification task, optimized the network under the supervision of self-predicted and online pseudo multi-class labels. Based on the [12], Tang et al. [13] leveraged the eligible neighbors as additional reference information to further boost the model performance in ranking accuracy. Recently, thanks to the self-supervised contrastive learning framework, many methods have pushed the fully unsupervised re-ID performance to a new height even outperforming UDA methods. Tang et al. [14] proposed to predict more robust multiple pseudo labels for each image for training.

B. Sampling Strategy

Instead of designing learning frameworks or loss functions, recent works [5], [15] showed that sampling strategies also play an important role in model performance. Random sampling is a widely used, simple and suitable sampling strategy for self-supervised instance discrimination task [1],

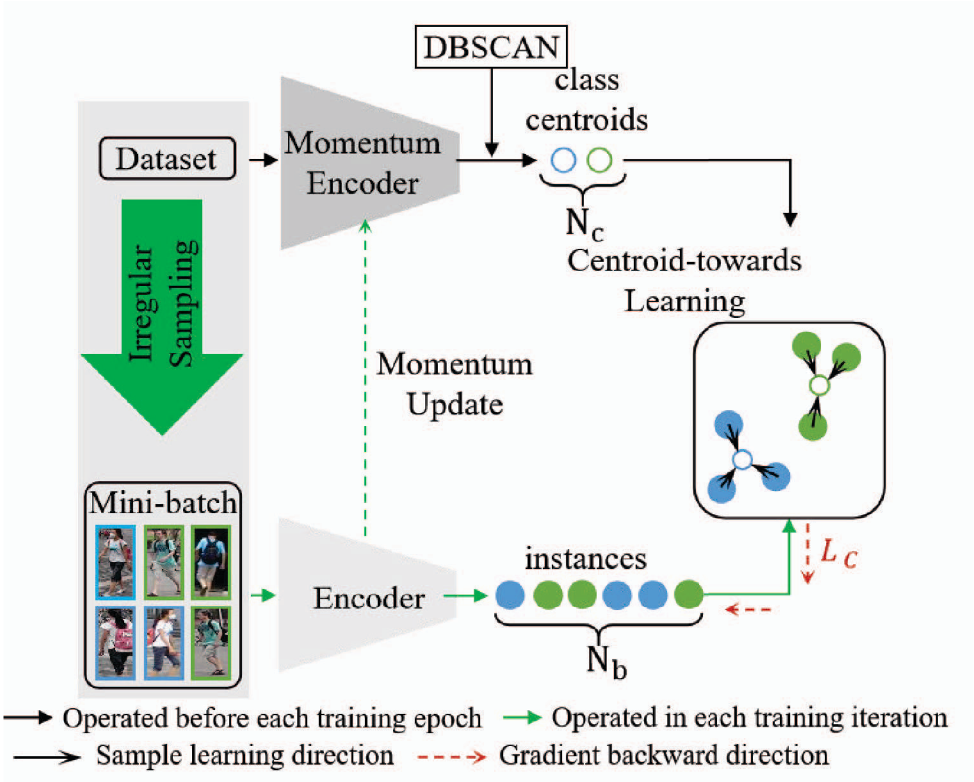
DBSCAN
class centroids
Dataset
Momentum Encoder
N_c
Centroid-towards Learning
Irregular Sampling
Momentum Update
Mini-batch
Encoder
instances
L_C
N_b
Operated before each training epoch
Operated in each training iteration
Sample learning direction
Gradient backward direction

Fig. 1. The illustration of the proposed fully unsupervised object re-ID framework. Before every training epoch, cluster algorithm DBSCAN [8] is used to roughly cluster every sample in the whole dataset into N_c classes. $C = \{c_1, ..., c_{N_c}\}$ represents centroids of N_c classes. In every training iteration, the encoder is fine-tuned according to C via centroids-towards learning and RHS learning, and the momentum encoder is updated by the encoder as Eq. (1) by momentum update [1].

[2], [16]. Random sampling randomly selects samples from the whole dataset to form each mini-batch, because the instance discrimination task considers each image as a distinct class for enforcing inter-class contrasting learning, as illustrated in Fig. 2(a).

Random sampling is also adopted in early re-ID works [12], [13], [17], however, it can not achieve satisfying performance. Random sampling can not ensure that every mini-batch contains inter-class samples. Random sampling leads to deteriorated over-fitting and harms the object re-ID performance because of neglecting intra-class learning [4], [5].

Therefore, recent researches [4], [6], [9], [18]–[20] adopted triplet sampling in re-ID task. They first roughly classify all samples into clustered inliers or unclustered outliers by clustering algorithm DBSCAN [8] or k-means. Then, P numbers of same class instances are selected to form mini-batch, intra-class and inter-class learning are perfromed simultaneously in every training iteration. Subsequently, Han et al. [5] proposed a Group Sampling strategy by addressing the deteriorated over-fitting problem in triplet sampling.

III. Proposed Method

The MoCo-based self-supervised contrastive learning framework [1] is adopted as the baseline framework in this work. The illustration of our proposed framework is shown in Fig. 1. Before every training epoch, unsupervised cluster algorithm DBSCAN [8] is used to roughly cluster every samples in the whole dataset into $C = \{c_1, ..., c_{N_c}\}$ classes. N_c denotes the numbers of clusters. Then, generated c_i fine-tunes the encoder model iteration by iteration until convergence.

In this section, we first describe the centroids-towards learning in Sec III.A. Then, we will introduce the proposed sampling strategy - Irregular Sampling (IS) in Sec III.B.

A. Centroids-towards learning

The objective of our work is to obtain a superior re-ID network, which can produce similar features for the same identity and produce distinct features for different identities. To achieve this goal, momentum contrast learning architecture MoCo [1] with InfoNCE loss [21] is used as the baseline to enforce centroids-towards learning. The framework of the proposed method is illustrated in Fig. 1.

The encoder and the momentum encoder are used to generate representations of instances and cluster centroids, respectively. We denote parameters of the Encoder as θ_e, and parameters of the Momentum Encoder as θ_{me}. θ_e is updated in each training iteration by gradient back-propagation. The momentum encoder, served as a robust encoder, updated by

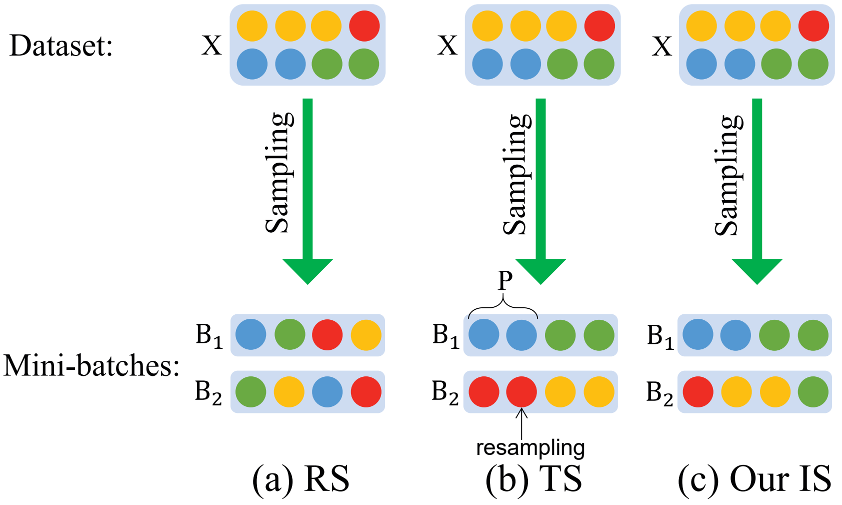

Fig. 2. The illustrations of sampling strategies. (a) Random Sampling (b) Triplet Sampling (c) Our proposed Irregular Sampling. The same color represents the samples sharing the same identity class.

θ_e with a momentum coefficient m after every iteration as follows,

$$\theta_{me} = m\theta_{me} + (1-m)\theta_e \tag{1}$$

Before each training epoch starts, given an unlabeled training dataset $X = \{x_1, ..., x_N\}$, all images representations $F_{me} = \{f_{me,1}, ..., f_{me,N}\}$ are extracted by the momentum encoder. Then, unsupervised dense-based clustering algorithm DBSCAN [8] clusters F_{me} into N_C numbers of clusters. After that, cluster centroids $C = \{c_0, ..., c_{N_C}\}$ are computed as the mean vector of all instances in the cluster. This clustering results are also used to split X into mini-batches by irregular sampling.

In each training iteration, given an irregular sampled mini-batch B, $F_e = \{f_{e,1}, ..., f_{e,N_B}\}$ are extracted by the encoder as representations of instances.

To pull intra-class instances close to their corresponding centroids and push other centroids away, the loss of centroids-towards learning L_C of an instance is designed based on InfoNCE loss [21] as follows,

$$L_C = -log\frac{exp(f_{e,i} \cdot c^+)/\tau}{exp(f_{e,i} \cdot C)/\tau} \tag{2}$$

, where $f_{e,i} \cdot c^+$ computes the distance between the instance x_i and its corresponding cluster centroid c^+, where $c^+ \in C$. $f_{e,i} \cdot C$ represents distances among x_i and all cluster centroids. τ is the temperature hyper-parameter.

B. Irregular Sampling (IS) Strategy for Input Preparation

The purpose of the sampling strategy is to split a whole dataset X into mini-batches $\{B_1, ..., B_{N_B}\}$ for contrastive training batch by batch. N_B denotes the numbers of batches, $N_B = \frac{|X|}{|B_i|}$. Previous works [4], [6], [9], [19], [22] utilized triplet sampling to enforce intra-class and inter-class learning simultaneously in every training iteration by sampling P numbers of same class samples in every mini-batch, as illustrated in Fig. 2(b). Different colors represent samples having different c_i classes. We denote the samples within the same cluster/class of c_i as $I(c_i)$, and the numbers of samples in $I(c_i)$ is represented as $|I(c_i)|$. If there is a large cluster ($|I(c_i)| \geq P$), only P instances are sampled. If there is a small cluster ($|I(c_i)| < P$), instances will be repeatedly sampled. P is a hyper-paremeter, which plays an important role in model performance. [5] demonstrated that larger P brings benefits in Memory-based re-ID framework by strengthening statistical stability of each class in a mini-batch. However, we found out that larger P harms model performance in MoCo-based re-ID framework, especially in $P = 16$.

To investigate the impact of P in the MoCo-based unsupervised re-ID framework, we perform experiments and report the results in Fig. 3. With the increase of P, triplet sampling harms the model performance consistently in three datasets, especially in $P = 16$. The dramatically declines are caused by the resampling in triplet sampling in two ways: 1) information leakage within a batch, and 2) training imbalance between small and large clusters.

1) Information leakage within a batch.: Repeatedly sampling in triplet sampling brings same and repeat patterns in a mini-batch, hence the model may learn to exploit such simple and repeat mini-batch information instead of learning correct representations of samples [1], [23]. To demonstrate this cheating behavior in triplet sampling, we further perform experiments "triplet sampling + perturbation factor" in Fig. 3. A perturbation factor $\sigma_p \sim N(0, 0.5)$ is added P to disturb the patterns in a mini-batch. More specifically, $P + \sigma_p$ same class samples are sampled in every mini-batch. Fig. 3 shows that the triplet sampling with perturbation factor makes good re-ID results in $P = 16$.

2) Training imbalance between small and large clusters: As mentioned in above, triplet sampling resamples samples from small cluster ($|I(c_i)| < P$) to form every mini-batch. Therefore, samples from smaller clusters have a higher pro-

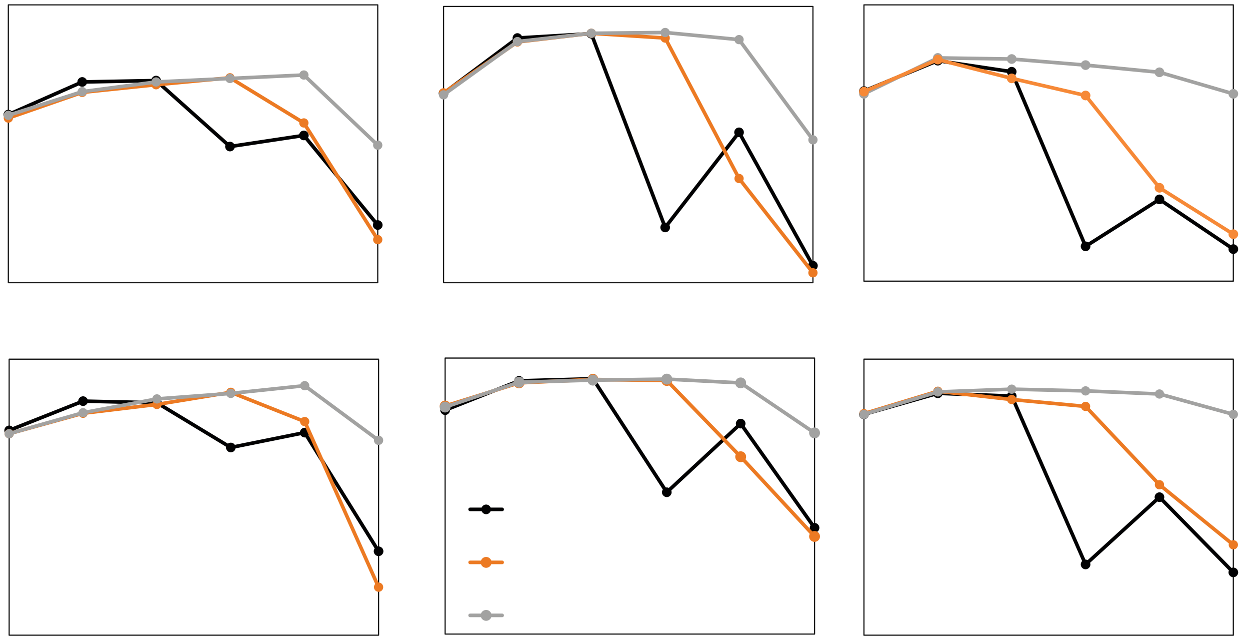

Fig. 3. The model performance of different sampling strategies in different P. The X-axis represents different numbers of positive samples P, and Y-axis represents the model performance. Graphs in the first-row report performance in mAP (%), and the second-row report performance in Rank-1 (%). Graphs in different columns report performance on different datasets.

Method	VeRi-776	
	mAP	R-1
SSML [24]	26.7	74.5
SpCL [4]	36.9	79.9
Ours	**39.9**	**85.2**

TABLE I. Experimental results of our proposed method and other fully unsupervised re-ID methods on vehicle re-ID datasets VeRi-776.

portion of updated than samples from large clusters in every training iteration. To investigate the impact of this imbalance problem, we further perform experiments "triplet sampling + w/o resampling" in Fig. 3. Without resampling from clusters, the steady, continued and consistent performance rise is observed with the increase of P.

The experimental results demonstrate the negative effect of resampling operation in triplet sampling. It is interesting to observe that without resampling indirectly breaks the same pattern in a mini-batch, as shown in Fig. 2(c). Therefore, here we introduce a simple but robust sampling strategy, called Irregular Sampling (IS). Different from triplet sampling, IS selects $|I(c_i)|$ numbers of instances for small clusters to avoid repeated sampling.

IV. Experiments

A. Datasets and Evaluation Metrics

The experiments are performed on three large-scale and mainstream datasets, i.e., one vehicle re-ID datasets, and two person re-ID datasets. **VeRi-776** [27] (VeRi) is a vehicle re-ID dataset, which has 20 cameras and 51,003 vehicle images of 775 identities in total. **Market-1501** [28] (Market) is a person re-ID dataset, which has 6 cameras and 32,217 person images of 1,501 identities in total. **MSMT17** [29] (MSMT) is a person re-ID dataset, which has 15 cameras and 126,441 person images of 4,101 identities in total.

Two evaluation metrics are used to measure model performance. The first one is Mean Average Precision (mAP) (%). Another one is the Cumulative Matching Characteristic (CMC) curve. The CMCs (%) of Rank-1 (R-1) is reported, which represents the probability of top-1 ranked gallery samples containing the query identity.

B. Implementation Details

ImageNet pre-trained ResNet-50 is used as the encoder and the momentum encoder. A batch normalization layer and an L_2-normalization layer are added after the last global pooling layer of ResNet-50 to generate 2048-dimensional features. The input images are resized to $256 \times 128 \times 3$. The size of training mini-batch N_b is 32. The network is trained by the Stochastic Gradient Descent (SGD) with a learning rate of 0.00055, 50

Method	Market-1501		MSMT17	
	mAP	Rank-1	mAP	Rank-1
BUC [11]	29.6	61.9	-	-
HCT [25]	56.4	80.0	-	-
MMCL [12]	45.5	80.3	11.2	35.4
DSCE [22]	61.7	83.9	15.5	35.2
SpCL [4]	79.1	88.1	19.1	42.3
CAP [19]	79.2	91.4	36.9	67.4
GS [5]	79.2	92.3	24.6	56.2
ICE (baseline) [6]	82.3	93.8	38.9	70.2
CCL [26]	82.1	92.3	27.6	56.0
HHCL [20]	**84.2**	93.4	-	-
Ours	83.4	**93.9**	**40.2**	**71.3**

TABLE II. Experimental results of our proposed method and other fully unsupervised re-ID methods on two person re-ID datasets. The top result is highlighted in bold.

Method	mAP	Rank-1
Random Sampling	55.3	76.5
Group Sampling [5]	76.0	91.0
Triplet Sampling	83.1	93.6
Irregular Sampling	**83.4**	**93.9**

TABLE III. Ablation study on different sampling methods. All results are obtained by our experiments using publicly available source code.

epochs in total. Hyper-parameters $m = 0.999$, $\tau = 0.05$, are used in all experiments for fair comparisons. Other hyper-parameters are selected for each datasets for achieving the best performance. $P = 20$ and $\tau = 0.15$ in VeRi-776, $P = 16$ and $\tau = 0.1$ in Market-1501, and $P = 8$ and $\tau = 0.1$ in MSMT17. The model performance with different P are reported in Fig. 3.

The experiments are performed on one NVIDIA Titan 1080Ti GPU with 11 GB of memory. The total training time is around 3 hours on Market-1501, and 6 hours on MSMT17 and VeRi-776.

C. Comparisons with The State-of-the-Arts in Three Datasets

The comparisons with the State-of-the-Arts fully unsupervised methods on one vehicle re-ID dataset VeRi-776 in Table I. We obtain mAP= 39.9% and Rank-1 $= 85.2\%$, which considerably outperforms SpCL. The superior performance indicates the effectiveness of our proposed Irregular sampling.

Comparisons are also performed in two person re-ID datasets, i.e., Market-1501 and MSMT17, which are reported in Table II. On Market-1501, our method achieves the best performance with mAP= 83.4% and Rank-1 $= 93.9\%$. Compared to the best MoCo-based re-ID method ICE, our method achieves good and competitive results. Specifically, our method outperforms ICE by 1.3% in mAP and 1.1% in Rank-1 in the largest and most difficult person re-ID datasets MSMT17.

D. Effectiveness of Irregular Sampling

We illustrate the model performance using triplet sampling and our proposed irregular sampling with different numbers of instances P in Fig. 3. It can be observed that P plays an important role in model performance. Small or large P indicates less or more instances belonging to the same class are selected in mini-batches, respectively. With the increase of P in triplet sampling, the model performance is increased first and then decreased rapidly. The performance increase is because selecting more positive instances helps the model learn more intra-class information and brings more statistical stability [5]. However, selecting more positive instances also causes a more serious imbalanced situation between small clusters and large clusters because more small clusters are re-sampled, as mentioned in Sec.III.

The above situation is not be observed after we remove the re-sampling operation. The performances do not decrease rapidly with the increase of P in irregular sampling. The experiments show that irregular sampling is a more effective and robust sampling strategy than triplet sampling.

We further compare our proposed irregular sampling with random sampling [1] and group sampling [5] in Table III. It can be clearly seen that random sampling has poor performance because it does select multiple positive instances in mini-batches, leading the model to neglect intra-class information. Group sampling can not achieve satisfying performance because it is designed based on Memory Bank architecture [2], [4]. Table III shows that group sampling is not suitable in MoCo-based architecture. Finally, irregular sampling achieves the best performance in MoCo-based architecture.

V. Conclusion

In this work, we research a fully unsupervised object re-ID method, which can be trained without using any labeled information. The current sampling strategies are analyzed. Based on the drawbacks of existing methods, we propose an effective and robust sampling strategy - irregular sampling. Experimental results on one vehicle and two person re-ID datasets show the effectiveness of the proposed sampling strategies - Irregular sampling.

Acknowledgment

This result was supported by ”Regional Innovation Strategy (RIS)” through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2021RIS-003).

References

[1] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, “Momentum contrast for unsupervised visual representation learning,” *arXiv preprint arXiv:1911.05722*, 2019.

[2] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, “Unsupervised feature learning via non-parametric instance-level discrimination,” *CoRR*, vol. abs/1805.01978, 2018. [Online]. Available: http://arxiv.org/abs/1805.01978

[3] T. S. Silva, “Exploring simclr: A simple framework for contrastive learning of visual representations,” *https://sthalles.github.io*, 2020.

[4] Y. Ge, F. Zhu, D. Chen, R. Zhao, and H. Li, “Self-paced contrastive learning with hybrid memory for domain adaptive object re-id,” in *Advances in Neural Information Processing Systems*, 2020.

[5] X. Han, X. Yu, G. Li, J. Zhao, G. Pan, Q. Ye, J. Jiao, and Z. Han, “Rethinking sampling strategies for unsupervised person re-identification,” 2021.

[6] H. Chen, B. Lagadec, and F. Bremond, “Ice: Inter-instance contrastive encoding for unsupervised person re-identification,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 14 960–14 969.

[7] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015.

[8] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *KDD*, 1996.

[9] S. Xuan and S. Zhang, “Intra-inter camera similarity for unsupervised person re-identification,” *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11 921–11 930, 2021.

[10] J. Peng, Y. Wang, H. Wang, Z. Zhang, X. Fu, and M. Wang, “Unsupervised vehicle re-identification with progressive adaptation,” *ArXiv*, vol. abs/2006.11486, 2020.

[11] Y. Lin, X. Dong, L. Zheng, Y. Yan, and Y. Yang, “A bottom-up clustering approach to unsupervised person re-identification,” in *AAAI Conference on Artificial Intelligence (AAAI)*, vol. 2, 2019, pp. 1–8.

[12] D. Wang and S. Zhang, “Unsupervised person re-identification via multi-label classification,” *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10 978–10 987, 2020.

[13] Q. Tang and K.-H. Jo, “Unsupervised person re-identification via nearest neighbor collaborative training strategy,” in *2021 IEEE International Conference on Image Processing (ICIP)*, 2021, pp. 1139–1143.

[14] Q. Tang, G. Cao, and K.-H. Jo, “Fully unsupervised person re-identification via multiple pseudo labels joint training,” *IEEE Access*, vol. 9, pp. 165 120–165 131, 2021.

[15] C.-Y. Wu, R. Manmatha, A. J. Smola, and P. Krahenbuhl, “Sampling matters in deep embedding learning,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[16] T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, “A simple framework for contrastive learning of visual representations,” *ArXiv*, vol. abs/2002.05709, 2020.

[17] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, “Invariance matters: Exemplar memory for domain adaptive person re-identification,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 598–607, 2019.

[18] Y. Ge, D. Chen, and H. Li, “Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification,” in *International Conference on Learning Representations*, 2020. [Online]. Available: https://openreview.net/forum?id=rJlnOhVYPS

[19] M. Wang, B. Lai, J. Huang, X. Gong, and X.-S. Hua, “Camera-aware proxies for unsupervised person re-identification,” in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

[20] Z. Hu, C. Zhu, and G. He, “Hard-sample guided hybrid contrast learning for unsupervised person re-identification,” *2021 7th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC)*, pp. 91–95, 2021.

[21] A. van den Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” *ArXiv*, vol. abs/1807.03748, 2018.

[22] F. Yang, Z. Zhong, Z. Luo, Y. Cai, Y. Lin, S. Li, and N. Sebe, “Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification,” *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4853–4862, 2021.

[23] Y. Wu and J. Johnson, “Rethinking ”batch” in batchnorm,” *ArXiv*, vol. abs/2105.07576, 2021.

[24] J. Yu and H. Oh, “Unsupervised vehicle re-identification via self-supervised metric learning using feature dictionary,” *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3806–3813, 2021.

[25] K. Zeng, M. Ning, Y. Wang, and Y. Guo, “Hierarchical clustering with hard-batch triplet loss for person re-identification,” *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13 654–13 662, 2020.

[26] Z. Dai, G. Wang, S. Zhu, W. Yuan, and P. Tan, “Cluster contrast for unsupervised person re-identification,” *ArXiv*, vol. abs/2103.11568, 2021.

[27] X. Liu, W. Liu, T. Mei, and H. Ma, “A deep learning-based approach to progressive vehicle re-identification for urban surveillance,” in *ECCV*, 2016.

[28] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable person re-identification: A benchmark,” *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1116–1124, 2015.

[29] L. Wei, S. Zhang, W. Gao, and Q. Tian, “Person transfer gan to bridge domain gap for person re-identification,” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 79–88, 2018.