Attribute Auxiliary Clustering for Person Re-identification

Ge Cao and Kanghyun Jo*

Department of Electrical, Electronic and Computer Engineering, University of Ulsan, Ulsan, 44610, Korea
{caoge,acejo}@ulsan.ac.kr

Abstract. The main objective of the person re-identification task is to retrieve the specific identity under multiple non-overlapping camera scenarios. Though unsupervised person re-ID has already achieved great performance and even surpasses some classic supervised re-ID methods, the existing methods pay much attention to training the neural networks with the memory-based idea which ignore the quality of the generated pseudo label. The quality of the clustering process does not only depend on the intra-cluster similarity but also on the number of clusters. In this paper, our approach employs an attribute auxiliary clustering method for person re-ID task. The proposed method could divide the generated cluster by the leveraged attribute label. Employed the attribute auxiliary clustering, the task changed from unsupervised case to weakly supervised case. The method is compared with state-of-the-art and analyzes the effectiveness caused by the variation of the cluster number. The proposed approach achieves great performance on the public Market-1501 datasets.

Keywords: weakly supervised person re-identification · attribute auxiliary clustering · cluster number variation

1 Introduction

The main objective of the person re-identification task is to retrieve the specific identity under multiple non-overlapping camera scenarios [1]. With the increasing requirements for video surveillance and the urge for lower label annotating costs, unsupervised person re-ID got more attention in the past few years. For dealing with the unsupervised person re-ID task, purely unsupervised re-ID [11], [20], [12], [16], [3] and the unsupervised domain adaptation are the widely applied method [2], [12], [22], [23].

In this paper, we focus on the purely unsupervised person re-ID task. The state-of-the-art methods [3] extracted feature embedding through neural network [13] and then employed the clustering algorithms, DBSCAN [4] commonly to generate the pseudo label for training samples. With the generated pseudo label, we can train as a supervised case. Finally, a contrastive loss [27] is employed for training. Though the existing method has already achieved great performance, it still didn't reach the upper bound of the baseline, where the upper bound means

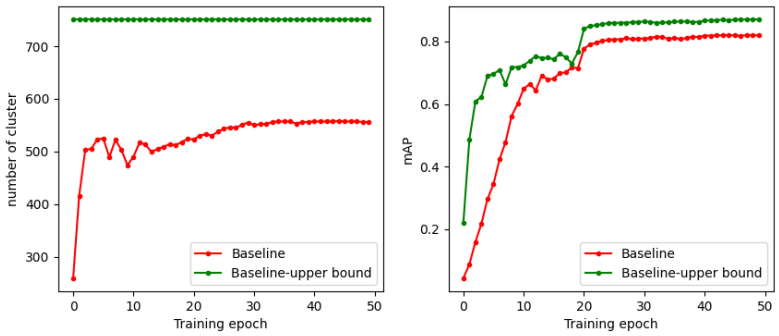

Fig. 1. The left and right subfigure shows the comparison of the number of clusters and performance between baseline and its upper bound (supervised)., respectively

the performance when the clustering process gains the ideal results (Assuming the clustering results are completely correct). In Fig. 1, we show the comparison of the number of clusters and performance between the baseline ClusterContrast [3] and its upper bound (supervised). The left subfigure shows that even after the training finished, the clustering process could only divide the training samples into around 500 clusters which is much less than the ideal number of clusters. Correspondingly, when we can get the ideal clustering results the upper bound obviously surpass the baseline by a large margin.

Fig. 2. Parts of clustering results selected from the final epoch's clustering results, where the samples of the same row are selected from the same cluster. Among them, the samples in the blue box and green box are captured from different identities.

The result demonstrates that the unsupervised method has not achieved the ideal performance and the key reason lies in the low quality of clustering quality. Fig. 2 displays parts of clustering results selected from the final epoch's clustering results, where the samples of the same row are selected from the same cluster. Among them, the samples in the blue box and green box are captured from different identities. The results show that the clustering could not recognize the highly similar vision features. But for human beings, we can easily find that the first four samples of the first row are captured from a male but the last four samples of the first row are captured from a female and in the same condition as the second row. In this paper, we generate the clustering results both in feature space and attribute space. The attributes of the identity annotate the sample at the semantic level. Our contribution could be summarized in three-fold:

- We leverage the attribute label and propose the attribute auxiliary clustering (AAC) method to explore the attribute auxiliary weakly supervised person re-ID task.
- The analysis of performance caused by cluster number variation is indicated in this paper.
- We comprehensively evaluate and compare the performance of AAC with state-of-the-art, which surpasses other weakly supervised person re-ID works.

2 Related Work

2.1 Unsupervised Person Re-ID Works

Despite the classic algorithm computing without deep learning, unsupervised person re-ID can be categorized into two situations. With the annotated label in the source domain, unsupervised domain adaptation (UDA) [2], [12], [22], [23] methods are the first category. Among them, ECN [22] firstly applied the memory bank idea [19] to store the features and update with the training process. SpCL [12] proposed a novel self-paced contrastive learning framework that gradually creates a more reliable cluster, which to refine the memory dictionary features. The second category is purely unsupervised person re-ID (USL) [11], [20], [12], [16], [3] which only focuses on the target dataset and does not leverage any labeled data. MMCL [11] employed the memory bank in the USL field and calculated the pseudo label with similarity. CAP [16] applied the cluster method DBSCAN to generate the pseudo label and construct the memory bank at cluster-level and proxy-level (detailed in camera id). ClusterContrast [3] summarized the mainstream contrastive learning-based USL method and mainly focused on controlling the cluster size for consistency in the training process. The proposed AAC is based on the ClusterContrast framework, and due to the leveraging of the attribute label, AAC is exploring the re-ID field under the weakly supervised case.

2.2 Attribute Auxiliary Person Re-ID

Thanks to the work and attribute annotation by Lin et al. [25], researchers are easier to train and learn the identity embedding with auxiliary attributes. GPS

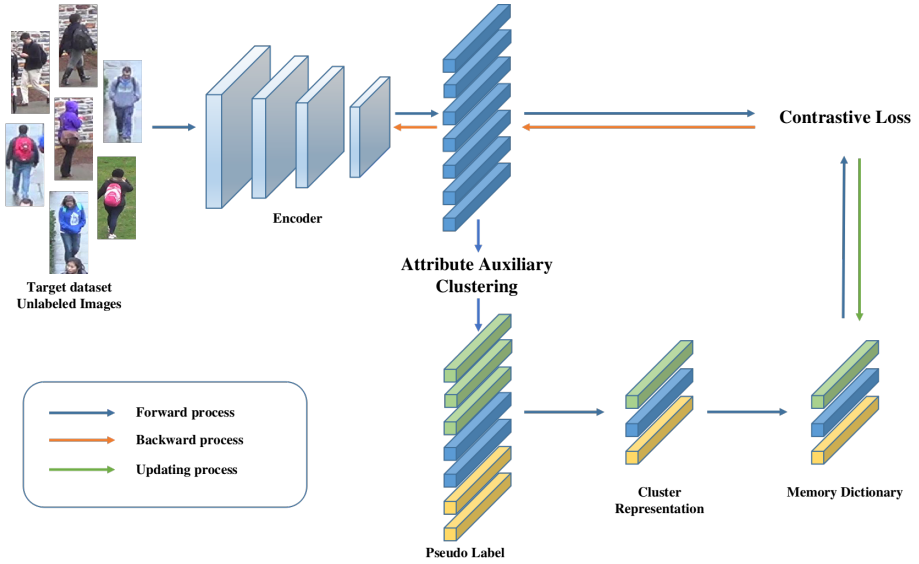

Fig. 3. The overview pipeline of the proposed method. The proposed attribute auxiliary clustering (AAC) method is applied for generating the pseudo label for the training samples. The ClusterNCE loss which is introduced in Eq. 1 is applied for the contrastive loss.

[24] constructs the relationship graph for identity attribute and human body part, which could represent the unique signature of the identity. The graph-based signature can also be employed in unsupervised cases [18]. TJ-AIDL [17] simultaneously trains with attribute level and feature level to transfer attribute and identity label information to the target domain. The proposed AAC re-allocated the clustering results rather than applying the attribute for training in the attribute-semantic space.

3 Methodology

The pipeline for purely unsupervised person re-ID is described in Section 3.1, which includes the re-ID problem formulation and training strategy followed [3]. The proposed attribute auxiliary clustering method is demonstrated in Section 3.2.

3.1 USL Person Re-ID pipeline

For the training process, given target dataset $X = \{x_1, x_2, ..., x_{N_t}\}$, we can extract discriminative feature embeddings $F = \{f_1, f_2, ..., f_{N_t}\}$ by the encoder network [13], where N_t denotes the number of training samples. The follow-up series of works employed for USL training is shown in Fig. 3. For the testing

Fig. 4. Parts of clustering results selected from the final epoch's clustering results, where the samples of the same row are selected from the same cluster. Among them, the samples in the blue box and red box are captured from different identities.

process, given query sample q and gallery samples $G = \{g_1, g_2, ..., g_{N_g}\}$, get the feature embedding f_q and $\{f_{g_1}, f_{g_2}, ..., f_{g_2}\}$ from the trained encoder network, then calculate the similarity between f_q and f_g, and finally rank the list.

In the training process, after extracting feature embedding from the encoder network, we employ the classic clustering algorithm DBSCAN [4] for generating the pseudo labels for training samples, which are denoted as $\{y_1, y_2, ..., y_{N_t}\}$. This work applies the ClusterNCE loss followed ClusterContrast [3] as the contrastive loss:

$$L = -log\frac{exp(f_q \cdot \phi_+/\tau)}{\sum_{k=0}^{K} exp(f_q \cdot \phi_k/\tau)} \tag{1}$$

where ϕ_+ is the positive cluster representation vector of q, and ϕ_k is negative unique representation vector of the k-th cluster. The cluster representation ϕ_k is initialized by Eq. 2:

$$\phi_k = \frac{1}{|N_k|} \sum_{f_i \in N_k} f_i \tag{2}$$

where N_k is the set of samples in the k-th cluster, and it is verified as the encoder network trains in the process. During the training process, we select K samples in P clusters and construct the training minibatch. The cluster representation vectors are updated by:

$$\phi_k \leftarrow m\phi_k + (1-m)q \tag{3}$$

where m is the momentum updating rate. And the above process followed [3] is framed as the baseline in this paper.

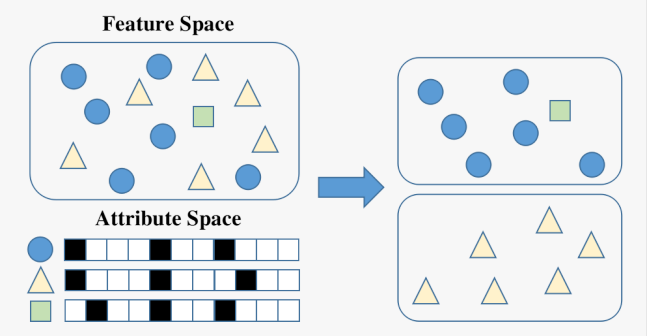

Fig. 5. Illustration of the proposed attribute auxiliary clustering (AAC) method. The samples in different shapes denote the sample captured from different identities. The samples are classified as the same cluster initially and re-clustered into the different clusters by applying AAC.

3.2 Attribute Auxiliary Clustering

Though baseline [3] has already gained state-of-the-art performance in the USL field, it still has plenty of space for improvement as shown in Fig. 1.

The upper bound of baseline: The upper bound means that we get the completely correct clustering results in every epoch. Obviously, the existing technology cannot reach that case, so the upper bound would happen when we directly apply the ground truth of the training label. For the baseline-upper-bound, we divide the training samples into N_t clusters directly applying the GT, so it is under the supervised case. The results are shown in Fig. 1 and Table. 1, which surpasses the baseline by a large margin. Due to the great potential for improving the clustering quality, we leverage the attribute label $A = \{a_1, a_2, ..., a_{N_t}\}$ for fine-tuning the clustering results. Operating with the clustering results of the final epoch of the baseline, there are mainly two cases. The first case is shown in Fig. 2, where a cluster contains samples captured from two or more identities and the samples from each identity could be separately clustered. And another case is shown in Fig. 4, where a cluster contains many clusters but the samples from some identities are just one or two, which cannot be individually clustered.

The process of AAC is shown in Fig. 5, where the samples in different shapes denote the sample captured from different identities. The samples are classified as the same cluster initially and re-clustered into the different clusters by applying AAC.

In the training process, given the training samples $\{x_1, x_2, ..., x_{N_t}\}$ and corresponding attribute label $\{a_1, a_2, ..., a_{N_t}\}$, we extract the feature embedding $\{f_{g_1}, f_{g_2}, ..., f_{g_{N_t}}\}$ from the encoder network. Then DBSCAN [4] is employed for generating the pseudo labels $\{y_1, y_2, ..., y_{N_t}\}$. The samples which have the same pseudo labels y_i are classified as the same cluster and some clusters contain different attribute labels a_i. For the samples which are in the same cluster and the

number of the samples with the same attribute label more than a threshold δ, we document their attribute label as $t_1, t_2, ..., t_K$, where K means the number of different attribute label in one cluster. The signal δ is set for avoiding generating some bad clusters with only a few samples which would be unbalanced distributed. So in the AAC algorithm, we will ignore the second case above.

We use the $y_i \rightarrow y_i'$ to denote the process that the training sample x_i should be re-clustered with new pseudo labels y_i':

$$\begin{aligned} & y_i \rightarrow y_i' \\ s.t. & \sum_{i=0, i \neq k}^{N_t} a_i \bigoplus t_k = 0 \end{aligned} \tag{4}$$

The discussions of the threshold δ and the start epoch for applying the AAC method are introduced in the ablation study.

4 Experiments

4.1 Datsets and Implementation

Datasets *Market-1501* [14] is a widely used public person re-ID dataset, which captured 12,936 samples with 751 identities in the training set, 3,368 and 15,913 samples captured from 750 identities for query and gallery set. The attribute label is provided by [25], which has 27 attributes for each training sample.

Implementation The ResNet50 [13] pre-trained on ImageNet [21] is employed for the encoder network. Followed [3], the feature embedding is 2048-d extracted by a global average pooling, batch normalization, and the L2-normalization layer.

The input of the samples is resized to 128×256 and processed by random horizontal flipping, padding, random cropping, and random erasing. The batch size is equal to 256 (16 samples from each identity). Adam is applied for the optimizer with 5e-4 of the weight decay. The initial learning rate is 5.5e-4 and reduced to ten times smaller every 20 epochs in a total of 60 epochs.

The maximum distance is set to 0.6 and the minimal number of clusters is set to 4 for the DBSCAN setting. The threshold δ is set to 5. and from the first epoch, we start to apply the AAC method.

4.2 Comparison with State-of-the-arts

We compare the proposed method with stat-of-the-arts. The method with attributes weakly supervised is few so we compare it with some SOTA USL papers. For Table. 1, the 'Setting' means the training case they applied and 'Auxiliary' means whether any auxiliary information is leveraged. And the mAp, rank-1 score, rank-5 score, and rank-10 score of the proposed AAC method surpasses the baseline [3] by 3.9%, 2.0%, 0.4%, and 0.6%, respectively.

Table 1. Comparison results with the state-of-the-arts on Market-1501 [14] dataset. In the table, AAC denotes the proposed attribute auxiliary clustering algorithm by this paper, GT denotes the ground truth, and the signal † denotes the results are tested under the same implementation with the proposed idea. The best results are bold in this table. Additionally, the upper bound of the baseline is shown as the maximum limit of unsupervised work.

Method	Reference	Setting	Auxiliary	Market1501			
				mAP	rank-1	rank-5	rank-10
LOMO [5]	CVPR15	USL	None	8.0	27.2	41.6	49.1
BOW [6]	ICCV15	USL	None	14.8	35.8	52.4	60.3
UDML [7]	CVPR16	USL	None	12.4	34.5	52.6	59.6
DECAMEL [8]	TPAMI18	USL	None	32.4	60.2	76.0	81.1
TJ-AIDL [17]	CVPR18	Weakly	Attribute	26.5	58.2	74.8	81.1
DBC [10]	BMVC19	USL	None	41.3	69.2	83.0	87.8
BUC [9]	AAAI19	USL	None	38.3	66.2	79.6	84.5
MMCL [11]	CVPR20	USL	None	45.5	80.3	89.4	92.3
SpCL [12]	NeurIPS20	USL	None	73.1	88.1	95.1	97.0
GCL [20]	CVPR21	USL	None	66.8	87.3	93.5	95.5
CAP [16]	AAAI21	USL	Camera ID	79.2	91.4	96.3	97.7
A2G [15]	Access21	Weakly	Attribute	71.6	87.4	95.2	97.2
ClusterContrast† [3]	ACCV22	USL	None	82.1	92.3	96.9	97.6
AAC	This paper	Weakly	Attribute	**86.0**	**94.3**	**97.9**	**98.5**
Baseline-upper	This paper	Supervised	GT	87.2	95.0	98.3	99.1

Fig. 6 shows the performance comparison and cluster number comparison among the baseline [3] (USL), AAC (weakly supervised), and baseline upper bound (supervised). The right subfigure shows that the performance of the proposed AAC surpasses the baseline a lot and is already close to the supervised upper bound. For the left subfigure which shows the cluster number variation with the training epoch, the cluster number increases very rapidly after applying AAC in some of the first epochs. It is caused by the poor clustering quality, as shown in Fig. 7, the clustering results of some of the first epochs are very bad for training, most clusters contain many samples captured from many identities. So when we apply the AAC idea with a small threshold δ, the cluster number would increase a lot and then decrease to the stable situation with the training.

4.3 Ablation Studies.

In this section, we introduce the extended experiments about the starting epoch for applying the AAC method and the effectiveness contributed by a changed number of clusters. As shown in Fig. 7, the samples are mostly clustered into wrong pseudo labels, so it is necessary for exploring whether the AAC should be applied in the initial epoch. Table. 2 demonstrates the performance when applying AAC in different start epochs with threshold δ=10, and Fig. 8 shows the

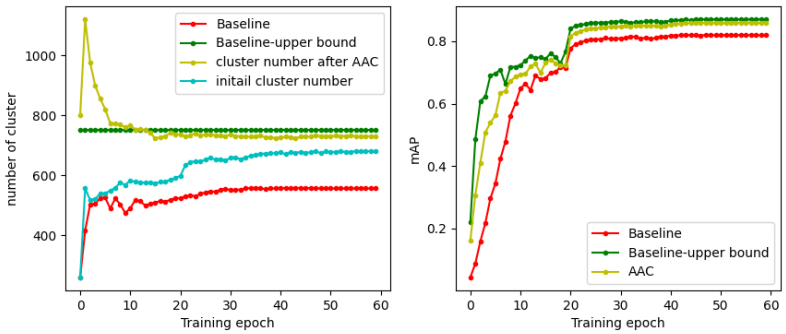

Fig. 6. The left subfigure shows the comparison of the number of clusters among the baseline, baseline upper bound (supervised), initial cluster number (before applying AAC), and cluster number after AAC (after applying AAC). The right subfigure shows the comparison of the mAP performance among the baseline, baseline upper bound, and the proposed AAC. The value of AAC is tested under the best performance.

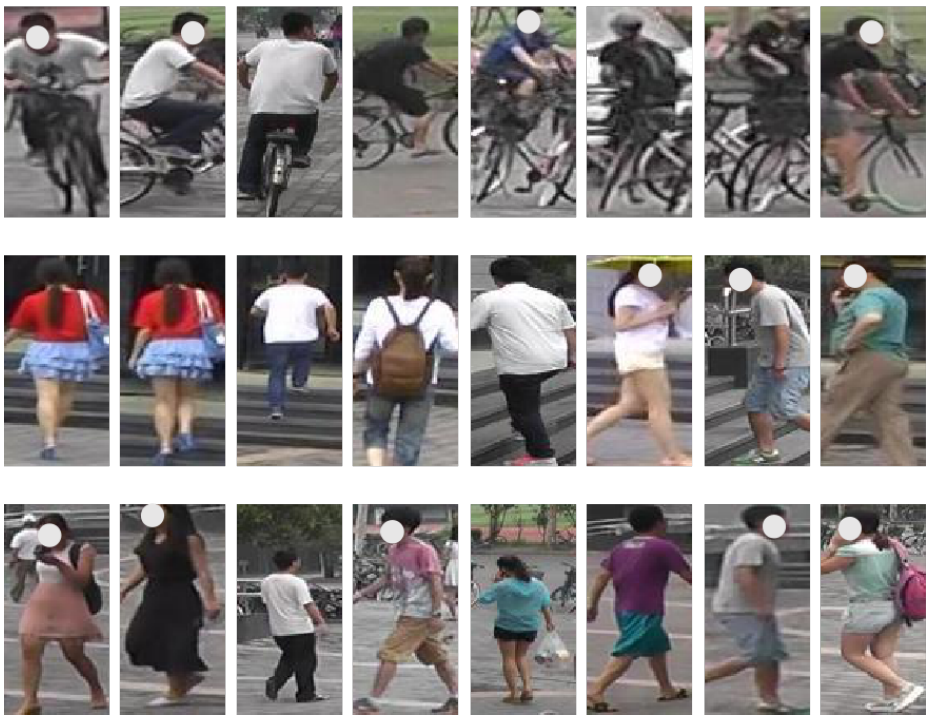

Fig. 7. Parts of clustering results selected from the first epoch's clustering results, where the samples of the same row are selected from the same cluster. Among them, the samples are mostly clustered into the wrong pseudo-label.

Table 2. Retrieval accuracy with different epochs for starting applying the proposed AAC method (δ=10 in this experiment). The best performance is bold.

Start Epoch	Market-1501			
	mAP	rank-1	rank-5	rank-10
0	**84.9**	**93.6**	**97.3**	**98.2**
5	84.6	93.6	97.4	98.3
10	84.2	93.8	97.5	98.2
15	83.2	92.9	96.9	97.8
20	83.0	92.6	96.7	97.7
25	82.7	92.7	96.7	97.6
30	82.6	92.4	96.6	97.7
35	82.2	92.3	96.5	97.5
40	82.0	92.3	96.5	97.6

performance when applying AAC with the different epochs. The results indicate that applying AAC during the training process achieves the best performance.

About the ablation study for the threshold δ, we test the performance from 4-10 for the Market-1501 dataset. The results do not have a linear pattern and we achieve the best performance with δ=5.

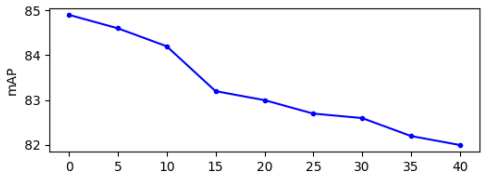

Fig. 8. The performance when applying AAC in different epochs.

Discussions: The effectiveness of applying AAC from the first epoch is best because of the low inter-class variations which caused a low initial cluster number in the Market-1501 dataset (the cluster numbers of some of the first epochs are much smaller than the total identity number).

5 Conclusions

This paper proposes the attribute auxiliary clustering method for weakly supervised person re-identification work. It re-allocates the pseudo label for training samples and effectively improves the performance and convergence speed compared with the baseline. The experiments show that the proposed idea achieves state-of-the-art.

Table 3. Retrieval accuracy with different epochs for starting applying the proposed AAC method. The best performance is bold.

δ	Market-1501			
	mAP	rank-1	rank-5	rank-10
3	85.8	94.3	97.8	98.8
4	85.9	94.3	97.9	98.5
5	**86.0**	**94.2**	**97.6**	**98.4**
6	85.5	94.1	97.6	98.5
7	85.4	93.7	97.3	98.3
8	85.4	93.6	97.3	98.1
9	85.7	93.8	97.8	98.7
10	83.0	92.6	96.7	97.7

Acknowledgements

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. Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., and Hoi, S. C. H., "Deep Learning for Person Re-identification: A Survey and Outlook", arXiv e-prints, 2020.
2. Liangchen Song, Cheng Wang, Lefei Zhang, Bo Du, Qian Zhang, Chang Huang, and Xinggang Wang. Unsupervised domain adaptive re-identification: Theory and practice. PR, 2020. 1, 2
3. Dai, Z., Wang, G., Yuan, W., Liu, X., Zhu, S., and Tan, P., "Cluster Contrast for Unsupervised Person Re-Identification", arXiv e-prints, 2021.
4. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X.; et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd.
5. S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person Re-identification by Local Maximal Occurrence Representation and Metric Learning," arXiv e-prints, p. arXiv:1406.4216, Jun. 2014.
6. L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," 12 2015, pp. 1116–1124
7. P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, and Y. Tian, "Unsupervised cross-dataset transfer learning for person reidentification," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1306–1315.
8. H. Yu, A. Wu, and W. Zheng, "Unsupervised person re-identification by deep asymmetric metric embedding," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 4, pp. 956–973, 2020.
9. Y. Lin, X. Dong, L. Zheng, Y. Yan, and Y. Yang, "A bottom-up clustering approach to unsupervised person re-identification," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 8738–8745, 07 2019.
10. G. Ding, S. H. Khan, and Z. Tang, "Dispersion based clustering for unsupervised person re-identification," in BMVC, 2019.

11. D. Wang and S. Zhang, “Unsupervised Person Re-identification via Multi-label Classification,” arXiv e-prints, p. arXiv:2004.09228, Apr. 2020.
12. 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. 1, 2, 3, 7, 8, 10, 11
13. He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In CVPR.
14. Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In ICCV, 2015. 2, 5, 6, 7, 8.
15. G. Tang, X. Gao, Z. Chen and H. Zhong, "Graph Neural Network Based Attribute Auxiliary Structured Grouping for Person Re-Identification," in IEEE Access, doi: 10.1109/ACCESS.2021.3069915.
16. Menglin Wang, Baisheng Lai, Jianqiang Huang, Xiaojin Gong, and Xian-Sheng Hua. Camera-aware proxies for unsupervised person re-identification. In AAAI, 2021. 2, 3, 4, 7, 8, 11.
17. Jingya Wang, Xiatian Zhu, Shaogang Gong, and Wei Li. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2275–2284, 2018.
18. G. Cao, Q. Tang and K. Jo, "Graph-based Attribute-aware Unsupervised Person Re-identification with Contrastive learning," 2022 International Workshop on Intelligent Systems (IWIS), Ulsan, Korea, Republic of, 2022, pp. 1-6, doi: 10.1109/IWIS56333.2022.9920894.
19. Wu, Z., Xiong, Y., Yu, S., and Lin, D., “Unsupervised Feature Learning via Non-Parametric Instance-level Discrimination”, arXiv e-prints, 2018.
20. Chen, H., Wang, Y., Lagadec, B., Dantcheva, A., and Bremond, F., “Joint Generative and Contrastive Learning for Unsupervised Person Re-identification”, arXiv e-prints, 2020.
21. J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.
22. Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, “Invariance Matters: Exemplar Memory for Domain Adaptive Person Re-identification,” arXiv e-prints, p. arXiv:1904.01990, Apr. 2019.
23. Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, and T. Huang, “Self-similarity Grouping: A Simple Unsupervised Cross Domain Adaptation Approach for Person Re-identification,” arXiv e-prints, p. arXiv:1811.10144, Nov. 2018.
24. B. X. Nguyen, B. D. Nguyen, T. Do, E. Tjiputra, Q. D. Tran and A. Nguyen, "Graph-based Person Signature for Person Re-Identifications," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021, pp. 3487-3496, doi: 10.1109/CVPRW53098.2021.00388.
25. Yutian Lin, Liang Zheng, Zhedong Zheng, Yu Wu, Zhilan Hu, Chenggang Yan, and Yi Yang. Improving person re-identification by attribute and identity learning. Pattern Recognition, 95:151–161, 2019.
26. H.-X. Yu, W.-S. Zheng, A. Wu, X. Guo, S. Gong, and J.-H. Lai, “Unsupervised Person Re-identification by Soft Multilabel Learning,” arXiv e-prints, p. arXiv:1903.06325, Mar. 2019.
27. Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748, 2018. 1, 2