CLUSTERING AND DYNAMIC SAMPLING BASED UNSUPERVISED DOMAIN ADAPTATION FOR PERSON RE-IDENTIFICATION

Jinlin Wu[1,2], *Shengcai, Liao*[3], *Zhen Lei*[1,2], *Xiaobo Wang*[4], *Yang Yang*[*,1,2], *Stan Z. Li*[1,2]

[1]CBSR&NLPR, Institute of Automation Chinese Academy of Sciences, Beijing, China, 100190
[2]University of Chinese Academy of Sciences, Beijing, China, 100049
[3] Inception Institute of Artificial Intelligence
[4] JD AI Research, Beijing, China
{jinlin.wu, zlei, szli, yang.yang}@nlpr.ia.ac.cn, scliao@ieee.org, wangxiaobo8@jd.com

ABSTRACT

Person Re-Identification (Re-ID) has witnessed great improvements due to the advances of the deep convolutional neural networks (CNN). Despite this, existing methods mainly suffer from the poor generalization ability to unseen scenes because of the different characteristics between different domains. To address this issue, a Clustering and Dynamic Sampling (CDS) method is proposed in this paper, which tries to transfer the useful knowledge of existing labeled source domain to the unlabeled target one. Specifically, to improve the discriminability of CNN model on source domain, we use the commonly shared pedestrian attributes (*e.g.*, gender, hat and clothing color *etc.*) to enrich the information and resort to the margin-based softmax (*e.g.*, A-Softmax) loss to train the model. For the unlabeled target domain, we iteratively cluster the samples into several centers and dynamically select informative ones from each center to fine-tune the source-domain model. Extensive experiments on DukeMTMC-reID and Market-1501 datasets show that the proposed method greatly improves the state of the arts in unsupervised domain adaptation.

Index Terms— Clustering, Dynamic Sampling, Pedestrian Attributes, A-Softmax

1. INTRODUCTION

Person Re-identification (Re-ID) is fundamental and of great practical value in surveillance video analysis, of which the core is to match the same pedestrian across multiple camera views. Recently, benefiting from the development of deep learning [1], the performance of Re-ID has obtained significant improvements by adopting supervised learning [2, 3, 4]. However, in real applications, it is expensive and time-consuming to obtain sufficient manually labeled data. Moreover, due to the diversity of data acquisition, the domain gap between different datasets is a critical issue. For example, the pedestrians from DukeMTMC-reID [5] usually wear long and thick clothes while those from Market-1501 [6] almost wear short and thin clothes. Besides the foreground gap, the backgrounds between these two datasets are also significantly different (see Fig.1). These factors lead to a sharp performance decrease when directly applying the trained model from the labeled source domain to the unlabeled target one (*e.g.*, Rank-1 rate declines from 76.2% when trained and evaluated on the Market-1501 to 36.1% when directly evaluated on the DukeMTMC-reID).

Fig. 1: Domain differences between the DukeMTMC-reID [5] and Market-1501 [6] datasets, including clothing styles, backgrounds and illuminations. Pedestrians in DukeMTMC-reID (upper row) usually wear long and thick clothes while the ones in Market-1501 (bottom row) generally wear short and thin clothes.

To address the domain gap, a large number of methods have been proposed. One group considers this issue as a style transferring task. The methods in [7, 8] adopt the generative adversarial nets [9] transferring the unlabeled target image style to labeled source image style, trying to eliminate the domain gap between the different datasets. However, the discriminant information might be lost in the transferring process, which leads to limited Rank-1 accuracy. Another group resort to the unsupervised domain transfer meth-

This work was supported by the National Key Research and Development Plan (Grant No.2016YFC0801003), the Chinese National Natural Science Foundation Projects Grant #61806203 and #61672521.

* Corresponding author.

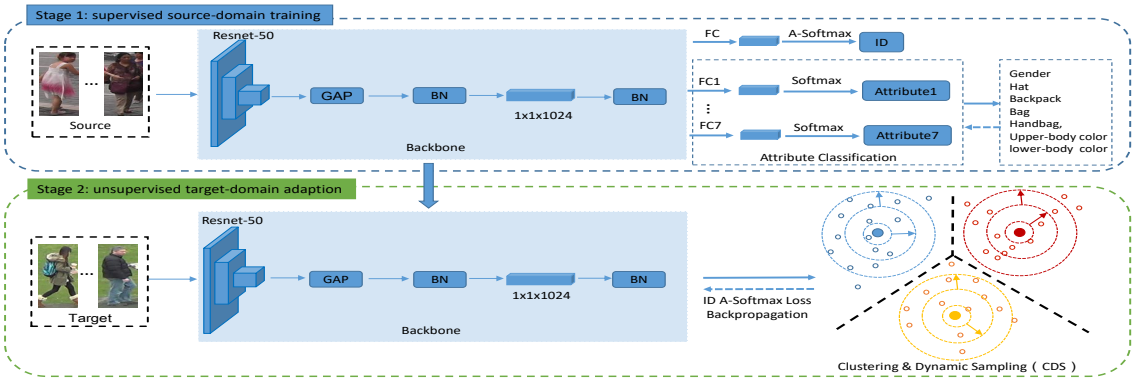


Fig. 2: The framework of the proposed Clustering and Dynamic Sampling (**CDS**) method: we firstly train our source-domain model with both identity classification and attribute classification. Seven attributes including gender, hat, backpack, bag, handbag, upper-body clothing color and lower body clothing color are employed to enhance the source training of the source-domain model. Then we iteratively cluster the target-domain samples and dynamically select informative samples in target domain to fine-tune the source-domain model.

ods. Approaches in [10, 11, 12, 13, 14] adopt unsupervised approaches to enhance the model's ability on the target unlabeled dataset. Wang *et al.* [10] propose an unsupervised transfer method called TJ-AIDL, transferring source domain knowledge to target domain by learning the attribute-semantic and identity-discriminative feature space. Fan *et al.* [14] propose a progressive unsupervised learning (PUL) approach obtaining pseudo labels to fine-tune the source-domain model. However, its performance is limited to the source-domain model's poor generalization and the fixed-threshold sampling.

In this paper, we propose a novel clustering and dynamic sampling (CDS) scheme to address the cross-domain unsupervised adaption issue. The framework of our CDS is shown in Fig. 2. Specifically, we firstly train the source-domain model with both identities and pedestrian attributes, under the supervision of the angular softmax loss (A-Softmax [15]). Generally speaking, that identities labels from different pedestrian datasets are non-overlapping, while the attributes [3] (*e.g.*, hat, clothing color, gender, bags and backpack) are commonly shared. In this work, we use attribute classification to improve the discriminability of the baseline source-domain model. We adopt the angular softmax loss (A-Softmax [15]) to supervise the training process. Then, to overcome the drawbacks of the fixed-threshold sampling in clustering centers, we employ a dynamic sampling strategy to select informative training samples on the target domain. To sum up, the contributions of this paper can be summarized as follows:

- We propose a novel Clustering and Dynamic Sampling (CDS) scheme to address unsupervised cross-domain adaptation task in person re-identification.
- We employ the commonly shared attributes and the margin-based A-Softmax loss to improve the discriminability of the source-domain model and adopt a dynamic threshold to select informative target-domain samples for unsupervised adaptation.
- Extensive experiments on the DukeMTMC-reID and Market-1501 datasets show that the proposed method largely improves the state of the arts in unsupervised domain adaptation.

2. RELATED WORK

2.1. Supervised Person Re-Identification

Most of existing Re-ID models are built by learning similarity or representation via cross-camera data [4, 16, 17, 3, 2, 18] in a supervised manner. In early works, Yi *et al.* [4] introduce part priors into deep convolutional network for similarity learning. In recent works, Hermans *et al.* [3] propose a multi-task training approach, combining ID classification and attributes classification tasks to improve the Re-ID performance. Fan *et al.* [2] introduce an angular margin to the softmax function, reducing intra-class variations and increasing inter-class variations simultaneously. These approaches achieve superior performance in a single domain task, but may not be effective for cross-domain person Re-ID.

2.2. Cross-Domain Person Re-Identification

To address the cross-domain issue, researchers shift their attention to the unsupervised methods, including unsupervised clustering [14], image-style transfer [7, 19, 8, 20] and unsupervised domain transfer [11, 10, 21]. To unsupervised clustering, Fan *et al.* [14] propose a clustering and self-pace training frame named PUL, which adopts the k-means algorithm to obtain pseudo labels and fine-tunes the source-domain model with these pseudo labels. However, in this work, the threshold for controlling training data sampling is

Fig. 3: Commonly shared attributes (*e.g.*, gender, backpack, upper-body color) on Market-1501 and DukeMTMC-reID.

fixed, in which a conservative value may result in losing correct cross-camera person pairs while an aggressive one may involve outliers. Besides, a poor source-domain may also hinder the performance of the latter self-pace training. For image-style transfer approaches, existing work try to bridge the domain gap through transferring the unlabeled image style to the labeled image style. For unsupervised domain transfer methods, Lin *et al.* [21] adopt the domain separation network to learn domain-shared representations.

3. PROPOSED METHODS

Notation. Given a labeled source dataset X^s and an unlabeled target dataset X^t, where $X^t = \{x_1^t, x_2^t, ..., x_N^t\}$ contains N images, the source-domain model ϕ_θ is pre-trained on the source dataset, where θ denotes the corresponding parameters. Since the target dataset is unlabeled, we consider the latent ground truth labels of X^t as $Y^t = \{y_1^t, y_2^t, ..., y_N^t\}$. While the pseudo labels predicted by the k-means algorithm are donated as $\hat{Y}^t = \{\hat{y_1^t}, \hat{y_2^t}, ..., \hat{y_N^t}\}$.

3.1. Supervised Learning on Source Domain

Attribute Knowledge. Person identies are usually non-overlapping across different Re-ID datasets, while pedestrian attributes are commonly shared. For instance, as shown in Fig. 3, the attributes female, backpack, red Upper-body are shared on the Market-1501 and DukeMTMC-reID. So it seems that the attribute knowledge is easier to transfer than identity knowledge. To this end, besides the identity classification, we additionally use seven common attributes (gender, hat, backpack, bag, handbag, upper-body clothing color, lower body clothing color) to improve the generalization ability of the source-domain model, so as to make a better starting point for the latter fine-tuning.

A-Softmax. For the loss function of the supervised learning, as we expect that the representations learned by the source-domain model are discriminative and are easy to transfer. Therefor we adopt the angular softmax loss (A-Softmax [15]) to train the source-domain model. Specifically, A-Softmax loss aims to increase the inter-class variations and meanwhile decrease the intra-class variations by an adding angular margin to the original softmax loss. The A-Softmax loss is formulated as follows:

$$\begin{aligned} L_{\text{A-Softmax}} &= -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{z_{y_i}}}{\sum_{j\neq y_i}^{C} e^{z_j} + e^{z_{y_i}}} \\ z_{y_i} &= \| x_i \| (-1)^k cos(m\varphi_{y_i,i}) - 2k \\ z_{y_j} &= \| x_j \| cos(\varphi_{y_j,j}), \end{aligned} \tag{1}$$

where $k \in [0, m-1]$ and m is the margin factor as a hyper parameter. For more details, please refer to the work [15].

3.2. Clustering and Dynamic Sampling

Clustering. On the target domain, since the samples are unlabeled, we firstly cluster them into several centers by using the well-known k-means algorithm. The formulation can be summarized as:

$$\min_{\hat{Y^t}, c_1, c_2, ..., c_K} \sum_{k=1}^{K} \sum_{\hat{y_i^t}=k} \| \phi_\theta(x_i^t) - c_k \|_2, \tag{2}$$

where K is the number of clusters pre-defined empirically, and c_i, $i \in \{1, 2, \ldots, K\}$ are the corresponding cluster centers.

Dynamic Sampling. Generally, samples in each cluster may contain noisy pseudo labels or outliers due to the limited ability of the initial source-domain model. Regarding this, we define a binary mask v_i and use a threshold λ to indicate whether a sample belongs to a certain cluster or not:

$$v_i = \begin{cases} 1, & cos(\phi_\theta(x_i^t), c_{x_i^t}) > \lambda, \\ 0, & otherwise, \end{cases} \tag{3}$$

where $c_{x_i^t}$ denotes the corresponding cluster center of the image x_i^t. From the definition, we can see that if the cosine similarity $cos(\phi_\theta(x_i^t), c_{x_i^t})$ is larger than a threshold λ, the sample x_i^t and its pseudo label $\hat{y_i^t}$ would be temporarily selected. Otherwise, it will be discarded. After that, we use the selected samples to fine-tune the source-domain model by optimizing the following objective:

$$\min_{\theta, \mathbf{w}} \sum_{i=1}^{N} v_i L(\hat{y_i^t}, f_w(\phi_\theta(x_i^t))), \tag{4}$$

where L is the A-Softmax loss function, and f_w is the classifier of the target domain.

For the threshold λ, a large value prefer to choose samples belonging to the same identity under the same camera. Such selected samples are relatively reliable, but they may be useless for cross-camera person retrieval. In contrast, a small value usually involves noisy samples or outliers. As a result, the model may be misled by them during the fine-tuning process. To solve this dilemma, we design a dynamic sampling strategy. Specifically, at the beginning of the training, since the model's discriminability is weak on the target

Algorithm 1 Clustering and Dynamic Sampling (CDS)

Input: Unlabeled dataset X; Upper bound U and lower bound L; Source domain pre-trained model ϕ_{θ_s}; Target domain classifier f_w; Relaxing rate η.
Output: Fine-tuned model ϕ_{θ_t};

```
Initial λ_0 ← U, θ_t ← θ_s ;
repeat
    extracting deep feature: f_i = φ_{θ_t}(x_i^t) for all x_i^t;
    L_2-normalization for deep feature f_i;
    k-means clustering;
    updating pseudo labels ŷ^t and cluster center C;
    for i = 1 to N do
        find the nearest cluster center c_{x_i^t} of x_i^t
        S = f_i · c_{x_i^t}
        if S > λ then
            selecting x_i^t: v_i = 1
        else
            discarding x_i^t: v_i = 0
        end if
    end for
    training < φ_{θ_t}, f_w > with the selected samples
    updating λ: λ ← λ − η · (U − L)
until (λ < L)
```

domain, it is easily misled by noisy samples or outliers. Thus in this case we set a larger threshold to cautiously select reliable samples, although these samples may belong to the same camera and the same ID. When the model's discriminability becomes stronger, we can set a lower threshold to select more informative samples including the cross-camera pairs to improve the model's cross-camera retrieval ability. To this end, we develop a dynamic sampling strategy by using a decreasing rate η to iteratively decrease the sampling threshold λ:

$$\lambda \leftarrow \lambda - \eta(U - L), \tag{5}$$

where a large threshold U is served as the upper bound and a small threshold L is used as the lower bound. The threshold λ is decreasing from the upper bound to the lower bound, as the model's discriminative ability is being enhanced gradually. For clarity, the proposed clustering and dynamic sampling (CDS) scheme is summarized in algorithm 1.

4. EXPERIMENTS

4.1. Datasets

We conduct extensive experiments on DukeMTMC-reID [5], Market-1501 [6] and CUHK03 [16] to evaluate our approach. Data statistics and evaluation settings are shown in table 1. On both Market-1501 and DukeMTMC-reID, we adopt the standard single-query protocol [6] as the testing protocol. On CUHK03, we follow the testing protocol of [22] to evaluate our method.

4.2. Implementation details

Our approach is implemented with Pytorch framework. The model structure and the training pipeline are shown in 2. The classifier f_w of target domain consists of one linear layer. The backbone is pre-trained with ID classification and attributes classification tasks on the source dataset. Seven pedestrian attributes (*i.e.*, gender, hat, backpack, bag, handbag, upper-body clothing color, lower body clothing color) are used in the source training process. A-Softmax is used as the loss function for ID classification task on both source training and target training. The m of A-Softmax is set to 3 in the experiments. The input images are resized to 288×188. We use Adam optimizer with the hyper-parameters($\epsilon = 10^{-8}, \beta_1 = 0.9, \beta_2 = 0.99, lr = 0.0001$). For the dynamic sampling, we set the dynamic decreasing rate to 1.5×10^{-3}. The upper bound U and lower bound L are set to 0.8 and 0.7 respectively. The batch size is set to 16.

	#All	#Train	#Test
Market-1501	1501/32688	751/12936	750/19732
DukeMTMC-reID	1812/36411	702/16522	702/19889
CUHK03	1467/14097	767/7365	700/6732

Table 1: Data splits (/#persons/#images) for experiments on the Market-1501 and DukeMTMC-reID datasets.

Strategy		Market->Duke		Duke->Market	
		mAP	Rank-1	mAP	Rank-1
Atributes	no	15.8	31.0	19.4	47.3
	yes	**19.9**	**37.2**	**22.4**	**50.1**
Fixed or Dynamic	0.85	33.0	57.3	32.9	64.6
	0.8	39.6	64.3	37.3	67.3
	0.7	35.6	58.8	34.4	64.8
	0.6	29.71	49.8	36.12	63.18
	DS	**42.7**	**67.2**	**39.9**	**71.8**
A-Softmax	no	39.3	65.5	37.6	68.9
	yes	**42.7**	**67.2**	**39.9**	**71.8**

Table 2: Ablation Experiments. The "Attributes" denotes whether using attributes on the source domain training. For the "Fixed or Dynamic", "Fixed" denotes using a fixed threshold during the sampling procedure and "Dynamic" means using dynamic sampling (DS) strategy. The "A-Softmax" denotes whether using the A-Softmax as loss function in the target training procedure.

4.3. Ablation Experiments

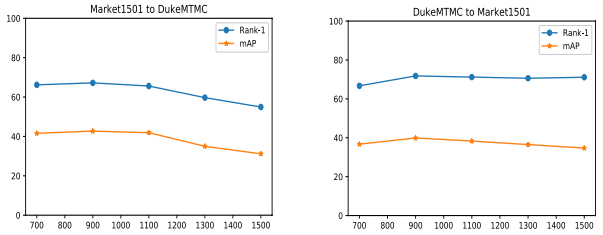
In this section, we analyze the impact of the three main strategies (Attributes Knowledge, A-Softmax, Dynamic Sampling) and cluster number K of CDS.

Attribute Knowledge. Attribute Knowledge makes a better starting point for the target fine-tuning. As shown in Table 2, training with attributes improves 6% Rank-1 accuracy for using the Market-1501 as the source domain, while improving 2.84% Rank-1 accuracy for using the DukeMTMC-reID as the source domain.

A-Softmax. A-Softmax loss is used on both the source domain training and target domain training. In Table 2, we show the impact of loss function on the target domain train-

Methods	DukeMTMC-reID ->Market-1501				Market-1501 ->DukeMTMC-reID			
	mAP	Rank-1	Rank 5	Rank-10	mAP	Rank-1	Rank 5	Rank-10
LOMO [23]	8.0	27.2	41.6	49.1	4.8	12.3	21.3	26.6
UMDL [12]	12.4	34.5	52.6	60.3	7.3	18.5	31.4	37.4
PTGAN [24]	-	38.6	-	66.1	-	27.4	-	50.7
PUL [14]	20.5	45.5	60.7	66.7	16.4	30.0	43.4	48.5
CAMEL [25]	26.3	54.5	-	-	-	-	-	-
SPGAN+LMP [7]	26.7	57.7	75.8	82.4	26.2	46.4	62.3	68.0
TJ-AIDL [10]	26.5	58.2	74.8	81.1	23.0	44.3	59.6	65.0
HHL [8]	31.4	62.2	78.8	84.0	27.2	46.9	61.0	66.7
CamStyle [26]	27.4	58.8	78.2	84.3	25.1	48.4	62.5	68.9
CamStyle + LMP [27]	30.4	64.7	80.2	85.3	27.7	51.7	67.0	72.8
ARN [11]	39.4	70.3	80.4	**86.3**	33.4	60.2	73.9	**79.5**
CDS	**39.9**	**71.6**	**81.2**	84.7	**42.7**	**67.2**	**75.9**	79.4

Table 3: Comparing CDS with other state-of-the-art methods on DukeMTMC-reID and Market-1501 datasets.



(a) (b)

Fig. 4: The impacts of the number of cluster centers. The blue curve denotes Rank-1 accuracy and the yellow curve denotes mAP.

ing. A-Softmax is 3.4% higher and 1.7% higher than softmax in mAP and Rank-1, respectively, when the target dataset is DukeMTMC-reID. While with the Market-1501 as the target dataset, A-Softmax is 2.3 % higher and 2.9% higher than softmax in mAP and Rank-1 accuracies, respectively.

Cluster Number. As shown in Fig 4, we evaluate the impact of the number of clusters K. $K = 900$ achieves the best performance both on Market-1501 and DukeMTMC-reID. From the results, we can see that the fluctuations are relatively stable, which means that the proposed CDS method is not sensitive to the number of clusters in a certain range. In the latter experiments, we set $K = 900$.

Dynamic Sampling. The impact of the sampling strategy is shown in Table 3. According to the PUL [14] framework, we set four initial thresholds 0.85, 0.8, 0.7, 0.6. From the results, we can see that $\lambda = 0.8$ achieves the best performance. However, CDS outperforms $\lambda = 0.8$ sampling in mAP and Rank-1 on Market-1501 to DukeMTMC-reID.

4.4. Comparing with the State-of-the-art Methods

The experimental results are displayed in Table 3 and Table 4. Comparing with the hand-crafted features LOMO [23], three image-style transform method [24, 7, 8] (*i.e.*, PTGAN, SPGAN+LMP and HHL), two unsupervised domain transfer methods TJ-AIDL and ARN, and two clustering methods PUL and CAMEL, the proposed CDS method outperforms these approaches in Rank-1, Rank-5 and mAP on both the Market-1501, DukeMTMC-reID and CUHK03 datasets. Comparing with the second best method ARN, the proposed CDS method outperforms ARN by 7% in Rank-1, 9.3% in mAP on the DukeMTMC-reID dataset, while achieving 1.3% and 0.5% gains in Rank-1 and mAP, respectively, on DukeMTMC-reID dataset. Comparing with the fixed sampling method PUL, the proposed CDS method outperforms it by 29.5% in Rank-1 and 19.4% mAP on the Market-1501, 26.3% in Rank-1 and 37.2% in mAP on the DukeMTMC-reID dataset. Together with the ablation experiments, we conclude that the improvement of the proposed CDS for cross-domain person Re-ID mainly benefits from the attribute guided training, dynamic sampling in clusters, and the A-Softmax loss function.

	Market->CUHK03			Duke->CUHK03		
metric	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5
PUL [14]	7.3	7.6	13.8	5.2	5.6	11.2
CDS	8.7	9.1	18.0	7.1	8.1	15.8

Table 4: Comparing CDS with other state-of-the-art methods on CUHK03

5. CONCLUSION

In this paper, we have proposed a novel Clustering and Dynamic Sampling (CDS) approach to address cross-domain person re-identification task. It uses the pedestrian attributes to improve the discriminability of the baseline model, the A-Softmax loss to train our model, and the dynamic sampling strategy to select informative samples from the clustering results. We have achieved the state-of-the-art results on both the DukeMTMC-reID and Market-1501 datasets under unsupervised domain adaptation. Particularly on the Market-1501,

we exceed the current state-of-the-art method ARN by 7% in Rank-1 accuracy and 9% in mAP.

6. REFERENCES

[1] K. He, X. zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[2] X. Fan, W. Jiang, H. Luo, and M. Fei, "Spherereid: Deep hypersphere manifold embedding for person re-identification," *arXiv:1807.00537*, 2018.

[3] Y. lin, L. Zheng, Z. Zheng, Y. Wu, and Y. Yang, "Improving person re-identification by attribute and identity learning," *arXiv:1703.07220*, 2017.

[4] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *ICPR*, 2014, pp. 34–39.

[5] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *ICCV*, 2017, pp. 3754–3762.

[6] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *ICCV*, 2015.

[7] W. Deng, L. Zheng, G. Kang, Y. Yang, Q. Ye, and J. Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person reidentification," in *CVPR*, 2018, vol. 1, p. 6.

[8] Z. Zhong, L. Zheng, S. Li, and Y. Yang, "Generalizing a person retrieval model hetero-and homogeneously," in *ECCV*, 2018, pp. 172–188.

[9] G. Ian J, P. Jean, M. Mehdi, B. Xu, W. David, O. Sherjil, C. Aaron C, and B. Yoshua, "Generative adversarial nets," pp. 2672–2680, 2014.

[10] J. Wang, X. Zhu, S. Gong, and W. Li, "Transferable joint attribute-identity deep learning for unsupervised person re-identification," *arXiv:1803.09786*, 2018.

[11] Y. Li, F. Yang, Y. Liu, Y. Yeh, X. Du, and Y. Wang, "Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification," *arXiv:1804.09347*, 2018.

[12] P. Peng, T. Xiang, Y. Wang, P. Massimiliano, S. Gong, T. Huang, and Y. Tian, "Unsupervised cross-dataset transfer learning for person re-identification," in *CVPR*, 2016, pp. 1306–1315.

[13] M. Li, X. Zhu, and S. Gong, "Unsupervised person re-identification by deep learning tracklet association," *arXiv:1809.02874*, 2018.

[14] H. Fan, L. Zheng, C. Yan, and Y. Yang, "Unsupervised person re-identification: Clustering and fine-tuning," *TOMM*, vol. 14, no. 4, pp. 83, 2018.

[15] W. Liu, Y. Wen, Z. Yu, M. Li, R. Bhiksha, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *CVPR*, 2017, vol. 1, p. 1.

[16] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *CVPR*, 2014, pp. 152–159.

[17] H. Shi, Y. Yang, X. Zhu, S. Liao, Z. Lei, W. Zheng, and S. Z. Li, "Embedding deep metric for person re-identification: A study against large variations," in *ECCV*. Springer, 2016, pp. 732–748.

[18] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S.Z. Li, "Salient color names for person re-identification," in *ECCV*. Springer, 2014, pp. 536–551.

[19] X. Qian, Y. Fu, W. Wang, T. Xiang, Y. Wu, Y. Jiang, and X. Xue, "Pose-normalized image generation for person re-identification," *arXiv:1712.02225*, 2017.

[20] J. Chang, L. Wang, G. Meng, S. Xiang, and C. Pan, "Deep adaptive image clustering," .

[21] S. Lin, H. Li, Chang-Tsun Li, and Alex Chichung Kot, "Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification," *arXiv:1807.01440*, 2018.

[22] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-reciprocal encoding," in *CVPR*, 2017, pp. 1318–1327.

[23] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *CVPR*, 2015, pp. 2197–2206.

[24] L. Wei, Shiliang Zhang, Wen Gao, and Qi Tian, "Person transfer gan to bridge domain gap for person re-identification," *arXiv:1711.08565*, 2017.

[25] H. Yu, A. Wu, and W. Zheng, "Cross-view asymmetric metric learning for unsupervised person re-identification," in *ICCV*, 2017.

[26] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "Camstyle: A novel data augmentation method for person re-identification," *TIP*, 2018.

[27] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "Camstyle: A novel data augmentation method for person re-identification," *TIP*, vol. 28, no. 3, pp. 1176–1190, 2019.