Unsupervised Person Re-identification Based on Clustering and Domain-Invariant Network

Yangru Huang[1], Yi Jin[1(✉)], Peixi Peng[2], Congyan Lang[1], and Yidong Li[1]

[1] School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China
{yrhuang,yjin,cylang,ydli}@bjtu.edu.cn
[2] Institute of Automation, Chinese Academy of Sciences, Beijing, China
peixi.peng@ia.ac.cn

Abstract. Person re-identification (Re-ID) is a task which aims to determine whether a pedestrian in a camera has emerged in other cameras. Earlier works stress importance of the supervised learning methods, however, creating labels by hand is too slow and expensive. Hence, supervised methods are always limited in real-world applications. To address the problem, we propose a novel domain adaptation framework for unsupervised person Re-ID. First, target data are clustered and selected to add relative reliable supervised information for target domain. Second, a novel domain adaptive network is designed to decompose the representations to person-related and domain-related part. The former aims at learning domain-invariant and discriminative representation by a adversarial loss and a Re-ID loss with the label smoothing regularization. And the latter further improve a model's ability of extracting domain-invariant features by separating the domain unique features. What's more, during learning representation for target domain, a labeled source data not only is utilized to initialize the model but also participate in the training as a beneficial supervision information to generalize the Re-ID model. Experimental results on Market-1501 and DukeMTMC-reID evidence the superior performance of the proposed model over state-of-the-art methods.

Keywords: Person Re-ID · Domain adaptation · Clustering

1 Introduction

Person re-identification (Re-ID) is a retrieval task of identifying the same person image captured from distinctively non-overlapping camera views. It has become an increasingly popular task in video surveillance due to its application and research significance. Despite the best efforts from many computer vision researchers, it remains an unsolved problem. It is difficult that the person image often undergoes dramatically changes in appearance and background due to changes in view angle, background clutter, illumination conditions and so on.

Recently, as the deep learning based technologies are developed, the performance of person Re-ID has been improved significantly by learning robust representations with invariance property [5,9,23] or learn an effective distance metric [7,20,21]. All of those approaches focus on the supervised way which requires a plenty of large-scale, high-quality annotated data. However, collecting and annotating data for every new task are extremely expensive and time-consuming. Therefore, the supervised methods may be limited in real-world scenarios.
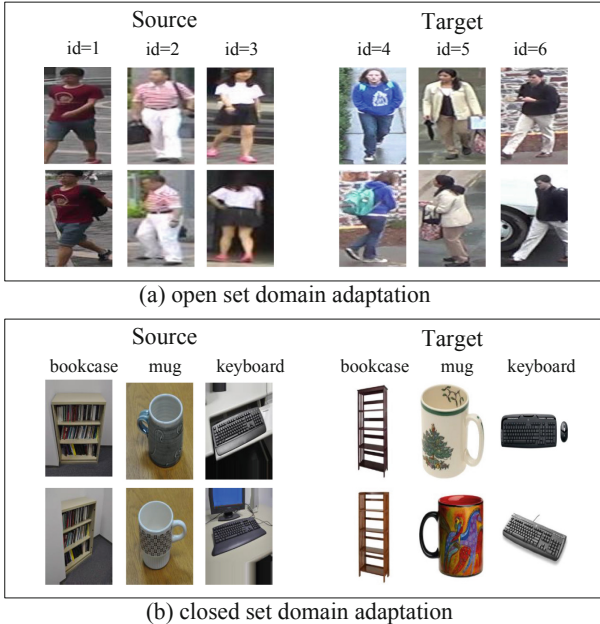

(a) open set domain adaptation

(b) closed set domain adaptation

Fig. 1. The difference of open set domain adaptation and closed set domain adaptation. For the former, target dataset contains absolutely different categories compared with the source dataset. For the latter, target dataset contains only images of the categories of the source dataset.

To address the aforementioned challenge, a common solution is the unsupervised domain adaptation [2,14] which attempts to transfer knowledge from labeled source data to unlabeled target data. However, the source dataset and the target dataset are drawn from two different distributions. Hence, if the model trained on the source dataset is directly used on the target dataset, the accuracy will decline dramatically. In addition, standard unsupervised domain adaptation method is denoted as the closed set problem which assume that the source and target domains contain the same set of classes. This assumption is not appropriate for domain adaptive person Re-ID which is an open set task. As shown in Fig. 1, open set domain adaptation includes images of unknown classes which

are not present in source domains. Therefore, most domain adaptation methods cannot be directly applied to person Re-ID task.

To make a person Re-ID model practical, existing methods solve the problem from two aspects. One way is to generate cross-domain source data which have similar style with target data [3,17]. Another way is to design a domain-shared model to learn the domain-invariant features from both source data and target data [10]. Different with previous works, in this paper, a novel method is designed to effectively transfer discriminative representation from a large number of labeled source dataset to unlabeled target dataset.

First, the model pre-trained on labeled source dataset is deployed to extract features for target unlabeled dataset. And then the clustering method is performed on the target features. However, it is obvious that the assigned pseudo labels may not be correct, so only relative reliable samples are selected to train model in a supervised way. The rest samples with unreliable labels in target data still are unlabeled but also take part in the training afterwards in an unsupervised way.

Second, to address the domain shift between datasets, a novel domain adaptive model is designed by decomposing the representations to person-related discriminative representations and domain-related representations, as shown in Fig. 2. For person-related discriminative representations, the domain-adversarial loss is performed on the source and the target data with relative reliable labels to match the feature space distributions of different datasets. Meanwhile, considering the confidence of target labels, the Re-ID loss with label smooth regularization is designed to recognize the identity of pedestrians. What's more, to reduce information loss, a decoder is designed to reconstruct feature maps from the person-related and domain-related features on source dataset and all target dataset.

To sum up, the contributions of our work are:

(1) Different from existing unsupervised Re-ID models, we propose to solve the Re-ID task by adapting the representation learning from the auxiliary labeled datasets.
(2) We are the first to integrate the cluster method with the domain-invariant model. The learning strategy facilitates the model to pay more attention to the domain-invariant and discriminative features at the same time effectively reduce the loss of information.
(3) Extensive experiments and ablation study on Market-1501 and DukeMTMC-reID demonstrate the proposed method is effective and can be applicable to unsupervised cross-dataset transfer learning problem.

2   Related Works

Supervised Person Re-ID Methods. Most of the existing works focus on the supervised method [7,11,15,26] which train a model based on a sufficient number of labeled images across cameras. Main stream works can be categorized into

two ways. One way is representation learning [11] which explores to design the discriminative features. Metric learning [7,15] is another powerful way to address the problem of person re-identification and it aims at learning the similarity of two images. However, directly deploying these trained methods to the real-world environment always leads to poor performance due to domain shift and the lack of label information.

Deep Domain Adaption. With the popularity of deep learning methods, more and more researchers use deep neural networks to enhance the performance of domain adaption. Yosinski *et al.* [18] demonstrate the generalizability of layers and find the first 3 layers of neural network learn mostly general features and higher layers learn higher levels of representations. Tzeng *et al.* [13] first propose the DDC method which fixed the first seven layers of AlexNet and add adaptive metrics to the previous layer of the classifier to solve the adaptive problem of deep network. Then Long [8] propose the DAN method to extend the DDC method. In contrast to the DDC method which only has one adaptive layer and a single kernel MMD, the DAN method add three adaptive layers at the same time and adopt a multi-kernel MMD measurement (mk-mmd) with better representation ability. Bousmalis *et al.* [2] propose a novel method, the Domain Separation Networks (DSN), for learning domain-invariant representations. However, in person Re-ID datasets, the source dataset and the target dataset have totally different identities, so traditional domain adaptation methods is not suitable to our task.

Domain Adaption in Person Re-ID. Although we have made a great progress in the field of supervised learning methods, it is inevitable to label these images manually, and the work is really expensive. Hence, it is necessary for us to further study unsupervised methods. Peng *et al.* [10] propose a multi-task dictionary learning model to transfer a learn a view-invariant representation from the labeled source dataset to the unlabeled target dataset. However, hand-craft methods always have a poor performance on large-scale dataset. Fan *et al.* [4] propose a method by clustering the unlabeled training set and using CNN fine tuning for iterative training. However, the method only use a labeled source data to initialize the model but ignore the labeled source data during the training of target domain. Recently, Generative Adversarial Networks (GAN) become more and more popular. In order to achieve cross-dataset classification tasks, there are many methods utilize GAN to transfer the style of different domain person images [3,17,26]. For example, Deng *et al.* are inspired by CycleGAN and apply it to generate images with similar target domain style. In [25], Zhong *et al.* introduce a Hetero-Homogeneous Learning (HHL) to learn camera-invariant network for target domain. However, the methods based on GAN is difficult to keep person identities during the progress of generating images.

3   Methodology

3.1   Problem Definition

In this section, we introduce some notations and definitions that are used in this paper. Assume the labeled source dataset $D_s = \{(I_i^s, y_i^s)\}_{i=1}^{N_s}$ including N_s
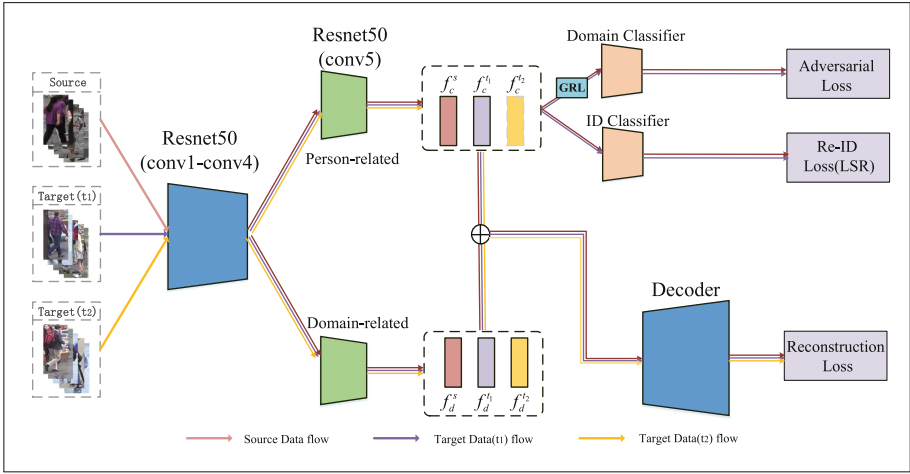
Fig. 2. Illustration of the proposed model. During training, the source data and all target data is input Resnet-50 (conv1-conv4) to extract features. And then, the model is decomposed to the person-related part and the domain-related part. For person-related part, only source and target (t_1) data is input to domain and ID classifier to learn domain-invariant discriminative features. For domain-related part, on the one hand, explicitly modeling what is unique to each domain is able to improve the ability of the model to extract domain-invariant features. On the other hand, for source, target (t_1) and target (t_2), domain-related part is combined with person-invariant part to reconstruct image for reducing information loss.

image samples and the unlabeled target dataset $D_t = \{I_i^t\}_{i=1}^{N_t}$ including N_t image samples, where I_i^s and I_i^t collected from different domains. The goal of the proposed domain-adaptive model is to make use of labeled source samples D_s to learn a model $M : I_i^t \mapsto y_i^t$ and make it equally effectual on the target samples D_s by learning the domain-invariant discriminative representations.

To minimize the discrepancy of the source dataset and target dataset effectively, a novel domain adaptation framework is designed. First, a model pretrained on source labeled data is utilized to extract features of target data. And the cluster method is adopted to generate weak labels for target samples and only those samples $D_{t_1} = \{I_i^{t_1}, y_i^{t_1}\}_{i=1}^{N_{t_1}}$ with more reliable labels are selected. It is a good way to supply the supervised information for the target dataset. The rest target data $D_{t_2} = \{I_i^{t_2}\}_{i=1}^{N_{t_2}}$ still are unlabeled. Second, source data, target data with weak labels D_{t_1}, target data without labels D_{t_2} are input the proposed domain-adaptive model (DAM). As shown in Fig. 2, Resnet-50 [6] is adopted as the backbone of the feature extraction module. As is known, in the neural network, the features extracted by the first several layers are general features. And with the deepening of network layers, the latter layers emphasise more on the specific features of learning tasks. To perform cross-dataset person Re-ID and further improve Re-ID performance, we keep the first several layers

and introduce 2 branches to learn the person-related and domain-related representations after Conv4 respectively. For the first branch, the features $(f_c^s, f_c^{t_1})$ of the source D_s and target data D_{t_1} are input to domain classifier and ID classifier to learn domain-invariant discriminative representations. Specially, the labels of target data D_{t_1} may deviate from the ground truth, thus, the cross entropy with the label smoothing regularization(LSR) is deployed. For the second branch, the domain-related features $(f_d^s, f_d^{t_1}$ and $f_d^{t_2})$ is combined with person-related features $(f_c^s, f_c^{t_1}$ and $f_c^{t_2})$ to reconstructed image for reducing the information loss.

3.2   Clustering

Firstly, the original model $\phi(\cdot; \theta)$ trained on labeled source dataset is utilised to initialize the parameters of target model. And then generating weak labels for target dataset by clustering based on such model. The idea is formulated as:

$$\min_{y, C_1, ..., C_J} \sum_{k=1}^{K} \sum_{y_i=k} \|\phi(x_i; \theta) - C_j\|^2 \tag{1}$$

where C_J is the cluster center of samples and y_i is the sample label. But not all generated weak labels are correct, in order to avoid erroneous labels making the model get stuck in a bad local optimum or oscillating, we merely select these more reliable samples which are more closer to the each ID class center C_j. To achieve this, a threshold is set, if the distance between sample and the corresponding cluster center is lower than the threshold, then x_i is selected as a reliable sample and is placed into target data D_{t_1} for supervised training; otherwise, x_i is placed into target dataset D_{t_2} for unsupervised training.

3.3   Learning

To alleviate domain shift for cross-dataset person Re-ID, a novel domain-adaptive model is designed. All data is input to Resnet-50 (conv1-conv4) to extract general features, and 2 branches (conv5) to learn person-related and domain-related parts respectively. For person-related part, the goal is to learn domain-shared and discriminative features related to person Re-ID by adversarial loss and Re-ID loss. Domain-related part is combined with person-related part to reconstruct image by reconstruction loss.

Adversarial Loss. Early methods for domain adaptation always try to find a common feature space. However, inspired by generative adversarial nets (GAN), more and more adversarial learning approaches have showed state-of-the-art performance for cross-dataset transfer learning. Hence, in the paper, adversarial learning [8] is adopted to make the features' distributions become more similar during training by confusing domain classification.

The adversarial loss trains the adversarial discriminator using a standard classification loss. During the forward propagation, the model leaves the input

unchanged, but during the backpropagation, the gradient is reversed by multiplying a negative scalar. Mathematically, the adversarial layer can be treated as a function $g(\cdot)$:

$$forward : g(x) = x$$
$$backward : g(x) = -x \tag{2}$$

The objective function of adversarial loss optimized by the stochastic gradient descent can be expressed as:

$$\mathcal{L}_{adv} = - \sum_{i=0}^{N_s + N_{t1}} d_i \log \hat{d_i} + (1 - d_i) \log(1 - \hat{d_i}) \tag{3}$$

where d_i denotes one hot encoding of domain labels, and $\hat{d_i}$ denotes the domain category prediction. Under the effect of adversarial loss, even if there are differences between the two domains, the outputs of feature extractor still are domain-invariant features.

Re-ID Loss. In order to make the model discriminative and preserve identity information, the Re-ID loss is designed to predict the output labels of source dataset and target data D_{t_1}. This part mainly focuses on learning the discriminative representations of pedestrians. Re-ID loss can be denoted as minimizing the negative log-likelihood of the ground truth class:

$$\mathcal{L}_{id} = - \sum_{i=0}^{N_s + N_{t_1}} q(i) \log p(i) \tag{4}$$

where $q(i)$ denotes the ground truth distribution and $p(i)$ is the predicted probability prediction.

However, considering the noise caused by weak labels or existing mislabeled samples in real data, we apply the label smoothing regularization (LSR) [12] to alleviate the influence of noisy samples by re-weighting the samples with weak labels. The LSR function can be written as:

$$q_{LSR}(i) = \begin{cases} \frac{\varepsilon}{N_s + N_{t_1}} & i \neq y \\ 1 - \varepsilon + \frac{\varepsilon}{N_s + N_{t_1}} & i = y \end{cases} \tag{5}$$

where $\varepsilon \in [0, 1]$ is a hyper-parameter denoting the confident of the ground truth. And the cross-entropy loss with $q_{LSR}(i)$ is expressed as:

$$\mathcal{L}_{id_{LSR}} = -(1 - \varepsilon) \log(p(y)) - \frac{\varepsilon}{N_s + N_{t_1}} \sum_{i=1}^{N_s + N_{t_1}} log(p(i)) \tag{6}$$

Reconstruction Loss. In order to reduce information loss in above procedure, a shared decoder is introduced to reconstruct the input sample by concatenating

the person-related and domain-related representations. The reconstruction loss is defined as:

$$\mathcal{L}_{rec} = \sum_{i=1}^{N_s} \left\| Decoder(f_c^s + f_d^s) \right\|_2^2 +$$
$$\sum_{i=1}^{N_{t_1}} \left\| Decoder(f_c^{t_1} + f_d^{t_1}) \right\|_2^2 + \sum_{i=1}^{N_{t_2}} \left\| Decoder(f_c^{t_2} + f_d^{t_2}) \right\|_2^2 \tag{7}$$

where $f_c^s, f_c^{t_1}, f_c^{t_2}$ denote the person-related representations of source, target D_{t_1} and target D_{t_2} respectively. And $f_d^s, f_d^{t_1}, f_d^{t_2}$ denote the domain-related representations of source, target D_{t_1} and target D_{t_2} respectively.

In short, the final integrated training objective can be written as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{id} + \alpha \mathcal{L}_{adv} + \beta \mathcal{L}_{rec} \tag{8}$$

where α and β are hyper-parameters and control the relative importance of each item. The model is trained by minimizing $\mathcal{L}_{total}$. The experiments is a iteration progress until the model is stable. In general, the number of iterations is typically < 5 in our experiment.

4   Experiment

4.1   Datasets

Table 1. Comparison with State-of-the-art Methods.

Methods	Duke → Market				Market → Duke			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
UMDL [10]	34.5	52.6	59.6	12.4	18.5	31.4	37.6	7.3
PUL [4]	45.5	60.7	66.7	20.5	30.0	43.4	48.5	16.4
CAMEL [19]	54.5	–	–	26.3	–	–	–	–
TJ-AIDL [16]	58.2	74.8	81.1	26.5	44.3	59.6	65.0	23.0
SPGAN+LMP [3]	57.7	75.8	82.4	26.7	46.4	62.3	68.0	26.2
CamStyle [26]	58.8	78.2	84.3	27.4	48.4	62.5	68.9	25.1
HHL [25]	62.2	78.8	84.0	31.4	46.9	61.0	66.7	27.2
Ours	**68.0**	**82.3**	**87.7**	**39.5**	**58.1**	**74.5**	**79.7**	**29.6**

For experiments, two widely used benchmark datasets Market-1501 [22] and DukeMTMC-reID [24] are chosen. The details are described as follows:

Market-1501 consists of 32,668 annotated bounding boxes of 1,501 identities which is collected in front of a supermarket at Tsinghua University. Images of

each identity are captured by at most six cameras and each annotated identity is present in at least two cameras. There are 19,732 images with 751 identities used for testing and 12,936 images with 750 identities used for training. Market-1501 dataset adopts Deformable Part Model (DPM) as pedestrian detector.

DukeMTMC-reID is sampled from video at 120 frames per image, resulting in 36,411 images. There are 1,404 people under more than two cameras, and 408 people under only one. It is composed of 16,522 training images of 702 identities, 2,228 query images of the other 702 identities and 17,661 gallery images.

A single-shot experiment setting is adopted. In each experiment, the source dataset is supposed to be labeled, and target dataset is supposed to be unlabeled. What's more, both rank-1 accuracy and mean Average Precision (mAP) are employed for person Re-ID evaluation.

4.2   Implementation Details

The model is implemented by using Pytorch. ResNet-50 [6] model is adopted with weights pre-trained on ImageNet as basic model. If Market is viewed as the source dataset with labels, Duke will be considered as the target datset without labels, and vice versa. During the progress of training, stochastic gradient descent with a momentum of 0.9 is adopted. And the learning rate is set to 0.001 and decay to 1×10^{-4} and 1×10^{-5} after 20 epochs and 120 epochs respectively. The maximum number of iterations is set to 200. For the label smoothing regularization(LSR), $\varepsilon = 0.1$. For the clustering step, standard kmeans clustering is adopted and k-means++ [1] is used to select initial cluster centers.

4.3   Comparison with State-of-the-Art Methods

We compare our approach with state-of-the-art methods when tested on Market and Duke, as shown in Table 1. From the results, it is evident that our model with the performance in Rank-1 accuracy = 68.0% (58.1%) and mAP = 39.5% (29.6%) is able to get better performance on Market (Duke) when compared with existing unsupervised methods including hand-craft method UMDL [10], based on weak labels method PUL [4], based on attribute-identity method TJ-AIDL [16], and based on GAN methods SPGAN [3], CamStyle [26] and HHL [25]. Specially, when the method is compared with PUL [4] which also use a labeled source data to initialize the model but ignore the labeled source data during training, our method leads to +22.5% (+38.0%) and +19.0% (+13.2%) improvement on Market (Duke) in Rank-1 and mAP, respectively. And when compared with current best results HHL [25], our model is 5.8% (11.2%) higher than HHL [25] on Market (Duke).

4.4   Ablation Studies

The Effectiveness of Adversarial Loss. As shown in Tables 2 and 3, when our method is compared to the direct transfer method, the loss gains 25.7%

Table 2. Methods comparison when tested on Market.

Methods	Duke → Market			
	Rank-1	Rank-5	Rank-10	mAP
Direct transfer	42.3	59.5	67.1	17.5
Ours w/o $\mathcal{L}_{adv}$	66.3	80.8	86.9	37.6
Ours w/o $\mathcal{L}_{id}(LSR)$	67.1	81.6	87.5	38.7
Ours w/o $\mathcal{L}_{rec}$	66.9	81.2	87.1	38.1
Ours	**68.0**	**82.3**	**87.7**	**39.5**

Table 3. Methods comparison when tested on Duke.

Methods	Market → Duke			
	Rank-1	Rank-5	Rank-10	mAP
Direct transfer	29.3	45.4	52.0	14.5
Ours w/o $\mathcal{L}_{adv}$	56.3	72.8	77.9	27.8
Ours w/o $\mathcal{L}_{id}(LSR)$	57.2	73.6	78.1	28.7
Ours w/o $\mathcal{L}_{rec}$	56.9	73.8	78.5	28.5
Ours	**58.1**	**74.5**	**79.7**	**29.6**

(28.8%) improvements in Rank-1 accuracy on Market (Duke). This indicates that adversarial loss is effective by confusion the domain classification, and the abundant unlabeled data have been well utilized by learning domain-invariant features.

The Effectiveness of Re-ID Loss with LSR. As Table 2 (Table 3) shown, when we introduce Re-ID loss with LSR, the result can improve by +0.9% (+1.8%) in Rank-1 on Market (Duke) respectively. This shows that Re-ID loss with LSR indeed helps for unsupervised Re-ID by re-weighting the samples with generated weak labels.

The Effectiveness of Reconstruction Loss. We observe that reconstruction loss is able to improve the Rank-1 by +1.1% (+1.2%) on Market (Duke). The experiment demonstrates that reconstruction loss is an effective way to reduce the loss of information.

4.5   The Impact of the Cluster Number

We evaluate the impact of the cluster number which is set to 300, 500, 700, 900, 1100. As shown in Fig. 3, our method is robust for the number of cluster. Even if the number of cluster is different from the actual situation, our method can still exceed many methods. Specially, when the cluster is close to the actual number of categories, our method have the best effect.
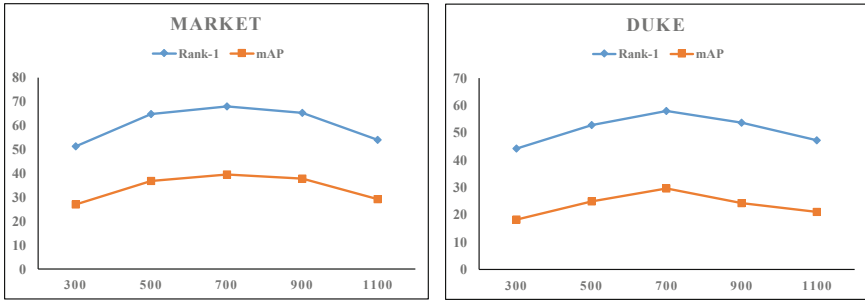
Fig. 3. The impact of cluster number on Market and Duke.

5   Conclusion

This paper proposes a novel unsupervised cross-dataset transfer learning method for Re-ID task. The proposed model aims to learn discriminative representation by leveraging the labeled dataset. To achieve that, we design a special domain adaptation framework for person Re-ID. In contrast to most existing approaches, our method combine cluster method and domain-invariant model to train the target model in an iterative way. We also show the importance of making full use of the relative reliable weak labels information on the target dataset. Extensive experiments demonstrate the effectiveness and robustness of the proposed model.

References

1. Arthur, D., Vassilvitskii, S.: k-means++: the advantages of careful seeding. In: ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035. Society for Industrial and Applied Mathematics (2007)
2. Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., Erhan, D.: Domain separation networks. In: NIPS, pp. 343–351 (2016)
3. Deng, W., Zheng, L., Kang, G., Yang, Y., Ye, Q., Jiao, J.: Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person reidentification. In: CVPR, vol. 1, p. 6 (2018)
4. Fan, H., Zheng, L., Yan, C., Yang, Y.: Unsupervised person re-identification: clustering and fine-tuning. ACM Trans. Multimedia Comput. Commun. Appl. **14**(4), 83 (2018)
5. Geng, M., Wang, Y., Xiang, T., Tian, Y.: Deep transfer learning for person reidentification. arXiv preprint arXiv:1611.05244 (2016)

6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
7. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017)
8. Long, M., Cao, Y., Wang, J., Jordan, M.I.: Learning transferable features with deep adaptation networks. arXiv preprint arXiv:1502.02791 (2015)
9. Matsukawa, T., Okabe, T., Suzuki, E., Sato, Y.: Hierarchical Gaussian descriptor for person re-identification. In: CVPR, pp. 1363–1372 (2016)
10. Peng, P., et al.: Unsupervised cross-dataset transfer learning for person re-identification. In: CVPR, pp. 1306–1315 (2016)
11. Sun, Y., Zheng, L., Yang, Y., Tian, Q., Wang, S.: Beyond part models: person retrieval with refined part pooling. arXiv preprint arXiv:1711.09349 (2017)
12. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: CVPR, pp. 2818–2826 (2016)
13. Tzeng, E., Hoffman, J., Darrell, T., Saenko, K.: Simultaneous deep transfer across domains and tasks. In: ICCV, pp. 4068–4076 (2015)
14. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: CVPR, vol. 1, p. 4 (2017)
15. Varior, R.R., Haloi, M., Wang, G.: Gated siamese convolutional neural network architecture for human re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 791–808. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_48
16. Wang, J., Zhu, X., Gong, S., Li, W.: Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: CVPR (2018)
17. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer GAN to bridge domain gap for person re-identification. In: CVPR (2018)
18. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: NIPS, pp. 3320–3328 (2014)
19. Yu, H.-X., Wu, A., Zheng, W.-S.: Cross-view asymmetric metric learning for unsupervised person re-identification. In: ICCV (2017)
20. Zhang, L., Xiang, T., Gong, S.: Learning a discriminative null space for person re-identification. In: CVPR, pp. 1239–1248 (2016)
21. Zhang, Y., Li, B., Lu, H., Irie, A., Ruan, X.: Sample-specific SVM learning for person re-identification. In: CVPR, pp. 1278–1287 (2016)
22. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: ICCV, pp. 1116–1124 (2015)
23. Zheng, L., Yang, Y., Hauptmann, A.G.: Person re-identification: past, present and future. arXiv preprint arXiv:1610.02984 (2016)
24. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In: ICCV, vol. 3 (2017)
25. Zhong, Z., Zheng, L., Li, S., Yang, Y.: Generalizing a person retrieval model hetero- and homogeneously. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11217, pp. 176–192. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01261-8_11
26. Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: Camera style adaptation for person re-identification. In: CVPR, pp. 5157–5166 (2018)