Asymmetric Mutual Learning for Unsupervised Cross-Domain Person Re-identification

Danyang Huang, Lei Zhang, Qishuai Diao, Wei Wu, and Zhong Zhou(✉)

State Key Laboratory of Virtual Reality Technology and Systems,
Beihang University, Beijing, People's Republic of China
zz@buaa.edu.cn

Abstract. Unsupervised domain adaptation in person re-identification is a challenging task. The performance of models trained on a specific domain generally degrades significantly on other domains due to the domain gaps. State-of-the-art clustering-based cross-domain methods inevitably introduce noisy labels. The negative effects of noisy labels gradually accumulate during iterative training. Besides, optimizing with conventional triplet loss could make the model stuck in local optima in the late stage of domain adaptation. To mitigate the effects of noisy labels, this paper proposes an asymmetric mutual learning framework which cooperates two models with asymmetric labels. The learned asymmetric information is helpful for the two models to complement with each other. Specifically, we propose a merging clusters algorithm to generate asymmetric labels. We also introduce a similarity weighted loss which can further adapt the model to target domain. Extensive experiments demonstrate that our approach outperforms the state-of-the-art methods on three popular person re-identification datasets.

Keywords: Person re-identification · Asymmetric mutual learning · Unsupervised · Cross-domain

1 Introduction

Person re-identification (re-id) aims to find the matched person in a candidate gallery given a query person image. Although existing supervised deep learning methods of person re-id have made great achievements, most of them require accurate labels which are time-consuming to annotate. Besides, these models perform poorly when the training dataset and the test dataset distribute in different domains. Unsupervised Domain Adaptation (UDA) approaches are proposed to alleviate above issues. UDA aims to transfer the knowledge learned on a source dataset with accurate identity labels to a target dataset without annotated labels. State-of-the-art UDA methods [3,15] alternatively generate pseudo labels on target domain with clustering algorithm and fine-tune the model with pseudo labels. Nevertheless, noisy labels are introduced into the iterative training since

D. N. Pham et al. (Eds.): PRICAI 2021, LNAI 13033, pp. 124–137, 2021.
https://doi.org/10.1007/978-3-030-89370-5_10

clustering algorithm can not classify images accurately. The noise will accumulate continuously and then hinder the improvement of the model. To address above issue, some recent works [4,5] adopt mutual learning framework to mitigate the negative effects of noise. Mutual learning framework can make remarkable improvement in cross-domain person re-id.

Mutual learning generally utilizes two collaborative models to solve a task together [4,5,14,17]. The two collaborative models usually start from different initial conditions. Diverse knowledge learned by two models can be combined in various ways to improve the discriminative capability of the whole network. For example, [17] utilizes KL divergence based loss to match the probability estimate of two peer networks. [5] makes the two models select the reliable samples from each other. Both of them use identical labels for two models, which restricts the diversity of information learned by the whole network and thus hinders the models from further adapting to the target domain. To address this issue, we propose an asymmetric mutual learning framework (AML) which uses asymmetric pseudo labels for two collaborative models. As shown in Fig. 1, one model uses original labels generated by clustering algorithm, the other uses the new labels augmented by our proposed algorithm of processing the original labels. When generating pseudo labels with clustering algorithm, images of the same person could be divided into different classes, these images will be separated further during iterative training. In light of this, we generate augmented pseudo labels by merging clusters based on k-nearest neighbors relationship. The augmented pseudo labels can make the model learn more generative information compared to original labels, while the model trained with original labels learns relatively discriminative information. Both augmented labels and original labels can be regarded as information complement to each other.

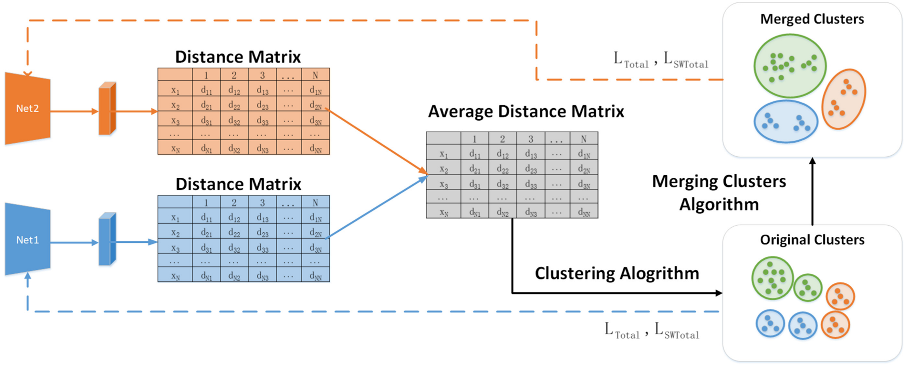


Fig. 1. The proposed asymmetric mutual learning framework (AML). L_{total} refers to the normal loss. $L_{SWtotal}$ refers to the similarity weighted loss. The distance matrix of two branches refers to the distance between features of all training images which is computed with re-ranking in [21]. The average distance matrix is the average of two distance matrix from two branches. Clustering algorithm takes the average distance as the input and generates original clustering results. Our proposed merging clusters algorithm merges part of the original clusters to get new labels.

Triplet loss is commonly used in person re-identification. It focuses on the difference between positive pairs and negative pairs. In the fully-supervised scenario, since the identity labels are accurate, the expansion of the gap between the distributions of positive pairs and negative pairs can enhance the discrimination ability of the model. However, the pseudo labels are inaccurate in unsupervised cross-domain scenario. The large gap between the inaccurate positive pairs and negative pairs makes the model stuck in local optima and hinders the model from further improving in the target domain. To address above issue, we utilize the triplets which become invalid due to the large gap between positive pairs and negative pairs. In this way, we propose a similarity weighted loss which can further bring dissimilar positive pairs closer despite the large gap mentioned above. We argue that similarity weighted loss allows the model to escape local optima and continue adapting to target domain in late training stage. The main contributions of our work are summarized as follows:

- We propose an asymmetric mutual learning framework (AML). AML utilizes asymmetric pseudo labels to optimize models on the target domain, which makes the whole network capable to learn more diverse information.
- We propose a similarity weighted loss which can further adapt the model to the target domain in late training stage. It mines dissimilar positive samples despite the difference between the distributions of positive pairs and negative pairs.
- To evaluate our method, we conduct experiments on three large-scale datasets. Experimental results show that our method outperforms state-of-the-art methods for unsupervised cross-domain person re-identification.

2 Related Work

Unsupervised Domain Adaptation. Existing UDA methods can be generally classified into three categories. The first category of UDA methods aims to improve the generalization ability of the model without training on target domain [6,10]. EANet [6] introduces pose segmentation as auxiliary information to enhance the generalization ability of the model. DIMN [10] improves the generalization ability by mapping an image directly into an identity classifier. The second category aims to reduce the domain gap between source domain and target domain with GAN [1,8]. Deng *et al.* [1] propose a similarity preserving generative adversarial network to transfer the image styles of source domain to target domain. Liu *et al.* [8] propose a framework consisting of an ensemble GAN and multiple factor GANs to do style transfer at image level and factor level. In the third category, clustering algorithms are adopted to generate pseudo labels on the target domain, and then pseudo labels are used to fine-tune the re-identification models. SSG [3] obtains multiple pseudo labels by clustering global and local features of persons respectively. Zhai *et al.* [15] present an augmented discriminative clustering method to enforce the discrimination ability of models in the target domain. Zhang *et al.* [16] propose a two-stage framework which consists of conservative stage and promoting stage, the conservative stage aims to capture the local structure of target-domain data, while the promoting stage aims to utilize of global information about the data distribution. The results of the first and second kinds of methods are generally poor compared to the third category. However, clustering-based algorithms are troubled by noisy labels and the results are still unsatisfactory compared to supervised approaches.

Supervised Mutual Learning. Mutual learning generally refers to the idea that two or more models learn from each other and stimulate each other. DML [17] utilizes a pool of networks to solve the task collaboratively rather than single network. Co-Teaching [5] makes two models select reliable samples for each other. Both of them were originally designed for supervised tasks. Different from them, we mainly focus on the unsupervised cross-domain task.

Unsupervised Mutual Learning. MMT [4] introduces mutual learning into cross-domain person re-identification and proposes an alternative training manner that combines hard pseudo labels and soft refined labels. Zhao *et al.* [18] propose a noise resistible mutual learning method which performs collaborative clustering and mutual instance selection during training. Most of the existing mutual learning works use symmetric structure, which makes the models learn similar information. Yang *et al.* [14] propose an asymmetric co-teaching framework (ACT) to make the models see hard examples.

We mainly focus on unsupervised mutual learning in this paper. Similar but different from above works, our proposed AML aims to combine generative information and discriminative information. Our work differs from ACT in the following two aspects: (1) Our work does mutual learning without complicated sample selection process, the two models interact in a simpler way. (2) While ACT mainly focuses on effective usage of unreliable outliers, our work makes two models learn more diverse information by utilizing reliable inliers effectively.

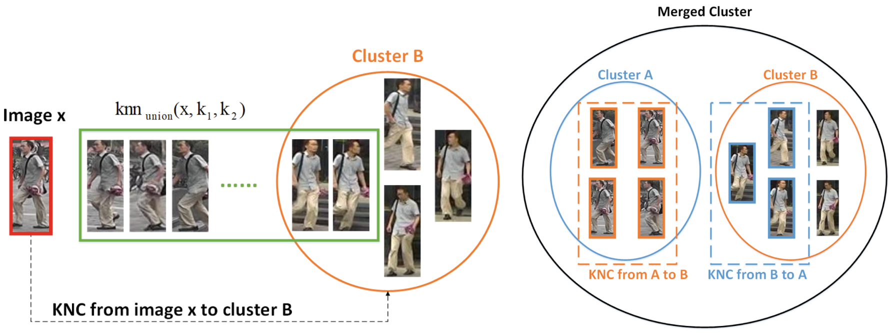
3 Proposed Method

3.1 Structure of Asymmetric Mutual Learning

Our proposed asymmetric mutual framework (AML) consists of two stages: (1) Supervised training in the source domain. (2) Unsupervised clustering-based adaptation to the target domain. In the supervised stage, we train two models with same architecture on the source dataset. In the unsupervised adaptation stage, we adapt the trained models to target domain with asymmetric pseudo labels as shown in Fig. 1. To generate asymmetric labels, we propose a merging clusters algorithm which will be discussed in Sect. 3.2. We train two models with normal triplet loss and cross-entropy loss at first, and then utilize similarity weighted loss in Sect. 3.3 to further adapt two models to target domain.

3.2 Merging Clusters Algorithm

Existing clustering algorithms generally need to set the number of clusters except those based on density. Density-based clustering algorithms can generate



(a) KNC from image to cluster (b) KNC from cluster to cluster

Fig. 2. Our proposed merging clusters algorithm. (a) We consider a image x has a KNC connection to cluster B if the union set of k_1 normal nearest neighbors and k_2 cross-camera nearest neighbors of x intersects with cluster B. (b) For two clusters A and B, we compute KNC connection between two classes according to Eq. 2 and merge then if both condition 3 and condition 4 satisfy.

the number of clusters by themselves. Since the number of clusters is usually unknown in UDA tasks, we adopt a density-based clustering algorithm [2] to cluster images. Density-based clustering algorithms generally consider points from the same continuous high-density region as a cluster. However, in cross-camera person re-identification scenario, the image distribution of the same person may be sparse due to the difference of pose and camera view. Thus the images belonging to the same person could be divided into different clusters. In contrast, k-nearest neighbors are less affected by the density, sparse points can also have k-nearest neighbors relationship. Accordingly, we propose a method to merge clusters by calculating k-nearest connection (KNC) between two clusters.

Given a data point x_a in cluster C_a, we look for two kinds of k-nearest neighbors of it. One kind is normal k-nearest neighbors $knn_{normal}(x_a, k_1)$ obtained by sorting distance matrix computed with [21]. The other kind is cross-camera k-nearest neighbors $knn_{crosscam}(x_a, k_2)$ which contains the nearest k_2 neighbors selected from samples of different cameras from x_a. Note that $knn_{crosscam}(x_a, k_2)$ is utilized to bridge the gap between images across cameras since the camera ID is easy to obtain in real scenes and has effective supervised information. As shown in Fig. 2(a), we consider that x_a is connected to cluster C_b if the union set of k_1 normal nearest neighbors and k_2 cross-camera nearest neighbors contains at least one sample in cluster C_b, i.e.,

$$KNC_{x_a->C_b} = \begin{cases} 1 & \text{if } |knn_{union}(x_a, k_1, k_2) \cap C_b| > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where $knn_{union}(x_a, k_1, k_2)$ denotes the union set mentioned above. Hence, as shown in Fig. 2(b), we define the asymmetric k-nearest connection (KNC) from cluster C_a to cluster C_b as:

$$KNC_{C_a->C_b} = \sum_{x_a \in C_a} KNC_{x_a->C_b}, \tag{2}$$

which represents the number of samples that have k-nearest connection (KNC) to cluster C_b in cluster C_a. Finally, we merge C_a and C_b if

$$\frac{KNC_{C_a->C_b}}{|C_a|} > thresh \tag{3}$$

and

$$\frac{KNC_{C_b->C_a}}{|C_b|} > thresh, \tag{4}$$

where *thresh* is a threshold that controls the proportion of $KNC_{C_a->C_b}$ to the number of samples in cluster C_a.

Our merging clustering algorithm tends to merge small clusters which usually do not contain all the images belonging to the same person. Although our algorithm merges some images belonging to different persons during the merging process, it should be noted that our purpose is not to improve the clustering accuracy. The key point is that the merged clusters contain relatively generative information compared to original clusters. Training with merged clusters can prevent the model from further separating some images belonging to the same person. Thus the two models can complement with each other, which is effective in mutual learning.

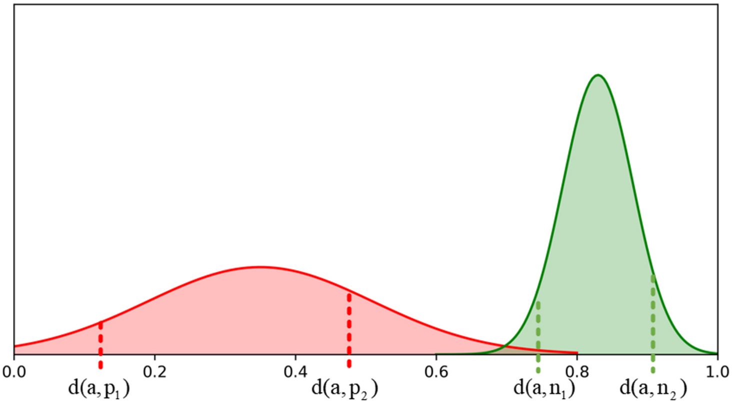


Fig. 3. A illustration of the motivation of similarity weighted loss. $d(a, p_1)$ and $d(a, p_2)$ denote the distance between anchor a and its positive samples, while $d(a, n_1)$ and $d(a, n_2)$ denote the distance between anchor a and its negative samples. When there is a large gap between the distributions of the distance of negative pairs and the distance of positive pairs, the triplet loss can not further narrow the distance between positive pairs.

3.3 Similarity Weighted Loss

Triplet loss and cross-entropy loss are two widely-used losses in person re-identification. The purpose of triplet loss is to bring positive pairs closer and push away the negative pairs. Typically, the triplet loss is defined as:

$$L_{Tri} = [d_p - d_n + \alpha]_+, \tag{5}$$

where d_p represents the distance between the anchor x^a and its positive samples x_i^p, d_n represents the distance between the anchor x^a and its negative samples x_i^n, α is the margin between d_p and d_n, $[x]_+$ means $max(x, 0)$. The triplet loss will expand the gap between d_p and d_n. When using triplet loss to fine-tune re-id model, the triplet loss tends to be zero at the end of training because d_n is much larger than d_p. However, it doesn't mean that d_p is nearly zero. As shown in Fig. 3, when d_n is too large, $[d_p - d_n + \alpha]_+$ can still be zero while d_p is a large value as long as $d_p \leq d_n - \alpha$. When L_{tri} is zero, the gradient of L_{tri} is zero, which makes the triplet invalid and the effect of d_p ignored. To address this issue, we adapt triplet loss to focus more on dissimilar positive pairs, which we call similarity weighted triplet loss. Our similarity weighted triplet loss is computed as:

$$L_{SWTri} = [d_p - s_p d_n + \alpha]_+, \tag{6}$$

where α is the margin between d_p and $s_p d_n$, s_p is the average cosine similarity of the anchor and its positive samples in a mini-batch, i.e., for an anchor x^a, its s_p is computed as:

$$s_p = \frac{1}{K} \sum_{i=1}^{K} cos(f(x^a), f(x_i^p)), \tag{7}$$

where K is the number of positive samples of anchor x^a in a mini-batch, $f(x^a)$ is the feature of anchor anchor x^a, x_i^p denotes the ith positive sample of x^a. For dissimilar positive samples, their s_p are smaller compared to similar positive samples. According to Eq. 6, dissimilar positive samples have lower weight of d_n, which means that L_{SWTri} is less likely to be zero while the positive pairs are not similar. Thus we argue that our adapted triplet loss can avoid the problem that the distance between the dissimilar positive samples can not be further narrowed in the late training period. To cooperate with similarity weighted triplet loss, we also design a similarity weighted cross-entropy loss:

$$L_{SWID} = \frac{1}{max(\beta, s_p)} L_{ID}, \tag{8}$$

where L_{ID} is the cross-entropy loss with label smoothing in [9], β is a factor controlling the range of similarity weight. L_{SWID} gives more weight to those samples which have low average cosine similarity with positive samples in a mini-batch compared to L_{ID}. Since L_{SWTri} could be larger for those dissimilar positive samples, L_{SWID} ensures that the proportion of triplet loss and cross-entropy loss will not change greatly. In summary, the normal total loss function is:

$$L_{total} = L_{Tri} + \lambda L_{ID}, \tag{9}$$

while the total similarity weighted loss is:

$$L_{SWtotal} = L_{SWTri} + \lambda L_{SWID}, \tag{10}$$

where λ is the balanced weight of cross-entropy loss.

4 Experiments

Market-1501 [19], DukeMTMC-reID [20] and MSMT17 [13] are three large-scale person re-identification datasets. We evaluate our method on four domain adaptation tasks: Duke-to-Market, Market-to-Duke, Market-to-MSMT17, Duke-to-MSMT17. We take Rank-1 accuracy and mean average precision (mAP) as evaluation metrics. As shown in Table 1, experimental results show that our method outperforms most of existing methods.

4.1 Datasets

Market-1501 [19]. The training set of Market-1501 contains 12936 annotated images of 751 person identities shot from 6 cameras in total. The testing set contains 3368 query images of 750 identities and 15913 gallery images of 751 identities.

DukeMTMC-reID [20]. The training set of DukeMTMC-reID contains 12936 annotated images of 751 person identities shot from 6 cameras in total. The testing set contains 3368 query images of 750 identities and 15913 gallery images of 751 identities.

MSMT17 [13]. As the largest and most challenging person re-ID dataset, MSMT17 contains 32621 images of 1041 person identities for training and 93820 images of 3060 identities for testing. In the testing set, 11659 images of 3060 identities are used for query and the gallery contains 82161 images of 3060 identities.

4.2 Implementation Details

Stage 1: Supervised Training in Source Domain. Previous works [3,6] have proved that focusing on local features can improve the cross-domain capabilities of the model. In view of this, we adopt PCB [12] to extract global features and local features of images and a semantic segmentation network to extract the masks of the upper and lower parts of the body. Hence, we apply the upper-part mask and lower-part mask to the global feature to get upper-part feature and lower-part feature which are used as local features. Then the global feature is used to calculate the triplet loss and all features are used to calculate the cross-entropy loss. We take ResNet-50 as backbone of PCB [12] and adopt SCHP [7] as our semantic segmentation network. SCHP is initialized with the weights trained on LIP dataset and does not update parameters during training. We adopt the Adam optimizer to optimize two re-id models separately. The learning rate is initially set to 3×10^{-4}, and decreased by 0.1 at the 35th epoch, 55th epoch and 70th epoch respectively. In addition, we use same warmup strategy following [9]. In the end of this stage, we get two feature extraction models with different weights.

Stage 2: Unsupervised Clustering-Based Adaptation to Target Domain. Given two models with different weights, we use them to extract features of person images. As mentioned in Sect. 3.2, we adopt DBSCAN [2] to cluster extracted global features, setting density radius $eps = 1.6 \times 10^{-3}$ and minimum size of a cluster to 4. The distance matrix between features is calculated separately using re-ranking in [21] and the average of them is given to DBSCAN [2]. With pseudo labels γ_{origin} generated by DBSCAN [2], we use the method in Sect. 3.2 to get the new pseudo labels γ_{new} with $thresh = 0.5$, $k_1 = 3$ and $k_2 = 15$. Then one of the two models is fine-tuned on target domain with γ_{origin} and the other with γ_{new}. Different from stage 1, the learning rate is initially set to 3×10^{-5} and decreased by 0.1 at the 10th epoch, and the warmup strategy is not used at this stage. Note that our proposed similarity weighted loss is not utilized until the training with Eq. 9 converges, since the proposed loss is to solve the problem that it is difficult to optimize the models in the late training period. In practice, we set β to 0.7 and λ to 0.01 when the model is transferred between Market1501 [19] and DukeMTMC-reID [20]. When the model is transferred to MSMT17 [13], we change β to 0.9 to get best result.

4.3 Comparison with State-of-the-Art Methods

In this section, we compare our proposed method with state-of-the-art unsupervised cross-domain methods for person re-identification including: (1) EANet [6] that uses auxiliary information (2) SPGAN [1], ATNet [8] and ECN [22] that use GANs (3) SSG [3], UDAP [11], PCB-R-PAST [16], ACT [14], AD-Cluster [15], MMT [4], NRMT[18] that use pseudo labels. Among above methods, ACT, MMT and NRMT adopt mutual learning for unsupervised cross-domain person re-identification, which is highly relevant to our work. Specifically, our proposed method combines asymmetric mutual learning with similarity weighted loss to improve performance of cross-domain person re-id.

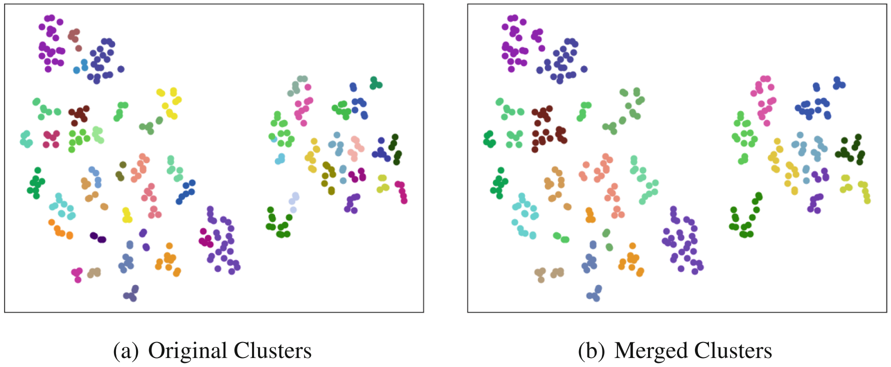
Table 1. Comparisons with state-of-the-art unsupervised cross-domain person re-id methods on Duke-to-Market, Market-to-Duke, Market-to-MSMT17, Duke-to-MSMT17.

Methods	Duke → Market		Market → Duke	
	mAP	Rank-1	mAP	Rank-1
SPGAN [1]	22.8	51.5	22.3	41.1
ATNet [8]	25.6	55.7	24.9	45.1
EANet [6]	35.8	66.1	36.0	56.1
ECN [22]	43.0	75.1	40.4	63.3
UDAP [11]	53.7	75.8	49.0	68.4
SSG^{++} [3]	68.7	86.2	60.3	76.0
PCB-R-PAST[16]	54.6	78.4	54.3	72.4
ACT [14]	60.6	80.5	54.5	72.4
Co-Teaching [5]	65.1	82.5	55.7	71.9
AD-Cluster [15]	68.3	86.7	54.1	72.6
MMT-500 [4]	71.2	87.7	63.1	76.8
NRMT [18]	71.7	87.8	62.2	77.8
Ours	**75.5**	**88.7**	**64.5**	**78.6**
Methods	Market → MSMT17		Duke → MSMT17	
	mAP	Rank-1	mAP	Rank-1
ECN [22]	8.5	25.3	10.2	30.2
SSG^{++} [3]	16.6	37.6	18.3	41.6
MMT-500 [4]	16.6	37.5	19.9	41.3
Ours	**19.4**	**46.8**	**22.2**	**51.5**

As shown in Table 1, our method outperforms all compared methods. For Duke → Market, our method outperforms state-of-the-art NRMT [18] by 3.8% in mAP and 0.9% in rank-1 accuracy. For Market → Duke, our method outperforms NRMT [18] by 2.3% in mAP and 0.8% in rank-1 accuracy. For Market → MSMT17, our method outperforms MMT-500 [4] by 2.8% in mAP and 9.3% in rank-1 accuracy. For Duke → MSMT17, our method outperforms MMT-500 [4] by 2.3% in mAP and 10.2% in rank-1 accuracy.

4.4 Ablation Study

In order to prove the effectiveness of our method, we create a baseline that optimize two models with original labels and normal loss function. As shown in Table 2, we perform ablation studies based on this baseline.



(a) Original Clusters

(b) Merged Clusters

Fig. 4. Visualization of some original clusters generated by DBSCAN and the corresponding merged clusters generated by our proposed algorithm.

Effectiveness of Asymmetric Labels. For better clarity, we visualize some original clusters and the corresponding merged clusters in Fig. 4. As shown in Fig. 4, small clusters are merged with their adjacent clusters. To show the effectiveness of new augmented labels generated by our proposed merging clusters algorithm, we train the two models with augmented labels, the result is denoted as “Baseline+Merged Clusters” in Table 2. As shown in the table, we improve the performance on Duke-to-Market by 9.7% in mAP and 6.6% in rank-1 accuracy with augmented labels. When testing on Market-to-Duke, “Baseline+Merged Clusters” surpass “Baseline” by 4.7% in mAP and 3.0% in rank-1 accuracy. To investigate the necessity of using asymmetric labels generated by our proposed merging clusters algorithm, we create mutual learning baseline models that only use original pseudo labels generated by DBSCAN [2]. As shown in Table 2, with asymmetric labels, we improve the performance by 11.3% in mAP and 6.8% in rank-1 accuracy compared to baseline on Duke-to-Market. Similarly, when the model is transferred from Market-1501 to DukeMTMC-reID, the performance gain becomes 5.8% in mAP and 3.1% in rank-1 accuracy. Besides, “AML” beats “Baseline+Merged Clusters” by 1.6% and 1.1% in mAP when testing on Duke-to-Market and Market-to-Duke respectively, which shows asymmetric labels performs better than symmetric augmented labels.

Effectiveness of Similarity Weighted Loss. To show the performance of similarity weighted loss, we train the baseline with similarity weighted loss after the training with normal loss converges, the result is denoted as “Baseline*” in Table 2. When testing on Duke-to-Market, “Baseline*” surpass “Baseline” by 1.6% in mAP and 0.9% in rank-1 accuracy. When testing on Market-to-Duke, “Baseline*” surpass “Baseline” by 2.3% in mAP and 1.3% in rank-1 accuracy. To prove the similarity weighted loss can work on AML, we also train the model with asymmetric labels by optimizing Eq. 10 after the training with Eq. 9 converges. As shown in Table 2, the combination of similarity weighted triplet loss and similarity weighted cross-entropy loss surpasses the combination of normal triplet

Table 2. Ablation studies of our proposed methods on Duke-to-Market and Market-to-Duke. “Direct Transfer” refers to directly applying the model trained on source domain to the target domain, “Baseline” refers to symmetric mutual learning with original labels and normal loss function L_{total}, “Baseline*” refers to symmetric mutual learning with similarity weighted loss function $L_{SWtotal}$, “Baseline+Merged Clusters” refers to symmetric mutual learning with augmented labels and L_{total}, “AML” denotes our proposed asymmetric mutual learning framework in Sect. 3 optimized with L_{total}, “AML*” stands for proposed AML enhanced by similarity weighted loss $L_{SWtotal}$.

Methods	Duke → Market		Market → Duke	
	mAP	rank-1	mAP	rank-1
Direct transfer	25.4	55.6	24.6	42.9
Baseline	62.5	81.5	56.8	74.1
Baseline*	64.1	82.4	59.1	75.4
Baseline+Merged Clusters	72.2	88.1	61.5	77.1
AML	73.8	88.3	62.6	77.2
AML*	**75.5**	**88.7**	**64.5**	**78.6**

loss and cross-entropy loss by 1.7% in mAP and 0.4% in rank-1 accuracy on Duke-to-Market. The performance testing on Market-to-Duke also boosts by 1.9% in mAP and 1.4% in rank-1 accuracy.

5 Conclusion

In this paper, we propose a novel asymmetric mutual learning framework for unsupervised cross-domain person re-identification. Our framework consists of two models which utilize asymmetric labels. We propose a merging clusters algorithm to generate new pseudo labels which contain different information from original pseudo labels. Furthermore, a similarity weighted loss is proposed to mine dissimilar positive samples so that the two models can continue adapting to target domain in late training stage. Comprehensive experimental results demonstrate that the performance of our approach outperforms the most of existing methods on three large-scale datasets. In the future, we will explore how to integrate camera information into the network more reasonably.

Acknowledgment. This work is supported by National Key R&D Program of China (Grant No. 2018YF B2100603) and National Natural Science Foundation of China (Grant No. 61872024).

References

1. Deng, W., Zheng, L., Ye, Q., Kang, G., Yang, Y., Jiao, J.: Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 994–1003 (2018)

2. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Knowledge Discovery and Data Mining, pp. 226–231 (1996)
3. Fu, Y., Wei, Y., Wang, G., Zhou, Y., Shi, H., Huang, T.S.: Self-similarity grouping: a simple unsupervised cross domain adaptation approach for person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 6112–6121 (2019)
4. Ge, Y., Chen, D., Li, H.: Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification. In: International Conference on Learning Representations (2020)
5. Han, B., et al.: Co-teaching: Robust training of deep neural networks with extremely noisy labels. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, pp. 8536–8546 (2018)
6. Huang, H., et al.: Eanet: enhancing alignment for cross-domain person re-identification. arXiv preprint arXiv:1812.11369 (2018)
7. Li, P., Xu, Y., Wei, Y., Yang, Y.: Self-correction for human parsing. arXiv preprint arXiv:1910.09777 (2019)
8. Liu, J., Zha, Z.J., Chen, D., Hong, R., Wang, M.: Adaptive transfer network for cross-domain person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7202–7211 (2019)
9. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1487–1495 (2019)
10. Song, J., Yang, Y., Song, Y.Z., Xiang, T., Hospedales, T.M.: Generalizable person re-identification by domain-invariant mapping network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 2019
11. Song, L., et al.: Unsupervised domain adaptive re-identification: Theory and practice. Pattern Recognition **102**, 107173 (2020)
12. Sun, Y., Zheng, L., Yang, Y., Tian, Q., Wang, S.: Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline). In: Proceedings of the European Conference on Computer Vision, pp. 480–496 (2018)
13. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer GAN to bridge domain gap for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern recognition, pp. 79–88 (2018)
14. Yang, F., et al.: Asymmetric co-teaching for unsupervised cross-domain person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 12597–12604 (2020)
15. Zhai, Y., et al.: Ad-cluster: augmented discriminative clustering for domain adaptive person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9021–9030 (2020)
16. Zhang, X., Cao, J., Shen, C., You, M.: Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 8222–8231 (2019)
17. Zhang, Y., Xiang, T., Hospedales, T.M., Lu, H.: Deep mutual learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4320–4328 (2018)
18. Zhao, F., Liao, S., Xie, G.-S., Zhao, J., Zhang, K., Shao, L.: Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12356, pp. 526–544. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58621-8_31

19. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1116–1124 (2015)
20. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3754–3762 (2017)
21. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1318–1327 (2017)
22. Zhong, Z., Zheng, L., Luo, Z., Li, S., Yang, Y.: Invariance matters: exemplar memory for domain adaptive person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern recognition, pp. 598–607 (2019)