2023 IEEE International Conference on Multimedia and Expo (ICME) | 978-1-6654-6891-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICME55011.2023.00267

# INTENSIFYING THE CONSISTENCY OF PSEUDO LABEL REFINEMENT FOR UNSUPERVISED DOMAIN ADAPTATION PERSON RE-IDENTIFICATION

1st Linfan Zha
*School of Compute Science and Technology*
*Anhui University*
Hefei, China
e21201072@stu.ahu.edu.cn

2nd* Yanming Chen
*School of Compute Science and Technology*
*Anhui University*
Hefei, China
cym@ahu.edu.cn

3rd Peng Zhou
*School of Compute Science and Technology*
*Anhui University*
Hefei, China
zhoupeng@ahu.edu.cn

4th Yiwen Zhang
*School of Compute Science and Technology*
*Anhui University*
Hefei, China
zhangyiwen@ahu.edu.cn

***Abstract*—Clustering-based unsupervised domain adaptation (UDA) for person re-identification aims to learn in the unlabeled target domain. However, the noise problem of clustering-based generated pseudo labels remains under-explored, and these wrong labels can mislead the feature learning process. In this paper, we propose a consistent and intensive pseudo label refinement method in which pseudo labels generated in two different feature spaces, local and global, refine each other to improve the pseudo label quality of the final clusters. Then we utilize a quantitative criterion to measure label inaccuracy and fine-tune the target domain to reduce the noise by an inaccuracy-guided pseudo label optimization scheme. On a strong benchmark, we demonstrate the superiority of the method with extensive experiments. Specifically, our method outperforms the baseline by 7.2% mAP on the Duke2Market task and 0.5% mAP on the Market2MSMT task over the state-of-the-art.**

***Index Terms*—person re-identification, domain adaptation, mutual refinement, label accuracy optimization**

## I. Introduction

Person re-identification is an crucial task to match character images across time, space and cameras. It is an indispensable tool in many applications in real life, such as character tracking in intelligent retail, image retrieval to find lost children, and public security. Supervisory approach has achieved impressive performance in this task, but the large amount of labeled data limits the real-world application of supervised methods. Due to this issue, unsupervised methods that learn the discriminative features for person retrieval from unlabeled data have recently received much attention. It has been proved that the existing

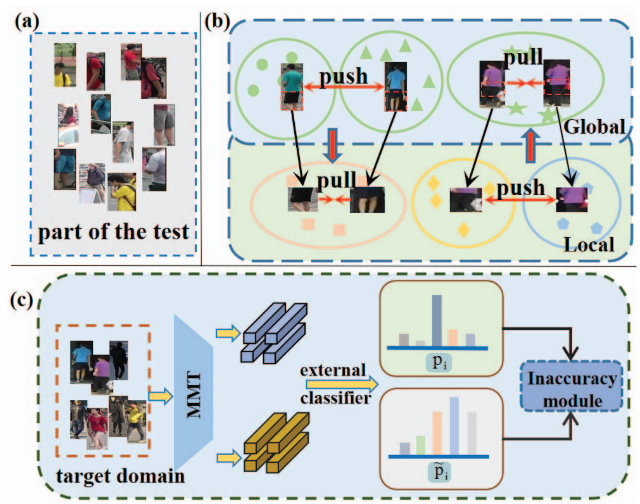


Fig. 1. (a) Selected images of the test dataset. (b) Example illustration of pseudo-label refinement based on local and global features. Left: local features are refined by global features. Right: Global features are refined by local features. (c) Inaccuracy-guided labeling optimization.

methods [1] can achieve significant performance when collecting training and testing data for the same application scenario. However, due to the inevitable domain gap, generalization between data sets cannot be performed well. Therefore, it is necessary for both academia and industry to study the re-identification of adaptive personnel in unsupervised fields. Training multiple backbones as teacher networks (e.g., dual ResNet in MMT [1], DenseNet and ResNet in MEB-Net [2] ) requires high computational costs. In addition, the labels refined by these methods only consider the global features,

*Corresponding author: Yanming Chen (cym@ahu.edu.cn).

This work is supported in part by the National Natural Science Foundation of China under Grand (61802001) and the Key Natural Science Foundation of Education Department of Anhui province (KJ2021A0046).

1945-788X/23/$31.00 ©2023 IEEE
DOI 10.1109/ICME55011.2023.00267

but ignore the local feature information, which is essential for the re-recognition of people, leading to the performance deficiency.Fig.1(a) shows some pedestrian images from the test dataset. It can be seen that under the camera, there are effects such as occlusion and illumination, and part of the test images show only partial information of the human body. If we include local information in the training process, this will definitely strengthen the re-recognition performance. If the generated pseudo labels match the ground truth exactly, the performance of the UDA Re-ID method will achieve the performance of the supervision method. Thus, enhancing consistency in different feature Spaces to improve the quality of pseudo labels could lead to huge performance gains. Meanwhile, due to divergence or domain gap between source and target domains, as well as the imperfect results of clustering algorithm, the pseudo labels assigned by clustering often contain noisy labels. Such noisy labels can mislead feature learning and impair the adaptive performance of the domain. Therefore, mitigating the negative effects of samples with unreliable pseudo labels is very important for the success of domain adaptation.

In order to solve the above two problems, this paper carries out consistency optimization from two directions respectively. (1) We propose to improve the quality of pseudo labeling by maintaining coherence between local and global features in the UDA person Re-id task. Fig.1(b) shows the mutual optimization of the clustering results between the global and local feature branches. Local features focus more on specific parts or details of a person. As can be seen from the figure, global and local features can correct each other for possible errors. We propose to refine the pseudo labels generated by different feature spaces. By this way a consistent feature space is formed. (2) Since the noisy labels inherent in such cross-domain pseudo labels can mislead the optimization of the network during the fine-tuning phase. In the optimization process, it is necessary to initially identify samples with noisy pseudo-labels and reduce their negative impact. After observing and analyzing the relationship between the characteristics of a large number of samples and the correctness of the pseudo labels. Based on uncertainty theory [3], the accuracy distribution of pseudo label with wrong and correct identities can be obtained. As illustrated in Fig.1(c), we propose an accuracy-guided label optimization strategy to select labels with excellent accuracy and stability.

Our main contributions can be summarized as follows:

- We propose a framework based on local and global features for mutual refinement of pseudo-labels(LGMR), and adopt mutual mean learning [1] and multi-branch network to ensure the consistency of feature space.
- We propose a pseudo label inaccuracy-guided mutual optimization strategy (IGMO) for selecting labels with high and stable accuracy.
- We conducted extensive experiments and achieved state-of-the-art performance on three benchmark datasets. Our proposed Intensified Consistency with Pseudo label Refinement (INCLR) framework boasts above-baseline performance in the Duke2Market task and a state-of-the-art approach in the Market2MSMT task.

## II. Related Work

**UDA person Re-ID.** Existing unsupervised domain adaptive person ReID methods can be divided into three broad categories. a) Domain translation based methods SPGAN [4] use Generative Adversarial Networks (GANs) to transform the image in the source domain to match the image style of the target domain. b) Memory bank-based methods are widely used in unsupervised representation learning, which helps to introduce contrast loss in general tasks. c) The cluster-based methods BUC [5] proposed a bottom-up clustering framework, which gradually grouped similar clusters. SpCL [6] adopts the intersection of strict and loose parameters in the clustering algorithm. MEB-Net [2] built three networks for mean learning. And our method is based on clustering, where the original features and the segmented features information are clustered separately, and then the consistency between different feature spaces is improved by mutual learning, which may lead to better performance.

**Pseudo label evaluation.** Quantifying and identifying the correctness of pseudo label can effectively eliminate the wrong pseudo label and improve the robustness of feature learning. Incorrect pseudo label samples are detrimental to the learning of robust feature embeddings for neural networks. Pseudo label evaluation becomes particularly important in order to quantify and identify the correctness of pseudolabeling. [3] and [7] developed an end-to-end framework to measure observation noise and mitigate the negative impact to better optimize the network. GLT [8] is a group-aware label transfer framework that corrects noisy labels explicitly while we select reliable pseudo-labels to progressively train the model, thus further correcting noisy labels implicitly. For clustering UDA person Re-ID, UNRN [9], $P^2$LR [10] are uncertainty-based methods to evaluate pseudo label and mitigate the negative impact of incorrect pseudo label. UNRN measures the output consistency between the average teacher and student models as uncertainty, while we softly assess the noise level of the sample for differences in the distribution between the teacher-student models as inaccuracy. $P^2$LR is a progressive label refinement approach that balances the negative effects of noisy labels through uncertainty-guided alternative optimization, while we progressively optimize pseudo labels through a quantitative criterion to measure label inaccuracy.

## III. Method

**Notation.** The goal of UDA person Re-ID is to adapt the trained model from a source domain $D_s = \{(x_i^s, y_i^s)|_{i=1}^{N_s}\}$to an unsupervised target domain $D_t = \{x_i|_{i=1}^{N_t}\}$, $N_s$ and $N_t$ denote the number of samples. $x_i^s$ denotes a sample and $y_i^s$ denotes its attached label in the source domain with supervisory information. $x_i$ denotes the sample in target domain without supervision.

Fig.2 shows an overview of the proposed Enhanced Consistency of Pseudo label Refinement (INCLR) method. In order to achieve a high degree of consistency in the feature space

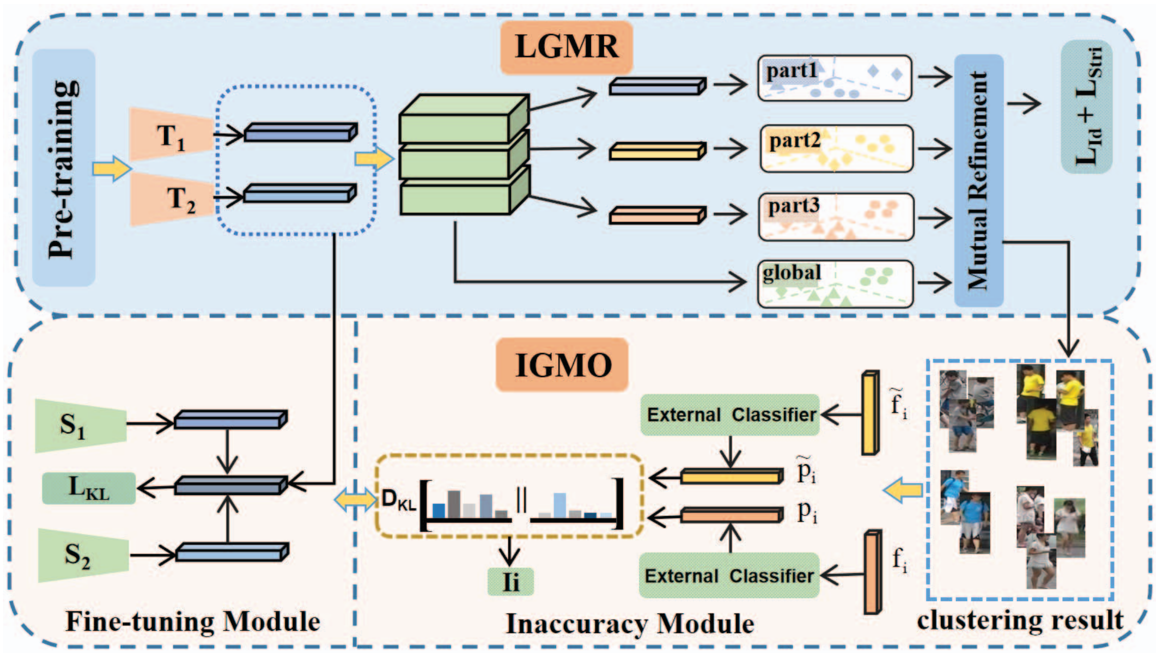


Fig. 2. The overall architecture of the proposed intensified pseudo label refinement method. The system consists of three parts, which are a pre-training module based on the MMT [1] model, a pseudo label mutual refinement framework (LGMR) used for local and global features, and a mutual optimization strategy (IGMO) guided by pseudo label accuracy. In the accuracy-guided pseudo label optimization strategy, the two processes of inaccuracy generation and model fine-tuning are alternately optimized.

and obtain high accuracy and stability of the pseudo label, it is implemented in two ways, specifically: (1) the input image is passed through two teacher models to generate feature vectors, and we use a multi-branch network structure to obtain three parts and global features. The features are independently clustered using the DBSCAN algorithm [11]. After refinement of the clustering results, the final global clustering results are obtained. (2) The pseudo-labels generated by the clustering are optimized again to discard the wrong labels and reduce the noise caused by the pseudo-labels to make the results more effective. Finally, fine-tuning is performed in the teacher-student model of the MMT [1] network. The two stages, fine-tuning and pseudo label refinement, are alternated for optimization.

### *A. Local-based optimization framework*

We adopt a fine-tuned architecture based on MMT networks with ResNet [6] as the backbone,and propose a pseudo-label refinement framework based on local features.Unlike most existing unsupervised domain adaptation methods,we use both local and global features to represent images. For a given image, our model first extracts the shared representation $F(x_i) \in R^{C\times H\times W}$, where C, H, and W are sizes of the channel, height and width of the feature map, respectively. After global average pooling (GAP), the global features are divided into $N_p$ local feature regions $R^{C\times \frac{H}{N_p}\times W}$. We adopt the concept of unsupervised method PPLR [12] to divide the global level into three localities. The local features are obtained by average pooling of each partition. The core idea of the mutual refinement framework is to maintain consistency between local and global features. In an unsupervised environment, We perform DBSCAN clustering [11] on the global feature set $\{f_i^g\}_{i=1}^{N_t}$ and assign pseudo labels. Following the standard protocol in the partial method in literature [13] , local and global features share the same pseudo label. We represent the pseudo label of the image $x_i$ as $y_i$ . For global features, cross entropy loss is calculated as follows:

$$\mathcal{L}_{gce} = -\sum_{i=1}^{N_t} y_i \cdot log(C_g^t(f_i^g)) \tag{1}$$

$C_g^t$ is a global feature classifier, which consists of a full connection layer and softmax function. $f_i^g$ denote the feature of local clustering. The prediction vector is obtained after the classifier calculation. Similarly, we train the local features using the cross-entropy loss by:

$$\mathcal{L}_{pce} = -\frac{1}{N_p}\sum_{i=1}^{N_t}\sum_{n=1}^{N_p} y_i \cdot log(C_{p_n}^t(f_i^{p_n})) \tag{2}$$

$C_{p_n}^t$ representing a classifier for a local feature region,and $f_i^{p_n}$ denote the feature of local clustering. We additionally add the softmax-triplet loss defined by:

$$\mathcal{L}_{Stri} = -\sum_{i=1}^{N_t} log\frac{e^{||f_i^g - f_{i,n}^g||}}{e^{||f_i^g - f_{i,p}^g||} + e^{||f_i^g - f_{i,n}^g||}} \tag{3}$$

Where $||\cdot||$ represents the $L_2$-norm, subscript $(i, p)$ and $(i, n)$ denote the hardest positive and negative samples respectively. Local feature and global feature can mutual optimize well, but due to the quality limitation of pseudo label, which is significantly noisy in the actual experiment. In the following section, we propose a method to eliminate unreliable pseudo label and better represent these two features.

### B. Inaccuracy-guided mutual optimization

For the clustering-based approach, the generated pseudo-labels are noisy. Our goal is to reduce the effect of noise by evaluating the accuracy of pseudo label. As shown in the Fig.3, we use the uncertainty criterion established based on uncertainty theory proposed by $P^2$LR [10], and samples with correct pseudo labels are associated with high accuracy. In contrast, samples with incorrect pseudo labels are associated with low accuracy, which has multiple peaks. Multiple peaks represent variable accuracy and each peak has a high probability.

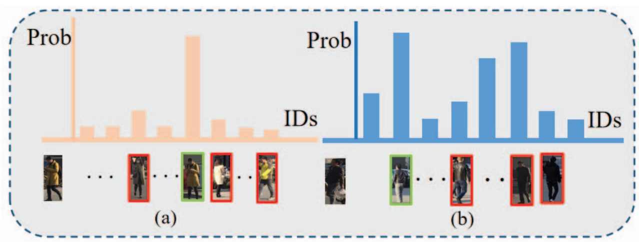


Fig. 3. Probability distribution with correct pseudo label (a) and incorrect pseudo label (b) in the first refinery step.

**Accuracy estimation.** Inspired by this, we use the distribution agreement between the average teacher model and the average student model as probabilistic accuracy to determine noise levels. Noise levels can be used as a quantitative measure of label inaccuracy. For each unlabeled sample $x_i$ in the target domain, we denote the feature extracted from the student model (as $S_1$, $S_2$ show in Fig. 2) as $f_i \in R^D$ of $D$ dimensional, and the features from the teacher model expressed as $\widetilde{f_i} \in R^D$ . Then utilize the external classifier $C_{v_t}$, Where $v_t \in R^{c\times d}$ is the parameterized weight of the classifier. Note that weights do not need to be trained separately and are generated dynamically. The following classifier is used to obtain the consistent distribution probability $\widetilde{p_i}$ in the teacher model and $p_i$ in the student model. Here $\beta$ is the temperature parameter.

$$p_i = C_{v_t}(f_i) = Softmax(\frac{v_t}{||v_t||} \cdot \frac{fi}{||f_i||} \cdot \beta) \tag{4}$$

$$\widetilde{p_i} = C_{v_t}(\widetilde{f_i}) = Softmax(\frac{v_t}{||v_t||} \cdot \frac{\widetilde{fi}}{||\widetilde{f_i}||} \cdot \beta) \tag{5}$$

We expect that the characteristics of the teacher-student model are inconsistent, where $Softmax(\cdot)$ represents the Softmax function of standardized similarity score. We use Kullback-Leibler (KL) divergence to measure the difference between the probability distributions of the two models as the inaccuracy $I_i$ of the image $x_i$ .

$$I_i = D_{KL}(\widetilde{p_i}||p_i) \tag{6}$$

**Optimize.** We observed that samples with wrong pseudo label generally have lower accuracy. For samples with low accuracy, we will reduce their contribution to the losses. We adopt the policy in [3], and denote $\mu_i = exp(-I_i)$ as the credibility weight to formulate an indeterminate oriented optimization based on KL divergence. We define the uncertainty-guided Kullback–Leibler (KL) divergence loss in a min-batch of $n_t$ target domain samples as:

$$\mathcal{L}_{KL} = -\frac{1}{n_t}\sum_{i=1}^{n_t} \mu_i logp(\widetilde{y}_i \| x_i) \tag{7}$$

where $p(\widetilde{y}_i \| x_i)$ denotes the probability that an image $x_i$ of being class $\widetilde{y}_i^t$ , where $\widetilde{y}_i^t$ denotes the pseudo ground-truth class (based on the pseudo label assigned after clustering). For a sample with high inaccuracy, a smaller weight is used to decrease its contribution to the overall loss in order to reduce its negative effect. The total loss of target domain data in the fine-tuning stage can be expressed as:

$$\mathcal{L}_{target} = \mathcal{L}_{gce} + \mathcal{L}_{pce} + \mathcal{L}_{Stri} + \lambda_{KL}\mathcal{L}_{KL} \tag{8}$$

$\lambda_{KL}$ is a weighting factor which prevents large inaccuracy.

## IV. Experiments

### A. Datasets and Evaluation Protocols

We evaluate the proposed method on datasets, i.e., Market-1501 [19], DukeMTMC-reID [20], MSMT17 [21]. The Market-1501 contains 32688 annotated images of 1501 person identities from 6 non-overlapping camera views. The DukeMTMC-reID dataset consists of 36411 images collected from 8 cameras. MSMT17 is a more challenging dataset consisting of 126441 images of 4101 identities captured from 15 different cameras. We adopt mean average precision (mAP) and cumulative matching characteristic (CMC) Rank-1/5/10(R1/R5/R10) accuracy for evaluation.

### B. Implementation Details

We employ ImageNet [22] pre-trained ResNet50 [23] as the backbone. We implement our model based on MMT [1] and train it on two Tesla P100-PCIE Gpus. We optimize the model using an optimizer with a weight decay of 5e-4. All the character images are resized to $256 \times 128$, The mini-batch size is 64, consisting of 16 pseudo-classes and 4 images of each class. Random cropping, flipping, and erasing are applied to data enhancement. Note that random erasure is not used during the number of sources training phase. We apply DBSCAN clustering algorithm and the clustering of Market, Duke and MSMT data sets is set as 500, 700 and 1500, respectively. We set the partial $N_p$ number as 3 empirically, and the temperature parameter $\beta$ of pre-training in $Eq.4$ and $Eq.5$ was set as 20. Let $\lambda_{KL} = 0.5$ and the total alternative optimization step $T$ is set to 100. In the pre-training stage of the source, the initial learning rate is set at $3.5 \times 10^{-4}$, which decreased by 1/10 in the 40th and 70th periods of the total 80 periods. In the target fine-tuning stage, the learning rate is fixed at $3.5 \times 10^{-4}$.

TABLE I
PERFORMANCE (%) COMPARISON WITH THE STATE-OF-THE-ART METHODS FOR UDA PERSON REID ON THE DATASETS OF MARKET-1501, DUKEMTMC-REID AND MSMT17. WE MARK THE RESULTS OF THE SECOND BEST BY UNDERLINE AND THE BEST RESULTS BY BOLD TEXT.

| Method | DukeMTMC→Market1501 | | | | Market1501→DukeMTMC | | | | DukeMTMC→MSMT17 | | | | Market1501→MSMT17 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mAP | R1 | R5 | R10 | mAP | R1 | R5 | R10 | mAP | R1 | R5 | R10 | mAP | R1 | R5 | R10 |
| SPGAN+LMP [4] (CVPR'18) | 26.7 | 57.7 | 75. | 82.4 | 26.2 | 46.4 | 62.3 | 68.0 | - | - | - | - | - | - | - | - |
| BUC [5] (AAAI'19) | 38.3 | 66.2 | 79.6 | 84.5 | 27.5 | 47.4 | 62.6 | 68.4 | - | - | - | - | - | - | - | - |
| ECN [14] (CVPR'19) | 43.0 | 75.1 | 87.6 | 91.6 | 40.4 | 63.3 | 75.8 | 80.4 | 8.5 | 25.3 | 36.3 | 42.1 | 10.2 | 30.2 | 41.5 | 46.8 |
| DAAM [15] (AAAI'20) | 67.8 | 86.4 | - | - | 63.9 | 77.6 | - | - | 20.8 | 44.5 | - | - | 21.6 | 46.7 | - | - |
| AD-Cluster [16] (CVPR'20) | 68.3 | 86.7 | 94.4 | 96.5 | 54.1 | 72.6 | 82.5 | 85.5 | - | - | - | - | - | - | - | - |
| NRMT [17] (ECCV'20) | 71.7 | 87.8 | 94.6 | 96.5 | 62.2 | 77.8 | 86.9 | 89 | 19.8 | 43.7 | 56.5 | 62.2 | 20.6 | 45.2 | 57.8 | 63.3 |
| MMT [1] (ICLR'20) | 71.2 | 87.7 | 94.9 | 96.9 | 65.1 | 78.0 | 88.8 | 92.5 | 22.9 | 49.2 | 63.1 | 68.8 | 23.3 | 50.1 | 63.9 | 69.8 |
| MEB-Net [2] (ECCV'20) | 76.0 | 89.9 | 96.0 | 97.5 | 66.1 | 79.6 | 88.3 | 92.2 | - | - | - | - | - | - | - | - |
| UNRN [9](AAAI'21) | 78.1 | 91.9 | 96.1 | 97.8 | 69.1 | 82.0 | 90.7 | 93.5 | 25.3 | 52.4 | 64.7 | 69.7 | 26.2 | 54.9 | 67.3 | 70.6 |
| SECRET [18] (AAAI'22) | 79.8 | 92.3 | - | - | 67.1 | 80.3 | - | - | - | - | - | - | 24.3 | 49.9 | - | - |
| $P^2$LR [10] (AAAI'22) | 81.0 | **92.6** | 97.4 | **98.3** | 70.8 | **82.6** | 90.8 | **93.7** | 29.9 | 60.9 | **73.1** | 77.9 | 29.0 | **58.8** | 71.2 | 76.0 |
| INCLR (ours) | **82.2** | **92.6** | **97.6** | **98.3** | **70.9** | **82.6** | **91.8** | 93.2 | **30.1** | **61.2** | 73 | **78.8** | **29.5** | 58.6 | **71.8** | **76.4** |

TABLE II
PERFORMANCE (%) OF DIFFERENT FEATURES AT INFERENCE TIME ON MARKET-1501 AND DUKEMTMC-REID.

| | DukeMTMC→Market1501 | | Market1501→DukeMTMC | |
|---|---|---|---|---|
| | mAP | R1 | mAP | R1 |
| INCLR-G | 79.9 | 92.5 | 67 | 80.1 |
| INCLR-part1 | 66.4 | 87.2 | 54.8 | 77.6 |
| INCLR-part2 | 64.3 | 85.3 | 52.0 | 75.9 |
| INCLR-part3 | 67.3 | 88.0 | 55.9 | 78.4 |
| INCLR-Joint | 82.2 | 92.6 | 70.9 | 82.6 |

### C. Comparison with the State-of-the-arts

We compare our proposed INCLR with the state-of-the-art advances in Table I for the four domain adaptation settings. Among the existing UDA person ReID methods, MMT [1] and MEB-Net [2] are all clustering-based methods.In order to make a fair comparison, We also adopt mean teacher to stabilize the training process and introduced a mutual average teaching of UDA person Re-ID. Compared with the baseline MMT, our proposed INCLR significantly improves the accuracy of UDA ReID by 11%, 5.8%, 6.6% and 6.8% for Duke→Market, Market→Duke, Market→MSMT and Duke→MSMT with mAP. Notably, UNRN and $P^2$LR utilize source data in the target fine-tuning phase and build external support memory to mine hard pairs. Our INCLR still achieves 4.2% and 3.9% mAP gains to UNRN and 0.2% and 0.5% map gains to $P^2$LR on the MSMT dataset. In general, our approach achieves state-of-the-art performance on all commonly used datasets, which validates the effectiveness of INCLR method.

### D. Ablation Studies

**Effectiveness of the mutual refinement procedure.** Since the proposed model can generate one global feature and three local features at the same time, multiple feature choices exist for inference: only the global feature and a combination of the global feature with all the local features. The combination can be implemented by modifying the weighted sum based on the calculation of the respective distances of the global and three different local features. Briefly, we use the same weight $\kappa$ to weight the three local features.

$$d_{i,j} = d_{i,j}^{global} + \kappa \cdot d_{i,j}^{part1} + \kappa \cdot d_{i,j}^{part2} + \kappa \cdot d_{i,j}^{part3} \quad (9)$$

The experimental results for each feature are shown in Table II. Local features alone can lead to a lot of poor performance. This makes sense because local features specifically represent only local information and are not like global features. The intuition is that a image is cut into three parts horizontally and then clustered separately, so that the clustering effect is more refined and specific, and more detailed information can be obtained. The weight of these three parts is set to the same parameter because the importance of the three parts (specifically, the detailed information of the person) cannot be determined, for example, different people's hairstyles and dresses will be very enriching, if the weight of the uppermost part is reduced and the diverse head information is ignored, the result is not correct and not significant enough. Therefore, it can be concluded that joint of local and global features can improve the performance.

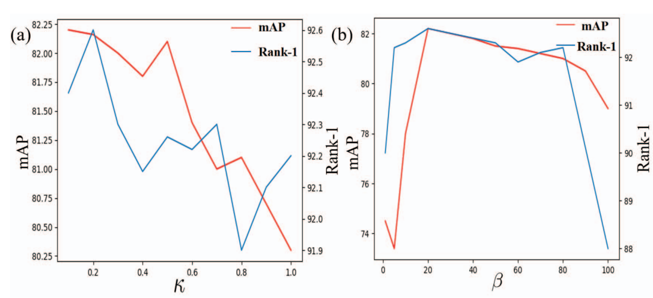


Fig. 4. (a)Sensitivity Analysis of Hyper-parameter $\kappa$. (b)Analysis of temperature parameter $\beta$. Here we pick the parameter value corresponding to the point where both mAP and Rank-1 can obtain the maximum value.

Fig.4(a) shows the experimental results with different weight parameters. $\kappa$ =0 means that only global features are used and $\kappa$ =1 means that all local features are equally important as global features. Results of different $\kappa$ show small fluctuation (mAP from 80.3 to 82.2 for Duke-to-Market, Rank-1 from 91.9 to 92.6). There experimental results can be concluded that the combined effect is best at $\kappa$ = 0.2, the achieved mAP is 82.2% and Rank-1 accuracy is 92.6%. And we set $\kappa$ to 0.2.

**Effectiveness of label refinement.** To analyze the effectiveness of our approach, we conducted extensive experiments on DukeMTMC-reID and Market-1501. We conducted separate experiments with the local feature-based pseudo label refinement module and the inaccuracy-based optimization pseudo label module. As show in Table III, the use of the separate modules shows a significant improvement in the results, respectively, which indicates that our two proposed modules are very meaningful. When the two modules are used in combination, we obtain significant performance improvements. Experimentally, our approach improves the mAP by 11% over the baseline, with a significant improvement.

TABLE III
ABLATION STUDY ON INDIVIDUAL COMPONENTS OF INCLR.

| **Method** | DukeMTMC→Market1501 | | | |
|---|---|---|---|---|
| | mAP | R1 | R5 | R10 |
| Module Pretraining | 28.1 | 56.0 | 71.6 | 77.5 |
| Baseline | 71.2 | 87.7 | 94.9 | 96.9 |
| Base+LGMR | 81.2 | 93.3 | 97.3 | 98.0 |
| Base+FAMO | 80.8 | 91.6 | 97.0 | 98.2 |
| Base+INCLR(ours) | 82.2 | 92.6 | 97.6 | 98.3 |

**Parameter analysis.** We explore the effect of temperature parameters on inaccuracy measurements in Fig.4(b). The achieved mAP and Rank-1 accuracy maintain high when $\beta$ is set to 20−60. When $\beta$ = 20, the achieved mAP is 82.2% and Rank-1 accuracy is 92.6%. When $\beta$ = 60, the achieved mAP is 81.4% and Rank-1 accuracy is 91.9%. Finally, we set the temperature parameter in the accuracy evaluation as $\beta$=20.

## V. CONCLUSION

In this paper, we propose an enhanced consistent pseudo label refinement framework for unsupervised domain-adapted peson re-identification. The method exploits the complementary relationship in two different feature spaces of the local feature space and the global feature space of the pedestrian images to mutually refine the pseudo label noise and introduces a label error-based quantization criterion to guide the pseudo label optimization to fine-tune the target domain for noise reduction. We have conducted extensive experimental and ablation studies on this, and the experiments show that the proposed method achieves state-of-the-art performance on a strong baseline and four benchmarks. The superiority of the proposed method is demonstrated.

## REFERENCES


[1] Y. Ge, D. Chen, and H. Li, “Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification,” in ICLR. 2020, OpenReview. net.

[2] Y. Zhai, Q. Ye, S. Lu, M. Jia, R. Ji, and Y. Tian, “Multiple expert brainstorming for domain adaptive person re-identification,” in EECV(7). 2020, vol. 12352 of Lecture Notes in Computer Science, pp. 594-611, Springer.

[3] A. Kendall, and Y. Gal, “What uncertainties do we need in bayesian deep learning for computer vision?,” in NIPS, 2017, pp. 5574-5584.

[4] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, “Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification,” in CVPR. 2018, pp. 994-1003, Computer Vision Foundation / IEEE Computer Society.

[5] Y. Lin, X. Dong, L. Zheng, Y. Yan, and Y. Yang, “A bottom-up clustering approach to unsupervised person re-identification,” in AAAI. 2019, pp. 8738-8745, AAAI Press.

[6] Y. Ge, F. Zhu, D. Chen, R. Zhao, and H. Li, “Self-paced contrastive learning with hybrid memory for domain adaptive object re-id,” in NeurIPS, 2020.

[7] J Chang, Z Lan, C Cheng, and Yichen Wei, “Data uncertainty learning in face recognition,” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, pp. 5710–5719.

[8] K Zheng, W Liu, L He, T Mei, J Luo, and Z Zha, “Group-aware label transfer for domain adaptive person re-identification,” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021b, pp. 5310–5319.

[9] K. Zheng, C. Lan, W. Zeng, Z. Zhang, and Z. Zha, “Exploiting sample uncertainty for domain adaptive person re-identification,” in AAAI. 2021, pp. 3538-3546, AAAI Press.

[10] J. Han, Y. Li, and S. Wang, “Delving into probabilistic uncertainty for unsupervised domain adaptive person re-identification,” in AAAI. 2022, pp. 790-798, AAAI Press.

[11] M. Ester, H. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in KDD. 1996, pp. 226-231, AAAI Press.

[12] Y. Cho, W. Kim, S. Hong, and S. Yoon, “Part-based pseudo label refinement for unsupervised person re-identification,” in CVPR. 2022, pp. 7298-7308, IEEE.

[13] F. Zheng, C. Deng, X. Sun, X. Jiang, X. Guo, Z. Yu, F. Huang and R. Ji, “Pyramidal person re-identification via multi-loss dynamic training,” in CVPR. 2019, pp. 8514-8522, Computer Vision Foundation / IEEE.

[14] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, “Invariance matters: exemplar memory for domain adaptive person re-identification,” in CVPR. 2019, pp. 598-607, Computer Vision Foundation / IEEE.

[15] Y. Huang, P. Peng, Y. Jin, J. Xing, C. Lang, and S. Feng, “Domain adaptive attention model for unsupervised cross-domain person re-identification,” CORR, vol. abs/1905. 10529, 2019.

[16] Y. Zhai, S. Lu, Q. Ye, X. Shan, J. Chen, R. Ji, and Y. Tian, “AD-cluster: Augmented discriminative clustering for domain adaptive person re-identification,” in CVPR. 2020, pp. 9018-9027, Computer Vision Foundation / IEEE.

[17] F. Zhao,S. Liao, G. Xie, J. Zhao, K. Zhang, and L. Shao, “Unsupervised domain adaptation with noise resistible mutual-training for person re-identification,” in ECCV(11). 2020,vol. 12356 of Lecture Notes in Computer Science, pp. 526-544, Springer.

[18] T. He, L. Shen, Y. Guo, G. Ding, and Z. Guo, “SECRET: self-consistent pseudo label refinement for unsupervised domain adaptive person re-identification,” in AAAI. 2022, pp. 879-887, AAAI Press.

[19] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable person re-identification: A benchmark,” in ICCV. 2015, pp. 1116-1124, IEEE Computer Society.

[20] E. Ristani, F. Solera, R. S. Zou, R. Cucchiara, and C. Tomasi, “Performance measures and a data set for multi-target, multi-camera tracking,” in ECCV Work-shops(2), 2016, vol. 9914 of Lecture Notes in Computer Science, pp. 17-35.

[21] L. Wei, S. Zhang, W. Gao, and Q. Tian, “Person transfer GAN to bridge domain gap for person re-identification,” in CVPR. 2018, pp. 79-88, Computer Vision Foundation / IEEE Computer Society.

[22] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei, “Imagenet: A large-scale hierarchical image database,” in CVPR. 2009, pp. 248-255, IEEE Computer Society.

[23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in CVPR. 2016, pp. 770-778, IEEE Computer Society.