Multi-Centroid Representation Network for Domain Adaptive Person Re-ID

Yuhang Wu[1*†], Tengteng Huang[2†], Haotian Yao[2], Chi Zhang[2],Yuanjie Shao[1],Chuchu Han[1], Changxin Gao[1], Nong Sang[1‡],
[1]Key Laboratory of Ministry of Education for Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology
[2]Megvii Technology
{wuyuhang,shaoyuanjie,hcc, cgao, nsang}@hust.edu.cn tengtenghuang@foxmail.com
{yaohaotian, zhangchi}@megvii.com

Abstract

Recently, many approaches tackle the Unsupervised Domain Adaptive person re-identification (UDA re-ID) problem through pseudo-label-based contrastive learning. During training, a uni-centroid representation is obtained by simply averaging all the instance features from a cluster with the same pseudo label. However, a cluster may contain images with different identities (label noises) due to the imperfect clustering results, which makes the uni-centroid representation inappropriate. In this paper, we present a novel Multi-Centroid Memory (MCM) to adaptively capture different identity information within the cluster. MCM can effectively alleviate the issue of label noises by selecting proper positive/negative centroids for the query image. Moreover, we further propose two strategies to improve the contrastive learning process. First, we present a Domain-Specific Contrastive Learning (DSCL) mechanism to fully explore intra-domain information by comparing samples only from the same domain. Second, we propose Second-Order Nearest Interpolation (SONI) to obtain abundant and informative negative samples. We integrate MCM, DSCL, and SONI into a unified framework named Multi-Centroid Representation Network (MCRN). Extensive experiments demonstrate the superiority of MCRN over state-of-the-art approaches on multiple UDA re-ID tasks and fully unsupervised re-ID tasks.

Introduction

Unsupervised Domain Adaptive person re-identification (UDA re-ID) is receiving increasing attention with the growing demand for practical video surveillance. The objective of UDA re-ID is to transfer knowledge learned from source domain with rich annotations to unlabeled target domain. Previous works usually tackle this problem by clustering (Ge, Chen, and Li 2020; Ge et al. 2020; Zhai et al. 2020a; Zheng et al. 2021a,b), which follow a two-step loop paradigm: (1) generating the pseudo labels of training samples from the target domain through clustering, (2) optimizing the model on the target domain with uni-centroid representation (*i.e.*,average feature or learnable weight of the cluster), under the supervision of the pseudo labels.

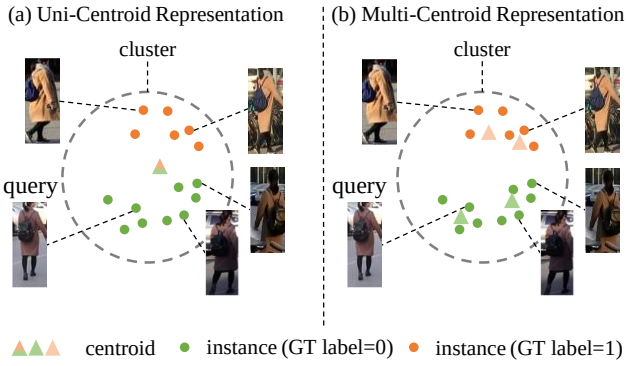

Figure 1: Comparison of traditional uni-centroid representation and our multi-centroid representation when the cluster is mixed with different identities. (a) The uni-centroid representation incorporates multiple identity information which is inappropriate. (b) Our multi-centroid representation provides multiple discriminative centroids, making it possible to select a suitable centroid as the positive sample that captures the same identity information with the query.

However, due to the imperfect results of the clustering algorithm, the pseudo labels always contain noises which are harmful to performance on the target domain. For example, as shown in Figure 1(a), instances belonging to two identities are incorrectly merged into a cluster and assigned the same pseudo label. In this case, traditional uni-centroid representation inevitably incorporates information from different identities, which would mislead the feature learning when the uni-centroid representation is used as the query's positive sample.

To alleviate the impact of label noises, we propose a Multi-Centroid Memory (MCM) to provide multiple centroids for each cluster. As Figure 1(b) shows, each centroid captures identity information within a local region of the cluster. This suggests that for each input query, we can se-

*This work was done when Yuhang Wu was an intern at Megvii Technology.
†equal contribution.
‡Corresponding author.

lect its reliable positive and negative samples from the centroids in the positive cluster and other negative clusters, respectively. However, for a specific query, its positive centroids may contain some incorrect ones that capture different identity information with it. Such centroids will hinder the feature learning when used as the positive samples. To reduce the effect of these false-positive centroids, we propose a matching mechanism between the query and each positive centroid to select a centroid as the positive sample. In general, the least similar positive centroid to the query is most likely be the false-positive centroid because of the unsatisfied inter-class separability and fixed clustering threshold, while the most similar one is not conducive to learning intra-class diversity. For striking a balance between the correctness and diversity, we select the moderate similar centroid as the positive sample. Besides, the inferior clustering may also lead to some false-negative centroids, damaging the intra-class compactness. We select the mean negative centroid of each negative cluster as the negative sample, which is more reliable than using all negative centroids.

In addition to considering the reliability of the positive and negative samples, we further considered their quality for feature learning. Some methods (Bai et al. 2021; Ge et al. 2020; Zheng et al. 2021a) use valuable source domain data for training. Here, we follow these methods and extend our MCM to the source domain. However, the cross-domain negative samples are quite easy for the query due to the huge domain gap between the source and target domains. These easy cross-domain negative samples contribute little to the optimization, and simply pushing them away from the query enlarges the domain gap. Given that, we propose Domain-Specific Contrastive Learning (DSCL) to fully mine intra-domain knowledge by only selecting the positive and negative samples from the query's domain for contrastive learning. Furthermore, inspired by the recent negative mining methods (Kalantidis et al. 2020; Zhong et al. 2021) that use interpolation in the latent space to synthesize more negative samples, we propose Second-Order Nearest Interpolation (SONI) to obtain additional hard negative samples for the query from the target domain. To ensure the synthetic negative samples are reliable and informative, SONI selects a set of nearest negative centroids and then uses each centroid as an anchor to interpolate with another nearest negative centroid that is nearest to it but has a different pseudo label. We integrate MCM, DSCL and SONI into a unified framework, Multi-Centroid Representation Network (MCRN), which provides each query with the positive and negative samples that are reliable and effective for contrastive learning.

Our contributions can be summarized as follows:

- We propose a Multi-Centroid Memory (MCM) to alleviate the label noise problem in previous UDA re-ID methods. By selecting reliable positive and negative centroids from MCM for each input query, the impact of label noises can be reduced.
- We further propose Domain-Specific Contrastive Learning (DSCL) and Second-Order Nearest Interpolation (SONI) to obtain negative samples that are not only reliable but also effective for contrastive learning, which significantly improve the learning process.
- Our integrated framework MCRN significantly outperforms state-of-the-art methods by a large margin on multiple UDA re-ID tasks. Besides, extensive experiments on fully unsupervised re-ID tasks consistently demonstrate the superiority of our approach over previous methods.

Related Work

Unsupervised domain adaptive (UDA) person re-ID

The existing methods can be categorized into two branches, *i.e.*,, domain translation-based methods (Deng et al. 2018; Wei et al. 2018; Zou et al. 2020) and clustering-based methods (Ge, Chen, and Li 2020; Ge et al. 2020; Zhai et al. 2020a; Zheng et al. 2021a,b). In this section, we mainly review clustering-based approaches since they are more related to our framework.

Clustering-based methods usually leverage the pseudo labels generated by clustering algorithms to optimize the network. However, it is quite challenging to assign correct pseudo labels to each unlabeled image due to the imperfect clustering results. MMT (Ge, Chen, and Li 2020) adopts a mutual mean-teaching framework to provide more robust soft labels. NRMT (Zhao et al. 2020) performs collaborative clustering and mutual instance selection by maintaining two networks during training. UNRN (Zheng et al. 2021a) introduces uncertainty estimation to explore the reliability of the pseudo label of each sample. SpCL (Ge et al. 2020) propose the self-paced learning strategy to obtain more reliable clustering results. GLT (Zheng et al. 2021b) uses a group-aware label transfer algorithm to online refine the pseudo labels. However, these works usually use average feature or learnable weight to represent a cluster, which is sensitive to label noises. Recently, a fully unsupervised re-ID method ClusterContrast (Dai et al. 2021) updates the cluster representation with the hardest positive instance feature in a batch, which is more robust than previous cluster representation. Different from these methods, we introduce multiple centroids to adaptively detect and represent potential multiple sub-classes in a cluster.

Contrastive Learning

The contrastive loss (Oord, Li, and Vinyals 2018) is widely used in unsupervised visual representation learning task (Chen et al. 2020; He et al. 2020; Tian, Krishnan, and Isola 2020) to learn discriminative feature representation by maximizing the similarity of augmented views generated from an identical instance. Recently, SpCL (Ge et al. 2020) successfully adapts contrastive loss to UDA re-ID task and propose a Unified Contrastive Loss (UCL) to distinguish the query from negative samples from both the source and target domain. Different from UCL, we propose a novel Domain-Specific Contrastive Learning (DSCL) mechanism which only selects informative negative samples from the same domain of query.

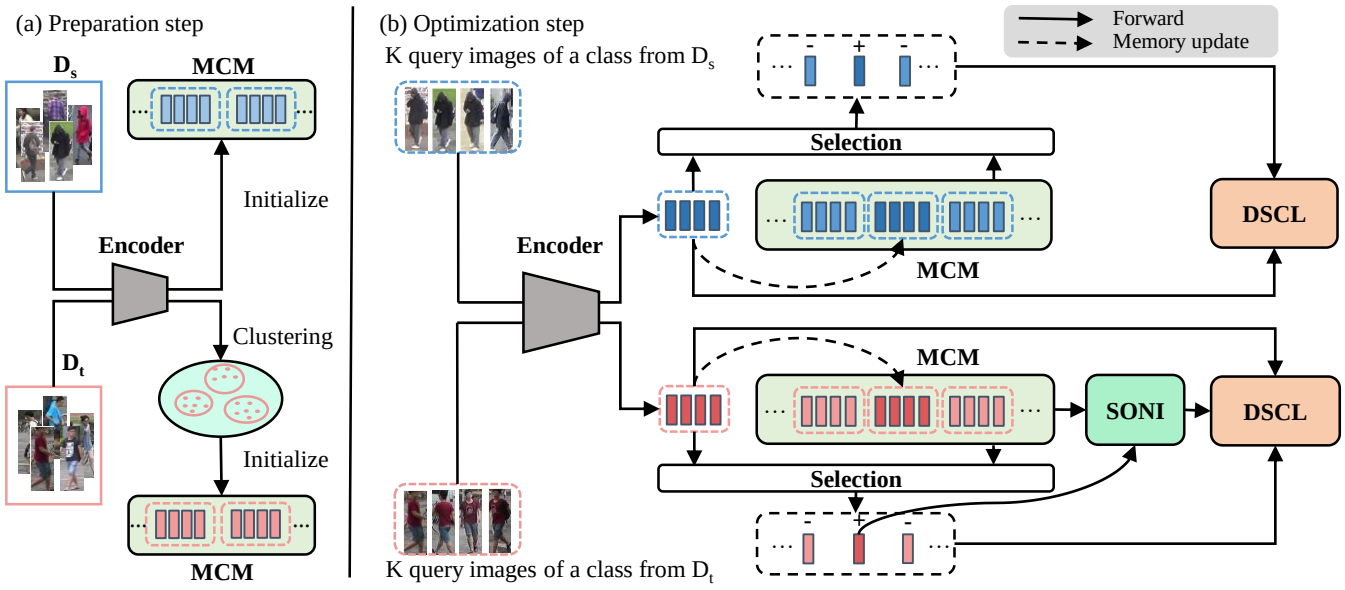

Figure 2: Illustration of the overall training pipeline of MCRN.

Hard Negative Mining

Mining hard negative samples plays an important role in boosting the performance of metric learning. The Embedding Expansion (Ko and Gu 2020) employs uniform interpolation between two positive and negative points to generate many synthetic points and then select the hardest pair as negative. NCD (Zhong et al. 2021) mixes the query in the novel classes with the samples in known classes to produce synthetic features, and then filters the hardest synthetic negatives by the cosine similarity with the query. MoCHi (Kalantidis et al. 2020) synthesizes hard negatives by interpolating between the query and hard negative samples or any two randomly selected hard negative samples. Unlike these methods, to ensure the reliability and quality of the synthesized samples, we synthesize negative samples by interpolation between two hard negative centroids that are close to each other but have different pseudo labels.

Proposed Method

Overview

Given a source domain D_s and a target domain D_t, the goal of UDA re-ID is to improve the model performance on D_t by transferring knowledge from D_s to D_t. $D_s = \left\{ \left(x_s^i, y_s^i\right)\big|_{i=1}^{N_s} \right\}$ consists of N_s labeled images, where $\left(x_s^i, y_s^i\right)$ denotes the i-th training sample and its associated label. And $D_t = \left\{ \left(x_t^i\right)\big|_{i=1}^{N_t} \right\}$ composes of N_t images without annotations.

In this paper, we propose a Multi-Centroid Representation Network (MCRN), which consists of an encoder and a novel Multi-Centroid Memory (MCM). Moreover, we introduce a new Domain-Specific Contrastive Learning (DSCL) objective and a Second-Order Nearest Interpolation (SONI) mechanism to jointly improve the feature learning process.

Figure 2 illustrates the overall training pipeline of MCRN, which alternates between two steps: preparation step in Figure 2(a) and optimization step in Figure 2(b). In the preparation step, we group unlabeled images from D_t into clusters using clustering algorithm (*e.g.*, DBSCAN (Ester et al. 1996)), based on the instance features extracted by the encoder. Then the Multi-Centroid Memory (MCM) is created and initialized, which is detailed later. In the optimization step, we carefully select positive/negative samples from MCM and generate more negative samples through SONI, followed by optimizing the encoder through DSCL between input queries and these samples. Note that MCM is dynamically updated during the optimization step.

Multi-Centroid Memory

Memory initialization. Each epoch begins with the preparation step. We first extract instance features for all the images from the source domain D_s and the target domain D_t. Then we group the unlabeled target domain images into n_t clusters through DBSCAN (Ester et al. 1996). In this process, we simply discard all the un-clustered instances. Then we build Multi-Centroid Memory (MCM) as a tensor in the shape of $M \times C$, where M denotes the total number of centroids in MCM and C denotes the dimension of feature channels. We set M equal to $K \times (n_s + n_t)$, where K represents the number of centroids for each class. n_s and n_t denote the number of ground-truth classes from D_s and pseudo classes (*i.e.*,clusters) from D_t, respectively. We initialize all the K centroids for a identical class as the mean feature of all the instance images from this class.

Memory update. In the optimization step, the centroids in MCM are continuously updated to detect and to represent potential multiple sub-classes with different identities, by continuously incorporating recent query features from the corresponding class. During training, MCM is update at the

end of each iteration. Concretely, each mini-batch is composed of P classes with K queries per class, sampled by the widely used PK sampling approach. Note that K is simply set identical to the number of centroids per class in MCM, which allows all of the centroids from the same class to be updated at the same pace. For the K queries from the same class, we search for a permutation of corresponding K centroids $\sigma \in \mathfrak{S}_K$ to establish best bipartite matching for query-centroid pairs with the highest overall similarity:

$$\hat{\sigma} = \underset{\sigma \in \mathfrak{S}_K}{\arg\max} \sum_{i}^{K} q^i \cdot c^{\sigma(i)} \tag{1}$$

where q^i denotes i-th sampled query of the class and $c^{\sigma(i)}$ indicates the i-th centroid in the permutation σ. We use Hungarian algorithm (Kuhn 1955) to efficiently compute the optimal permutation $\hat{\sigma}$. Based on the matched query-centroid pairs, we update the centroids in an exponential moving average manner as follows:

$$c^{\hat{\sigma}(i)} \leftarrow m c^{\hat{\sigma}(i)} + (1-m)\, q^i \tag{2}$$

where m is the momentum coefficient. We set m to a relatively small value (*e.g.*, 0.2) so that the centroid can absorb more identity information from the query feature.

Reliable centroids for contrastive learning

For each query in the mini-batch, MCM provides K positive candidates and $(n_s + n_t - 1)K$ negative candidates. How to obtain reliable and informative positive/negative samples plays an important role in the following contrastive learning.

Reliable positive samples. For a query, some of the K positive candidates in MCM may capture different identity information (*i.e.*, false-positive centroid), due to incorrect clustering results. To obtain a reliable positive sample, we rank the K positive candidates in ascending order according to their cosine similarity with the query. A natural choice is to select the candidate with the largest similarity as the positive sample. However, the most similar candidate usually incorporates the query feature in previous updates and is thus less informative for learning intra-class diversity. Besides, the least similar candidate is more likely to be an outlier. Therefore, we select the candidate ranked in the median (*i.e.*, $\lceil \frac{K}{2} \rceil$), which we call the moderate positive centroid, as the positive sample.

Reliable negative samples. A naive choice is simply taking all the $(n_s + n_t - 1)K$ negative candidates as negative samples. However, images with the same identity may be incorrectly split into multiple clusters due to unsatisfactory clustering results, resulting in false-negative candidates. Pushing the query and these false-negative candidates away would bias the feature learning. However, it is quite difficult to find and exclude possible false-negative candidates. To alleviate this problem, we represent each cluster as the mean feature of its K centroids and take the mean feature (named mean negative centroid) as a negative sample. In this way, we can obtain $(n_t + n_s - 1)$ in total negative samples from all the clusters except the one that the query falls in.

Notably, we use the same selection strategy for the query from D_s to address the issue of possible annotation errors.

Domain-Specific Contrastive Learning

Previous work SpCL (Ge et al. 2020) employs a Unified Contrastive Learning (UCL) to push samples from different classes away and pull those within the same class together. All the negative samples from the source and target domains are considered, no matter which domain the query comes from. UCL can be formulated as follows:

$$L_U = -\log \frac{\exp\left(\frac{1}{\tau} q \cdot c^+\right)}{\sum_{i=1}^{n_s} \exp\left(\frac{1}{\tau} q \cdot c_s^i\right) + \sum_{j=1}^{n_t} \exp\left(\frac{1}{\tau} q \cdot c_t^j\right)} \tag{3}$$

where q is the query in the mini-batch. c_s^i and c_t^j are the selected centroid of the i-th source-domain class and the j-th target-domain pseudo class, respectively. c^+ is the moderate positive centroid of the positive class and τ indicates the temperature coefficient.

However, due to the significant domain gap, it is quite easy for a model to distinguish the query from those negative centroids from a different domain. Such negative samples cannot provide effective information to learn discriminative representations. Besides, simply pushing them away from the query enlarges the domain gap. Thus, we propose Domain-Specific Contrastive Learning (DSCL) which push query away from negative samples in the same domain:

$$L_{D_s} = -\log \frac{\exp\left(\frac{1}{\tau} q_s \cdot c^+\right)}{\sum_{i=1}^{n_s} \exp\left(\frac{1}{\tau} q_s \cdot c_s^i\right)} \tag{4}$$

$$L_{D_t} = -\log \frac{\exp\left(\frac{1}{\tau} q_t \cdot c^+\right)}{\sum_{i=1}^{n_t} \exp\left(\frac{1}{\tau} q_t \cdot c_t^i\right)} \tag{5}$$

where q_s and q_t denote the query from source and target domains, respectively. Focusing on distinguishing the pairs from the same domain, DSCL can fully mine domain-specific semantic information and improve the generalization capability.

Second-Order Nearest Interpolation

We further propose a novel interpolation mechanism called Second-Order Nearest Interpolation (SONI) to synthesize abundant and informative negative samples. For each query from D_t, SONI interpolates between two hard negative centroids in MCM that are close to each other but belong to different pseudo classes. As indicated by its name, SONI involves two nearest neighbor searching processes. First, we collect the top $\gamma = \alpha n_t$ nearest negative centroids into a set $H = \left\{h^j|_{j=1}^{\gamma}\right\}$, where α is a hyper-parameter that controls the number of synthetic negative samples and n_t is the number of pseudo classes. In this process, we use the moderate positive centroid to select the hard negative centroids since it is a reliable representation for the query. Then we search for the nearest negative neighbor $\tilde{h^i} \in H$ for each centroid $h^i \in H$. We interpolate between $h^i \in H$ and $h^j \in H$ to obtain a synthetic negative sample s^i.

$$s^i = \beta h^i + (1-\beta)\, \tilde{h^i} \tag{6}$$

where β is randomly sampled from a uniform distribution in the range of $[0.2, 0.5]$ in each iteration.

We reformulate Equation (5) for DSCL as the follows to incorporate the negative samples generated by SONI:

$$L^*_{D_t} = -\log \frac{\exp\left(\frac{1}{\tau} q_t \cdot c^+\right)}{\sum_{i=1}^{n_t} \exp\left(\frac{1}{\tau} q_t \cdot c_t^i\right) + \sum_{j=1}^{\gamma} \exp\left(\frac{1}{\tau} q_t \cdot s^j\right)} \tag{7}$$

where s^j is the j-th synthetic negative sample for the query.

Overall Loss

Each mini-batch consists of n encoded source-domain queries $Q_s = \left\{q_s^i|_{i=1}^n\right\}$ and n encoded target-domain queries $Q_t = \left\{q_t^i|_{i=1}^n\right\}$. The overall optimization goal is as follows:

$$L_{total} = \frac{1}{n} \sum_{q_s \in Q_s} L_{D_s} + \frac{1}{n} \sum_{q_t \in Q_t} L^*_{D_t} \tag{8}$$

Experiments

Datasets and Evaluation Metrics

We evaluate our method on three person re-ID datasets, including Market-1501 (Zheng et al. 2015), DukeMTMC-reID (Ristani et al. 2016) and MSMT17 (Wei et al. 2018). Rank-1/5/10 (R1/R5/R10) of Cumulative Matching Characteristic (CMC) and mean average precision (mAP) are adopted for evaluation.

Training details of MCRN

Baseline. We use SpCL (Ge et al. 2020) as our uni-centroid baseline and follow its most settings. ResNet-50 (He et al. 2016) pretrained on ImageNet is used as the backbone for our encoder. We adopt domain-specific BNs (Chang et al. 2019) for narrowing the domain gap. DBSCAN (Ester et al. 1996) clustering followed by a self-paced strategy (Ge et al. 2020) is adopt for generating pseudo labels. For a fair comparison of uni-centroid and multi-centroid settings, we make two modifications to the original SpCL. First, we reinitialize the memory bank at the beginning of every epoch, while SpCL only initializes once at the first epoch. Second, we simply discard un-clustered instances while SpCL keeps them.

Training details. Each mini-batch consists of 64 source domain images and 64 target domain images, with 4 images per ground-truth/pseudo class (*i.e.*, K is set to 4). All training images are resized to 256×128 and various data augmentations are applied, including random cropping, random flipping and random erasing (Zhong et al. 2020). Adam optimizer is utilized to optimize the encoder with a weight decay of 0.0005. The initial learning rate is set to 0.00035 and is decayed by 1/10 every 20 epochs in the total 50 epochs. The momentum coefficient m in Equation 2 is set to 0.2, and the temperature coefficient τ in the contrastive losses is set to 0.05. α in SONI is set to 0.03. We implement our approach using the Pytorch (Paszke et al. 2019) framework and use four NVIDIA RTX-2080TI GPUs for training.

Ablation Studies

Superiority of multi-centorid representation. In Table 1, we compare the performance of uni-centroid (**Baseline**) and multi-centroid (**MCM**) representation method. Notably, both of them adopt UCL (Equation 3) as the learning objective and the only difference between them is the representation of each class. The result shows that our MCM significantly surpasses the baseline by considerable margins. As shown in Table 1, MCM outperforms baseline by 8.4%/4.4%, 8.7%/5.9%, 11.1%/15.8% and 9.9%/14.5% in terms of mAP/R1 on four UDA tasks, clearly demonstrating the superiority of multi-centroid representation over traditional uni-centroid representation.

Effectiveness of DSCL. We further conduct experiments by replacing UCL with DSCL (**MCM** v.s. **MCM+DSCL**). As Table 1 shows, DSCL brings the model with consistent performance gain on all tasks. Specially, mAP/R1 is improved by 0.9%/1.2% and 1.6%/2.0% on Duke→MSMT and Market→MSMT tasks, respectively. Following (Bai et al. 2021), we compare the domain distance, which is measured by the cosine distance between the average feature of two domains. As is shown in Figure 3, DSCL reduces the distance between source and target domains, indicating the effectiveness of DSCL in reducing the domain gap.

Effectiveness of SONI. As is shown in Table 1, SONI yields a general improvement on all the four UDA tasks. For example, mAP/R1 is increased by 3.5%/4.7% and 4.2%/6.5% on Duke→MSMT and Market→MSMT tasks, respectively. These results demonstrate the effectiveness of SONI in providing informative and beneficial negative samples. Besides, SONI is complementary to DSCL. When combine them together, our approach achieves superior results of mAP 83.8%, 71.5%, 35.7% and 32.8% on these tasks, respectively.

Design Choices

Strategies for selecting positive samples. Besides the moderate positive centroid, we further present two alternatives for selecting positive samples, *i.e.*, selecting the most or the least similar centroid. We call these three strategies **Moderate**, **Most**, and **Least** for short. As is shown in Table 2, **Moderate** consistently outperforms **Most/Least** by a large margin, yielding an improvement of mAP 25.2%/1.6% and 26.2%/1.4% on Duke→Market and Market→Duke tasks, respectively. It might be surprising that **Most** leads to heavily degraded performance. We assume the reason is that the most similar centroid is likely to absorb the query feature in previous updates and thus is less informative for learning intra-class diversity.

Strategies for selecting negative samples. We further compare two strategies for selecting negative samples, including 1) mean negative centorid (**Mean** for short) and 2) all negative centroids (**All** for short). For a query, **All** simply uses all centroids from a negative class, while **Mean** only takes the mean centroid. As is shown in Table 2, **Mean** leads to considerable and general improvements over **All**, which indicates that **Mean** can effectively filter bad samples and alleviate the issue of false-negative samples.

Methods	Duke→Market		Market→Duke		Duke→MSMT		Market→MSMT	
	mAP	R1	mAP	R1	mAP	R1	mAP	R1
Baseline	73.1	87.8	62.1	77.6	19.1	44.5	18.2	43.0
MCM	81.5	92.2	70.8	83.5	30.2	60.3	28.1	57.5
MCM+DSCL	82.4	92.8	71.4	84.0	31.1	61.5	29.7	59.5
MCM+SONI	82.9	93.2	70.8	83.9	33.7	65.0	32.3	64.0
MCRN	**83.8**	**93.8**	**71.5**	**84.5**	**35.7**	**67.5**	**32.8**	**64.4**

Table 1: Ablation studies of our proposed components. Baseline: uni-centroid baseline based on SpCL (Ge et al. 2020).

Positive	Negative	Duke→Market		Market→Duke	
		mAP	R1	mAP	R1
Most	Mean	56.3	76.0	44.6	58.7
Least	Mean	79.9	91.2	69.4	82.8
Moderate	All	81.0	91.4	70.0	83.3
Moderate	Mean	**81.5**	**92.2**	**70.8**	**83.5**

Table 2: Comparison of different strategies for selecting positive/negative samples.

Methods	Duke→Market		Market→Duke	
	mAP	R1	mAP	R1
QNNI	76.5	89.4	66.1	80.9
RNNI	83.4	93.5	70.4	83.6
SONI	**83.8**	**93.8**	**71.5**	**84.5**

Table 3: Comparison with different strategies for synthesizing negative samples in MCRN.

MCM on the source data. We ablate the effect of MCM on the source domain (MCM-S). On Duke→Market, MCM-S outperforms Baseline (in Table 1) by +1.8%/+1.1% in terms of mAP/R1. We assume the reason is that MCM can increase the diversity of learned representation for each class. Therefore, we use MCM on both source and target domains.

Alternatives of interpolation approach. Besides SONI, we evaluate another two interpolation methods proposed in MoCHi (Kalantidis et al. 2020). One is an interpolation between the query and its nearest negative sample (QNNI for short). The other is an interpolation between two samples randomly selected from the top-γ nearest negative samples (RNNI for short). As is shown in Table 3, SONI performs best among these three interpolation approaches. Since samples generated by QNNI incorporates the query feature, these samples are usually too hard to differentiate by the model and harmful for optimization. Instead of using two random negative samples adopted in RNNI, SONI interpolates between two negative samples which are semantically similar but have different pseudo labels, which is beneficial to obtain more informative samples.

The number of centroids for each class. We conduct experiments by varying K from 2 to 6 with an interval of 1 to investigate how the number of centroids K for each class influence the UDA performance. As is shown in Table 4, the UDA performance is continuously boosted when K is increased from 2 to 4. With larger K, the UDA performance reaches a plateau and no obvious gain is observed. Hence, we set $K = 4$ as our default setting.

Value of K	Duke→Market				
	2	3	4	5	6
mAP	75.7	83.2	**83.8**	83.5	83.5
R1	89.3	93.0	**93.8**	93.4	93.6

Table 4: Influence of the number of the centroids K for each class.

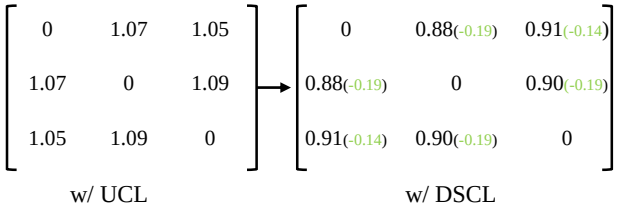

Figure 3: Pair-wise cosine distances among three domains: Market, Duke and MSMT.

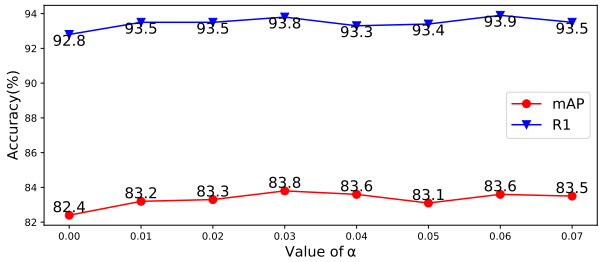

Figure 4: Performance of MCRN on Duke→Market task with different value of α.

The number of synthetic negative samples. We generate $\gamma = \alpha n_t$ synthetic hard negative samples through SONI. To explore the effect of the number of synthetic negative samples, we vary α from 0 to 0.07 with an interval of 0.01 and present the results in Figure 4. As is shown, α in the range of $[0.01, 0.07]$ consistently outperforms $\alpha = 0$, indicating the effectiveness of synthetic negative samples. Besides, with the increase of α, the UDA performance first increases and then reaches a plateau, with the best performance achieved at 0.03. Hence, we set $\alpha = 0.03$ as our default setting.

Methods	Reference	DukeMTMC→Market1501				Market1501→DukeMTMC			
		mAP	R1	R5	R10	mAP	R1	R5	R10
AD-Cluster (Zhai et al. 2020a)	CVPR 20	68.3	86.7	94.4	96.5	54.1	72.6	82.5	85.5
MMT (Ge, Chen, and Li 2020)	ICLR 20	71.2	87.7	94.9	96.9	65.1	78.0	88.8	92.5
NRMT (Zhao et al. 2020)	ECCV 20	71.7	87.8	94.6	96.5	62.2	77.8	86.9	89.5
MEB-Net (Zhai et al. 2020b)	ECCV 20	76.0	89.9	96.0	97.5	66.1	79.6	88.3	92.2
DG-Net++ (Zou et al. 2020)	ECCV 20	61.7	82.1	90.2	92.7	63.8	78.9	87.8	90.3
SPCL (Ge et al. 2020)	NIPS 20	76.7	90.3	96.2	97.7	68.8	82.9	90.1	92.5
HGA (Zhang et al. 2021a)	AAAI 21	70.3	89.5	93.6	95.5	67.1	80.4	88.7	90.3
UNRN (Zheng et al. 2021a)	AAAI 21	78.1	91.9	96.1	97.8	69.1	82.0	90.7	93.5
GCL (Chen et al. 2021)	CVPR 21	75.4	90.5	96.2	97.1	67.6	81.9	88.9	90.6
GLT (Zheng et al. 2021b)	CVPR 21	79.5	92.2	96.5	97.8	69.2	82.0	90.2	92.8
RDSBN+MDIF (Bai et al. 2021)	CVPR 21	81.5	92.9	**97.6**	98.4	66.6	80.3	89.1	92.6
MCRN	This paper	**83.8**	**93.8**	97.5	**98.5**	**71.5**	**84.5**	**91.7**	**93.8**
Methods	Reference	DukeMTMC→MSMT17				Market1501→MSMT17			
		mAP	R1	R5	R10	mAP	R1	R5	R10
MMT (Ge, Chen, and Li 2020)	ICLR 20	23.3	50.1	63.9	69.8	22.9	49.2	63.1	68.8
DG-Net++ (Zou et al. 2020)	ECCV 20	22.1	48.8	60.9	65.9	22.1	48.4	60.9	66.1
SpCL (Ge et al. 2020)	NIPS 20	26.5	53.1	65.8	70.5	26.8	53.7	65.0	69.8
UNRN (Zheng et al. 2021a)	AAAI 21	26.2	54.9	67.3	70.6	25.3	52.4	64.7	69.7
HGA (Zhang et al. 2021a)	AAAI 21	26.8	58.6	64.7	69.2	25.5	55.1	61.2	65.5
GLT (Zheng et al. 2021b)	CVPR 21	27.7	59.5	70.1	74.2	26.5	56.6	67.5	72.0
GCL (Chen et al. 2021)	CVPR 21	29.7	54.4	68.2	74.2	27.0	51.1	63.9	69.9
RDSBN+MDIF (Bai et al. 2021)	CVPR 21	33.6	64.0	75.6	79.6	30.9	61.2	73.1	77.4
MCRN	This paper	**35.7**	**67.5**	**77.9**	**81.6**	**32.8**	**64.4**	**75.1**	**79.2**

Table 5: Comparison with state-of-the-art UDA person re-ID methods on common UDA benchmarks.

Methods	Reference	Market1501		DukeMTMC		MSMT17	
		mAP	R1	mAP	R1	mAP	R1
SpCL (Ge et al. 2020)	NIPS 20	73.1	88.1	65.3	81.2	19.1	42.3
GCL (Chen et al. 2021)	CVPR 21	66.8	87.3	62.8	82.9	21.3	45.7
RLCC (Zhang et al. 2021b)	CVPR 21	77.7	90.8	69.2	83.2	27.9	56.5
MCRN	This paper	**80.8**	**92.5**	**69.9**	**83.5**	**31.2**	**63.6**

Table 6: Comparison with state-of-the-art fully unsupervised person re-ID methods on person re-ID datasets.

Comparison with the State-of-the-arts

Performance under the UDA re-ID setting. We compare our proposed MCRN with the state-of-the-art UDA re-ID methods on four domain adaptation tasks in Table 5. Our method significantly outperforms the second best UDA re-ID methods by 2.3%, 2.3%, 2.1% and 1.9% in mAP on these tasks, respectively. With the same base configuration as SpCL, our method outperform SpCL by 7.1%, 2.7%, 9.2% and 6.0% in terms of mAP on these tasks, respectively. The comparison with MMT (Ge, Chen, and Li 2020) and UNRN (Zheng et al. 2021a) are valuable, since they adopt a teacher-student framework which consists of two identical models while our method can outperform them with only a single model.

Performance under the fully unsupervised re-ID setting. Our proposed method can be easily generalized to fully unsupervised re-ID tasks. We compare our method with other state-of-the-art approaches for unsupervised re-ID in Table 6. As is shown, our method remarkably surpasses the state-of-the-art fully unsupervised person re-ID methods on Market, Duke and MSMT datasets, which validates the effectiveness of our method once again. Specially, our MCRN outperforms the second best method RLCC (Zhang et al. 2021b) by 3.1%/1.7% and 3.3%/7.1% in mAP/R1 on Market and MSMT datasets, respectively.

Conclusion

In this work, we propose a unified framework, Multi-Centroid Representation Network (MCRN), to address the unsupervised domain adaptive person re-ID task. To alleviate the impact of label noises, we propose a Multi-Centroid Memory (MCM) to capture more identity information and select reliable positive/negative samples for each input query. In order to learn more discriminative feature representation, we propose Domain-Specific Contrastive Loss (DSCL) to fully explore intra-domain information and Second-Order Nearest Interpolation (SONI) to enrich informative hard negative samples for the query from the target domain. Extensive experiments have demonstrated the effectiveness of our framework.

Acknowledgements

This work was supported by the National Key R&D Plan of the Ministry of Science and Technology (Project No. 2020AAA0104400), the Project of the National Natural Science Foundation of China No. 61876210, the Fundamental Research Funds for the Central Universities No.2019kfyXKJC024, and the 111 Project on Computational Intelligence and Intelligent Control under Grant B18024.

References

Bai, Z.; Wang, Z.; Wang, J.; Hu, D.; and Ding, E. 2021. Unsupervised Multi-Source Domain Adaptation for Person Re-Identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12914–12923.

Chang, W.-G.; You, T.; Seo, S.; Kwak, S.; and Han, B. 2019. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7354–7362.

Chen, H.; Wang, Y.; Lagadec, B.; Dantcheva, A.; and Bremond, F. 2021. Joint Generative and Contrastive Learning for Unsupervised Person Re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2004–2013.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607. PMLR.

Dai, Z.; Wang, G.; Zhu, S.; Yuan, W.; and Tan, P. 2021. Cluster Contrast for Unsupervised Person Re-Identification. *arXiv preprint arXiv:2103.11568*.

Deng, W.; Zheng, L.; Ye, Q.; Kang, G.; Yang, Y.; and Jiao, J. 2018. 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*, 994–1003.

Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X.; et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, 226–231.

Ge, Y.; Chen, D.; and Li, H. 2020. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. *arXiv preprint arXiv:2001.01526*.

Ge, Y.; Zhu, F.; Chen, D.; Zhao, R.; and Li, h. 2020. Self-paced Contrastive Learning with Hybrid Memory for Domain Adaptive Object Re-ID. In *Advances in Neural Information Processing Systems*, volume 33, 11309–11321. Curran Associates, Inc.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9729–9738.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Kalantidis, Y.; Sariyildiz, M. B.; Pion, N.; Weinzaepfel, P.; and Larlus, D. 2020. Hard negative mixing for contrastive learning. *arXiv preprint arXiv:2010.01028*.

Ko, B.; and Gu, G. 2020. Embedding expansion: Augmentation in embedding space for deep metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7255–7264.

Kuhn, H. W. 1955. The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2): 83–97.

Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32: 8026–8037.

Ristani, E.; Solera, F.; Zou, R.; Cucchiara, R.; and Tomasi, C. 2016. Performance measures and a data set for multi-target, multi-camera tracking. In *European conference on computer vision*, 17–35. Springer.

Tian, Y.; Krishnan, D.; and Isola, P. 2020. Contrastive multiview coding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, 776–794. Springer.

Wei, L.; Zhang, S.; Gao, W.; and Tian, Q. 2018. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 79–88.

Zhai, Y.; Lu, S.; Ye, Q.; Shan, X.; Chen, J.; Ji, R.; and Tian, Y. 2020a. Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9021–9030.

Zhai, Y.; Ye, Q.; Lu, S.; Jia, M.; Ji, R.; and Tian, Y. 2020b. Multiple expert brainstorming for domain adaptive person re-identification. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*, 594–611. Springer.

Zhang, M.; Liu, K.; Li, Y.; Guo, S.; Duan, H.; Long, Y.; and Jin, Y. 2021a. Unsupervised Domain Adaptation for Person Re-identification via Heterogeneous Graph Alignment. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(4): 3360–3368.

Zhang, X.; Ge, Y.; Qiao, Y.; and Li, H. 2021b. Refining Pseudo Labels with Clustering Consensus over Generations for Unsupervised Object Re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3436–3445.

Zhao, F.; Liao, S.; Xie, G.-S.; Zhao, J.; Zhang, K.; and Shao, L. 2020. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. In *European Conference on Computer Vision*, 526–544. Springer.

Zheng, K.; Lan, C.; Zeng, W.; Zhang, Z.; and Zha, Z.-J. 2021a. Exploiting Sample Uncertainty for Domain Adaptive

Person Re-Identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(4): 3538–3546.

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

Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; and Tian, Q. 2015. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, 1116–1124.

Zhong, Z.; Fini, E.; Roy, S.; Luo, Z.; Ricci, E.; and Sebe, N. 2021. Neighborhood Contrastive Learning for Novel Class Discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10867–10875.

Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; and Yang, Y. 2020. Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 13001–13008.

Zou, Y.; Yang, X.; Yu, Z.; Kumar, B. V.; and Kautz, J. 2020. Joint disentangling and adaptation for cross-domain person re-identification. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, 87–104. Springer.