Uncertainty-Aware Clustering for Unsupervised Domain Adaptive Object Re-Identification

Pengfei Wang, Student Member, IEEE, Changxing Ding, Member, IEEE, Wentao Tan, Mingming Gong, Member, IEEE, Kui Jia, Member, IEEE, and Dacheng Tao, Fellow, IEEE

Abstract—Unsupervised Domain Adaptive (UDA) object re-identification (Re-ID) aims at adapting a model trained on a labeled source domain to an unlabeled target domain. State-of-the-art object Re-ID approaches adopt clustering algorithms to generate pseudo-labels for the unlabeled target domain. However, the inevitable label noise caused by the clustering procedure significantly degrades the discriminative power of Re-ID model. To address this problem, we propose an uncertainty-aware clustering framework (UCF) for UDA tasks. First, a novel hierarchical clustering scheme is proposed to promote clustering quality. Second, an uncertainty-aware collaborative instance selection method is introduced to select images with reliable labels for model training. Combining both techniques effectively reduces the impact of noisy labels. In addition, we introduce a strong baseline that features a compact contrastive loss. Our UCF method consistently achieves state-of-the-art performance in multiple UDA tasks for object Re-ID, and significantly reduces the performance gap between unsupervised and supervised Re-ID. In particular, the performance of our unsupervised UCF method in the MSMT17→Market1501 task is better than that of the fully supervised setting on Market1501. The code of UCF is available at https://github.com/Wang-pengfei/UCF.

Index Terms—Domain adaptation, object re-identification, unsupervised learning.

I. INTRODUCTION

The goal of object re-identification (Re-ID) is to retrieve object images belonging to the same identity across different camera views. Due to its broad range of potential applications, (e.g., smart retail), Re-ID research has experienced explosive growth in recent years [1]–[15]. Most existing approaches achieve remarkable performance when the training and testing data are drawn from the same domain. However, due to the presence of significant domain gaps, Re-ID models trained on source datasets typically exhibit clear performance drops when directly applied to the target datasets. Unsupervised Domain Adaptive (UDA) object Re-ID is therefore proposed to adapt the model trained on the source image domain with identity labels to the target image domain without the need for identity annotations. Unlike the traditional UDA setting, which assumes that both domains share the same classes, UDA in object Re-ID is a more challenging open-set problem, in that the two domains have totally different identities (classes).

State-of-the-art methods [16]–[24] adopt clustering algorithms to generate pseudo-labels for the target domain. At the beginning of each epoch, a clustering algorithm is applied on the features extracted from the current model to generate pseudo-labels for each image. The current model is then updated via re-training with the pseudo-labels. These two steps alternate so that the model gradually adapts to the target data. While pseudo-label approaches have achieved promising results, there are still two major challenges to deal with. First, due to the domain gap, the current model is not an optimal feature extractor for the target domain; second, the unsupervised nature of the clustering makes it difficult to obtain the real identity labels, even given the optimal feature extractor. The obtained pseudo-labels therefore usually contain a certain level of noise, which undermines the final Re-ID performance.

In this paper, we propose an uncertainty-aware clustering framework (UCF) to handle the above problem from two perspectives. First, we identify and decompose unreliable clusters using a novel hierarchical clustering algorithm. Due to the domain shift, the Re-ID model has limited discriminative power in the target domain; as a result, inter-class distances may vary dramatically. This means that images of visually similar identities may be grouped into the same cluster, the size of which tends to be large. To handle this problem, we first adopt a clustering algorithm, such as DBSCAN [25], to perform coarse clustering. We then calculate the silhouette coefficient [26], which measures both the tightness and separation of each cluster. For clusters with small silhouette coefficient, we further perform fine-grained clustering within the cluster. In this way, unreliable clusters can be decomposed into several smaller ones.

Second, we identify images with unreliable pseudo-labels using a novel uncertainty-aware collaborative instance selection method. Specifically, we adopt a deep network and its...
temporally averaged model, i.e., the mean-Net [27], to cluster images in the target domain, respectively. Since these two models have different learning capabilities, their clustering results will be different. We then evaluate whether each instance is located in similar clusters across the two networks. If a large number of overlapping samples exist in the two clusters, the clustering result of this instance is considered to be reliable. Finally, we only adopt instances with reliable labels for model training, which reduces the impact of noise in the pseudo-labels.

Through joint hierarchical clustering and reliable sample selection, our UCF framework can effectively reduce the adverse effects of noisy pseudo-labels. We further propose a compact contrastive loss for UDA Re-ID. Recent approaches [23], [24], [28] typically adopt contrastive loss for model training. However, these losses require all image features in the target domain to be stored in the memory bank. This may result in two problems: first, this strategy consumes a lot of memory; second, only the features of a small number of images are updated in each iteration. These problems become especially serious for large-scale Re-ID datasets, such as MSMT17 [29]. To solve the above mentioned problems, we propose an improved contrastive loss using a class-level memory bank, which stores one single feature vector for each class rather than the features of all images.

Our main contributions can be summarized as follows: 1) We propose a strong baseline that adopts an improved contrastive loss using compact class-level memory banks; 2) We design a hierarchical clustering scheme to improve the quality of clustering, which decomposes unreliable clusters from coarse to fine; 3) We introduce a novel collaborative clustering method to identify images with unreliable pseudo-labels, which significantly relieves the impact of noise in pseudo-labels; 4) Our approach outperforms state-of-the-art methods by large margins on many UDA tasks for Re-ID.

The remainder of this paper is organized as follows. We first review the related works in Section II. Then, we describe the proposed UCF in more detail in Section III. Extensive experimental results on three benchmarks are reported and analyzed in Section IV, after which the conclusions of the present work are outlined in Section V.

II. RELATED WORKS

We review the literature in three parts: 1) unsupervised domain adaptive (UDA) object Re-ID, 2) contrastive learning, and 3) deep learning with noisy labels.

A. UDA Object Re-ID

Existing UDA approaches for object Re-ID can be roughly divided into two categories: pseudo-label-based methods [16]–[21], [28], [30] and domain translation-based methods [29], [31]–[33]. Domain translation-based methods transfer labeled images in the source domain to the style of the target domain images, then use these transferred images and the inherited ground-truth labels for model training. However, a gap inevitably arises between the translated image and the real target domain image, which affects the performance of these approaches. Pseudo-label-based methods group unannotated images using clustering algorithms and then train the network with pseudo-labels generated by clustering. For example, Li et al. [23] employed both visual and temporal similarity cues to promote the quality of pseudo-labels. Some recent works [34], [35] introduced the co-teaching [36] framework into UDA object Re-ID. In order to make better use of training samples, [34] and [35] designed new co-teaching frameworks that mine possibly useful samples from the outliers after clustering. However, they ignore the noise in pseudo-labels for samples in clusters.

Recently, some methods have been proposed that attempt to solve the label noise problem. [37] models the uncertainty of the pseudo labels based on the prediction difference between the primary and auxiliary classifiers. It then sets small weights to samples with high uncertainty to reduce their impact. Different from [37], we model the uncertainty of the pseudo labels according to the consistency of collaborative clustering in the clustering process. Ge et al. [17] proposed generating more robust soft labels via mutual mean-teaching. However, the classifier trained with noisy labels forms the foundation for soft label generation, which hinders the improvement of Re-ID performance. Zhao et al. [22] introduced the Noise Resistible Mutual-Training (NRMT) approach, as shown in Fig. 1(b), which removes triplets that are considered to be unreliable. The reliability of a triplet is measured with reference to the distance between the features of the triplet samples extracted by two networks. Ge et al. [24] proposed the SpCL approach. As shown in Fig. 1(a), it regards all instances in one unreliable cluster as outliers with reference to their proposed reliability criterion. However, removing all images in a cluster may waste samples with reliable pseudo-labels. Motivated by the above observation, we propose an uncertainty-aware clustering framework (UCF) for UDA tasks. Our method is more fine-grained. It first improves the clustering quality, and then removes unreliable instances rather than the complete clusters or triplets.

B. Contrastive Learning

As a promising paradigm of unsupervised learning, contrastive learning has lately achieved state-of-the-art performance in unsupervised visual representation learning. Recently, contrastive learning methods combined with data augmentation strategies achieved great successes, such as SimCLR [38], MoCo [39], and BYOL [40]. These methods treat each instance as a class represented by a feature vector and data pairs are constructed through data augmentations. These methods
treat each instance as a class, which yields poor results for the domain adaptive object Re-ID task, because the intra- and inter-class similarity on the unlabeled target domain cannot be measured accurately. Some recent works [21], [23], [24], [28] have introduced improved contrastive loss to domain adaptation. For example, the SpCL approach [24] includes a unified contrastive loss, which jointly distinguishes source-domain classes, target-domain clusters, and un-clustered instances. One common drawback of these methods is the need to store all instance features, which requires a large amount of memory. Compared with the existing method [24], our proposed contrastive loss has two advantages. First, we only need to store class prototypes in the memory bank rather than storing features of all samples, meaning that our approach brings in less memory cost. Second, each feature in the memory bank can be updated frequently within one epoch, which enables more accurate contrastive loss computation.

C. Deep Learning With Noisy Labels

Many studies have attempted to effectively train deep neural networks in the presence of noisy labels for close-set classification problems. Some recent works have introduced a sample selection approach that selects data with reliable labels for training [41], [42]. Notably, the small loss trick, which regards samples with small training loss as clean, has demonstrated powerful ability. However, the small loss trick is not suitable to select clean samples in UDA object Re-ID task. This is because the number of target domain clusters (classes) changes through re-clustering during the training process. Moreover, recent studies suggest various ways in which additional performance gain can be achieved by maintaining two networks to avoid accumulating sampling bias [36], [43]. For example, co-teaching [36] works by training two deep models simultaneously, where each network selects the small-loss instances as reliable samples for the other one. These methods focus primarily on the close-set problems with pre-defined classes, which cannot be generalized to our open-set object Re-ID task with completely unknown classes on the target domain.

III. METHODOLOGY

In this section, we present the details of our uncertainty-aware clustering framework (UCF), which reduces the effects of the noisy pseudo-labels in clustering-based Unsupervised Domain Adaptation (UDA). Our key idea is to select samples with reliable pseudo-labels in the target domain for model training purposes. To this end, we propose hierarchical clustering and uncertainty-aware collaborative instance selection methods to reduce the adverse effects of noisy pseudo-labels, and therefore improves the ability of model to learn cross-domain discriminative representations. In addition, we propose a strong baseline with a new contrastive loss using compact class-level memory banks.

Formally, we denote the source domain data as $D_s = \{(x^s_i, y^s_i)_{i=1}^{N_s}\}$, where $x^s_i$ and $y^s_i$ denote the $i$-th training instance and its annotation, respectively. The target-domain data without ground-truth labels are denoted as $D_t = \{x_i\}_{i=1}^{N_t}$. $N_s$ and $N_t$ denote the sample size in the source and target domains, respectively.

A. Supervised Pre-Training for Source Domain

In the first stage of UCF, we train the Re-ID model $F(\cdot | \theta)$ with the labeled source dataset $D_s$ using the cross-entropy loss and the triplet loss [44]; here, $\theta$ denotes parameters of the deep network. The pre-trained Re-ID model has the basic discriminability for domain adaptation. Then we adopt this pre-trained network $F(\cdot | \theta)$ to extract the features of the target domain images. Following the existing clustering-based UDA methods [20], [23], [24], we use DBSCAN [25] and Jaccard distance to cluster the extracted features into $K$ clusters before each epoch. We consider each cluster as a class and assign the same pseudo label for the instances belonging to the same cluster.

B. Uncertainty-Aware Clustering Framework

1) Hierarchical Clustering: As explained in Section I, images of visually similar identities may be grouped into the same cluster, which introduces significant noise to the pseudo-labels. Different from [45] that proposed two measures to evaluate the reliability of individual samples, we measure the reliability of clusters. We then identify unreliable clusters based on such uncertainty. Moreover, we also propose a hierarchical clustering (HC) method that conducts fine-grained clustering in these unreliable clusters to improve the clustering quality in the following.

Intuitively, a reliable cluster should be compact and independent from other clusters. This means that the distances between instances in the same cluster should be small, and the distance between different clusters should be large. To measure the reliability of one cluster, we first calculate the silhouette coefficient [26] for each of its instances. Specifically, the silhouette coefficient for the $i$-th instance in the $k$-th cluster is formulated as follows:

$$S(f_i^k) = \frac{b(f_i^k) - a(f_i^k)}{\max\{a(f_i^k), b(f_i^k)\}} \in [-1, 1],$$

where $f_i^k$ denotes the feature of the instance, $a(f_i^k)$ represents the average distance between the $i$-th instance and all the other instances in the $k$-th cluster. Moreover, $b(f_i^k)$ represents the average distance between the instance and all instances in the nearest cluster, which can be calculated as follows:

$$a(f_i^k) = \frac{1}{|I_k| - 1} \sum d_j(f_i^k, f_j^k), f_j^k \in I_k, i \neq j,$$

$$b(f_i^k) = \min_{k \neq l} \left\{ \frac{1}{|I_l|} \sum d_j(f_i^k, f_j^l) \right\}, f_j^l \in I_l,$$

where $d_j(\cdot, \cdot)$ represents the Jaccard distance, $I_k$ denotes the set of samples belonging to the $k$-th cluster, and $| \cdot |$ denotes the number of features in a cluster. Since the Jaccard distance between each pair of samples has been calculated during DBSCAN clustering, this step hardly increases time consumption. Finally, we calculate the average silhouette coefficient for the
When $S(I_k) < 0$, the intra-class distance surpass the inter-class distance. This usually indicates unreliable clustering from an object Re-ID perspective. We adopt a threshold of $\alpha$ to select these unreliable clusters. As shown in Fig. 2(a), we do not change the reliable ($S(I_k) > \alpha$) clusters, but we decompose an unreliable cluster into several smaller ones. In more detail, we use DBSCAN with the maximum neighbor distance $d$ for coarse clustering and then measure the reliability of each cluster. Within each unreliable cluster, we use DBSCAN with the maximum neighbor distance of $2/\beta d$ for fine-grained clustering. Since the number of samples in each unreliable cluster is limited, this step only adds a small amount of time consumption.

2) Uncertainty-Aware Collaborative Instance Selection: Although hierarchical clustering improves the quality of clustering, there are still inevitably noisy pseudo-labels in many clusters. Some existing works \cite{46, 47} introduced the uncertainty to evaluate the reliability or quality of a sample. They then incorporate the uncertainty to re-weight the loss of each sample. They model the uncertainty based on the class predictions in the training stage. Different from these methods, we model the uncertainty of pseudo labels based on the consistency of collaborative clustering in the clustering process. In order to identify individual instances with noisy pseudo-labels, we propose an uncertainty-aware collaborative instance selection (UCIS) method, which adopts a deep network and its temporally averaged (mean-Net) \cite{27} model to cluster the samples in the target domain separately. The parameters of the two models at iteration $T$ are denoted as $\theta$ and $E^{(T)}[\theta]$, respectively. $E^{(T)}[\theta]$ is obtained as follows:

$$E^{(T)}[\theta] = \sigma E^{(T-1)}[\theta] + (1 - \sigma) \theta,$$

where $\sigma$ is a temporal ensemble momentum coefficient whose value is within the range $[0, 1)$.

After hierarchical clustering, we obtain the fine-grained clustering results of the two models. We regard the clustering result of one instance as reliable if it is located in two similar clusters across the two models. The similarity of the two clusters is evaluated according to their overlap. More specifically, we propose the following metric to measure the clustering uncertainty of one instance $x_i$:

$$U(x_i) = \frac{|I_k \cap I_{\text{mean}}|}{|I_k|},$$

where $I_k$ and $I_{\text{mean}}$ denote the clusters containing $x_i$ by the two models, respectively. A larger $U(x_i)$ indicates larger overlap between $I_k$ and $I_{\text{mean}}$.

The value of $U(x_i)$ is therefore able to reflect the reliability of the pseudo-label for $x_i$. We set $\beta \in [0, 1]$ as a threshold to select instances with reliable pseudo labels. As shown in Fig. 2(b), in each epoch, we only preserve instances for model training where $U(x_i)$ is larger than $\beta$. The network gradually gains more discriminative power as training proceeds, which gradually improves the clustering quality. As a result, the number of discarded outliers decreases dramatically during the training process.

Here we adopt mean-Net to select instances with reliable pseudo labels in offline clustering. Some methods train two networks together for close-set UDA problems \cite{36}, where the two networks select reliable samples for each other. This strategy may not work well in our framework. This is because UCF selects samples deemed reliable by both networks based on uncertainty, which requires the two networks to have different discriminative power. However, as empirically proved in \cite{17}, the two networks will obtain similar discriminative power if they are trained with exactly the same supervision signals. Therefore, mean-Net is a better choice in our framework.

C. A Strong Baseline for Clustering-Based UDA

Fig. 3 illustrates the structure of our method. Aside from the commonly used cross-entropy loss, we propose to use the following contrastive loss. Given the feature $f$ of one target domain instance, our proposed contrastive loss is formulated as follows:

$$L_c^i(\theta) = -\log \frac{\exp((f, c^+)/\tau)}{\sum_{k=1}^{K} \exp((f, c_k)/\tau)},$$

where $c^+$ stands for the positive class prototype corresponding to $f$, $\tau$ is a temperature factor, and $\langle \cdot, \cdot \rangle$ denotes the inner product between two feature vectors. The loss value is low when $f$ is similar to $c^+$ and dissimilar to all the other class prototypes.

1) Memory Initialization: Each cluster is regarded as one class. The class-level memory bank $\{c_1, \ldots, c_K\}$ is initialized with the mean feature of each cluster. Note that a stop-gradient
Fig. 3. Model architecture of UCF in the training stage. UCF adopts a deep network and its temporally averaged model, i.e., the mean-Net, to cluster images in the target domain. After that, the proposed novel hierarchical clustering scheme and the uncertainty-aware collaborative instance selection method are used to select images with reliable labels for model training. This effectively reduces the impact of noisy labels. Step (a) and step (b) are performed alternately. Note that the parameters of mean-Net model are not updated during back propagation. In the testing stage, the mean-Net model is adopted for inference. (Best viewed in color.)

### TABLE I

| Dataset         | # type | # train IDs | # train images | # test IDs | # query images | # cameras | # total images |
|-----------------|--------|-------------|----------------|------------|---------------|-----------|---------------|
| Market-1501     | real   | 751         | 12,936         | 750        | 3,368         | 6         | 32,217        |
| CUHK03          | real   | 767         | 7365           | 700        | 1,400         | 2         | 14,097        |
| MSMT17          | real   | 1,041       | 32,621         | 3,060      | 11,659        | 15        | 126,441       |
| PersonX         | synthetic | 410       | 9,840           | 856        | 5,136         | 6         | 45,792        |
| VeRi-776        | real   | 575         | 37,746         | 200        | 1,678         | 20        | 51,003        |
| VehicleID       | real   | 13,164      | 113,346        | 800        | 5,693         | -         | 221,763       |
| VehicleX        | synthetic | 1,362       | 192,150        | -          | -             | 11        | 192,150       |

IV. EXPERIMENTS

#### A. Datasets and Evaluation Protocol

Following [24], we conduct extensive experiments on multiple large-scale Re-ID benchmarks, including two real-world person datasets and one synthetic person dataset, as well as two real-world vehicle datasets and one synthetic vehicle dataset. We evaluate our proposed method on both the mainstream real → real adaptation tasks and the more challenging synthetic → real adaptation tasks in person and vehicle Re-ID problems. The details of these datasets are summarized in Table I.

1) **Person Re-ID Datasets:** Market-1501 [50], CUHK03 [51], and MSMT17 [29] are real-world person image datasets that are widely used in domain adaptive tasks. MSMT17 includes more images that were captured in more challenging scenarios. The synthetic PersonX database [52] was constructed based on the Unity engine [56], [57] with manually designed challenges, including random occlusion, resolution and illumination changes.

2) **Vehicle Re-ID Datasets:** To verify the generalization ability of our method on different kinds of objects, we conduct experiments with the real-world VeRi-776 [53], VehicleID [54], and the synthetic VehicleX datasets. VehicleX [55] was also generated by the Unity engine [56], [57]. Moreover, this dataset adopted SPGAN [31] to translate the styles of the synthetic images to those of real-world images. The real-world vehicle data...
from VeRi-776 [53] and VehicleID [54] datasets were used to train SPGAN.

3) Evaluation Protocol: In our experiments, only ground-truth IDs of the source-domain datasets are provided for training. Experiments are conducted in line with the official evaluation protocol for each database. We adopt the widely used top-1/5/10 and mean Average Precision (mAP) as evaluation metrics. Moreover, following [17], [58], [59], the mean-Net is adopted for inference for both the baseline and our UCF method.

B. Implementation Details

We implement our framework in PyTorch [60]. We adopt ResNet-50 [61] as the backbone of the feature extractor and initialize the model with the parameters pre-trained on ImageNet [62]. After Layer4 of the ResNet-50 model, we add one Generalized-Mean (GeM) pooling [63] layer, one 1-D batch normalization [64] layer, and one L2-normalization layer. The L2-normalization layer produces 2048-dimensional feature vectors. Following [65], we perform data augmentation via random erasing, cropping, and flipping. For both source-domain pre-training and target-domain fine-tuning, we consistently construct a mini-batch with 64 person images of 16 identities. The person and vehicle images are resized to $256 \times 128$ and $224 \times 224$ pixels, respectively. To achieve faster convergence, we adopt embeddings of cluster centroids to initialize the weights of the classifiers. The momentum coefficients in (9) and (5) are set to 0.2 and 0.999, respectively. For DBSCAN, following [20], [23], [24], the hyper-parameter $d$ is set to 0.6 and the minimal number of neighbors in a core point is set to 4. Following [24], [28], the temperature $\tau$ in (7) is set as 0.05. The threshold $\alpha$ in hierarchical clustering is set to 0.0. The uncertainty threshold $\beta$ is set to 0.8. The ADAM method is adopted for optimization. The initial learning rate is set to 0.00035 and is decreased by multiplying by 0.1 on the 50-th epoch. The training lasts until the 80-th epoch.

C. Comparison With State-of-The-Art Methods

We compare the performance of UCF with state-of-the-art methods on multiple domain adaptation tasks, including real $\rightarrow$ real and more challenging synthetic $\rightarrow$ real tasks. The performance of these methods is tabulated in Table II, Table III, and Table IV, respectively. “Oracle” stands for the Re-ID performance in the fully supervised setting. It is clear that UCF significantly outperforms all state-of-the-art methods on both person and vehicle datasets with a plain ResNet-50 backbone.

1) Results on Real $\rightarrow$ Real UDA Person Re-ID Tasks: We compare the performance of UCF with state-of-the-art methods on six UDA settings in Table II. It is clear that UCF consistently outperforms existing approaches by large margins on all these benchmarks. In particular, UCF outperforms MMT [17] by 12.4%, 6.4%, 11.9%, 11.4%, 9.9%, and 8.2% in terms of mAP on these six tasks. It is worth noting that both UCF and MMT adopt mean-Net during the training stage. Moreover, UCF surpasses SpCL [24] by as much as 8.0% and 12.4% in terms of mAP and top-1 accuracy in the Market1501 $\rightarrow$ MSMT17 task, respectively. Finally, UCF significantly outperforms one very recent method named GLT [71] by 8.3% in terms of the mAP accuracy, for Market-1501 $\rightarrow$ MSMT17 task. The above experimental results clearly demonstrates the effectiveness of UCF. UCF also significantly bridges the gap between the unsupervised and fully-supervised settings. For example, UCF achieves 94.6% top-1 accuracy and 85.5% mAP on the MSMT17 $\rightarrow$ Market1501 task, meaning that it surpasses the performance of “Oracle” on the Market-1501 database by 0.5% in top-1 accuracy and 2.8% in mAP, respectively. In addition to the reliable pseudo labels generated by UCF, another possible reason is that MSMT17 is larger than Market1501; therefore, supervised pre-training on MSMT17 provides better model initialization before domain adaptation.

2) Results on Synthetic $\rightarrow$ Real UDA Person Re-ID Tasks: Compared with the real $\rightarrow$ real UDA re-ID tasks, the synthetic $\rightarrow$
real UDA tasks are usually more challenging due to the dramatic domain gap. As shown in Table III, UCF outperforms state-of-the-art methods by large margins. For example, UCF beats the SpCL [72] method by 4.1% in terms of top-1 accuracy and 6.7% in terms of mAP on the PersonX → Market-1501 task. It is also worth noting that the performance of UCF in synthetic → real tasks still exceeds that of SpCL in real → real tasks. Specifically, UCF achieves 92.1% top-1 accuracy and 80.5% mAP on the PersonX → Market1501 task, which outperform SpCL [72] on the MSMT17 → Market-1501 task by 2.4% in terms of top-1 accuracy and 3.0% in terms of mAP.

Although these results are promising, there is still a clear gap between UCF and “Oracle” on large-scale datasets such as MSMT17. This motivates us to develop more robust clustering and pseudo label generation methods in the future.

3) Results on Vehicle Re-ID Datasets: As Table IV shows, the performance of UCF surpasses that of SpCL by 4.8% in top-1 accuracy and 1.6% in mAP on the VehicleID → VeRi-776 task. Moreover, UCF outperforms SpCL by 3.1% in top-1 accuracy and 1.7% in mAP on the VehicleX → VeRi-776 task. These experimental results further demonstrate the effectiveness of UCF for object Re-ID.

D. Ablation Studies

We systematically investigate the effectiveness of each key component of UCF: namely, the strong baseline, hierarchical clustering (HC), and uncertainty-aware collaborative instance selection (UCIS), respectively. Experiments are conducted on real → real and more challenging synthetic → real adaptation tasks, specifically Market1501 → MSMT17 and PersonX → MSMT17. The results are summarized in Table V. “Source Pretrain” represents the Re-ID model trained in the source domain and tested directly in the target domain.

1) Effectiveness of the Strong Baseline: We build our baseline with the cross-entropy loss and our new contrastive loss, both of which are described in Section III-C. We first evaluate the performance when only classification loss or contrastive loss is used. As shown in Table V, the two settings achieve 28.5% and 26.8% mAP respectively for the Market1501 → MSMT17 task. In addition, as shown in Table VI, our new contrastive loss outperforms the conventional instance-level contrastive loss by 2.5% and 3.3% mAP on the two UDA tasks, respectively. When the two loss functions are used together, we obtain a strong baseline. For example, compared with “Source Pretrain” in Table V, our baseline promotes the top-1 accuracy by 39.1% and 46.2%, as well as mAP by 23.2% and 23.5%, on the two UDA tasks, respectively. These results prove that our baseline is simple but effective.

2) Effectiveness of the Hierarchical Clustering: Compared with our baseline, the hierarchical clustering method consistently yields performance gains. For example, “Baseline w/ HC” outperforms the baseline in terms of top-1 accuracy by 1.7% and 2.1%, as well as mAP by 23.2% and 23.5%, on the two UDA tasks, respectively. These results prove that our baseline is simple but effective.
3) **Effectiveness of the Uncertainty-Aware Collaborative Instance Selection:** When the baseline is equipped with the UCIS module, the performance of both UDA tasks is promoted. In particular, UCIS improves the top-1 accuracy of the baseline by 2.9% and 2.4%, as well as mAP by 1.7% and 1.4%, on the two tasks, respectively. The above results demonstrate the necessity of reducing the impact of noisy labels, as well as the effectiveness of our method.

4) **Effectiveness of the UCF Framework:** Finally, with both the HC and UCIS modules, our full model achieves better performance than using either of the modules alone. The above comparisons justify the effectiveness of each key component in our framework.

Furthermore, we test the performance of SpCL [72] based on our strong baseline. We equip SpCL with a hybrid memory to save target-domain cluster centroids and target-domain un-clustered instance features. Experimental results are summarized in Table VII. It is shown that UCF still outperforms SpCL by 2.6% and 1.5% in terms of top-1 accuracy and mAP on Market1501→MSMT17 task, respectively. The above experimental results justify the effectiveness of each key component in our framework.

5) **Ablation Studies on Other Clustering Algorithms:** First, we repeat the ablation studies using K-means as the clustering method. The optimal number of clusters is 1000. Since our hierarchical clustering method is designed for DBSCAN and cannot be applied to K-means, we perform ablation studies on the Uncertainty-aware Collaborative Instance Selection (UCIS) only. The experimental results for the Market1501→MSMT17 task are summarized in Table VIII. It is shown that UCIS achieves significantly better performance than our baseline.

Second, we conduct ablation studies using HDBSCAN on the Market1501→MSMT17 task. As shown in Table IX, UCIS consistently outperforms the baseline model. These experimental results further demonstrate the effectiveness of UCIS for object Re-ID.

6) **Performance Comparison Between Different Baselines:** We compare the performance of these two baselines. As shown in Table X, our proposed contrastive loss boosts the top-1 accuracy by 4.2% and the mAP by 3.1% on Market1501→MSMT17 task. In contrast, the triplet loss brings in a small promotion in performance. The result on triplet loss is consistent with that in [17].

7) **Comparisons in Computational Cost:** We compare the computational cost of UCF with two state-of-the-art methods, namely, SpCL [72] and MMT [17]. Comparisons are conducted on four Titan V GPUs. To facilitate fair comparison, all the other experimental settings are consistent with the descriptions in their original papers. The results are summarized in Table XI.

### TABLE VI

**Performance Comparison Between Our Class-Level Contrastive Loss and Instance-Level Contrastive Loss**

| Methods | Market1501→MSMT17 | PersonX→MSMT17 |
|---------|--------------------|-----------------|
|         | mAP | top-1 | top-5 | top-10 | mAP | top-1 | top-5 | top-10 |
| Instance-level contrastive loss | 24.3 | 52.5 | 64.1 | 69.1 | 19.1 | 45.3 | 56.2 | 61.4 |
| Class-level contrastive loss (ours) | 26.8 | 55.0 | 66.5 | 71.5 | 22.4 | 48.8 | 60.3 | 64.9 |

### TABLE VII

**Performance Comparison Between UCF and SpCL [24] With Our Strong Baseline**

| Methods | Market1501→MSMT17 | PersonX→MSMT17 |
|---------|--------------------|-----------------|
|         | mAP | top-1 | top-5 | top-10 | mAP | top-1 | top-5 | top-10 |
| Strong baseline | 31.6 | 61.7 | 72.3 | 76.4 | 26.2 | 55.0 | 67.0 | 71.6 |
| Strong baseline+SpCL [72] | 33.3 | 63.5 | 74.0 | 78.6 | 27.3 | 56.8 | 68.0 | 73.4 |
| Ours(full) | 34.8 | 66.1 | 76.6 | 80.6 | 28.3 | 58.2 | 69.7 | 74.3 |

### TABLE VIII

**Ablation Studies on Each Key Component of UCF Using the K-Means Clustering Algorithm**

| Methods | Market1501→MSMT17 |
|---------|--------------------|
|         | mAP | top-1 | top-5 | top-10 |
| Classification loss | 17.2 | 42.6 | 55.7 | 63.7 |
| Contrastive loss | 18.6 | 46.9 | 57.2 | 64.3 |
| Strong baseline | 24.7 | 55.9 | 66.0 | 71.7 |
| UCIS | 28.2 | 59.4 | 70.6 | 75.0 |

### TABLE IX

**Ablation Studies on Each Key Component of UCF Using the HDBSCAN Clustering**

| Methods | Market1501→MSMT17 |
|---------|--------------------|
|         | mAP | top-1 | top-5 | top-10 |
| Classification loss | 18.1 | 45.7 | 56.9 | 62.6 |
| Contrastive loss | 16.8 | 39.4 | 50.5 | 56.1 |
| Strong baseline | 21.2 | 48.7 | 60.6 | 65.2 |
| UCIS | 23.8 | 50.4 | 62.2 | 67.5 |

### TABLE X

**Performance of Different Baselines**

| Methods | Market1501→MSMT17 |
|---------|--------------------|
|         | mAP | top-1 | top-5 | top-10 |
| cross-entropy loss | 28.5 | 57.5 | 69.6 | 74.3 |
| +triplet loss | 29.0 | 57.8 | 70.2 | 74.9 |
| +contrastive loss | 31.6 | 61.7 | 72.3 | 76.4 |

### TABLE XI

**Comparisons in Computational Cost With SpCL [24] and MMT [17]**

| Methods | Market1501→MSMT17 |
|---------|--------------------|
|         | clustering | Tram/epoch | Total hours |
| SpCL [24] | 25.2s | 301.6s | 4.2h |
| MMT [17] | 40.4s | 345.6s | 4.8h |
| UCF | 52.4s | 145.8s | 3.2h |
TABLE XII
TRAINING STABILITY OF UCF ON THE MARKET-1501→MSMT17 TASK

| Methods | Market1501→MSMT17 | mAP | top-1 | top-5 | top-10 |
|---------|-------------------|-----|-------|-------|--------|
| UCF     | 34.25±0.86        | 65.6±1.28 | 76.2±1.03 | 80.9±0.74 |

Fig. 4. Comparisons on the Normalized Mutual Information (NMI) scores of clusters during the training process on the Market1501→MSMT17 task.

From Table XI, we can see that our clustering method costs more time than SpCL and MMT because of the hierarchical clustering and uncertainty-aware collaborative instance selection. However, our method has clear less overall training hours. This is mainly due to two reasons: First, we do not use triplet loss during training, which requires more training time compared with contrastive loss; second, our model converges faster than SpCL [72] and MMT [17].

8) The Training Stability of UCF: We repeat the experiments on the Market-1501→MSMT17 task for 10 times. The results are summarized in Table XII. The confidence interval is consistently less than 1.5%, which indicates that the performance of UCF is stable.

9) Analysis of the Quality of Pseudo Labels: In Fig. 4, we illustrate the improvement in the quality of pseudo labels. Following SpCL [24], we illustrate the Normalized Mutual Information (NMI) [73] scores of clusters during training on the Market1501→MSMT17 task. NMI [73] is an index that measures the accuracy of the clustering results. It can accordingly be observed that, compared with the baseline, the quality of the pseudo-labels is significantly improved when the proposed techniques are applied.

E. Parameter Analysis

We tune the hyper-parameters on the Market-1501→MSMT17 task, then directly apply the chosen hyper-parameters to all the other tasks.

1) Maximum Neighbor Distance $d$ for DBSCAN: DBSCAN is one of the most popular clustering algorithms in the UDA Re-ID literature. For DBSCAN, the maximum neighborhood distance $d$ is an important hyperparameter. As demonstrated in Fig. 5, we find that the value of $d$ may considerably affect the performance of state-of-the-art methods. In particular, a larger value of $d$ may result in a dramatic performance drop; this is because the pseudo-labels will contain more noise as the value of $d$ increases. In comparison, the performance of UCF is significantly more robust. This is because UCF successfully improves the clustering quality and removes samples with unreliable pseudo-labels.

2) Cluster Reliability Threshold $\alpha$ for Hierarchical Clustering: $\alpha$ is a threshold on $S(I_k)$. According to the definition of $S(I_k)$, a negative value of $S(I_k)$ means that intra-class distance surpasses inter-class distance. This usually indicates unreliable clustering from an object Re-ID perspective. As demonstrated in Fig. 6, our framework achieves the optimal performance when $\alpha$ is set to 0.0 on the MSMT17→Market-1501 task, which is consistent with our above analysis. When $\alpha$ is larger than 0.0, the top-1 accuracy and mAP gradually decrease. This is because some reliable clusters will be forced to be decomposed, resulting in more noisy pseudo-labels and therefore performance degradation.

3) Uncertainty Threshold $\beta$ for Collaborative Instance Selection: As described in Section III, we require an uncertainty threshold $\beta$ to select samples with reliable pseudo-labels. In Fig. 7, we investigate the effect of different values of $\beta$. The performance of UCF is generally robust to the value of $\beta$ while the best performance is achieved when $\beta$ is set to 0.8. When $\beta$ is set
to a higher value such as 0.9, the performance of UCF decreases; this may be because some samples with reliable pseudo-labels are also discarded when the threshold is higher. The performance of UCF also reduces when $\beta$ is set to a smaller value, such as 0.6; this may be because samples with noisy pseudo-labels cannot be identified when the threshold is low.

F. Qualitative Comparisons

In Fig. 8, we utilize t-SNE [74] to visualize the clustering results by “Baseline,” “Baseline w/ HC,” “Baseline w/ UCIS,” and the “UCF” model for the Market-1501→MSMT17 task, respectively. We have the following observations.

First, as illustrated in Fig. 8(a), due to the limited discriminative power of the Re-ID model in the target domain, many visually similar images may be grouped into the same cluster. The size of such clusters is often large. Second, as illustrated in Fig. 8(b), when the proposed hierarchical clustering (HC) method is utilized, the unreliable clusters in Fig. 8(a) are decomposed into multiple smaller ones. Third, as shown in Fig. 8(c), the uncertainty-aware collaborative instance selection (UCIS) method identifies instances with unreliable pseudo labels, which are represented using the gray color in the figure. Finally, combining UCIS and HC can achieve the best clustering results, which proves that the two modules are complementary. The above visualization results are consistent with the results in the experimentation section.

V. Conclusion

In this work, we propose an uncertainty-aware clustering framework (UCF) to tackle the problem of noisy pseudo labels in clustering-based UDA object Re-ID tasks. UCF handles the label noise problem on two levels. First, a novel hierarchical clustering scheme is proposed to promote the clustering quality; second, an uncertainty-aware collaborative instance selection method is introduced to select images with reliable labels for model training. These two techniques significantly relieve the noise in pseudo-labels and consequently improve the quality of deep feature learning. Our UCF method significantly outperforms state-of-the-art object Re-ID methods on many domain adaptation tasks. Although these results are very promising, there is still a clear gap in performance between UCF and supervised learning on large-scale datasets such as MSMT17. In the future, we will develop more powerful and efficient clustering and pseudo label generation methods.

REFERENCES

[1] Z. Zhang, C. Lan, W. Zeng, and Z. Chen, “Densely semantically aligned person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 667–676.
[2] W. Li, X. Zhu, and S. Gong, “Harmonious attention network for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 2285–2294.
[3] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, “Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline),” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 480–496.
[4] J. Xu, R. Zhao, F. Zhu, H. Wang, and W. Ouyang, “Attention-aware compositional network for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 2119–2128.
[5] H. Liao et al., “A strong baseline and batch normalization neck for deep person re-identification,” IEEE Trans. Multimedia, vol. 22, no. 10, pp. 2597–2609, Oct. 2020.
[6] L. Wei, S. Zhang, H. Yao, W. Gao, and Q. Tian, “Glad: Global-local-alignment descriptor for scalable person re-identification,” IEEE Trans. Multimedia, vol. 21, no. 4, pp. 986–999, Apr. 2019.
[7] C. Ding, K. Wang, P. Wang, and D. Tao, “Multi-task learning with coarse priors for robust part-aware person re-identification,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 3, pp. 1474–1488, Mar. 2022.
[8] C. Yan et al., “Beyond triplet loss: Person re-identification with fine-grained difference-aware pairwise loss,” IEEE Trans. Multimedia, to be published, doi: 10.1109/TMM.2021.3095652.
[9] B. Jiang, X. Wang, A. Zheng, J. Tang, and B. Luo, “Ph-gcn: Person retrieval with part-based hierarchical graph convolutional network,” IEEE Trans. Multimedia, to be published, doi: 10.1109/TMM.2021.3095789.
[10] K. Wang, P. Wang, C. Ding, and D. Tao, “Batch coherence-driven network for part-aware person re-identification,” IEEE Trans. Image Process, vol. 30, pp. 3405–3418, 2021.
[11] X. Gong et al., “Lag-net: Multi-granularity network for person re-identification via local attention system,” IEEE Trans. Multimedia, vol. 24, pp. 217–229, 2022.
[12] C. Wan, Y. Wu, X. Tian, J. Huang, and X.-S. Hua, “Concentrated local part discovery with fine-grained part representation for person re-identification,” IEEE Trans. Multimedia, vol. 22, no. 6, pp. 1605–1618, Jun. 2020.
[13] C. Zhao et al., “Deep fusion feature representation learning with hard mining center-triplet loss for person re-identification,” IEEE Trans. Multimedia, vol. 22, no. 12, pp. 3180–3195, Dec. 2020.

[14] P. Zhang, J. Xu, Q. Wu, Y. Huang, and X. Ben, “Learning spatial-temporal representation for car walking tracklet for person re-identification in the wild,” IEEE Trans. Multimedia, vol. 23, pp. 3562–3576, 2021.

[15] Y. Xie, H. Wu, F. Shen, J. Zhu, and H. Zeng, “Object re-identification using teacher-like and light students,” in Proc. Brit. Mach. Vis. Conf., 2021.

[16] L. Song et al., “Unsupervised domain adaptive re-identification: Theory and practice,” Pattern Recognit., vol. 102, 2020, Art. no. 107173.

[17] Y. Ge, D. Chen, and H. Li, “Mutual mean-teaching: Pseudo label refinement for unsupervised domain adaptation on person re-identification,” in Proc. Int. Conf. Learn. Representation, 2020, pp. 1–15.

[18] Y. Zhai et al., “Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 9021–9030.

[19] Y. Fu et al., “Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification,” in Proc. IEEE Int. Conf. Comput. Vis., 2019, pp. 6112–6121.

[20] X. Zhang, J. Cao, C. Shen, and M. You, “Self-training with progressive augmentation for unsupervised cross-domain person re-identification,” in Proc. IEEE Int. Conf. Comput. Vis., 2019, pp. 8222–8231.

[21] D. Wang and S. Zhang, “Unsupervised person re-identification via multi-label classification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 10981–10990.

[22] F. Zhao et al., “Unsupervised domain adaptation with noise resistant mutual-training for person re-identification,” in Proc. Eur. Conf. Comput. Vis., 2020, pp. 1–14.

[23] J. Li and S. Zhang, “Joint visual and temporal consistency for unsupervised domain adaptive person re-identification,” in Proc. Eur. Conf. Comput. Vis., 2020, pp. 1–14.

[24] 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 Proc. Adv. Neural Inf. Process. Syst., to be published, doi: 10.1109/TMM.2021.3082687.

[25] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” J. Comput. Appl. Math., vol. 20, pp. 53–65, 1987.

[26] A. Tarvainen and H. Valpola, “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results,” in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 1195–1204.

[27] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, “Invariance matters: Exemplar memory for domain adaptive person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 598–607.

[28] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in Proc. Data. Mining Knowl. Discov., 1996, pp. 226–231.

[29] P. J. Rousseuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," J. Comput. Appl. Math., vol. 20, pp. 53-65, 1987.

[30] Z. Tang et al., “Pantri: Pose-aware multi-task learning for vehicle re-identification using highly randomized synthetic data,” in Proc. IEEE Int. Conf. Comput. Vis., 2019, pp. 211–220.

[31] Y. Xie, H. Wu, F. Shen, J. Zhu, and H. Zeng, “Object re-identification using highly randomized synthetic data,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 594–611.

[32] Y. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, “Momentum contrast for unsupervised visual representation learning,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 9729–9738.

[33] J.-B. Grill et al., “Bootstrap your own latent: A new approach to self-supervised learning,” in Proc. Adv. Neural Inf. Process. Syst., 2020, pp. 21271–21284, arXiv:2006.07733.

[34] X. Yu, Y. Yao, L. Zheng, and T. Gedeon, “Simulating content consistent vehicle datasets with attribute descent,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 608–617.

[35] Z. Yang, X. Yang, S. X. Yu, and D. Lin, “Unsupervised feature learning via non-parametric instance discrimination,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 3733–3742.

[36] A. Paszke et al., “Pytorch: An imperative style, high-performance deep learning library,” in Proc. PyTorch, 2014, pp. 152–159.

[37] X. Sun, W. Liu, T. Mei, and H. Ma, “A deep learning-based approach to progressive vehicle re-identification for urban surveillance,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 869–884.

[38] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in Proc. Int. Conf. Mach. Learn., 2020, pp. 1597–1607.

[39] K. He, Y. Fan, Y. Wu, S. Xie, and R. Girshick, “Momentum contrast for unsupervised visual representation learning,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 9729–9738.

[40] J. Deng et al., “Imagenet: A large-scale hierarchical image database,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2009, pp. 248–255.

[41] F. Radenovic G. Tolias, and O. Chum, “Fine-tuning cnn image retrieval with no human annotation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 7, pp. 1655–1668, Jul. 2019.

[42] J. Deng et al., “ImageNet: A large-scale hierarchical image database,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2009, pp. 248–255.

[43] E. Malach and S. Shalev-Shwartz, “Decoupling ‘when to update’ from ‘how to update’”, in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 960–970.

[44] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “Simple framework for contrastive learning of visual representations,” in Proc. Int. Conf. Mach. Learn., 2020, pp. 1597–1607.
Pengfei Wang (Student Member, IEEE) received the B.S. degree in 2019 from the South China University of Technology, Guangzhou, China, where he is currently working toward the M.S. degree with the School of Electronic and Information Engineering. His current research interests include computer vision, deep learning, and person re-identification.

Changxing Ding (Member, IEEE) received the Ph.D. degree from the University of Technology Sydney, Ultimo NSW, Australia, in 2016. He is currently a Professor with the School of Electronic and Information Engineering, South China University of Technology. He has authored or coauthored more than 30 articles in prominent journals and conferences, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (T-PAMI), the IEEE TRANSACTIONS ON IMAGE PROCESSING (T-IP), International Journal of Computer Vision (IJCV), CVPR, ECCV, and MICCAI. His research interests include visual relationship detection, person re-identification, and face recognition.

Wentao Tan received the bachelor’s degree in 2020 from the South China University of Technology, Guangzhou, China, where he is currently working toward the Postgraduate degree. His research focuses on person re-identification.

Mingming Gong (Member, IEEE) is currently a Lecturer (Assistant Professor) in data science with the School of Mathematics and Statistics, the University of Melbourne, Parkville VIC, Australia. His research interests include causal reasoning, machine learning, and computer vision. He has authored and coauthored more than 40 research papers on top venues, such as ICML, NeurIPS, ICLR, AISTATS, CVPR, ICCV, ECCV, IJCV, T-IP, T-NNLS, with more than ten oral/spotlight presentations. He was the recipient of the Discovery Early Career Researcher Award from Australian Research Council in 2020.

Kui Jia (Member, IEEE) received the Ph.D. degree in computer science from the Queen Mary University of London, London, U.K., in 2007. He was with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, Chinese University of Hong Kong, Hong Kong, Advanced Digital Science Center, University of Illinois at Urbana-Champaign, Champaign, IL, USA, and the University of Macau, Macau, China. He is currently a Professor with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China, and he is the Director of Geometric Perception and Intelligence Research Lab. His recent research interests include theoretical deep learning and its applications in vision and robotic problems, including deep learning of 3D data and deep transfer learning. He is an Associate Editor for the TIP and TSMC.

Dacheng Tao (Fellow, IEEE) is currently the Inaugural Director of the JD Explore Academy and a Senior Vice President of JD.com. He is also an Advisor and Chief Scientist of the Digital Science Institute, University of Sydney, Camperdown NSW, Australia. He mainly applies statistics and mathematics to artificial intelligence and data science, and his research is detailed in one monograph and more than 200 publications in prestigious journals and proceedings at leading conferences. He was the recipient of the 2015 Australian Scopus-Eureka Prize, 2018 IEEE ICDM Research Contributions Award, and 2021 IEEE Computer Society McCluskey Technical Achievement Award. He is a Fellow of the Australian Academy of Science, AAAS, and ACM.