Pattern Recognition 124 (2022) 108514

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Unsupervised cross-domain person re-identification by instance and distribution alignment☆

Xu Lan[a], Xiatian Zhu[b,*], Shaogang Gong[a]

[a] Queen Mary, University of London, London E1 4NS, UK
[b] Vision Semantics Limited, London E1 4NS, UK

ARTICLE INFO

Article history:
Received 23 August 2020
Revised 22 December 2021
Accepted 23 December 2021
Available online 27 December 2021

Keywords:
Unsupervise person re-identification
Domain adaptation

ABSTRACT

Most existing person re-identification (re-id) methods assume supervised model training on a *separate* large set of training samples from the target domain. While performing well in the training domain, such trained models are seldom generalisable to a new independent unsupervised target domain *without* further labelled training data from the target domain. To solve this scalability limitation, we develop a novel Hierarchical Unsupervised Domain Adaptation (HUDA) method. It can transfer labelled information of an existing dataset (a source domain) to an unlabelled target domain for unsupervised person re-id. Specifically, HUDA is designed to model jointly global distribution alignment and local instance alignment in a two-level hierarchy for discovering transferable source knowledge in unsupervised domain adaptation. Crucially, this approach aims to overcome the under-constrained learning problem of existing unsupervised domain adaptation methods. Extensive evaluations show the superiority of HUDA for unsupervised cross-domain person re-id over a wide variety of state-of-the-art methods on four re-id benchmarks: Market-1501, DukeMTMC, MSMT17 and CUHK03.

1. Introduction

Person re-identification (re-id) aims to match the identity of person bounding boxes captured by disjoint camera views [1]. Most existing re-id methods rely heavily on *supervised learning* [2–9], assuming that the model training and test data are drawn from the same camera network, i.e. the same domain. However, such trained models suffer from significant performance degradation when deployed to an unlabeled target domain due to the domain shift problem [10].

In reality, we often have *no* access to a large number of *manually* labelled matching person image pairs for every camera pair as required by supervised learning methods, in order to effectively learn a feature representation and a matching function for each camera pair. Such large human labelling is both costly and not always available, due to a quadratic number of camera pairs in each surveillance domain. Existing supervised learning methods have limited cross-domain usability. To overcome this limitation, a number of approaches have been proposed, including (1) hand-crafting features [11,12], (2) image adaptation (synthesis) [13–17], (3) feature adaptation [18–21], (4) unsupervised learning [22–24], and (5) joint feature adaptation and unsupervised learning [15,25–28].

In this study, we focus on the *feature adaptation* approach for unsupervised cross-domain person re-id. The key idea is to align *feature statistics* between source and target training data. In doing so, re-id discriminative knowledge from the labelled source data can be transferred into the unlabelled target data. Existing feature adaptation methods typically rely on cross-domain alignment of *global feature distributions* [19,20]. This however suffers from an *under-constrained optimisation* problem, yielding suboptimal re-id models. We address this issue by discovering transferable source knowledge at both the local *instance* and global *distribution* levels. This idea leads to a *Hierarchical Unsupervised Domain Adaptation* (HUDA) model. This is a non-trivial learning task due to the lack of direct correlations between source and target person identities. To solve this problem, we formulate a new cross-domain cross-class association learning algorithm.

We make three contributions in this study: **(1)** We propose a novel idea of exploring instance-wise localised source knowledge for unsupervised cross-domain person re-id. It addresses the limitations of existing global feature distribution adaptation based methods. To our best knowledge, this is the first attempt of leveraging instance level association between different classes in unsupervised *feature adaptation* across domains. **(2)** We formulate a *Hierarchical Unsupervised Domain Adaptation* (HUDA) method.

☆ Fully documented templates are available in the elsarticle package on CTAN.

* Corresponding author.
E-mail address: xiatian.zhu@surrey.ac.uk (X. Zhu).

https://doi.org/10.1016/j.patcog.2021.108514
0031-3203/

HUDA is designed particularly to discover both localised source knowledge at the instance level and the global feature distribution knowledge across domains in model learning. **(3)** We analyse the underlying feature representations required for domain adaptation model learning in the context of *closed-set* supervised learning (e.g. softmax cross-entropy loss) vs. *open-set* unsupervised learning (e.g. Maximum Mean Discrepancy) and interpret their roles in optimising *open-set and cross-class* person re-id. Extensive evaluations demonstrate the superiority of HUDA over a variety of state-of-the-art models for unsupervised cross-domain person re-id on four benchmarks: Market-1501 [8], DukeMTMC [13,29], MSMT17 [6], and CUHK03 [4].

2. Related work

Most existing person re-id methods require *supervised* learning on a large labelled training dataset collected for every camera pair [2,3,7–9,30]. They assume that the training and test data are sampled from the same domain and have limited cross-domain generalisation. As a result, they have poor scalability to large scale re-id deployments in real-world when a large labelled training set is unavailable. While reducing the labelling effort, semi-supervised learning [31,32] approaches still need some cross-camera pairwise labels which may not be available inherently.

Recently, unsupervised domain adaptation (UDA) methods have demonstrated increasing significance in solving cross-domain re-id deployments [6,14,15,19,20]. The existing UDA models fall into two categories: (1) image adaptation (synthesis) [13,14,16], and (2) feature adaptation [19,20]. The *first* approach is often built on Generative Adversarial Networks (GANs) [33]. The main idea is to transform the labelled source domain images into the style of the unlabelled target domain while attempting to preserve the person identity information. In doing so, the source class labels can be used for supervised learning on the synthetic imagery. The *second* approach adopts a global feature distribution alignment strategy. This assumes that the model discrimination is related to global feature distribution statistics. Representative methods for feature distribution alignment include [34–38]. They all aim at minimizing the distribution discrepancy between the source and target domain in a shared feature space. Specifically, Tzeng et al. [34] and Long et al. [35,39] minimize the Maximum Mean Discrepancy (MMD) metric to align the global distribution between source and target domain. Another useful metric to be minimized is the cross-domain feature covariance matrix [36]. Imposing manifold regularization along with MMD metric is also shown to be effective by preserving the neighboring structures of training data sets [38].

Conceptually, both feature and image adaptation approaches are based on global data distribution alignment, with the former using the images (pixels) and the latter using the feature representations. One of their common weaknesses is that they all suffer from a *highly under-constrained learning* problem. That is, both do not consider instance level alignment to enable explicit fine-grained source knowledge adaptation. Recently, CR-GAN [17] proposes a novel instance-guided context rendering scheme which transfers the person identities of source domain into diverse target domain contexts to enable supervised re-id model learning in the unlabelled target domain. This can be regarded as instance alignment in the image space. However, CR-GAN is unfriendly to be integrated with global feature distribution level alignment due to their complex dual conditional image generator scheme. The proposed HUDA addresses this limitation by formulating a unified model for simultaneous global (distribution alignment) and local (instance alignment) knowledge transfer and adaptation across domains.

Our experiments show clearly the added benefits from modelling both levels of knowledge adaptation between the labelled source and the unlabelled target domains. In comparison to UDA, unsupervised deep learning [22] provides an *orthogonal* strategy. It aims to self-mine re-id discriminative information from the unlabelled training data in the target domain. It is generally beneficial to model performance by combining different strategies, for instance, integrating feature adaptation with image generation [15,25] or unsupervised learning [26].

3. Unsupervised hierarchical adaptation

Problem statement. For unsupervised cross-domain person re-id, we have a *supervised* (labelled) source dataset (domain) $D^s = \{\boldsymbol{I}_i^s, y_i^s\}_{i=1}^{K^s}$, consisting of K^s person bounding box images $\boldsymbol{I}_i^s$ each with the corresponding *identity* label $y_i^s \in \mathcal{Y} = \{1, \cdots, K_{\text{id}}^s\}$, i.e. a total of K_{id}^s different persons in the source domain. Meanwhile, we assume a set $D^t = \{\boldsymbol{I}_i^t\}_{i=1}^{K^t}$ of K^t *unsupervised* (unlabelled) training data randomly sampled from the target domain with unknown and non-overlapping identity labels. Using D^t is for model domain adaptation. The ***goal*** is to learn a feature representation optimal for the unlabelled target domain ID class discrimination by transferring the identity discriminative information learned from a labelled source domain.

Approach overview. To solve the aforementioned problem, we present a *Hierarchical Unsupervised Domain Adaptation* (HUDA) model. It can jointly perform *global feature distribution alignment* and *local instance alignment* between the source and target domains by end-to-end deep learning. This is uniquely characterised by *more fine-grained* knowledge transfer during unsupervised domain adaptation. This is crucial for person re-id since a key objective is to capture subtle discrimination of different persons with high appearance similarity. A large number of pedestrians observed in open surveillance scenes can appear visually alike. Aligning only global distributions across domains is *incapable* of capturing critical fine-grained instance-level information which is significant for re-id. With a joint modelling, fine-grained instance alignment enriches global distribution alignment. This provides a stronger constraint for unsupervised domain adaptation in a two-level hierarchy, whilst addressing the under-constrained problem. An overview of HUDA is depicted in Fig. 1.

3.1. Person re-identification model

To build a re-id model $\boldsymbol{\theta}^{\text{tar}}$ (Fig. 1(c_1, c_2)), we use ResNet-50 [40] as backbone. We discard the last 1,000-dim fully-connected (FC) layer and add one FC layer (i.e., the classifier) with K_{id}^s-dim output. Given labelled *source* training data D^s, we train the model by a discriminative loss function $\mathcal{L}_{\text{re-id}} = \mathcal{L}_{\text{ce}} + \lambda_{\text{tri}}\mathcal{L}_{\text{tri}}$ where $\mathcal{L}_{\text{ce}}$ and $\mathcal{L}_{\text{tri}}$ denote the softmax Cross Entropy loss and the triplet loss, respectively. We empirically set the weight parameter $\lambda_{\text{tri}} = 0.3$.

Discussion. A trained re-id model by the above formulation is suitable *only* for the source domain deployment, therefore having limited generalisation. To adapt the model to an independent target domain, we perform unsupervised domain adaptation by a HUDA model. In HUDA, *unlabelled* target domain data are used as a bridge for transferring source domain knowledge. Our model consists of two parts: (1) global distribution alignment, and (2) local instance alignment.

3.2. Global distribution alignment

The Global Distribution Alignment (GDA) component of HUDA aims to adapt holistic statistical information between the source and target domains (Fig. 1(d)). Due to the disjoint nature of source and target identity classes (i.e. an open-set recognition setting), GDA seems improper and has been shown to be ineffective for generic open-set object classification [41,42]. Nonetheless, person

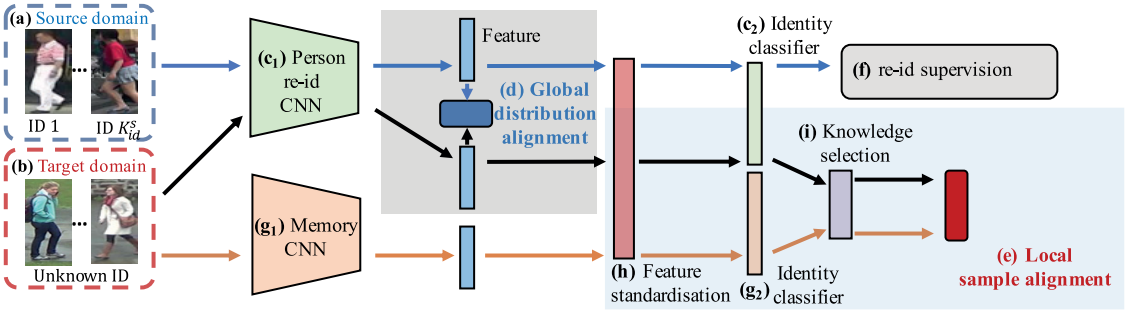


Fig. 1. Overview of *HUDA*.Given **(a)** *supervised* source domain and **(b)** *unlabelled* target training person imagery data, we aim to learn **(c_1, c_2)** a re-id model generalisable to the target domain. To this end, the proposed HUDA model jointly conducts**(d)** Global Distribution Alignment (GDA) and **(e)** Local Instance Alignment (LIA) in an end-to-end network learning architecture subject to **(f)** source re-id supervision. Cross-domain adaptation by the GDA alone is highly under-constrained. We address this by introducing the LIA for more fine-grained unsupervised domain adaptation with the stronger constraint. In re-id, there is often no identity class overlap between the source and target domains. Motivated by our primitive attribute viewpoint, we leverage cross-class association to discover and exploit *reliably transferable* knowledge for domain adaptation. This is achieved by the proposed LIA through incrementally building **(g_1, g_2)** a knowledge memory network to cumulatively memorise the past learned knowledge throughout training and simultaneously offer target domain instance-specific local knowledge for high quality adaptation from the labelled source domain to the unlabelled target domain. To further improve the knowledge quality, we introduce **(h)** a feature normalization layer to accelerate the model training and **(i)** a knowledge selection mechanism for more reliable domain adaptation.

re-id is rather different from generic object recognition, since it is a fine-grained matching problem.

A counter-intuitive phenomenon in re-id. Essentially, person re-id aims to derive a feature representation for pairwise similarity based matching and ranking. The training and testing person identity classes are totally *disjoint*. Such *cross-class* (i.e. open-set recognition) nature between training and testing is *universal* and *intrinsic* to the problem. Consider that the learning target is for optimal *pairwise matching*, early deep re-id models reasonably use *pairwise loss functions* (including the triplet ranking loss involving positive and negative pairs) for model training [4,43,44]. Subsequent works empirically find that the softmax Cross-Entropy (CE) loss, which is commonly used for training *closed-set* multi-class classification models, is similarly effective, even without the complexity of pairing samples [5]. This selection (presumably occasional) is actually *not* as intuitive as the pairwise counterparts, because the CE loss is conventionally considered effective *only* for *closed-set* recognition [45], so it would have been "*ineffective*" for cross-class learning as re-id. That being said, this traditional wisdom is *against* the wide practices. Interestingly, this counter-intuitive phenomenon lacks proper interpretation in the literature.

The essence to cross-class recognition in re-id. We provide an explanation to the above phenomenon as follows. By learning re-id feature representation for pairwise similarity matching, we consider the *fundamental key* is to derive *a set of primitive patterns (attributes)* which are formally composited of individual feature dimensions or some dimension combinations. They are useful to distinguish different person appearance and largely *independent* of any person identity classes including training classes. That is, these primitive attributes can describe arbitrary person appearance due to their massive combination space, which is the *essence* for them to possess cross-class recognition capability. Therefore, the essential learning objective is to obtain such a set of class independent primitive attributes, rather than a pairwise similarity matching function (previous understanding). Consequently, it is not necessarily to limit the learning objective to pairwise loss functions; The CE loss function can be similarly effective since the learning of classifiers also results in a set of primitive attributes optimal for multi-class discrimination. These loss functions are *functionally* similar in this primitive attribute viewpoint. This naturally interprets the *mysterious* efficacy of the CE loss for re-id.

Cross-domain in re-id. Unlike the generic object class classification with distinct appearance difference [41,42], person re-id handles uniquely fine-grained identity discrimination with similar holistic person appearance. This suggests that a large proportion of primitive attributes can be shared across domains, i.e. overlapped in the distribution. Specifically, the feature representations contain more primitive attributes shared over domains. Together with cross-class interpretation, GDA navigates cross-domain person re-id learning.

GDA formulation. Due to highly complex distributions of visually ambiguous and diverse re-id image data, it is difficult to select a suitable parametric model for such a distribution. We adopt a non-parametric representation to characterising re-id visual data statistics. In particular, we exploit the Maximum Mean Discrepancy (MMD) [46] to measure the feature dissimilarity between the source and target domains for distribution alignment:

$$\mathcal{L}_{\text{mmd}^2} = \left\| \frac{1}{n_s}\sum_{i=1}^{n_s}\phi(\boldsymbol{f}_{s,i}) - \frac{1}{n_t}\sum_{j=1}^{n_t}\phi(\boldsymbol{f}_{t,j}) \right\|^2_{\mathcal{H}} \tag{1}$$

where $\boldsymbol{f}_s \in \mathbb{R}^{n_s \times d}$ and $\boldsymbol{f}_t \in \mathbb{R}^{n_t \times d}$ specify the feature vectors of n_s source and n_t target images in each mini-batch, and d is the feature dimension. We further enforce non-linearity by using a mapping function $\phi(\cdot)$ to project the feature samples into a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ [47]. By the kernel trick, we design the GDA loss by reformulating Eq. (1) as:

$$\begin{aligned}\mathcal{L}_{\text{gda}} = &\frac{1}{n_s^2}\sum_{i=1}^{n_s}\sum_{i'=1}^{n_s} k(\boldsymbol{f}_{s,i}, \boldsymbol{f}_{s,i'}) \\ &+ \frac{1}{n_t^2}\sum_{j=1}^{n_t}\sum_{j'=1}^{n_t} k(\boldsymbol{f}_{t,j}, \boldsymbol{f}_{t,j'}) - \frac{2}{n_s n_t}\sum_{i=1}^{n_s}\sum_{j=1}^{n_t} k(\boldsymbol{f}_{s,i}, \boldsymbol{f}_{t,j})\end{aligned} \tag{2}$$

We adopt the common Gaussian kernel function:

$$k(\boldsymbol{f}_{s,i}, \boldsymbol{f}_{t,j}) = \exp\left(-\frac{||\boldsymbol{f}_{s,i} - \boldsymbol{f}_{t,j}||_2^2}{2\sigma^2}\right) \tag{3}$$

where σ is the kernel bandwidth. To reduce the selection bias and enable to automatically identify an optimal kernel, we deploy a predefined set of kernels with $\sigma \in \{1, 5, 10\}$.

3.3. Local instance alignment

To enrich GDA based cross-domain adaptation by cross-class discriminative learning necessary for person re-id, we further introduce Local Instance Alignment (LIA) to explore instance level fine-grained discriminative learning (Fig. 1(e)). Specifically, we

want to progressively discover and adapt *reliably transferable* source information specific to individual target samples during training. The key idea is learning to associate target samples with visually similar source data for guiding cross-domain knowledge transfer. The intuition is that, re-id of target instances can benefit ("borrow" information) from a model discriminatively trained by labelled source instances if the target and source instances are visually aligned (similar).

The association in LIA is often across identity classes between domains. Inspired by our primitive attribute viewpoint, we classify the target person images into the source identity classes. Specifically, given an unlabelled target person image sample $\boldsymbol{I}^t$, we predict a class probability vector for it in the source domain class-label space:

$$\boldsymbol{p}(\boldsymbol{I}^t) = \{p(1|\boldsymbol{I}^t), p(2|\boldsymbol{I}^t), \cdots, p(K^{\mathrm{s}}_{\mathrm{id}}|\boldsymbol{I}^t)\} \tag{4}$$

This classification indicates how visually similar a target person image is measured against all the source classes. It encodes the *cross-domain transferable knowledge* we aim to extract for unsupervised domain adaptation.

3.3.1. Source knowledge discovery

In a unified design, the source and target domain model learning shares a single network trained *simultaneously*. A faster training on the source data is essential for ensuring the knowledge quality. Consider deep learning using mini-batches of training samples as a stochastic learning process, the feature distribution changes per batch. This may complicate and slow down the unsupervised domain adaptation process, because the model needs to repeatedly and continuously adapt to new distributions throughout the training process.

Feature normalization. To address the above problem, we enforce that the model always outputs the feature representations in a fixed distribution. Specifically, we standardise the re-id feature representations (the average pooling of the last conv layer of ResNet-50). This performs a per-dimension normalisation on the per-batch feature vectors from both domains (Fig. 1(h)), as follows:

$$\hat{\boldsymbol{f}} = \frac{\boldsymbol{f} - \mathbb{E}[\boldsymbol{f}]}{\sqrt{\mathbb{V}[\boldsymbol{f}] + \epsilon}} \tag{5}$$

where $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ denote the per-dimension expectation and variance of feature values per batch. The small constant $\epsilon > 0$ is for ensuring numerical stability. Given this, we use the standardised features $\hat{\boldsymbol{f}}$ for re-id deployment in test.

Remarks Feature normalization has been used elsewhere, e.g. Sparsifying Features [48], and Batch Normalisation (BN) [49]. In this study, we investigate its potential for unsupervised domain adaptation in person re-id. The key differences are: Compared to BN that introduces two extra free parameters for scaling and shift in order to preserve the identity transform respectively, our method does not have such requirements. BN is used to normalise the layer inputs, whereas our model is applied to the model output. In contrast to [48], our method does not improve the feature sparsity nor constrain the internal layer outputs.

Knowledge memory network. To project a target instance into the source identity class space, a straightforward way is to apply the current up-to-date deep model. However, this is not ideal. The reason is as follows. In stochastic deep learning, the in-training model updates at each iteration. This may cause the model performance to temporally deteriorate on samples of the past mini-batches, due to the nature of *catastrophic forgetting* [50]. As target domain samples are randomly sampled, it is possible that the up-to-date model has degraded in recent updates when assessing some target samples of the current batch.

To further improve the knowledge quality, we propose to incrementally memorise the source information learned per mini-batch during training. In particular, we establish a *knowledge memory network* (Fig. 1(g_1, g_2)) $\boldsymbol{\theta}^{\mathrm{mem}}$ in identical architecture as the target model, and we exploit it to obtain the knowledge in the form of class posterior probability. Formally, this knowledge memory network $\boldsymbol{\theta}^{\mathrm{mem}}$ is updated along with the target model $\boldsymbol{\theta}^{\mathrm{tar}}$ at each iteration τ by exponential moving average as:

$$\boldsymbol{\theta}^{\mathrm{mem}}_{\tau} = \alpha\boldsymbol{\theta}^{\mathrm{mem}}_{\tau-1} + (1-\alpha)\boldsymbol{\theta}^{\mathrm{tar}}_{\tau} \tag{6}$$

where α is the smoothing coefficient hyper-parameter. We set $\alpha = 0.99$ empirically. In doing so, the discriminative information derived from each mini-batch is absorbed and memorised into $\boldsymbol{\theta}^{\mathrm{mem}}$, so that the memory model serves as a stronger knowledge extractor as compared to the up-to-date target model. That is, in mini-batch training we exploit the $\boldsymbol{\theta}^{\mathrm{mem}}_{\tau}$ as the replacement of $\boldsymbol{\theta}^{\mathrm{tar}}_{\tau}$ to obtain the posterior probability vector (Eq. (4)) for each unlabelled target sample in the source domain class space.

Remarks. The proposed memory network is inspired by the neuron memory mechanism [51]. This is due to that the memorising capacity of deep networks is often incomplete and limited in representing knowledge experienced in the past learning iterations. However, unlike [51], our method uses a network for memory organisation without the need for extra components to customise the network structure and designing particular knowledge representations for access operations. LSTM [52] is a family of deep models with a memory mechanism for learning sequential data. Nonetheless, it is not suited for our problem due to several reasons: (1) If we consider the iterative model update as a sequential process over training iterations, this will give a huge input dimension (*e.g.*, 4.6×10^7 CNN parameters) and many temporal steps (thousands of training mini-batches). Both challenge the ability of LSTM. (2) There is no ground-truth for training such a LSTM network in the re-id model parameter space. Algorithmically, building our memory network is similar to the notion of mean-teacher in semi-supervised learning [53], but the two address different goals. Our method seeks a reliable cross-class knowledge extraction in training. In contrast, mean-teacher aims to improve label prediction on unlabelled data from the same domain in a closed-set classification setting.

3.3.2. Source knowledge transfer

The aim of source knowledge transfer is to enhance the generalisation of the target model $\boldsymbol{\theta}^{\mathrm{tar}}$ in the target domain. To this end, we consider the richer memorised knowledge in the memory network that is relevant to target domain samples. However, the underlying transferable knowledge between source and target domains is *unknown* a priori. It is sub-optimal to blindly transfer all memory knowledge with all target samples. To address this, we design a knowledge selection mechanism (Fig. 1(i)) for more reliable adaptation on individual samples.

Knowledge selection. In unsupervised cross-domain re-id, not all target person images can be associated with some source identity classes with high confidence. This is due to the cross-class nature between independent domains with entirely different person classes. Given that source knowledge is expressed in a probability form, one intuitive way to measure the knowledge transferability and reliability is to use the maximum likelihood:

$$\mathcal{ML}(\boldsymbol{I}^t) = \max(\{p(1|\boldsymbol{I}^t), p(2|\boldsymbol{I}^t), \cdots, p(K^{\mathrm{s}}_{\mathrm{id}}|\boldsymbol{I}^t)\}) \tag{7}$$

With this, we can then deploy a thresholding strategy for knowledge selection by choosing those target samples satisfying that the corresponding $\mathcal{ML}(\boldsymbol{I}^t)$ exceeds a pre-defined threshold u. We denote the selected target samples as $\tilde{\boldsymbol{I}}^t$. In cross-class context, it is often that most $\mathcal{ML}(\boldsymbol{I}^t)$ values are not high. Hence, a mild threshold value is preferred to ensure sufficient source-target associations. Too small threshold values, on the other hand, may lead

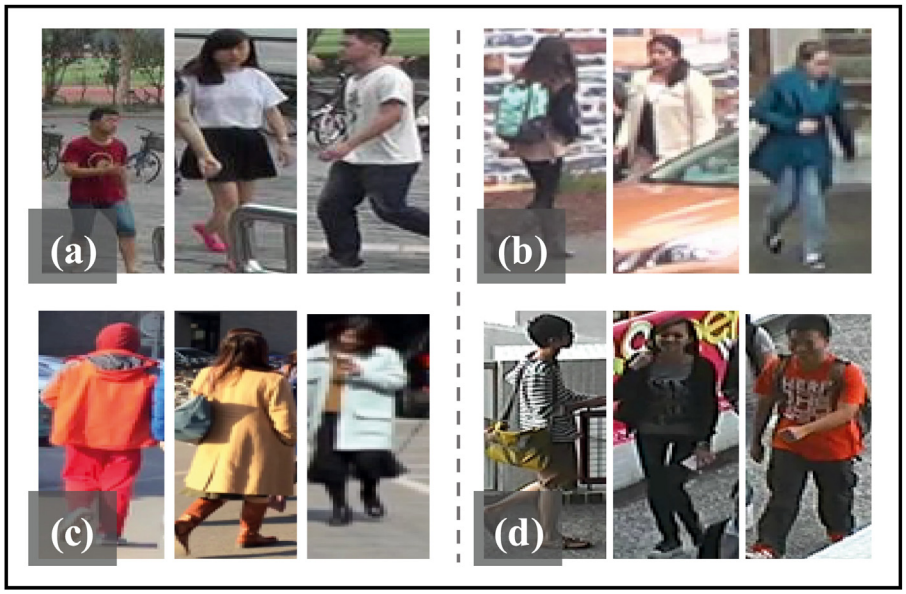


Fig. 2. Example person images from (a) Market-1501, (b) DukeMTMC, (c) CUHK03, (d) MSMT-17.

to adapting non-transferable knowledge with negative effects. We empirically find that setting $u = 0.3$ is satisfactory.

Knowledge transfer. Once we have the selected knowledge, the next is to transfer it into the target model, i.e. knowledge domain adaptation. To accomplish this, we align the knowledge memory model and the target model in their predictions of selected target samples $\tilde{\boldsymbol{I}}^t$ by exploiting the Kullback-Leibler (KL) divergence written as:

$$\mathcal{L}_{\text{lia}} = \sum_{j=1}^{K_{\text{id}}^s} p_(j|\tilde{\boldsymbol{I}}^t, \boldsymbol{\theta}^{\text{mem}}) \log \frac{p_(j|\tilde{\boldsymbol{I}}^t, \boldsymbol{\theta}^{\text{mem}})}{p_(j|\tilde{\boldsymbol{I}}^t, \boldsymbol{\theta}^{\text{tar}})} \tag{8}$$

3.4. Overall model loss formulation

Given the re-id and HUDA loss functions, we obtain the final objective function for model training as:

$$\mathcal{L} = \mathcal{L}_{\text{re-id}} + \lambda_{\text{gda}}\mathcal{L}_{\text{gda}} + \lambda_{\text{lia}}\mathcal{L}_{\text{lia}} \tag{9}$$

where λ_{gda} and λ_{lia} are the relative importance parameters. We set $\lambda_{\text{gda}} = 1$ and $\lambda_{\text{lia}} = 1$ in our experiments. The whole model can be trained end-to-end subject to the loss function of Eq. (9) by the stochastic gradient descent algorithm.

4. Experiments

Datasets. For evaluation, We used four person re-id benchmarks with distinct camera viewing conditions. (Fig. 2). The ***Market-1501*** [8] contains 32,668 images of 1501 identities (ID) captured by 6 cameras. We used the standard 751/750 train/test ID split. The ***DukeMTMC*** [13,29] consists of 36,411 labelled images of 1404 IDs from 8 camera views. We adopted the same 702/702 ID split as [13]. The ***CUHK03*** [4] provides 14,096 images of 1467 IDs from 6 camera views. We used the detected images as the source as [15]. The ***MSMT-17*** [6] is a largest person re-ID benchmark thus far. contains 126,411 person images from 4101 IDs captured from 15 camera views. We adopted the standard 1041/3060 train/test ID split.

Performance metrics. We adopted the Cumulative Matching Characteristic (CMC) and mean Average Precision (mAP) as the model performance measurements.

Model parameter setting. In this context, no target domain supervision is available for hyper-parameter cross-validation. We hence used a *single* set of empirical parameter setting for HUDA (including λ_{tri} for $\mathcal{L}_{\text{re-id}}$, α in Eq. (6), u for Eq. (7), λ_{gda} and λ_{lia} in Eq. (9)) in all the experiments.

Implementation details. We performed all the experiments in PyTorch [54]. We used ResNet-50 as person re-id c_1 and memory network g_1. TThe identity classifier c_2/g_2 consists of one fully-connection layer at the shape of $d \times K_{\text{id}}$, where d is the feature dimension and K_{id} is the number of training identity classes in the source domain. We used a triplet loss to enhance identity discriminative learning with the cross-entropy loss. To train a re-id model, we deployed SGD with the momentum set to 0.9, the weight decay to 0.0005, and the mini-batch size of 64 (32 source plus 32 target samples), the epoch number to 60. All input images were resized to 256 $\times$ 128 and subtracted by ImageNet mean. We applied data augmentation for the target and memory networks independently in training, including random cropping, random flipping, and colour jitter. In test time, we used the Euclidean distance as the re-id matching metric.

4.1. Comparisons to the state-of-the-art methods

For a fine-grained evaluation, we compared five types of existing methods: **(a)** two hand-crafted feature models (LOMO [7], BoW [8]); **(b)** four image adaptation models (PTGAN [6], SPGAN+LMP [14], ATNet [16]), CR-GAN[17]); **(c)** six feature adaptation models (UMDL [18], CAMEL [55], PUL [56], TJ-AIDL [19], MMFA [20], MAR [21]); **(d)** four unsupervised deep learning method (TAUDL [22], SSG [23], PCB-R-PAST [24], UDA [57]); **(e)** seven hybrid methods (HHL [15], ECN [25], PAUL [26]), MMT-500 [27], MMT-700 (IBN-ResNet-50) [27], PDA-Net [28], CR-GAN+TAUDL [17]), **(f)** one semi-supervised method (SSG++ [23]). We made three HUDA based hybrid models: (i) Taking TAUDL [22] as unsupervised learning, termed as HUDA+TAUDL, (ii) Further taking PCB [26] for part based classification as in PCB-R-PAST [24], termed as HUDA+TAUDL(PCB), (iii) Following MMT-700 (IBN-ResNet) [27] we use clustering driven unsupervised learning (*i.e.*, SSG [23]) to produce pseudo labels and adopt IBN-ResNet-50 as feature backbone, termed as HUDA+SSG. We evaluated three transfer scales in source data size: (1) large: MSMT17⇒Market, (2) medium: Market1501⇔DukeMTMC, (3) small: CUHK03⇒Market.

Table 1
Results on Market-1501⇔DukeMTMC.

Source→Target	Duke→Market				Market→ Duke			
Metric (%)	R1	R5	R10	mAP	R1	R5	R10	mAP
LOMO [7]	27.2	41.6	49.1	8.0	12.3	21.3	26.6	4.8
BOW [8]	35.8	52.4	60.3	14.8	17.1	28.8	34.9	8.3
PTGAN [6]	38.6	-	66.1	-	27.4	-	50.7	-
SPGAN+LMP [14]	57.7	75.8	82.4	26.7	46.4	62.3	68.0	26.2
ATNet [16]	55.7	73.2	79.4	25.6	45.1	59.5	64.2	24.9
CR-GAN[17]	64.5	79.8	85.0	33.2	56.0	70.5	74.6	33.3
TAUDL [22]	63.7	-	-	41.2	61.7	-	-	43.5
SSG [23]	80.0	90.0	92.4	58.3	69.3	80.2	83.1	53.4
PCB-R-PAST [24]	78.4	-	-	54.6	72.4	-	-	54.3
UDA [57]	75.8	89.5	93.2	53.7	68.4	80.1	83.5	49.0
SSG+[23]	86.2	94.6	96.5	68.7	76.0	85.8	89.3	60.3
UMDL [18]	34.5	52.6	59.6	12.4	18.5	31.4	37.6	7.3
CAMEL [55]	54.5	-	-	26.3	-	-	-	-
PUL [56]	45.5	60.7	66.7	20.5	30.0	43.4	48.5	16.4
TJ-AIDL [19]	58.2	74.8	81.1	26.5	44.3	59.6	65.0	23.0
MMFA [20]	56.7	75.0	81.8	27.4	45.3	59.8	66.3	24.7
HUDA (Ours)	68.5	82.9	87.1	37.6	52.3	65.4	68.7	30.2
HHL [15]	62.2	78.8	84.0	31.4	46.9	61.0	66.7	27.2
ECN [25]	75.1	87.6	91.6	43.0	63.3	75.8	80.4	40.4
MMT-500 [27]	86.8	94.6	96.9	71.2	78.0	88.8	92.5	65.1
MMT-700(IBN) [27]	91.1	96.5	98.2	74.5	**81.8**	91.2	93.4	68.7
PDA-Net [28]	75.2	86.3	90.2	47.6	63.2	77.0	82.5	45.1
CR-GAN+TAUDL [17]	77.7	89.7	92.7	54.0	68.9	80.2	84.7	48.6
HUDA+TAUDL [22]	78.8	90.2	93.4	57.6	70.4	82.5	86.2	51.2
HUDA+TAUDL (PCB) [22]	81.0	91.1	93.5	59.3	73.1	83.7	87.2	54.5
HUDA+SSG [57]	**91.4**	**96.7**	**98.5**	**74.7**	81.5	**91.5**	**93.7**	**69.0**

Table 2
Results on MSMT17/CUHK03⇒Market-1501.

S→T	MSMT→Market				S→T	CUHK→Market			
Metric(%)	R1	R5	R10	mAP	Metric(%)	R1	R5	R10	mAP
MAR	67.7	81.9	-	40.0	HHL	42.7	57.5	64.2	23.1
PAUL	68.5	82.4	87.4	40.1	SPGAN	42.3	-	-	19.0
HUDA	**72.3**	**85.2**	**89.2**	**42.4**	**HUDA**	**49.7**	**62.8**	**67.7**	**27.9**

Evaluation on DukeMTMC ⇔Market-1501. Table 1 shows the comparisons between HUDA and 22 state-of-the-art methods. We have the following observations. **(1)** Hand-crafted feature methods [7,8] produce the poorest performance, due to weak representations. **(2)** Image adaptation methods [6,14,15] yield fairly strong re-id rates. but weaker than the best feature adaptation counterparts, *e.g.*, HUDA. **(3)** Interestingly, unsupervised re-id methods(TAUDL [22], SSG [23], PCB-R-PAST [24], UDA [57]) achieve competitive performance without using any labelled source data. **(4)** For the feature adaptation models, HUDA outperforms all the competitors [18–20,55,56]. This suggests strongly the modelling superiority of our method over the state-of-the-art counterparts. **(5)** Unsupervised domain adaptation alone (*e.g.*, CR-GAN, HUDA) is clearly inferior than those hybrid models (MMT variants, PDA-Net), as expected. When integrated with unsupervised learning, our HUDA can reach the best overall results. This indicates the superior complementary of our model with previous unsupervised learning methods. **(6)** Some hybrid methods (MMT variants, HUDA+SSG) even surpass the semi-supervised learning method SSG++, showing the joint effectiveness of unsupervised learning and domain adaptation.

Evaluation on MSMT17/CUHK03 ⇒Market-1501. We further tested the domain adaptation with large and small scale transfer. Table 2 compares the performance of HUDA to 4 state-of-the-art alternative methods with reported re-id results. Overall, we have similar observation as above. For MSMT17⇒Market-1501, as a feature adaptation method, HUDA even surpasses the hybrid competitor PAUL. In the case of small scale transfer on CUHK03⇒Market-1501, HUDA consistently outperforms all strong competitors. This test validates the superiority of HUDA in varying cross-domain adaptation scenarios.

4.2. Further analysis and discussions

We conducted a series of component analysis for HUDA using DukeMTMC⇔Market-1501.

HUDA design. We tested the significance of HUDA and its components (GDA and LIA). Table 3 shows that: (1) Without HUDA, the model suffers clearly the domain gap, e.g. large performance drop. (2) GDA *Only* gives significant performance boost. This validates our *primitive attribute* interpretation. (3) LIA *Only* also yields similar re-id rate gain. This verifies the idea of our local alignment and the proposed design. (3) When GDA and LIA are jointly exploited (i.e. full HUDA), model performance is further increased. This validates good complementary of GDA and LIA, as well as our motivation of integrating them into a single formulation.

Cross-class association between domains. Recall that we classify the unlabelled target person images into the source identity class space in a cross-class manner. This aims to associate target persons with visually similar source people in the LIA process (see Fig. 4). We examined the effectiveness of this association. Specifically, we measured the proportion of target person images highly associated to any source identity classes with the maximum likelihood above the threshold u. We tracked this measurement *with* and *without* the LIA. We observed from Fig. 3 that, the proposed association scheme significantly improves the cross-domain alignment at the fine-grained instance level. LIA makes the most target persons associated to the relevant source identities with similar

Table 3
HUDA design analysis. GDA: Global Distribution Alignment. LIA: Local Instance Alignment.

Source→Target	Duke→ Market				Market→Duke			
Metric(%)	R1	R5	R10	mAP	R1	R5	R10	mAP
w/o HUDA	55.2	74.3	81.3	27.1	41.8	57.6	63.2	22.3
GDA Only	61.8	77.9	83.6	32.4	46.8	62.6	68.8	26.5
LIA Only	61.9	78.3	83.8	32.9	44.3	59.4	65.5	24.1
Full HUDA	**68.5**	**82.9**	**87.1**	**37.6**	**52.3**	**65.4**	**68.7**	**30.2**

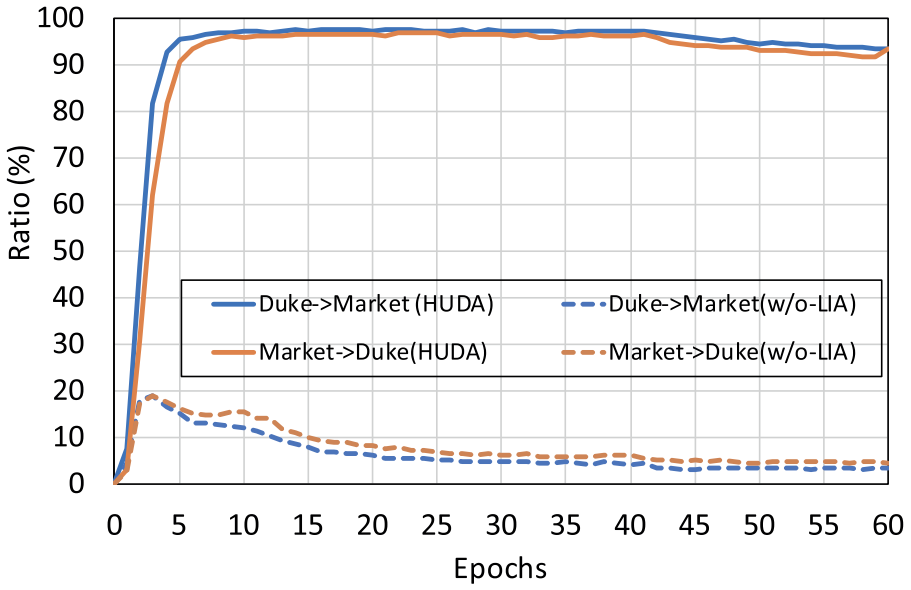


Fig. 3. The proportion of target training samples that is highly associated with source classes during model training.

Fig. 4. Association of *target* DukeMTMC persons to *source* Market-1501 identity classes. **(a)** The pairs of source and target persons extracted automatically by cross-domain cross-class association. The associated persons show strong visual similarities. **(b)** The target person images associated to a source person have either the *same* identity (when in the same domain) or *similar* visual appearance (when cross-domain). **(c)** Cross-domain associations can be distracted by background clutters.

Table 4
Examination of feature normalization (FN).

Source→Target		Duke→ Market				Market→Duke			
Model	FN	R1	R5	R10	mAP	R1	R5	R10	mAP
w/o HUDA	✗	55.2	74.3	**81.3**	27.1	41.8	57.6	63.2	22.3
w/o HUDA	√	**56.9**	**74.2**	80.1	**28.4**	**42.1**	**57.9**	**63.3**	**22.5**
HUDA	✗	61.5	77.2	82.9	32.3	44.5	57.6	64.0	24.6
HUDA	√	**68.5**	**82.9**	**87.1**	**37.6**	**52.3**	**65.4**	**68.7**	**30.2**

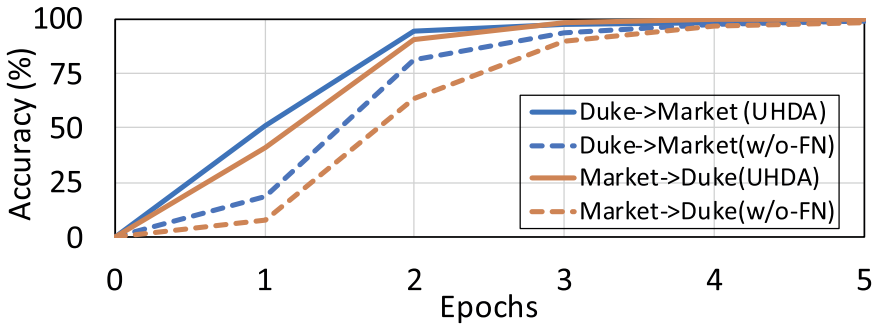


Fig. 5. Effect of the feature normalization (FN) to the model convergence on the source domain data.

Table 5
Examination of knowledge selection (KS).

Source→Target	Duke→ Market				Market→Duke			
KS	R1	R5	R10	mAP	R1	R5	R10	mAP
✗	65.5	79.1	84.7	34.7	48.3	63.5	67.9	27.5
√	**68.5**	**82.9**	**87.1**	**37.6**	**52.3**	**65.4**	**68.7**	**30.2**

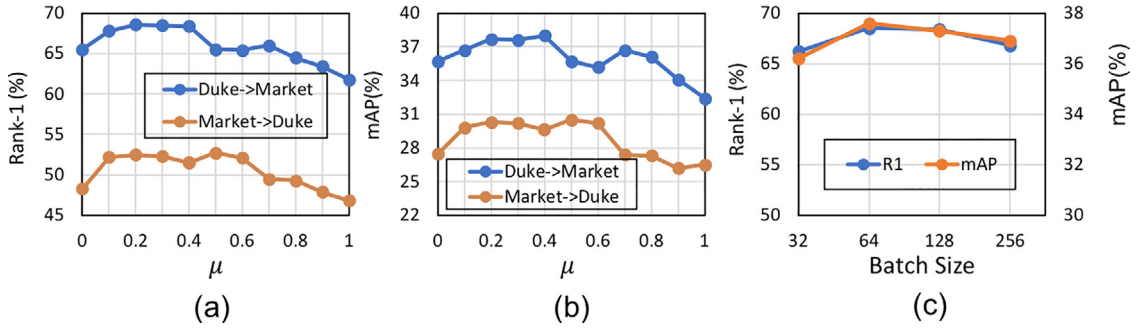


Fig. 6. Effect of controlling the knowledge reliability in cross-domain transfer in (a) Rank-1 and (b) mAP rates, and (c) the effect of batch size on Duke→ Market.

appearance. This indicates that GDA is *under-constrained*. Not every target sample can be associated with a visually similar source identity by HUDA. This is reasonable due to the independent nature between source and target domains. The rising association rate of HUDA *without* LIA in the beginning of training is due to inaccurate predictions by the *immature* in-training model.

Feature normalization. We evaluated the effect of feature normalization (FN) on unsupervised domain adaptation *with* and *without* HUDA. Table 4 shows that FN is significant for effective cross-domain knowledge transfer in HUDA context, validating our design consideration. This is because, the cross-domain association becomes reliable and effective for unsupervised domain adaptation, only when the model learns sufficiently discriminative information from the source labels. Besides, FN slightly helps the baseline without HUDA, suggesting a generic usefulness. We further tested the impact of FN on the model performance convergence on the source domain data. We chose the memory network that is used for knowledge extraction. Fig. 5 shows that FN is clearly beneficial for accelerating the model learning speed on the source labelled data.

Table 6
Domain adaptation (DA) effects on the source domain.

Dataset	Market				Duke			
Metric(%)	R1	R5	R10	mAP	R1	R5	R10	mAP
Before DA	86.6	94.7	97.0	67.5	**77.4**	**88.5**	**91.7**	**59.5**
SPGAN	59.9	78.7	84.5	34.3	53.9	70.9	76.5	32.4
HUDA	**87.0**	**95.0**	**97.1**	**67.8**	77.1	87.9	91.4	59.3

Knowledge selection. We tested the performance benefit from knowledge selection (KS). The KS is controlled by setting a threshold u on the maximum likelihood in the source class space (Eq. (7)). We compared the re-id accuracy rates on the target domain *with* and *without* the thresholding based (u) selection. Table 5 and Fig. 6(a,b) support the significance of knowledge selection for more reliable unsupervised domain adaptation. The optimal selections lie in the range of $[0.1, 0.4]$, validating our consideration that a mild threshold value u would be used. Note that not the *entire* ($u=0$) source knowledge are equally relevant and reliably transferable to the target domain. Adapting unsuitable source

information can hurt the model generalisation. Besides, the performance is clearly inferior when *no* local knowledge adaptation is considered ($u=1$), validating our modelling motivation.

Batch size. The mean embeddings of two probability distributions in MMD metric are calculated for the source and target domains respectively within each mini-batch during training. The default batch size is 64. We further compared more batch sizes [32,64,128,256] on Duke→ Market. Fig. 6(c) indicates that a good range for batch size is around 64.

Source domain performance. Unlike the image adaptation methods [14], HUDA avoids the need for re-id model fine-tuning for target domain. This helps maintain the model performance on the source domain. Table 6 shows that HUDA can preserve well model performance on the source data *after* domain adaptation. In contrast, SPGAN suffers significantly due to losing much of original discrimination ability in fine-tuning.

5. Conclusion

We presented a novel HUDA person re-id model for more discriminative domain adaptation from a labelled source domain to an unlabelled target domain. HUDA is designed for simultaneous global distribution alignment and local instance alignment. It addresses the limitations of existing unsupervised domain adaptation re-id models where only global distribution alignment is considered. Extensive evaluations validate the advantages of HUDA over state-of-the-art models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

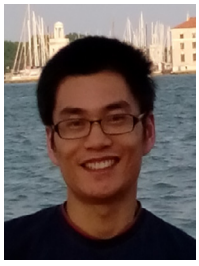
This work was partly supported by the China Scholarship Council, Vision Semantics Limited, the Royal Society Newton Advanced Fellowship Programme (NA150459), and Innovate UK Industrial Challenge Project on Developing and Commercialising Intelligent Video Analytics Solutions for Public Safety (98111-571149).

References

[1] S. Gong, M. Cristani, S. Yan, C.C. Loy, Person Re-Identification, Springer, 2014.
[2] W. Li, X. Zhu, S. Gong, Harmonious attention network for person re-identification, CVPR, 2018.
[3] Y. Sun, L. Zheng, Y. Yang, Q. Tian, S. Wang, Beyond part models: person retrieval with refined part pooling, ECCV, 2018.
[4] W. Li, R. Zhao, T. Xiao, X. Wang, DeepReID: deep filter pairing neural network for person re-identification, CVPR, 2014.
[5] T. Xiao, H. Li, W. Ouyang, X. Wang, Learning deep feature representations with domain guided dropout for person re-identification, CVPR, 2016.
[6] L. Wei, S. Zhang, W. Gao, Q. Tian, Person transfer GAN to bridge domain gap for person re-identification, CVPR, 2018.
[7] S. Liao, Y. Hu, X. Zhu, S.Z. Li, Person re-identification by local maximal occurrence representation and metric learning, CVPR, 2015.
[8] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: a benchmark, ICCV, 2015.
[9] C. Wang, Q. Zhang, C. Huang, W. Liu, X. Wang, Mancs: a multi-task attentional network with curriculum sampling for person re-identification, ECCV, 2018.
[10] G. Csurka, Domain adaptation for visual applications: a comprehensive survey, in: arXiv, 2017.
[11] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, CVPR, 2010.
[12] D.S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, V. Murino, Custom pictorial structures for re-identification, BMVC, 2011.
[13] Z. Zheng, L. Zheng, Y. Yang, Unlabeled samples generated by GAN improve the person re-identification baseline in vitro, arXiv, 2017.
[14] W. Deng, L. Zheng, G. Kang, Y. Yang, Q. Ye, J. Jiao, Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person reidentification, CVPR, 2018.
[15] Z. Zhong, L. Zheng, S. Li, Y. Yang, Generalizing a person retrieval model hetero- and homogeneously, ECCV, 2018.
[16] J. Liu, Z.-J. Zha, D. Chen, R. Hong, M. Wang, Adaptive transfer network for cross-domain person re-identification, CVPR, 2019.
[17] Y. Chen, X. Zhu, S. Gong, Instance-guided context rendering for cross-domain person re-identification, CVPR, 2019.
[18] P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, Y. Tian, Unsupervised cross-dataset transfer learning for person re-identification, CVPR, 2016.
[19] J. Wang, X. Zhu, S. Gong, W. Li, Transferable joint attribute-identity deep learning for unsupervised person re-identification, in: arXiv, 2018.
[20] S. Lin, H. Li, C.-T. Li, A.C. Kot, Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification, in: arXiv, 2018.
[21] H.-X. Yu, W.-S. Zheng, A. Wu, X. Guo, S. Gong, J.-H. Lai, Unsupervised person re-identification by soft multilabel learning, CVPR, 2019.
[22] M. Li, X. Zhu, S. Gong, Unsupervised person re-identification by deep learning tracklet association, ECCV, 2018.
[23] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, T.S. Huang, Self-similarity grouping: a simple unsupervised cross domain adaptation approach for person re-identification, ICCV, 2019.
[24] X. Zhang, J. Cao, C. Shen, M. You, Self-training with progressive augmentation for unsupervised cross-domain person re-identification, ICCV, 2019.
[25] Z. Zhong, L. Zheng, Z. Luo, S. Li, Y. Yang, Invariance matters: exemplar memory for domain adaptive person re-identification, CVPR, 2019.
[26] Q. Yang, H.-X. Yu, A. Wu, W.-S. Zheng, Patch-based discriminative feature learning for unsupervised person re-identification, CVPR, 2019.
[27] Y. Ge, D. Chen, H. Li, Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification, ICLR, 2020.
[28] Y.-J. Li, C.-S. Lin, Y.-B. Lin, Y.-C.F. Wang, Cross-dataset person re-identification via unsupervised pose disentanglement and adaptation, CVPR, 2019.
[29] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, ECCV Workshop, 2016.
[30] Y.-C. Chen, X. Zhu, W.-S. Zheng, J.-H. Lai, Person re-identification by camera correlation aware feature augmentation, IEEE TPAMI, 2017.
[31] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, J. Bu, Semi-supervised coupled dictionary learning for person re-identification, CVPR, 2014.
[32] H. Wang, X. Zhu, T. Xiang, S. Gong, Towards unsupervised open-set person re-identification, ICIP, 2016.
[33] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, NeurIPS, 2014.
[34] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, T. Darrell, Deep domain confusion: maximizing for domain invariance, in: arXiv, 2014.
[35] M. Long, Y. Cao, J. Wang, M.I. Jordan, Learning transferable features with deep adaptation networks, in: arXiv, 2015.
[36] B. Sun, K. Saenko, Deep coral: correlation alignment for deep domain adaptation, ECCV, 2016.
[37] H. Yan, Y. Ding, P. Li, Q. Wang, Y. Xu, W. Zuo, Mind the class weight bias: weighted maximum mean discrepancy for unsupervised domain adaptation, CVPR, 2017.
[38] Y. Li, L. Cheng, Y. Peng, Z. Wen, S. Ying, Manifold alignment and distribution adaptation for unsupervised domain adaptation, in: 2019 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2019, pp. 688–693.
[39] M. Long, H. Zhu, J. Wang, M.I. Jordan, Deep transfer learning with joint adaptation networks, in: arXiv, 2016.
[40] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, CVPR, 2016.
[41] P. Panareda Busto, J. Gall, Open set domain adaptation, ICCV, 2017.
[42] K. Saito, S. Yamamoto, Y. Ushiku, T. Harada, Open set domain adaptation by backpropagation, ECCV, 2018.
[43] E. Ahmed, M. Jones, T.K. Marks, An improved deep learning architecture for person re-identification, CVPR, 2015.
[44] A. Hermans, L. Beyer, B. Leibe, In defense of the triplet loss for person re-identification, in: arXiv, 2017.
[45] A. Bendale, T.E. Boult, Towards open set deep networks, CVPR, 2016.
[46] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, JMLR (2012).
[47] A. Gretton, K.M. Borgwardt, M. Rasch, B. Schölkopf, A.J. Smola, A kernel method for the two-sample-problem, NeurIPS, 2007.
[48] Ç. Gülçehre, Y. Bengio, Knowledge matters: importance of prior information for optimization, JMLR (2016).
[49] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: arXiv, 2015.
[50] M. McCloskey, N.J. Cohen, Catastrophic interference in connectionist networks: the sequential learning problem, Psychology of learning and motivation, Elsevier, 1989.
[51] J. Weston, S. Chopra, A. Bordes, Memory networks, ICLR, 2014.
[52] F.A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: continual prediction with LSTM, Neural Comput. 12 (1999) 2451–2471.
[53] A. Tarvainen, H. Valpola, Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results, NeurIPS, 2017.
[54] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, Automatic differentiation in pytorch, in: arXiv, 2017.
[55] H.-X. Yu, A. Wu, W.-S. Zheng, Cross-view asymmetric metric learning for unsupervised person re-identification, ICCV, 2017.
[56] H. Fan, L. Zheng, Y. Yang, Unsupervised person re-identification: clustering and fine-tuning, in: arXiv, 2017.

[57] Y.-J. Li, F.-E. Yang, Y.-C. Liu, Y.-Y. Yeh, X. Du, Y.-C. F. Wang, Adaptation and re-identification network: an unsupervised deep transfer learning approach to person re-identification, in: arXiv, 2018.

Xu Lan is working toward the PhD degree at the Queen Mary University of London. His research interests include computer vision and machine learning.



Xiatian Zhu was a Computer Vision Researcher at Vision Semantics Limited, London, UK. He received his Ph.D. from Queen Mary University of London. He won The Sullivan Doctoral Thesis Prize 2016, an annual award representing the best doctoral thesis submitted to a UK University in computer vision. His research interests include computer vision and machine learning.

Shaogang Gong is Professor of Visual Computation at Queen Mary University of London (since 2001), a Fellow of the Institution of Electrical Engineers and a Fellow of the British Computer Society. He received his D.Phil (1989) in computer vision from Keble College, Oxford University. His research interests include computer vision, machine learning and video analysis.