Contents lists available at ScienceDirect

Expert Systems With Applications

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

Human-in-the-loop cross-domain person re-identification

Rita Delussu [*], Lorenzo Putzu, Giorgio Fumera

Department of Electrical and Electronic Engineering, University of Cagliari – Piazza d'Armi, 09123 Cagliari, Italy

ARTICLE INFO

Keywords:
Person re-identification
Human in the loop
Relevance feedback
Online domain adaptation

ABSTRACT

Person re-identification is a challenging cross-camera matching problem, which is inherently subject to domain shift. To mitigate it, many solutions have been proposed so far, based on four kinds of approaches: supervised and unsupervised domain adaptation, direct transfer, and domain generalisation; in particular, the first two approaches require target data *during system design*, respectively labelled and unlabelled. In this work, we consider a very different approach, known as *human-in-the-loop*HITL), which consists of exploiting user's feedback on target data processed *during system operation* to improve re-identification accuracy. Although it seems particularly suited to this application, given the inherent interaction with a human operator, HITL methods have been proposed for person re-identification by only a few works so far, and with a different purpose than addressing domain shift. However, we argue that HITL deserves further consideration in person re-identification, also as a potential alternative solution against domain shift. To substantiate our view, we consider *simple* HITL implementations which do not require model re-training or fine-tuning: they are based on well-known relevance feedback algorithms for content-based image retrieval, and of novel versions of them we devise specifically for person re-identification. We then conduct an extensive, cross-data set experimental evaluation of our HITL implementations on benchmark data sets, and compare them with a large set of existing methods against domain shift, belonging to the four categories mentioned above. Our results provide evidence that HITL can be as effective as, or even outperform, existing ad hoc solutions against domain shift for person re-identification, even under the *simple* implementations we consider. We believe that these results can foster further research on HITL in the person re-identification field, where, in our opinion, its potential has not been thoroughly explored so far.

1. Introduction

Person re-identification (Re-Id) consists of matching images of a person of interest across different non-overlapping cameras. This is a challenging computer vision task due to different poses, low camera resolution, illumination changes, occlusions and differences in image background, and also because it is usually an open-world problem (Leng et al., 2020). Early approaches were based on manually defined pedestrian descriptors and similarity measures (e.g., Farenzena et al., 2010), and on metric learning. Current approaches use Convolutional Neural Networks (CNNs) as feature extractors, and in some cases also for metric learning, and always require a training phase (Ye et al., 2022).

Re-Id is inherently a *cross-scene* task, since the pedestrian in the query image has to be searched among images acquired by *different* cameras. Accordingly, benchmark data sets contain images from at least two different cameras (Ye et al., 2022). However, in the standard experimental setting for supervised methods (including CNN-based ones), training and testing data are taken from the *same* set of cameras: this can lead to overestimating performance with respect to real application scenarios where a Re-Id system has to be deployed on target scenes which are *unknown* during design. Additionally, benchmark data sets turned out to be affected by a considerable bias (Genç et al., 2019), similarly to other computer vision tasks (Khosla et al., 2012). In fact, under a *cross-data set* evaluation procedure which is more representative of real cross-scene scenarios, i.e., training on a given data set and testing on a *different* one (direct transfer approach, DT), the performance of supervised methods significantly decreases (Genç et al., 2019; Song et al., 2020).

The above issue is a specific case of the well-known *domain shift* (DS) problem in machine learning, that occurs when a model trained on a *source* domain is deployed on a different, although related, *target* domain. To address DS and to improve the performance, with respect to DT, several approaches proposed in the machine learning field have also been applied to Re-Id. Among them, domain adaptation (DA) and unsupervised DA (UDA) require target data (labelled according to the person's identity, for DA) for model fine-tuning or re-training, either during design (if the target domain is known and target data can be

[*] Corresponding author.
E-mail addresses: rita.delussu@unica.it (R. Delussu), lorenzo.putzu@unica.it (L. Putzu), fumera@unica.it (G. Fumera).

https://doi.org/10.1016/j.eswa.2023.120216
Received 29 June 2022; Received in revised form 17 April 2023; Accepted 17 April 2023
Available online 23 April 2023

collected at that stage) or even after deployment. In contrast, domain generalisation (DG) approaches (Ye et al., 2022) do not use target data: they aim at improving the generalisation capability of the model on any target domain using several source domains for training, without modifying the model after deployment. Since DG does not use target data, it usually achieves lower accuracies than those attained by DA and UDA. In particular, its effectiveness depends on how much the source domains represent the target one. Indeed, building a model capable of generalising to any unseen target domain is considered one of the more difficult problems in machine learning (Zhou, Liu, Qiao, Xiang, & Loy, 2023).

In this work, we address DS in Re-Id under the different perspective of the *human-in-the-loop* (HITL) approach, extending our preliminary work (Delussu et al., 2020). HITL is being studied in the machine learning field since a long time, as an approach to leverage the synergy between human and machine capabilities (Mosqueira-Rey, Hernández-Pereira, Alonso-Ríos, Bobes-Bascarán, & Fernández-Leal, 2022), in several applications including computer vision (Deng et al., 2016). It generally consists in exploiting human feedback on the behaviour of a given learning-based system, typically during operation, to improve its performance in the first place, and could also allow promoting trustworthiness (Holzinger, 2021). In particular, HITL appears well suited to Re-Id systems since they inherently include interaction with a user *during operation*, i.e., during the person image search process. Despite this, HITL methods for Re-Id have been proposed so far by only a few works (e.g., Liu et al., 2013; Wang et al., 2016), and none of them looked at it as a possible solution against DS.

We believe that the potential of HITL has not been thoroughly explored so far in the specific Re-Id task. In particular, we argue that HITL can be a valid alternative to existing DA, UDA and DG approaches to address DS in Re-Id, since the inherent interaction with the user during Re-Id systems' operation allows obtaining human feedback on *target* data, which can be exploited to improve re-identification accuracy on the target domain. Moreover, this can be achieved without demanding additional effort in a preliminary, offline phase, e.g., for collecting and possibly manually annotating target data as required by many DA and UDA solutions. This can be very useful in challenging scenarios where no target data (not even unlabelled) can be collected for model training or refinement; this is the same scenario addressed by DG, with the difference that HITL *can* exploit target data (acquired during operation). Moreover, similarly to DG, HITL can also be implemented without any model refinement step.

Based on the above premises, this work's main goal and contribution is to investigate the effectiveness of HITL in Re-Id against DS, as an alternative solution to the aforementioned DA, UDA and DG approaches. To this aim, we start by considering *simple* implementations of HITL based on representative relevance feedback (RF) algorithms that are a well-known, effective solution for *content-based image retrieval* (CBIR) problems which Re-Id is a particular instance of Vezzani, Baltieri, and Cucchiara (2013) and Ye et al. (2022), and that have not been considered by previous work on HITL Re-Id. Additionally, we devise novel versions of the considered RF algorithms tailored explicitly to Re-Id. We also adopt a feedback protocol more suited to RF than the ones used by existing HITL Re-Id methods. In summary, after the underlying Re-Id system returns to the user the gallery images ranked by decreasing similarity to the query image, our HITL methods ask the user to indicate all true matches (if any) in the top-K ranks (for a given K value), and exploit this feedback to re-rank the whole gallery using RF algorithms, aiming at pushing other images of the query individual toward the top ranks; the user can repeat this process for several iterations.

We point out that our RF-based HITL methods are very general and can be implemented on top of *any* Re-Id model. We then carry out an extensive empirical analysis of our HITL implementations, and a comparison with state-of-the-art DA, UDA, DT and DG methods, on three benchmark Re-Id data sets, under a cross-data set setting which is representative of the cross-scene application scenario of interest. Our results confirm that, even under a simple RF-based implementation, HITL can outperform DT and DG methods, which do not use target data for model training or fine-tuning, at the cost of a very limited user effort (feedback) during operation; it can also achieve comparable or even better performance than the one attained by DA and UDA.

In summary, this work provides the following contributions: (i) We reconsider the HITL approach to Re-Id as a potential, additional solution to DS beside DA, UDA and DG; (ii) We implement HITL methods based on existing RF algorithms for CBIR, as well as on novel versions of such algorithms we specifically devise for Re-Id; (iii) We provide extensive empirical evidence of the effectiveness (in terms of re-identification accuracy) and efficiency (in terms of the required amount of human feedback and of the lack of model refinement requirements) of the HITL approach to Re-Id, in comparison with DA, UDA and DG.

We believe that these results can reawaken interest on, and foster further research on the HITL approach in the Re-Id field, where, in our opinion, its potential has not been thoroughly explored so far.

The rest of this work is organised as follows. In Section 2 we review related work. In Section 3 we motivate our RF-based HITL implementation for Re-Id, and present the considered RF algorithms and their novel versions. Section 4 describes the experimental set-up, whereas results are reported in Section 5. A discussion in Section 6 concludes this paper.

2. Related work

In this section, we survey the literature on Re-Id, focusing on DT, DA, UDA and DG methods proposed to address DS, and to HITL methods.

2.1. Direct transfer methods

Although DT is usually not effective for supervised methods, some works proposed to improve its performance by extracting generalisable features from a source domain, i.e., robust to domain shift, using specific CNN models, e.g., OSNet (Zhou et al., 2019) and ADIN (Yuan et al., 2020). For instance, **OSNet** extracts omni-scale features through scale-specific streams which are then combined using different weights.

2.2. Domain adaptative and unsupervised domain adaptive methods

DA consists in pre-training a model on labelled source data and fine-tuning it on labelled target data. For instance, DA-ReID (Xu et al., 2021) considers the reliability of lower-body, upper-body and global features in case of occlusions. The fine-tuning step aims to improve the identification of pedestrians' lower and upper body in the target domain.

UDA assumes that target data can be collected either during design, or even after deployment, to refine a model previously trained on a source domain, and exploits them without requiring manual annotation (Feng et al., 2021; Ge et al., 2020; Qi et al., 2019; Song et al., 2020; Wu et al., 2019; Zhang et al., 2019).

Method based on pseudo-labels. Several UDA methods are based on pseudo-labels automatically assigned by the source model either to *all* target images, e.g., MTNet (Chen et al., 2023b), Theory&Practice (Song et al., 2020), JL (Feng et al., 2021), MMT (Ge et al., 2020), or to a *subset* of target images for which they are considered reliable, e.g., PAST (Zhang et al., 2019), CACHE (Liu, Ge, Sun, & Hou, 2022). For instance, the Mutual Tri-training Network (**MTNet**) framework (Chen et al., 2023b) combines self-paced learning and mutual learning in order to gradually increase the discrimination capability of the model and reduce noisy pseudo-labels. The Mutual Mean Teaching (**MMT**) model (Ge et al., 2020) collaboratively learns two CNNs on the source and target domains to improve the prediction of pseudo-labels, and selects the best performing one (on validation data) as a feature extractor for the inference step. The Joint Learning (**JL**) model (Feng et al., 2021) uses an embedding CNN followed by two different branches for source and target images which share a module aimed at improving the prediction of the k-nearest neighbours of each target image. **PAST** (Zhang et al., 2019) uses a self-training with progressive augmentation framework consisting of two stages, "conservative" and "promoting", which are used iteratively for model optimisation. In the first stage, a subset of reliable target images is selected and is used in the second stage together with the corresponding pseudo-labels for model updating. Complementary Attention-driven Contrastive learning with Hard-sample Exploring (**CACHE**) model (Liu et al., 2022) aims at improving the discrimination capability of the model by integrating multiple feature sub-spaces and the compactness of clusters, investigating the hard samples for each centroid.

Methods based on synthetic image generation. Another UDA approach is to augment the training set by generating synthetic images, using either generative adversarial networks (GANs) (Ainam et al., 2021; Verma et al., 2023; Wei et al., 2018; Zhang et al., 2020b; Zhong et al., 2019; Zhou et al., 2021) or other methods (Chen et al., 2023a; Chong et al., 2021; Song, Liu, & Jin, 2022; Tang, Xue, & Chen, 2022; Zou et al., 2020). The simplest approach consists in augmenting the *source* images, e.g., **MDJL** (Chen et al., 2023a) shuffles the colour channels to enrich data diversity and to improve the feature representation. However, *source* images are mostly modified according to the style of the target domain, e.g., DPCFG (Song et al., 2022), IPES-GAN (Tang et al., 2022), MLMS (Tang et al., 2022), SILC (Ainam et al., 2021), PT-GAN (Wei et al., 2018), ECN (Zhong et al., 2019), AAAN (Zhang et al., 2020b), CVSE (Zhou et al., 2021), STReID (Chong et al., 2021); other methods modify *target* images according to the style of *different* target cameras, e.g., DAL (Zhang et al., 2020a); some methods modify both source and target images, e.g., DGNet++ (Zou et al., 2020). For instance, Individual-Preserving and Environmental Switching cyclic generation network (**IPES-GAN**) (Tang et al., 2022) aims at reducing the domain gap between source and target data by generating images (with a target pose) through GAN and swapping backgrounds in a cycling manner. Adaptive Attention-Aware Network (**AAAN**) (Zhang et al., 2020b) belongs to the first group: it generates target images by first learning camera-style and camera-invariant features. Dual-Alignment Learning (**DAL**) (Zhang et al., 2020a) belongs to the second group; it optimises the model using mutual information between the target and generated samples. **DG-Net**++ (Discriminative and Generative Network) (Zou et al., 2020) is a method belonging to the third group: it generates target images by swapping the appearance of source and target ones and consequently by augmenting the number of samples in both domains; to improve the feature representation, an adversarial training strategy is used.

Other methods. Different UDA approaches have also been proposed, e.g., CASCL (Wu et al., 2019), TALM-IRM (Li et al., 2021) and D-MMD (Mekhazni et al., 2020). For instance, Camera-Aware Similarity Consistency Learning (**CASCL**) (Wu et al., 2019) aims at learning consistent distributions of similarity scores for intra- and cross-camera image matching, through several loss functions. A coarse-to-fine consistency learning strategy is used to improve consistency, which considers similarity in feature space and the top-ranked neighbours of a given image.

2.3. Domain generalisation methods

DG is a recently proposed approach against DS, which has also been applied to Re-Id, e.g., MMFA-AAE (Lin et al., 2021), DomainMix (Wang, Liao, Zhao, Kang, & Shao, 2021) and OSNet (see Section 2.1). DG does not use target data or any knowledge of the target domain(s). For instance, **MMFA-AAE** (Multi-task Mid-level Feature Alignment with an Adversarial Auto Encoder) uses an adversarial learning strategy to extract domain-invariant features by using a decoder that mitigates the influence of domain-specific information. **DomainMix** pre-trains a feature extractor using labelled synthetic images and unlabelled images from the source domains; a discriminator between synthetic and real images is then learnt and the feature extractor is simultaneously refined to obtain domain-invariant features.

2.4. Person re-identification with a human in the loop

Existing HITL methods for Re-Id have been proposed mainly to improve the accuracy of a model during operation, or to collect labelled target data for model training or refinement after deployment, before operation. We point out that none of them has been proposed to address DS, nor has been evaluated in a *cross-domain* setting exploiting user's feedback on gallery images returned in response to queries by the same user *during system operation*.

Most of the existing methods exploit the inherent interaction with the user *during system operation*, and require feedback on the top-K images retrieved in response to a user's query, for a given K (Ali et al., 2010; Hirzer et al., 2011; Liu et al., 2013; Navaneet et al., 2020; Wang et al., 2016). Retrieved images are then re-ranked based on the user's feedback, to push true matches toward the top ranks. The feedback required from the user can be of different types, and can be exploited in several ways.

In Ali et al. (2010), the operator is shown the top-K ranked images together with other gallery images selected by an active learning algorithm, and is asked to select any subset of true and false matches; the image similarity measure is then updated by a distance metric learning algorithm based on such feedback, and the gallery is re-ranked. Several feedback rounds can be carried out. The method of Hirzer et al. (2011) asks the user whether the query identity is present in the top K ranks; if not, a discriminative model of the query identity is learnt and is used to re-rank the gallery images. Similarly, if the query identity is not present among the top K ranks, the Post-rank OPtimisation (POP) method (Liu et al., 2013) asks the operator to select a single "strong negative", i.e., a false match showing an individual very different from the query, and optionally a few weak negatives; this feedback is used to learn a post-rank function to re-rank the gallery, based on an affinity graph that describes pairwise image similarity between gallery images. The method of Navaneet et al. (2020) uses a slightly different approach, by sequentially processing the gallery images coming from different cameras: the operator is asked to select a single true match from the first gallery, then this feedback is exploited to rank the images of the second gallery, and so on.

The operator's feedback collected by the previous methods during operation, for a given query, is used to refine the ranking of gallery images only with respect to the *current* query. Other works proposed *incremental* strategies to improve the underlying Re-Id systems based on the operator's feedback collected over a *sequence* of queries. The Human Verification Incremental Learning method (Wang et al., 2016) uses a Mahalanobis distance metric between feature vectors of image pairs, initially set as the Euclidean distance; for a given query the operator is asked to select a true match, if any, or a strong negative among the top-K gallery images; the distance metric is then incrementally updated using an online metric learning algorithm. Several feedback rounds can be carried out for each query; the updated distance metric is then used for the next round or the next query. In Wang et al. (2018) the HITL approach is used to collect during system operation pairs of target images that the operator is asked to label, and are then used for model updating by incremental learning; however, the collected images are not related to the queries selected by the operator, but are automatically selected, randomly or by an active learning algorithm, from the image stream acquired during system operation. The Deep Reinforcement Active Learning (DRAL) method (Liu et al., 2019) aims instead at collecting *offline* a training set of target images, to fine-tune a model previously trained on a different source domain, *before* starting to use a Re-Id system. To minimise manual annotation effort, active and reinforcement learning algorithms are used to select a small amount of pairs of target images, which the user is asked to label as the same or different identity. Therefore, differently from the above methods, DRAL does not ask for user's feedback during system *operation*, i.e., on gallery images retrieved in response to queries selected by the user (e.g., during a real investigation), but involves the user in an offline manual labelling session on the target cameras.

All HITL methods described above, whether requiring online or offline feedback, are relatively complex and mostly devised for ad hoc Re-Id systems. In the next section, we present our HITL methods, focusing on cross-scene application scenarios. Our methods are based on well-known RF algorithms, and are simpler than existing HITL Re-Id methods, also because they do not require model refinement. They are also model-agnostic, i.e., they can be implemented on top of any Re-Id system.

3. Relevance feedback for human-in-the-loop person re-identification

In this work, we focus on challenging, cross-scene application scenarios where a Re-Id system has to be deployed on target camera views that were unknown during design. In this context, we argue that the HITL approach is a further alternative to address the resulting DS, beside the approaches specifically proposed to this aim, i.e., DA, UDA, DT and DG, for the following reasons: (i) Re-Id systems inherently include an interaction with a human operator, e.g., a forensic investigator.[1] (ii) Exploiting the user's feedback on the gallery images returned in response to a query by the same user to re-rank them can be seen as a form of *adaptation* of a Re-Id system (e.g., its distance metric) to the target domain, which is carried out *online*, i.e., during system operation; if this is carried out without re-training or fine-tuning the source model, HITL can be considered as a specific kind of online DA (Royer et al., 2015); its effectiveness can be further improved by an *incremental* adaptation process, i.e., by continuously updating the Re-Id system over the sequence of queries selected by the operator, still without refining the source model. (iii) The operator's feedback allows to collect *labelled* target data, contrary to DT and DG; however, contrary to DA, this can be achieved with a *limited* annotation effort, if the required feedback consists of selecting true or false matches among the top-ranked gallery images returned by a Re-Id system in response to a user query: indeed, matching the selected query with top-ranked gallery images, to check whether they show the same individual or not, is already part of the user's task during system operation.[2]

As mentioned in Section 2.4, existing HITL methods for Re-Id have not been devised nor evaluated as a possible solution against DS. They are also relatively complex since they are based on, e.g., training a query-specific discriminative classifier (Hirzer et al., 2011), building and processing a graph representing pairwise image similarity between gallery images (Liu et al., 2013), or solving a complex optimisation (online metric learning) problem after each *single* feedback on a gallery image (Wang et al., 2016). Additionally, previous work disregarded the fact that Re-Id is a particular case of CBIR task, and therefore the HITL approach can also be implemented using simpler, well-known RF algorithms.[3]

Based on the above motivations, the main objective of this work is to investigate the effectiveness of the HITL approach to Re-Id, specifically against DS, as a further solution beside DA, UDA and DG. To this aim, we chose to focus on a *simple* implementation of HITL, which is based on representative RF algorithms, does not require model refinement on target data annotated through user's feedback, and is model-agnostic, i.e., can be used on top of *any* Re-Id system. In particular, based on an analysis of the peculiarities of Re-Id with respect to generic CBIR tasks, we also develop novel versions of the considered RF algorithms specifically tailored to the Re-Id task, including *incremental* versions aimed at improving online DA capability. Finally, we adopt a feedback protocol more suited to RF than the ones used by existing HITL Re-Id methods.

Fig. 1 shows a general scheme of the proposed HITL Re-Id implementation based on RF algorithms, in a cross-domain scenario. A source model is first trained on images from a source domain. After system deployment, during operation (i.e., a Re-Id session on a target domain), the source model is used to extract features from the query images selected by the operator, and from gallery images. The latter are then automatically ranked based on their similarity to the query, and are shown to the operator starting from the most similar one. The operator can then provide feedback (same or different identity with respect to the query) on the top-ranked images, which is exploited by an RF algorithm to re-rank the gallery. The user can choose to repeat the feedback and re-ranking steps for several iterations. More detailed schemes, depending on the specific RF algorithm, are reported in the next section.

This work extends a previous conference paper by the authors (Delussu et al., 2020), with the following additional contributions: (i) we analyse the peculiarities of the Re-Id task with respect to generic CBIR tasks; (ii) we consider an additional RF algorithm (Passive-Aggressive) for better coverage of existing RF approaches; (iii) we develop novel versions of the considered RF algorithms specifically tailored to Re-Id, based on the analysis mentioned above; (iv) we carry out experiments on an additional, larger benchmark data set (MSMT17); (v) we extend our experimental comparison, previously limited to two UDA methods, to a much larger set of methods, including DA, DT and DG.

3.1. Relevance feedback algorithms

CBIR systems aim at retrieving from large databases digital images that are similar to a given query image in terms of visual and semantic content, where similarity is mainly evaluated as the Euclidean distance in a given feature space. Retrieved images are shown to the user as a ranked list. One of the main issues of CBIR is the *semantic gap* between the similarity perceived by the user (semantic level) and the similarity computed by the machine (feature level). RF is one of the mechanisms introduced to improve the effectiveness of CBIR systems by reducing the semantic gap. RF algorithms typically ask the user to provide feedback on a subset of top-ranked retrieved images, as relevant (positive) or not (negative) to the query; retrieved images are then re-ranked accordingly, aiming at increasing the number of relevant ones in the top ranks. This process can be repeated either for a fixed number of iterations/rounds, or until the user decides not to engage in further rounds.

The existing RF approaches can be divided into four main categories. One of the first approaches consists in computing a new query vector exploiting the feature vectors of relevant and non-relevant images (Lin et al., 2015). A similar approach is based on *distance* or *similarity learning*: instead of updating the query, it optimises the distance metric used to compute image similarity (Wu et al., 2019). A third approach consists in estimating the posterior probability distribution of the relevant and non-relevant images using nearest neighbour (NN) methods, and in using it as a similarity measure (Giacinto, 2007). A fourth approach views RF as a two-class *pattern classification* task, and uses the sets of relevant and non-relevant images obtained through user's feedback to train existing or ad hoc machine learning algorithms (Piras et al., 2013).

[1] Note that in Re-Id the feedback is provided by a human expert and can therefore be considered reliable, which is not guaranteed by methods involving crowd-sourcing, e.g., Deng et al. (2016).

[2] Providing feedback can be made convenient using ad hoc, user-friendly graphical interfaces.

[3] A few RF algorithms have been considered only for very limited experimental comparisons with existing HITL Re-Id methods (Liu et al., 2013; Wang et al., 2016).

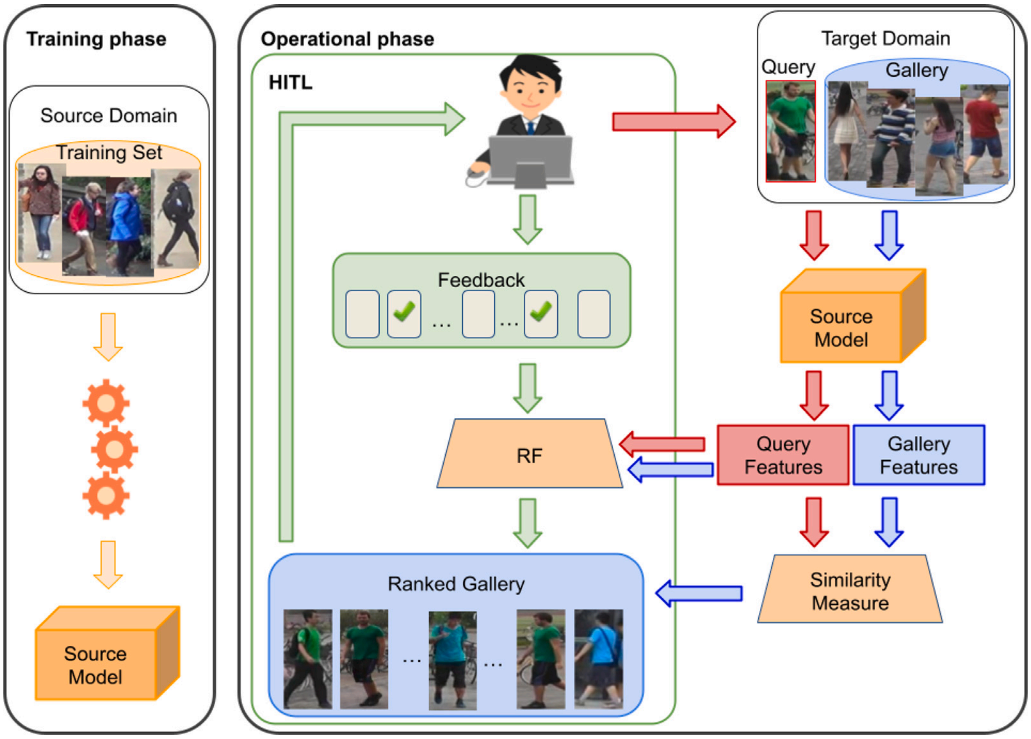

Fig. 1. Generic schema of the proposed HITL approach based on RF for a cross-domain Re-Id system (see text).

In the following, we focus on three RF algorithms characterised by low complexity, and representative of the different categories: the classical Rocchio or Query Shift (QS) algorithm (Lin et al., 2015), Relevance Score (RS) (Giacinto, 2007) and Passive-Aggressive (PA) (Piras et al., 2013).

QS is one of the earliest RF methods based on moving the query in feature space. It assumes that positive images form a single cluster in feature space, and that the feature vector x_{q} of the original query could lie (relatively) far from it. Accordingly, after the user's feedback on a set P of positive images and a set N of negative ones, QS computes a new query vector $\overline{x}_{\mathrm{q}}$ by moving x_{q} toward the Euclidean centre of P and farther from N, such that in the next round a larger number of positive images can be found in the top ranks:

$$\overline{x}_{\mathrm{q}} = x_{\mathrm{q}} + \frac{1}{|P|}\sum_{\mathrm{p}\in P} x_{\mathrm{p}} - \frac{1}{|N|}\sum_{\mathrm{n}\in N} x_{\mathrm{n}} \ . \tag{1}$$

RS is a state-of-the-art algorithm that computes a *relevance score* for each retrieved image. Contrary to QS, it assumes that relevant images can be spread over several clusters, and considers each positive image in P, as well as the query, as the centre of a positive cluster. Accordingly, the relevance score $s_{\mathrm{NN}}(x_{\mathrm{i}})$ of each retrieved image x_{i} is computed based on its distance to the nearest positive and nearest negative neighbouring images in P and N, $x_{\mathrm{p}}^{\mathrm{NN}}$ and $x_{\mathrm{n}}^{\mathrm{NN}}$:

$$s_{\mathrm{NN}}(x_i) = \frac{\|x_i - x_{\mathrm{n}}^{\mathrm{NN}}\|}{\|x_i - x_p^{\mathrm{NN}}\| + \|x_i - x_{\mathrm{n}}^{\mathrm{NN}}\|} \ , \tag{2}$$

where $\|\cdot\|$ is a given metric in feature space, typically the Euclidean distance. The relevance score increases as the distance from the nearest positive image decreases compared to the distance from the nearest negative one. Similarly to RS, **PA** computes a score for each retrieved image, using however a discriminant approach based on the assumption that positive and negative images are linearly separable in feature space:

$$s(x_i) = w \cdot x_i \ , \tag{3}$$

where w is a query-specific weight vector that should provide higher scores for positive images than for negative ones. Accordingly, it is obtained as the solution to the following optimisation problem:

$$w = \arg\max_{v} \sum_{\forall p\in P}\sum_{\forall n\in N} v(x_p - x_n) \ . \tag{4}$$

To this aim, an online, iterative learning process is used Piras et al. (2013): at each iteration, a pair of images $x_{\mathrm{p}} \in P$ and $x_{\mathrm{n}} \in N$ is randomly extracted, and w is updated accordingly. Since P and N are typically imbalanced, with $|P| < |N|$, positive images are drawn more than once.

3.2. Adapting relevance feedback algorithms to person re-identification

RF algorithms, including the ones considered in Section 3.1, have been originally devised to address the semantic gap, i.e., the mismatch between the result provided by a machine and the one sought by the user (Lin et al., 2015). We point out that this gap could be present at different levels: between high-level image content and its feature representation, introduced by the machine, and between the high-level image content and the image label, which is introduced by the user. The latter kind of semantic gap is typical of CBIR systems, i.e., the query chosen by the user could be *misplaced* or *marginally representative* of the concept sought. As an example, in a category-level CBIR system, each image can contain objects or scenes relevant to *different* possible concepts of interest for different users, e.g., an image of a dog can also contain a car. Moreover, if a category-level CBIR system includes images of dogs of different breeds, and the user's query is an image of a dog, the real user's intention (i.e., is the user interested in generic dog images or in dogs of the same specific breed?) cannot be "understood" by the CBIR system, unless an RF process is used.

Under this viewpoint, Re-Id is a specific kind of CBIR task, i.e., a fine-grained *instance-level* retrieval task. Indeed the gallery is (ideally) made up of tight *bounding boxes*, each one containing a *single* pedestrian, and queries should be formulated as well as tight bounding boxes containing the *specific* individual of interest. This implies that the query is not *marginal* as it could be in generic CBIR systems. Therefore, the only kind of semantic gap that has to be addressed by RF, if it used in a Re-Id system, is the one between the high-level image content

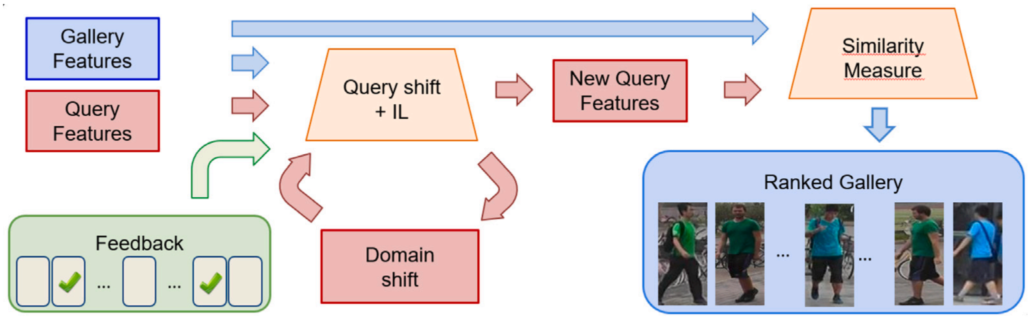

Fig. 2. Schema of the proposed QS+IL approach.

and its feature representation. In particular, this gap can be due to differences in camera perspective, scene illumination and background, and pedestrian pose, and is even more likely to occur in the cross-scene scenario considered in this work. Based on the above considerations, we propose the following novel versions of the QS, RS and PA algorithms specifically devised for Re-Id.

First, note that the effect of QS and PA can be seen as adapting the *representation* of the original query to the user's intent, either by moving it in feature space (QS) or by assigning different weights to each feature (PA). However, they operate on each query independently from the other ones. Instead, we devised *incremental learning* versions of them that exploit the whole sequence of user queries to compute the query shift (QS+IL), or the weights of the hyperplane that separates the query identity from the other ones (PA+IL). This can be seen as continuously updating the feature space; accordingly, QS+IL and PA+IL can be considered online DA methods, where target images are used only during system operation.

In QS+IL we replace the new query vector of the original QS formulation with a vector $\overline{\overline{x}}_q^{(i)}$, which is computed as follows. We start from the query shift vector computed by the original QS (see Eq. (1)), which measures the amount of shift of a given query toward its positive cluster:

$$\Delta_q^{(i)} = \bar{x}_q^{(i)} - x_q^{(i)} . \tag{5}$$

We exploit it to compute a *domain shift* vector $\overline{\Delta}_q^{(i)}$ that incrementally estimates over the sequence of user queries the amount of shift needed to move them closer to their positive clusters. At the first feedback round of the first query, $x_q^{(1)}$, the new query vector $\overline{\overline{x}}_q^{(1)}$ is computed as in the original QS by Eq. (1). For each subsequent feedback round over any query $x_q^{(i)}$, the domain shift vector is first computed as:

$$\overline{\Delta}_q^{(i)} \leftarrow \begin{cases} \overline{\overline{x}}_q^{(i-1)} - x_q^{(i-1)}, & \text{first feedback round} \\ & \text{of } i\text{–th query } (i > 1), \\ \overline{\overline{x}}_q^{(i)} - x_q^{(i)}, & \text{otherwise,} \end{cases} \tag{6}$$

where $\overline{\overline{x}}_q^{(i-1)}$ or $\overline{\overline{x}}_q^{(i)}$ denote the new query computed in the *previous* step. Then, $\overline{\overline{x}}_q^{(i)}$ for the *current* step is computed as:

$$\overline{\overline{x}}_q^{(i)} \leftarrow (1-\gamma)\bar{x}_q^{(i)} + \gamma\overline{\Delta}_q^{(i)} , \tag{7}$$

where $\bar{x}_q^{(i)}$ is obtained by the original QS, Eq. (1), and $\gamma \in [0,1]$ is a coefficient that weighs the contribution of the two components. In other words, differently from the original QS version, in our QS+IL version, the query-specific shift computed by QS is combined with a shift vector summarising the "history" of the previous queries and the previous feedback rounds (if any) of the current one (see Fig. 2).

In PA+IL we compute a weight vector $\overline{\overline{w}}^{(i)}$ for the ith query as follows. For the first feedback round of the first query $x_q^{(i)}$, $\overline{\overline{w}}^{(i)}$ is computed as in the original PA. For each subsequent feedback round over any query, first, the weight vector computed in the *previous* round, denoted as $\overline{w}^{(i)}$, is considered:

$$\overline{w}^{(i)} \leftarrow \begin{cases} \overline{\overline{w}}^{(i-1)}, & \text{for the first feedback round} \\ & \text{of the } i\text{th query } (i > 1), \\ \overline{\overline{w}}^{(i)}, & \text{otherwise.} \end{cases} \tag{8}$$

Then, the one of the *current* step is computed as:

$$\overline{\overline{w}}^{(i)} \leftarrow (1-\gamma)w^{(i)} + \gamma\overline{w}^{(i)} , \tag{9}$$

where $w^{(i)}$ is the weight vector computed by the original PA, and $\gamma \in [0,1]$ is again a coefficient that weighs the contribution of the two weight vectors. Similarly to QS+IL, the query-specific weights computed by PA are combined with a weight vector, from now on a *domain weight*, summarising the "history" of the previous queries and the feedback rounds of the current one (see Fig. 3).

RS is based on the assumption that in CBIR systems positive images can be spread over different clusters, some of which may be relatively farther from the query than some negative clusters (see Section 3.1). Accordingly, the query and each positive image selected by the user are considered the centres of distinct clusters. A high score is given even to images far from the cluster defined by the query but close to at least one other positive cluster. However, for the reasons discussed above, we can assume that in Re-Id the query is not marginal, and therefore images of the same identity and of different identities form two distinct (positive and negative) clusters (see Fig. 4). Accordingly, in our modified version of RS, named M-RS, the score is computed with respect to the centroids μ_P and μ_N of the positive and negative clusters:

$$s_\mu(x_i) = \frac{\|x - \mu_N\|}{\|x_i - \mu_P\| + \|x_i - \mu_P\|} . \tag{10}$$

Note that M-RS has a lower processing cost than RS.

3.3. Feedback protocol

As pointed out at the beginning of this section, existing HITL methods for Re-Id that fit the online application scenario considered in this work (i.e., exploiting user's feedback on ranked gallery images obtained in response to user's queries during operation) ask the operator to select a *single* true/false match at each feedback round, to re-rank gallery images (Hirzer et al., 2011; Liu et al., 2013; Wang et al., 2016). However, it is known that RF algorithms benefit from more feedback per round, which was also confirmed in Re-Id tasks by our preliminary results (Delussu et al., 2020). In particular, the higher the number of feedback is, the lower the number of rounds required to push positive gallery images toward the top ranks can be. Accordingly, in this work, we adopt the multi-feedback protocol of Delussu et al. (2020), consisting in asking the operator to select *all* the true matches in the top-K ranks, for a given K value; the remaining images in such ranks are *automatically* labelled as negative. This means that K images (positive and negative) are used at each RF round. In practice, the value of K can also be chosen *dynamically* (per query, and even per single

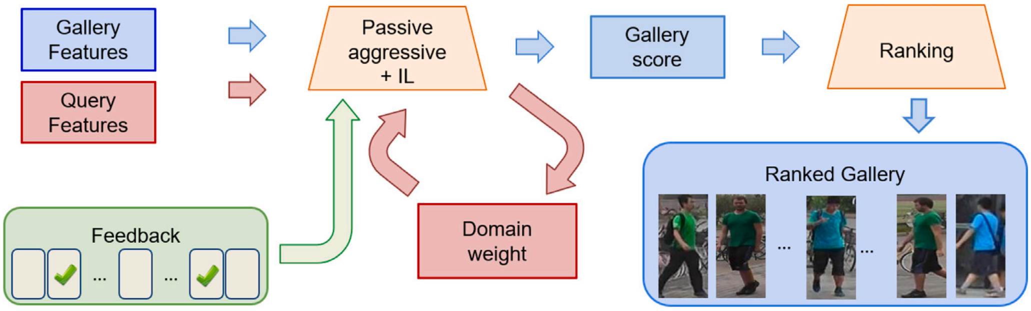

Fig. 3. Schema of the proposed PA+IL approach.

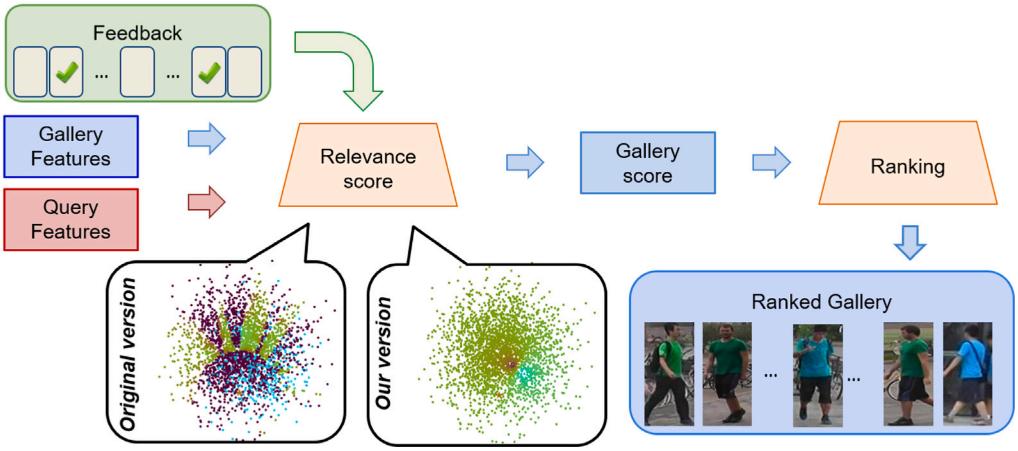

Fig. 4. Schema of the original RS algorithm and of our M-RS version, highlighting their differences.

feedback round) by the user, as the highest rank up to which he or she is willing to inspect the retrieved images.

Note that our multi-feedback protocol is specifically suited to investigation scenarios in which the user needs to retrieve *all* the images of an individual of interest, e.g., to reconstruct his or her movements. It may seem that our protocol requires the user a higher effort than the single-feedback one. However, as discussed in Delussu et al. (2020), real-world applications involve large camera networks that produce very large template galleries, and therefore it becomes less likely to find several positive matches, or even a single one in the first round, in the few top ranks (Wang et al., 2016). Moreover, when no true match is present in the top ranks, to determine this fact the user needs to inspect *all* the corresponding gallery images under both the single- and multi-feedback protocols; however, in this case the single-feedback protocol also requires the user to select a *strong* negative, which demands an additional effort; we also point out that several different strong negative images may exist, and that the choice among them is inevitably subjective, which may affect the resulting performance.

4. Experimental set-up

Our experiments are aimed at evaluating our HITL Re-Id implementation, based on the RF algorithms described in Sections 3.1 and 3.2, in the specific *cross-domain* setting, and at comparing it with state-of-the-art methods against DS based on the DA, UDA, DT and DG approaches. We point out that our goal is not to outperform these approaches, but to assess whether and to what extent the considered HITL solution is a further, valid solution to DS with respect to them, in terms of both effectiveness (re-identification accuracy) and efficiency (required amount of human feedback and lack of model refinement).

Data sets. We considered three widely used benchmark data sets: Market-1501 (Zheng et al., 2015), DukeMTMC-reID (Gou et al., 2017), and MSMT17 (Wei et al., 2018) (for short, Market, Duke and MSMT), whose characteristics are summarised in Table 1. Market consists of 32,668 images showing 1501 identities, acquired from six cameras placed in front of a supermarket. It is split into 751 identities for training (12,936 images) and 750 for testing (the remaining images). The gallery comprises 19,732 images, of which 15,913 belong to query identities. Duke contains 36,411 images (bounding boxes) of 1404 identities, captured from eight cameras in a campus. They are subdivided into 16,522 images of 702 identities for training, and the remaining images of the other 702 identities for testing. The number of gallery images is 17,661. MSMT consists of 126,441 images and 4101 identities captured from 15 cameras. They are split into 32,621 images of 1401 identities for training (2373 images are used for validation), and the remaining images and identities for testing. The gallery contains 82,161 images. We simulated cross-scene application scenarios characterised by DS through **cross-data set** experiments: each data set was used in turn as the source domain, and each of the remaining ones as the target domain.

Table 1
Number of identities (#ID), images (#IM) and cameras (#CAM) in the data sets.

Data set	#IDs/#IM			#CAM
	Training set	Query set	Gallery	–
Market	751/12 936	750/3368	751/15 913	6
Duke	702/16 522	702/2228	1110/17 661	8
MSMT	1041/30 248	3060/11 659	3060/82 161	15

Methods used for comparison. We carried out an extensive comparison with all state-of-the-art DT, DA, UDA and DG methods described in Section 2. Whenever available, we used the source code provided by the authors with the recommended parameter settings (Ge et al., 2020; Zhang et al., 2019; Zhong et al., 2019); otherwise, we reported the results from the respective papers. With regard to DG methods, the following data sets were used in the respective papers as source and target domains: DomainMix used RandPerson and Market or MSMT as sources, and MSMT or Market as target; MMFA-AAE (Ye et al., 2022) used CUHK02, CUHK03, Market-1501, DukeMTMC-reID and CUHK-SYSU as source domains, and MSMT as the target; OSNet used Market-1501, DukeMTMC-reID and CUHK03 as the source domains, and MSMT as the target.

We point out that a comparison with existing HITL methods for Re-Id dealing with the considered application scenario (Liu et al., 2013; Wang et al., 2016) (see Section 2.4) turned out to be not possible since their source code is not publicly available, and the lack of details in their relative manuscripts hindered re-implementation. Moreover, the different experimental setting used in that works (i.e., no cross-data set experiments were carried out) does not allow a direct comparison with the results reported in the respective papers.

Implementation of RF algorithms. As in Wang et al. (2016), we carried out three feedback rounds for each RF algorithm and set $K = 50$. For a fair comparison with DS methods, we used the whole query sets of the target data sets. Given the considerable size of query sets (see Table 1), we simulated the user feedback using the ground truth identity labels (Navaneet et al., 2020). Although this disregards potential errors in the user's feedback, we point out that they are unlikely in the considered scenario, where the user is a specialist (e.g., an investigator) and is also asked to give feedback on a *limited* amount of retrieved images. For the three original RF algorithms, we adopted the authors' recommended parameter settings (Giacinto, 2007; Lin et al., 2015; Piras et al., 2013). For the proposed QS+IL (Eq. (7)) and PA+IL (Eq. (9)) we used a fixed γ value of 0.5, according to the fact that in the considered application scenario target data are not available to refine the source model, including parameter setting. However, in Section 5.3 we evaluate how the value of γ affects the performance of QS+IL and PA+IL.

Baseline model for RF algorithms. For RF algorithms, we used as a feature extractor a ResNet-50 network pre-trained on ImageNet, and then fine-tuned on the training partition of the source data set. During training, we used horizontal flip and random crop with a probability of 0.5 to reduce over-fitting. Stochastic Gradient Descent was used for optimisation with momentum 0.9 and weight decay 5×10^{-4}; the learning rate was set to 0.00035. It is worth noting that our HITL approach is model-agnostic, and does not require any change of the source model after deployment.

Performance measures. We considered the Cumulative Matching Curve (CMC) at ranks $k = 1, 5, 10, 20$, and the mean Average Precision (mAP):

$$CMC(k) = \sum_{r=1}^{k} P(r), \quad mAP = \frac{1}{Q}\sum_{q=1}^{Q} \text{AveP}(q),$$

where $P(r)$ is the fraction of queries for which the gallery image of the correct identity (or the top-ranked image, if more than one) is found at rank r, Q is the total number of queries and AveP(q) is the average precision for a given query q. Note that in real applications, more than one image of the query identity may be present in the gallery (this is the case of the considered data sets). In this case, operators may be interested in retrieving all of them, e.g., forensic investigators may want to reconstruct the movements of a suspect individual across all the available video cameras. Under this scenario, the mAP measure gives a more complete account of the performance of a Re-Id system, since the CMC curve only considers the top-ranked image of the query identity.

5. Experimental results

We first evaluate the performance of our HITL approach attained using the original RF algorithms and using our modified versions; then, we compare it against state-of-the-art DA, UDA, DT and DG methods.

5.1. Evaluation of RF algorithms

Table 2 reports the overall results obtained by the baseline (i.e., the source model) and by the HITL method based on the RF algorithms (QS, QS+IL, RS, M-RS, PA and PA+IL), implemented on top of the same baseline. For completeness, we also report results attained under a fully supervised learning setting (denoted by "supervised"), i.e., training and testing on the same *target* data set, without using the HITL approach.

All RF algorithms outperformed the source model in terms of both CMC curve and mAP, in all target data sets. In particular, they attained a remarkable improvement since the first round. In most cases, the highest improvement was achieved by PA, except for the case where MSMT was the target data set: in this case, the highest improvement was achieved by PA+IL. A similar trend can be observed after the third round.

Notably, our modified RF algorithm versions often outperformed the original ones, especially when MSMT was the target data set. An exception is represented by PA+IL, which did not outperform PA when Market or Duke were the target data sets. This is mainly related to the used γ value (0.5), which has not been tuned on the target data sets and turned out to be not optimal for small galleries. Nevertheless, the analysis performed in Section 5.3 demonstrates that with a suitable γ value, the incremental RF versions can outperform the original ones. On the other hand, the highest improvements were attained by QS+IL over QS, except when Market was the source data set and MSMT the target one.

It is also worth noting that, for all target domains, using our RF-based HITL approach on the model trained on the source domain outperformed the supervised version of the same model trained on the target domain (see the rows labelled as "supervised" in Table 2), in terms of mAP, rk-1 and rk-5. This means that, even when a large amount of labelled target data is available for model training, the considered HITL solution can be more effective, despite the fact it does not involve model training or fine-tuning on target data, but only exploits the operator's feedback on a *much smaller* amount of target data.

As expected, for a given target data set the performance of a given RF algorithm depends on the source data set. This is evident by a comparison with the baseline (Table 2). To better highlight this behaviour, Fig. 5 shows the performance of RF algorithms after the third feedback round on each target data set. Results suggest that using a source data set with better quality and a larger variability, including a larger number of cameras (orange lines in Fig. 5), can be beneficial to the generalisation capability of the underlying model.

Returning to the comparison between the proposed versions of RF algorithms and the original ones, we observe that the former performed very well in the most unfavourable situations, i.e., when there are very few or even no positive images among the top-K ones. To highlight this trend, Fig. 6 shows different plots with the positions of the first ten true matches within the ranked list, after each feedback round, for the original and the modified RF algorithms, on two different queries (for the sake of brevity, we consider a single cross-data set experiment, with Market as the source and Duke as the target). The two queries are shown on the left of Fig. 6, and correspond to two very different cases: many positive images (query 1) and no positive image (query 6) present in the top-K ranks ($K = 50$). The latter case represents a query identity that the source model considers dissimilar to the other positive images, for various reasons related to the underlying feature representation.

All RF algorithms, except for QS, were capable of bringing all ten positive images to the first ten ranks for query 1. M-RS, PA and PA+IL achieved this result since the first round. For query 6, instead, the improvements were less marked (in the case of RS the results even worsened) and required more feedback rounds. Still, we point out that all the modified RF algorithms performed better than their original versions, and in particular, M-RS was capable of bringing all ten positive images to the first ten ranks. This is a relevant result, especially considering that in this case no true match was present among the top-50 images, which were therefore considered automatically as non-relevant (negatives).

Table 2
Results attained by RF algorithms at each round (R) of cross-data set experiments. Best results in each column are highlighted in bold.

RF	R	Source: Market - Target: Duke					Source: Duke - Target: Market					Source: Market - Target: MSMT				
		mAP	rk-1	rk-5	rk-10	rk-20	mAP	rk-1	rk-5	rk-10	rk-20	mAP	rk-1	rk-5	rk-10	rk-20
Baseline		30.88	53.41	67.06	72.49	77.24	30.44	62.2	78.24	83.31	87.65	6.44	19.08	29.14	34.39	40.22
QS	1	47.81	73.92	80.16	82.23	84.61	47.23	83.64	89.07	90.32	92.1	12.5	38.01	43.83	46.5	49.76
	2	51.83	77.87	83.35	84.92	86.27	50.89	85.72	91.21	92.16	92.99	14.39	41.96	47.99	49.96	51.81
	3	53.08	78.77	84.07	85.41	86.62	52.04	86.07	91.45	92.61	93.23	15.05	43.0	49.22	50.88	52.38
QS+IL	1	46.65	72.53	79.85	81.69	84.25	46.17	83.17	88.84	90.59	91.89	15.83	44.0	52.11	55.42	59.0
	2	54.0	79.67	83.98	85.73	87.43	53.57	87.74	91.57	92.37	93.08	18.71	51.17	57.31	59.12	60.89
	3	57.73	82.23	86.27	87.25	88.11	57.29	89.64	92.4	92.87	93.56	20.3	53.89	58.86	60.24	61.59
RS	1	55.08	79.98	82.36	83.75	85.1	57.48	89.22	90.32	91.06	91.63	16.39	45.36	48.7	50.4	52.38
	2	66.08	87.21	88.15	88.91	89.68	69.09	92.43	93.05	93.35	93.65	23.59	55.28	57.09	58.35	59.76
	3	72.5	90.48	91.16	91.47	92.19	75.81	94.06	94.51	94.63	94.74	28.84	61.68	62.61	63.28	64.34
M-RS	1	55.7	80.61	84.92	86.54	87.97	57.9	89.1	91.39	92.28	93.05	16.08	44.79	49.66	52.27	55.01
	2	68.53	89.18	90.62	91.38	92.5	71.29	93.02	94.18	94.6	95.67	24.49	57.37	60.21	61.82	63.68
	3	75.06	92.32	93.22	93.4	**93.85**	78.8	95.28	95.61	**95.87**	**96.23**	30.18	63.64	65.36	66.4	67.88
PA	1	58.24	81.06	85.5	87.34	88.78	61.08	88.9	92.19	93.47	94.6	17.19	46.15	51.71	54.58	57.93
	2	73.86	90.75	91.52	92.06	92.68	79.27	94.89	95.16	95.37	95.67	28.99	61.53	63.03	64.23	65.77
	3	**79.88**	**93.0**	**93.31**	**93.49**	93.81	**84.91**	**95.69**	**95.78**	95.81	96.17	35.58	66.69	67.15	67.69	68.49
PA+IL	1	11.28	10.77	24.82	37.34	53.64	10.45	10.45	26.07	40.56	57.36	31.98	60.56	62.81	64.24	66.53
	2	56.31	80.61	82.18	83.39	84.78	59.84	89.96	90.77	91.27	92.04	39.24	74.41	75.43	76.2	77.23
	3	64.77	85.64	85.91	86.4	86.76	69.7	92.43	92.49	92.52	92.73	**42.6**	**77.49**	**77.72**	**77.91**	**78.33**
RF	R	Source: MSMT - Target: Duke					Source: MSMT - Target: Market					Source: Duke - Target: MSMT				
		mAP	rk-1	rk-5	rk-10	rk-20	mAP	rk-1	rk-5	rk-10	rk-20	mAP	rk-1	rk-5	rk-10	rk-20
Baseline		28.49	47.17	67.01	74.33	79.94	38.33	69.77	83.94	88.18	91.66	8.09	24.7	36.44	42.23	48.15
QS	1	67.9	88.15	91.02	91.79	93.0	57.63	89.9	93.05	94.03	95.4	15.37	46.62	52.83	55.51	58.77
	2	75.46	92.32	93.94	94.21	94.7	62.16	92.13	94.69	95.13	95.9	17.64	51.14	57.67	59.48	61.41
	3	77.82	93.27	94.61	94.79	95.24	63.52	92.19	94.92	95.4	95.96	18.45	52.23	58.74	60.42	62.12
QS+IL	1	64.05	86.71	89.45	90.57	91.61	56.53	89.55	93.11	94.09	95.13	19.45	52.83	61.26	64.91	68.59
	2	75.06	92.32	92.95	93.27	93.76	64.74	93.32	95.16	95.58	96.14	22.94	61.03	66.8	68.48	70.43
	3	79.0	93.85	94.48	94.61	94.84	68.68	94.48	95.61	95.9	96.32	24.83	63.84	68.47	69.65	71.11
RS	1	68.16	89.5	90.66	91.07	91.52	66.47	92.9	93.94	94.21	94.69	20.05	53.93	57.16	58.95	61.11
	2	78.21	92.15	92.73	92.91	93.13	78.02	95.25	95.55	95.78	96.2	28.55	63.93	65.59	66.75	68.09
	3	82.93	94.12	94.3	94.52	94.61	83.84	96.44	96.7	96.94	97.15	34.47	69.58	70.47	71.14	72.03
M-RS	1	68.24	88.91	90.48	90.93	91.52	68.02	92.9	94.06	94.92	95.55	19.85	53.08	58.14	60.79	63.56
	2	78.82	92.15	93.22	93.49	93.9	80.61	95.64	96.11	96.59	97.21	29.39	65.66	68.18	69.59	71.46
	3	83.87	94.34	94.75	95.06	95.2	86.88	97.12	97.33	97.6	97.89	35.78	71.7	73.36	74.11	75.39
PA	1	69.85	89.27	90.8	91.61	92.5	71.55	93.29	95.43	96.2	96.88	21.06	54.87	60.2	62.83	65.91
	2	82.36	93.67	94.25	94.48	94.66	86.44	97.15	97.3	97.48	97.71	34.0	68.94	70.42	71.41	72.86
	3	**86.94**	**95.06**	**95.29**	**95.42**	**95.42**	**91.01**	**97.71**	**97.77**	**97.8**	**97.95**	41.14	73.44	73.97	74.33	74.97
PA+IL	1	11.46	6.06	18.99	33.17	57.32	11.3	8.64	25.95	40.32	59.03	35.37	63.64	65.82	67.42	70.05
	2	65.24	87.21	88.11	89.09	89.68	67.08	92.67	93.2	93.5	93.94	43.48	79.28	80.0	80.58	81.58
	3	74.54	90.04	90.48	90.62	90.84	76.37	94.27	94.39	94.42	94.6	**47.07**	**81.9**	**82.06**	**82.23**	**82.62**
Supervised		76.1	89.5	95.0	96.5	97.5	85.7	95.4	98.7	99.1	99.4	43.4	71.3	83.4	87.1	90.1

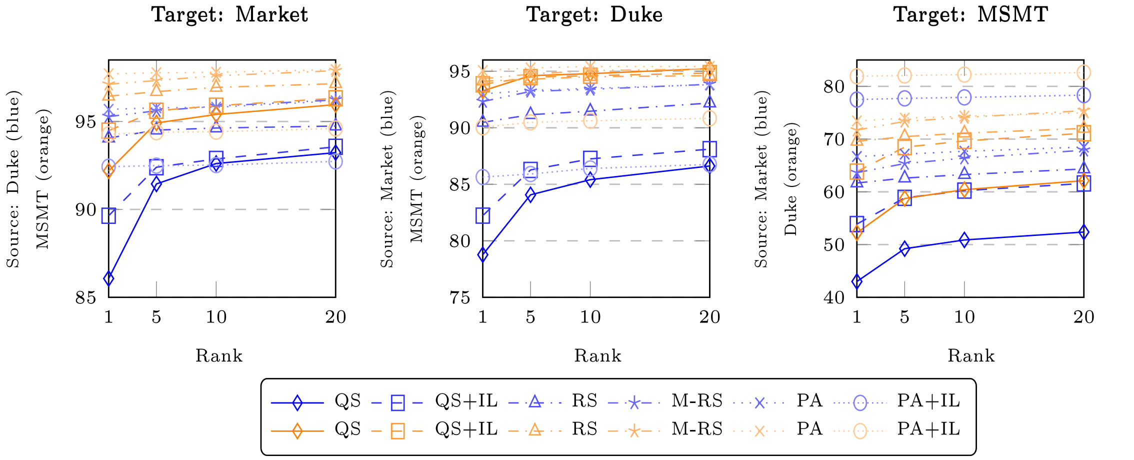

Fig. 5. Cross-data set CMC curves of RF algorithms after the third feedback round. Each plot corresponds to a different target data set. Blue and orange lines correspond to two different source data sets.

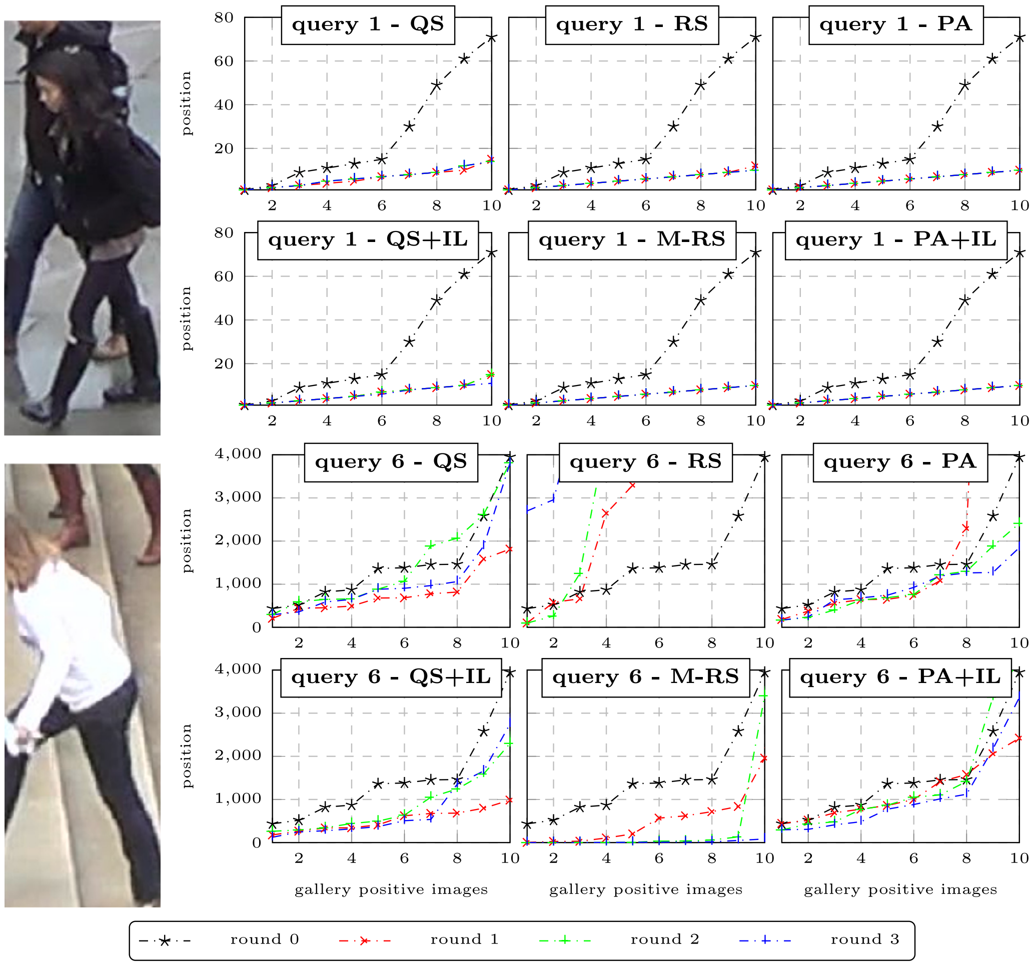

Fig. 6. Ranks of the top ten true matches for queries 1 and 6 of Duke (target), using a model trained on Market (source), for the original and the modified RF algorithms, at each feedback round (round 0 denotes the results before RF).

5.2. Comparison between RF algorithms and DS methods

In the following, we focus only on the proposed variants of RF algorithms, since they outperformed the original ones as shown above. We compare them with state-of-the-art methods based on the DA, UDA, DT and DG approaches (Section 2). We start with DT and DG methods, which, similarly to our RF-based HITL implementations, do not use target data for model training or fine-tuning; we then consider DA and UDA methods that *do* use target data to this aim.

Comparison with DT and DG. The comparison with DT methods is reported in Table 3, for six source/target combinations (the results of RF-based HITL methods are taken from Table 3). As expected, for a given target data set, the performance of DT methods strongly depends on the source one; generally, a better performance is attained when a larger source data set is used. As expected, RF-based HITL methods always outperformed DT ones, and always by a considerable amount.

The comparison with DG methods is reported in Fig. 7. Only the mAP and rank-1 CMC accuracy are considered in this case, since these are the only measures reported in the respective papers.

Although the considered DG methods used a different number of source data sets (see Section 4), only OsNet and MMFA-AAE exhibited comparable performances. This might be due to the fact that a subset of source data sets used by MMFA-AAE was employed to train OsNet. In contrast, the lower performances obtained by DomainMix can be due to the use of a synthetic and a real data set during training. Although the use of synthetic data may in principle be helpful, the resulting performance can be affected by the domain gap between real and synthetic images.

Fig. 7 shows that RF-based HITL methods also outperformed DG methods both in rank-1 CMC and in mAP, despite using a single source data set. The highest improvement was attained when Duke was the source domain; for instance, PA+IL outperformed MMFA-AAE by about 36% in rankk-1 and by about 26% in mAP, whereas a slightly lower improvement was achieved when Market was the source domain. A similar trend can be observed for the other RF-based HITL methods. The above results show that a limited amount of operator's feedback during operation (i.e., $K = 50$) can be very beneficial to improve the Re-Id performance of models trained on one source domain.

Comparison with DA and UDA. Table 3 reports the performance of DA and UDA methods. We further subdivided the latter into methods based on synthetic image generation (UDA-IG, see Section 2.2) and methods based on different approaches (UDA-other).

All the RF-based HITL methods significantly outperformed DA methods in all target domains in terms of both mAP and CMC curve, even if the former did not use target data to refine the source model. In most cases, at least one of the RF-based HITL methods also outperformed UDA methods in terms of mAP (except when Duke is the source and Market is the target domain) and of rank-1 accuracy. This shows that the considered RF-based HITL methods, even the simplest ones (QS and QS+IL), can attain a similar or even better re-identification accuracy than DA and UDA methods.

To sum up, the above results provide evidence that, despite its simplicity, the considered HITL approach to Re-Id based on RF algorithms can attain a competitive performance with respect to DA, UDA, DT and DG approaches, in the considered cross-domain scenario, which confirms that it is a further, valid alternative against DS. Our results

Table 3
Cross-data set performance of methods against DS used for comparison. The best result in each column is highlighted in bold while the second best is underlined.

Type	Method	Source: Market - Target: Duke					Source: Duke - Target: Market					Source: Market - Target: MSMT				
		mAP	rk-1	rk-5	rk-10	rk-20	mAP	rk-1	rk-5	rk-10	rk-20	mAP	rk-1	rk-5	rk-10	rk-20
DT	ADIN (Yuan et al., 2020)	–	–	–	–	–	27.4	57.2	73.0	80.0	–	–	–	–	–	–
	OSNet (Zhou et al., 2019)	30.5	52.4	66.1	71.2	–	30.6	61.0	77.0	82.5	–	8.2	23.5	34.5	40.2	–
DA	DAReID (Xu et al., 2021)	30.3	51.1	–	–	–	33.0	61.7	–	–	–	–	–	–	–	–
UDA-other	MTNet (Chen et al., 2023b)	–	–	–	–	–	–	–	–	–	–	32.1	59.6	72.2	77.0	–
	CACHE (Liu et al., 2022)	71.7	83.5	91.4	93.9	–	83.1	93.4	97.5	98.2	–	31.3	58.0	69.8	74.5	–
	JL (Feng et al., 2021)	67.9	81.3	89.2	91.5	–	78.6	90.6	96.3	97.8	–	21.2	47.3	60.3	65.5	–
	TALM-IRM (Li et al., 2021)	41.34	63.53	76.62	–	–	39.95	73.08	86.34	–	–	11.24	30.87	43.53	–	–
	Theory&Practice (Song et al., 2020)	49.0	68.4	80.1	83.5	–	53.7	75.8	89.5	93.2	–	–	–	–	–	–
	MMT[a] (Ge et al., 2020)	64.8	82.7	90.2	92.3	**94.0**	74.8	91.9	97.0	**97.9**	**98.9**	14.5	36.4	49.2	54.8	60.7
	D-MMD (Mekhazni et al., 2020)	46.0	63.5	78.8	83.9	–	48.8	70.6	87.0	91.5	–	13.5	29.1	46.3	54.1	–
	PAST[a] (Zhang et al., 2019)	54.26	78.41	86.49	89.09	91.43	54.62	84.29	92.99	95.43	96.97	–	–	–	–	–
	CASCL (Wu et al., 2019)	30.5	51.5	66.7	71.7	–	35.6	64.7	80.2	85.6	–	–	–	–	–	–
UDA IG	MDJL (Chen et al., 2023a)	62.8	78.6	86.6	88.7	–	59.8	80.3	87.4	89.9	–	13.4	34.3	44.5	50.6	–
	IPES-GAN (Verma et al., 2023)	32.9	53.5	69.1	73.1	–	33.6	64.1	79.3	83.1	–	5.9	18.4	28.9	34.4	–
	DPCFG (Song et al., 2022)	73.7	85.7	92.8	**94.4**	–	**85.4**	94.2	**97.8**	**98.7**	–	36.9	65.3	76.0	**79.8**	–
	MLMS (Tang et al., 2022)	65.1	79.1	–	–	–	74.5	89.7	–	–	–	25.9	52.5	–	–	–
	STReID (Chong et al., 2021)	29.2	52.3	65.9	71.1	–	31.6	62.3	79.1	84.4	–	–	–	–	–	–
	SILC (Ainam et al., 2021)	50.3	68.5	80.2	85.4	88.6	61.8	80.7	90.1	93.0	95.6	10.9	27.8	38.1	45.8	–
	AAAN (Zhang et al., 2020b)	58.4	70.7	82.4	85.0	–	67.6	84.8	92.6	94.8	–	15.1	30.8	39.3	43.8	–
	CVSE (Zhou et al., 2021)	56.1	75.3	82.9	85.4	–	63.2	84.1	92.8	95.0	–	–	–	–	–	–
	DAL (Zhang et al., 2020a)	57.3	75.2	84.3	87.1	–	68.6	86.4	94.6	96.4	–	16.9	42.9	56.1	61.7	–
	DG-Net++ (Zou et al., 2020)	63.8	78.9	87.8	90.4	–	61.7	82.1	90.2	92.7	–	22.1	48.4	60.9	66.1	–
	ECN[a] (Zhong et al., 2019)	43.0	67.9	80.1	83.8	86.8	41.7	73.5	86.9	90.9	94.2	7.5	21.7	32.3	37.8	44.0
	PT-GAN (Wei et al., 2018)	–	–	–	–	–	–	–	–	–	–	2.9	10.2	–	24.4	–
RF	QS+IL	57.73	82.23	86.27	87.25	88.11	57.29	89.64	92.4	92.87	93.56	20.3	53.89	58.86	60.24	61.59
	M-RS	**75.06**	**92.32**	**93.22**	93.4	93.85	78.8	**95.28**	95.61	95.87	96.23	30.18	63.64	65.36	66.4	67.88
	PA+IL	64.77	85.64	85.91	86.4	86.76	69.7	92.43	92.49	92.52	92.73	**42.6**	**77.49**	**77.72**	77.91	**78.33**
Type	Method	Source: MSMT - Target: Duke					Source: MSMT - Target: Market					Source: Duke - Target: MSMT				
		mAP	rk-1	rk-5	rk-10	rk-20	mAP	rk-1	rk-5	rk-10	rk-20	mAP	rk-1	rk-5	rk-10	rk-20
DT	ADIN (Yuan et al., 2020)	39.1	60.7	74.7	–	–	30.3	59.1	75.4	–	–	–	–	–	–	–
	OSNet (Zhou et al., 2019)	52.7	71.1	83.3	86.4	–	43.3	70.1	84.1	88.6	–	10.2	30.3	42.2	47.9	–
UDA-other	MTNet (Chen et al., 2023b)	–	–	–	–	–	82.7	93.0	97.1	**98.6**	–	–	–	–	–	–
	CACHE (Liu et al., 2022)	72.6	84.2	92.0	94.4	–	83.6	92.8	96.8	97.9	–	31.0	57.8	70.4	75.3	–
	JL (Feng et al., 2021)	–	–	–	–	–	–	–	–	–	–	24.6	53.5	65.2	70.2	–
	TALM-IRM (Li et al., 2021)	48.67	68.35	80.89	–	–	42.94	74.55	87.55	–	–	14.22	39.04	51.51	–	–
	MMT[a] (Ge et al., 2020)	68.1	85.3	91.6	93.2	94.4	72.3	90.9	96.0	97.6	**98.6**	17.9	44.0	56.9	62.2	67.4
	D-MMD (Mekhazni et al., 2020)	51.6	68.8	82.6	87.1	–	50.8	72.8	88.1	92.3	–	15.3	34.4	51.1	58.5	–
	CASCL (Wu et al., 2019)	37.8	59.3	73.2	77.8	–	35.5	65.4	80.6	86.2	–	–	–	–	–	–
UDA IG	MDJL (Chen et al., 2023a)	–	–	–	–	–	–	–	–	–	–	17.1	40.3	51.2	56.3	–
	IPES-GAN (Verma et al., 2023)	–	–	–	–	–	–	–	–	–	–	6.5	20.6	31.0	36.4	–
	DPCFG (Song et al., 2022)	–	–	–	–	–	–	–	–	–	–	37.6	66.8	77.3	81.1	–
	MLMS (Tang et al., 2022)	–	–	–	–	–	–	–	–	–	–	31.4	60.9	–	–	–
	SILC (Ainam et al., 2021)	–	–	–	–	–	–	–	–	–	–	12.6	33.1	45.2	48.0	–
	AAAN (Zhang et al., 2020b)	–	–	–	–	–	–	–	–	–	–	17.5	35.4	44.5	48.5	–
	DAL (Zhang et al., 2020a)	–	–	–	–	–	–	–	–	–	–	15.4	40.4	53.7	59.5	–
	DG-Net++ (Zou et al., 2020)	58.2	75.2	73.6	86.9	–	64.6	83.1	91.5	94.3	–	22.1	48.8	60.9	65.9	–
	ECN[a] (Zhong et al., 2019)	43.0	68.6	79.9	83.2	86.5	44.6	77.5	89.6	93.1	95.1	8.9	25.3	36.3	42.1	47.9
	PT-GAN (Wei et al., 2018)	–	–	–	–	–	–	–	–	–	–	3.3	11.8	–	27.4	–
RF	QS+IL	79.0	93.85	94.48	94.61	94.84	68.68	94.48	95.61	95.9	96.32	24.83	63.84	68.47	69.65	71.11
	M-RS	**83.87**	**94.34**	**94.75**	**95.06**	**95.2**	**86.88**	**97.12**	**97.33**	97.6	97.89	35.78	71.7	73.36	74.11	75.39
	PA+IL	74.54	90.04	90.48	90.62	90.84	76.37	94.27	94.39	94.42	94.6	**47.07**	**81.9**	**82.06**	**82.23**	**82.62**

[a] Denotes results reproduced by the authors.

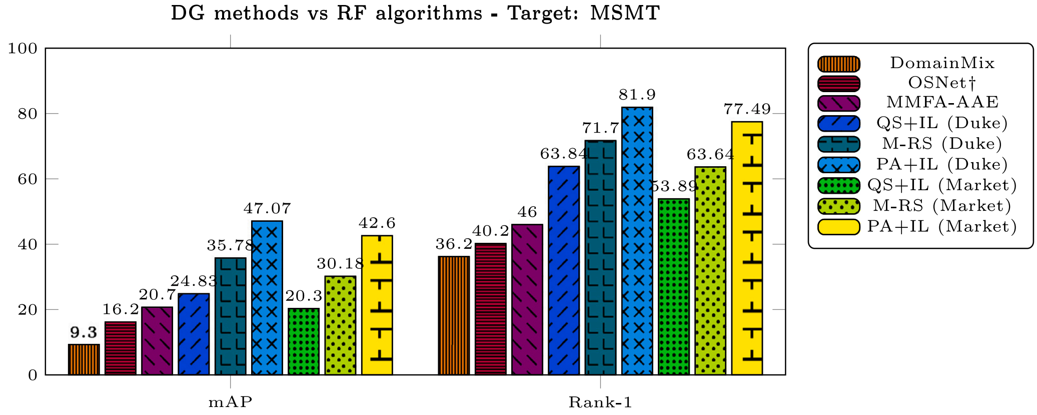

Fig. 7. Comparison of mAP (left) and rank-1 CMC (right) attained by DG methods (DomainMix, OSNet, MMFA-AAE) and by our RF algorithms. † results taken from the appendix of https://arxiv.org/pdf/1910.06827.pdf.

also point out that the effectiveness of UDA and DG methods, as well as RF-based HITL methods, depend on how much the source domain is representative of the target one (see Table 2). Indeed, when MSMT is the target data set, which is very different from Market and Duke, the performance of UDA methods decreases. Nevertheless, RF-based HITL methods still provide a significant improvement with respect to the baseline also on MSMT.

Finally, as mentioned in Section 4, a direct comparison with existing HITL methods for Re-Id dealing with the considered cross-domain scenario (see Section 2.4) was not possible. Nevertheless, for the sake of completeness, we report that the method of Wang et al. (2016) attained a 0.78 rank-1 accuracy on Market, which is the only data set in common with this work, although a smaller number of queries (300) and only a subset of the gallery (1000 images) was used therein.

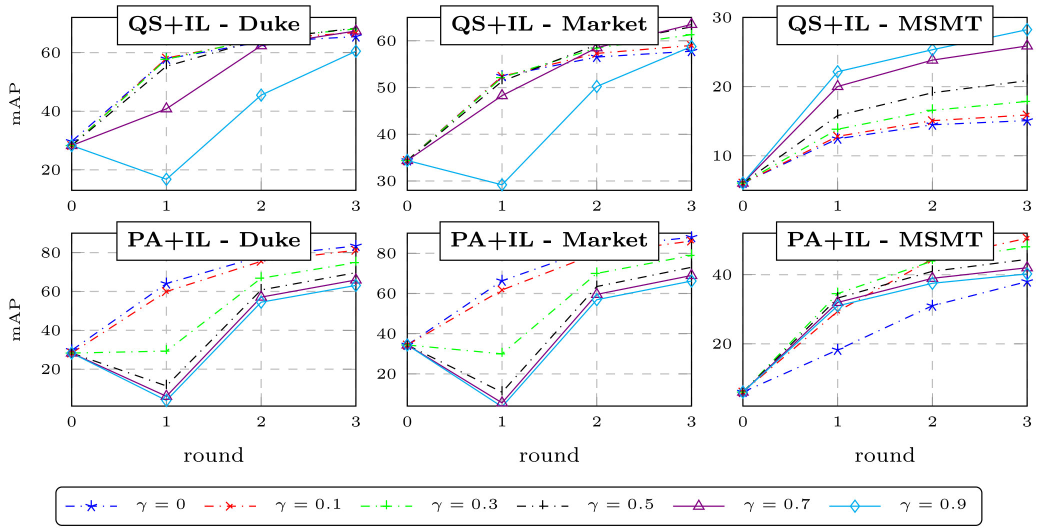

Fig. 8. Average mAP attained on each target data set over the two corresponding source data sets by QS+IL and PA+IL, using different values of γ, at each feedback round (round 0 denotes the performance before RF).

5.3. Parameter analysis of incremental RF algorithms

Our incremental RF algorithms QS+IL and PA+IL use a hyper-parameter $\gamma \in [0,1]$ to weigh the contribution of their components related to the current query and to previous ones. As explained in Section 4, in the above experiments we did not tune γ but set it to a default value of 0.5; it is nevertheless interesting to evaluate how γ affects the performance of QS+IL and PA+IL, to derive guidelines on how to set it in real applications. To this aim, we evaluated the average mAP attained on each target data set over the two corresponding source data sets, for $\gamma = 0.1, 0.3, 0.5, 0.7, 0.9$ (note that $\gamma = 0$ corresponds to the original QS and PA). Fig. 8 reports the results.

Both QS+IL and PA+IL exhibited a very similar behaviour. In particular, on the MSMT target data set, they outperformed the respective, original RF algorithms (QS and PA, $\gamma = 0$) for all the considered γ values, often by a large amount; moreover, for QS+IL, the higher the γ value, the higher the mAP, whereas no clear trend emerged under this viewpoint for PA+IL. Instead, on the Market and Duke target data sets, QS+IL and PA+IL generally attained a similar mAP as QS and PA for low γ values, and a lower mAP for higher values; often, a notable performance gap can also be observed at the first feedback round for some γ values; however, QS+IL outperformed QS at the third round on Market, for several γ values. It can also be seen that the performance tends to increase over different rounds for all γ values, with the exceptions mentioned above for the first round.

Overall, based on these results, our choice of $\gamma = 0.5$ seems reasonable as a default one for unknown target domains, i.e., where no target data is available for validation. The results on Duke and Market, especially the ones obtained by PA+IL, also suggest that a lower γ value may be better for small galleries. At the same time, the results on MSMT suggest that a higher γ value may be better for very large galleries, typical of real applications. We also point out that for both QS+IL and PA+IL, the online DA component (i.e., the second term of Eqs. (7) and (9)) may require several queries and feedback rounds before providing a significant contribution; therefore, we envisage that changing the value of γ over time, starting from lower values and gradually increasing it, could be an effective solution in practical applications.

5.4. Accuracy-user effort trade-off and amount of target data involved

We remind the reader that, in the considered application scenario, the HITL approach aims at adapting a Re-Id system to the target domain unseen during design, based on a limited amount of feedback provided by the operator on target data processed during operation. Similarly to previous work (Wang et al., 2016), the above experiments are based on $K = 50$, which can be considered a rather small value for investigation scenarios involving a large number of surveillance videos, and thus very large template galleries (Wang et al., 2016). However, in principle, the higher the value of K, the higher the effectiveness of RF-based HITL methods, due to the higher number of feedback, but at the same time the higher the operator's effort. To better evaluate this trade-off, we repeated the same experiments of Section 5.1 for $K = 20{,}100$. The results are reported in Table 4, together with a comparison with the ones attained by $K = 50$ taken from Table 2; for the sake of brevity, we only consider the results after the third feedback round. As expected, the performance of all RF-based HITL methods increased for increasing values of K, in all target domains. It can be seen that the performance gap between $K = 20$ and $K = 50$ is about 5% in mAP for QS and QS+IL, whereas for the other RF algorithms it exceeds 10%. Notably, by increasing K from 20 to 100 the highest improvement was attained when MSMT was the target domain, corresponding to the largest template gallery among the considered data sets: in this case the average improvement was about 5% in both mAP and CMC curve. In the other cases (i.e., Market or Duke as the target), the average improvement was about 4% in mAP and 3% in the CMC curve.

Let us now evaluate the user effort as a function of K. We argued in Section 3.3 that, in real-world applications, the user effort required by our multi-feedback protocol is not significantly higher compared to the single-feedback protocol used by previous work (Wang et al., 2016), for the same value of K. To evaluate it on the considered data sets, whose gallery sets are likely to be much smaller than the ones of real applications, for each RF-based HITL method Table 5 shows the average number of true matches present in the top-K ranks at each feedback round for a given query, which corresponds to the amount of feedback required to the user. Here we consider $K = 20, 50, 100$.

Note that, for a given target data set, the number of true matches in the first feedback round depends only on the source model and is, therefore, identical for all RF-based HITL methods. The highest number of true matches (user's feedback) almost always occurred in the first round; the only exception is PA+IL when MSMT was the target data set. In subsequent rounds, the number of feedback tended to decrease. In most cases, less than five and four true matches were present in the second and third rounds, respectively (note that a true match selected in a round does not have to be selected in subsequent rounds if it remains among the top-K ranks). Over all three rounds, the average

Table 4
Comparison between the cross-data set performance attained by RF algorithms after the 3rd round, for $K = 20$, $K = 50$ and $K = 100$.

RF	K	Source: Market - Target: Duke					Source: Duke - Target: Market					Source: Market - Target: MSMT				
		mAP	rk-1	rk-5	rk-10	rk20	mAP	rk-1	rk-5	rk-10	rk-20	mAP	rk-1	rk-5	rk-10	rk-20
QS	20	49.92	76.08	80.79	82.23	84.43	48.54	84.03	87.71	88.9	91.21	13.19	38.81	42.77	44.47	48.04
	50	53.08	78.77	84.07	85.41	86.62	52.04	86.07	91.45	92.61	93.23	15.05	43.0	49.22	50.88	52.38
	100	54.31	79.26	86.31	88.06	89.09	53.5	86.28	93.2	94.48	95.34	16.12	45.3	53.56	55.85	57.4
QS+IL	20	53.41	78.73	82.09	84.02	86.49	52.5	86.55	88.84	89.85	92.19	17.39	48.69	52.17	54.07	56.97
	50	57.73	82.23	86.27	87.25	88.11	57.29	89.64	92.4	92.87	93.56	20.3	53.89	58.86	60.24	61.59
	100	60.06	84.34	88.6	89.59	90.48	60.41	90.62	94.54	95.13	95.69	22.17	57.02	63.03	64.76	66.25
RS	20	61.83	85.28	86.62	87.43	88.29	63.76	90.35	91.03	91.48	92.31	20.39	51.39	53.16	54.57	56.55
	50	72.5	90.48	91.16	91.47	92.19	75.81	94.06	94.51	94.63	94.74	28.84	61.68	62.61	63.28	64.34
	100	78.47	93.09	93.58	93.76	93.94	82.66	96.17	96.29	96.44	96.64	35.55	68.9	69.48	69.89	70.76
M-RS	20	67.24	87.75	89.0	89.68	90.8	69.05	92.13	93.38	94.0	95.22	22.65	54.06	57.18	59.16	61.75
	50	75.06	92.32	93.22	93.4	93.85	78.8	95.28	95.61	95.87	96.23	30.18	63.64	65.36	66.4	67.88
	100	79.08	94.7	95.42	95.56	95.83	82.9	96.76	97.33	97.42	97.51	35.61	69.83	71.52	72.14	72.79
PA	20	70.79	89.05	90.53	90.98	91.97	75.44	92.99	93.44	93.88	94.98	25.61	55.83	57.98	59.45	62.02
	50	79.88	93.0	93.31	93.49	93.81	84.91	95.69	95.78	95.81	96.17	35.58	66.69	67.15	67.69	68.49
	100	84.38	95.29	95.33	95.38	95.42	89.45	97.33	97.39	97.39	97.45	42.54	72.85	73.12	73.32	73.6
PA+IL	20	55.17	78.5	80.39	81.51	83.44	57.7	87.2	87.65	88.24	89.85	32.76	70.23	71.11	71.87	73.4
	50	64.77	85.64	85.91	86.4	86.76	69.7	92.43	92.49	92.52	92.73	42.6	77.49	77.72	77.91	78.33
	100	71.72	90.22	90.31	90.44	90.71	78.23	94.89	94.89	94.95	94.98	50.37	83.6	83.73	83.77	83.85
RF	K	Source: MSMT - Target: Duke					Source: MSMT - Target: Market					Source: Duke - Target: MSMT				
		mAP	rk-1	rk-5	rk-10	rk20	mAP	rk-1	rk-5	rk-10	rk-20	mAP	rk-1	rk-5	rk-10	rk-20
QS	20	73.76	90.8	91.38	91.61	92.55	59.6	90.41	92.37	93.14	94.71	16.45	48.5	52.69	54.4	57.63
	50	77.82	93.27	94.61	94.79	95.24	63.52	92.19	94.92	95.4	95.96	18.45	52.23	58.74	60.42	62.12
	100	81.68	95.96	97.08	97.26	97.31	65.05	92.58	96.26	96.82	97.09	19.74	54.28	63.14	65.25	66.88
QS+IL	20	72.75	89.0	89.81	90.53	91.34	63.52	91.75	93.11	94.0	95.25	21.74	58.81	62.5	64.29	66.76
	50	79.0	93.85	94.48	94.61	94.84	68.68	94.48	95.61	95.9	96.32	24.83	63.84	68.47	69.65	71.11
	100	82.35	96.18	96.86	96.99	97.08	71.45	95.58	96.97	97.24	97.39	26.95	67.33	72.59	74.11	75.33
RS	20	71.5	86.85	87.52	87.75	88.33	72.48	93.53	94.3	94.74	95.34	25.16	60.13	61.88	63.17	64.96
	50	82.93	94.12	94.3	94.52	94.61	83.84	96.44	96.7	96.94	97.15	34.47	69.58	70.47	71.14	72.03
	100	87.48	96.05	96.27	96.36	96.5	89.69	97.89	97.98	98.1	98.16	42.06	75.95	76.51	76.88	77.5
M-RS	20	73.84	88.11	89.18	89.59	89.99	78.12	94.63	95.43	95.99	96.44	27.51	62.83	65.53	67.33	69.45
	50	83.87	94.34	94.75	95.06	95.2	86.88	97.12	97.33	97.6	97.89	35.78	71.7	73.36	74.11	75.39
	100	86.91	96.14	96.63	96.72	96.77	90.36	98.31	98.55	98.63	98.72	41.69	77.51	78.84	79.37	79.95
PA	20	76.98	89.72	90.08	90.62	90.98	83.06	95.55	95.87	96.2	96.82	30.75	65.13	66.64	67.84	69.82
	50	86.94	95.06	95.29	95.42	95.42	91.01	97.71	97.77	97.8	97.95	41.14	73.44	73.97	74.33	74.97
	100	90.71	97.13	97.26	97.26	97.31	94.26	98.57	98.57	98.63	98.66	48.15	78.72	78.94	79.06	79.23
PA+IL	20	64.21	83.12	84.02	84.78	85.5	65.2	90.17	90.74	91.18	92.34	36.91	75.51	76.24	77.17	78.61
	50	74.54	90.04	90.48	90.62	90.84	76.37	94.27	94.39	94.42	94.6	47.07	81.9	82.06	82.23	82.62
	100	81.19	94.03	94.08	94.12	94.17	83.9	96.67	96.67	96.73	96.76	54.54	86.29	86.46	86.51	86.61

Table 5
Average feedback count at each RF round (R) on each target (T) data set.

RF/round	Source: Market - Target: Duke									Source: Duke - Target: Market									Source: Market - Target: MSMT								
	k = 20			k = 50			k = 100			k = 20			k = 50			k = 100			k = 20			k = 50			k = 100		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
QS	6.9	2.5	0.7	8.9	2.9	0.8	10.2	3.1	0.8	6.1	2.2	0.7	7.8	2.7	0.7	9.2	2.9	0.8	4.3	1.7	0.5	5.7	2.3	0.7	6.8	2.7	0.8
QS+IL	6.94	2.39	1.11	8.86	2.82	1.26	10.21	3.00	1.22	6.10	2.13	1.07	7.83	2.62	1.19	9.16	2.83	1.27	4.32	2.63	0.76	5.74	3.61	1.00	6.85	4.22	1.13
RS	6.94	3.10	1.51	8.86	3.85	1.82	10.21	4.18	1.85	6.10	3.24	1.71	7.83	4.01	2.04	9.16	4.15	1.99	4.32	2.39	1.42	5.74	3.66	2.46	6.85	4.52	3.21
M-RS	6.94	3.55	1.94	8.86	4.32	2.19	10.21	4.69	2.02	6.10	3.37	2.22	7.83	4.39	2.41	9.16	4.71	2.12	4.32	2.55	1.86	5.74	3.92	2.89	6.85	4.93	3.52
PA	6.94	3.46	2.31	8.86	4.61	2.34	10.21	4.92	2.18	6.10	3.38	2.65	7.83	4.79	2.71	9.16	5.20	2.14	4.32	2.49	2.15	5.74	4.16	3.50	6.85	5.55	4.31
PA+IL	6.94	0.37	2.82	8.86	1.27	3.11	10.21	2.31	2.71	6.10	0.32	2.89	7.83	1.27	3.43	9.16	2.42	2.94	4.32	5.51	1.53	5.74	9.28	2.13	6.85	12.03	2.37
RF/round	Source: MSMT - Target: Duke									Source: MSMT - Target: Market									Source: Duke - Target: MSMT								
	k = 20			k = 50			k = 100			k = 20			k = 50			k = 100			k = 20			k = 50			k = 100		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
QS	5.82	5.78	1.57	8.60	5.90	1.46	9.88	6.19	1.40	6.93	2.62	0.76	8.83	3.087	0.84	10.15	3.22	0.81	4.69	1.99	0.62	6.22	2.70	0.81	7.49	3.14	0.92
QS+IL	5.82	5.17	1.79	8.209	5.8	1.79	9.88	5.809	1.66	6.93	2.53	1.18	8.83	3.01	1.31	10.15	3.10	1.27	4.69	3.02	0.87	6.22	4.14	1.15	7.49	4.85	1.29
RS	5.82	5.05	1.70	8.60	5.51	1.91	9.88	5.89	1.83	6.93	3.51	1.81	8.83	4.31	2.09	10.15	4.37	1.87	4.69	2.69	1.58	6.22	4.25	2.77	7.49	5.22	3.68
M-RS	5.82	5.27	1.99	8.60	5.87	2.08	9.88	6.21	1.80	6.93	3.73	2.29	8.83	4.73	2.33	10.15	4.88	1.94	4.69	2.89	2.06	6.22	4.53	3.19	7.49	5.66	3.81
PA	5.82	5.21	2.38	8.60	5.97	2.29	9.88	6.41	1.97	6.93	3.83	2.57	8.83	5.14	2.42	10.15	5.27	1.86	4.69	2.85	2.37	6.22	4.786	3.78	7.49	6.21	4.55
PA+IL	5.82	0.51	4.18	8.20	1.69	4.26	9.88	2.90	3.52	6.93	0.29	3.07	8.83	1.18	3.60	10.15	2.34	2.93	4.69	5.87	1.62	6.22	9.84	2.23	7.49	12.54	2.42

number of feedback was less than 11, 16 and 18, respectively for $K = 20$, $K = 50$ and $K = 100$.

If we compare these values with the number of target images that are required by DA or UDA methods to attain similar performances, we see that the former are orders of magnitude lower. Indeed, reported results for DA and UDA have been attained using up to about 12,000 images (Market) and 30,000 (MSMT, see Table 1) albeit unlabelled for UDA. This means that the considered RF-based HITL approach allows the underlying Re-Id system to adapt to the target domain with comparable or better performance than the one that would be attained by DA and UDA (if target data had been available during design), at a very limited annotation (feedback) effort by the user, and without retraining or fine-tuning the source model. Furthermore, we point out again that, despite being relatively large, the template galleries of the considered benchmark data sets are likely to be much smaller than the ones of real application scenarios involving dozens of surveillance cameras and hours of video recordings: in this case, the number of true matches in the top-K ranks is likely to be even smaller than the one observed in our experiments. In particular, our results show that $K = 100$ is a suitable value for a Re-Id system in investigation scenarios: it requires a reasonably small user's effort, and the slight additional effort with respect to the other considered values of K is well rewarded by the increase in performance, as can be seen from Table 4.

6. Discussion and conclusions

In many challenging real-world application scenarios, Re-Id systems have to be used on target scenes different from the ones used for training, and the resulting domain shift can severely affect their performance. In this work, we revisited the HITL approach to Re-Id, originally proposed with different purposes, arguing that it can be a further effective solution against domain shift, besides existing solutions based on domain adaptation (either supervised or unsupervised) and domain generalisation, without requiring model refinement on target data. Under this viewpoint, HITL can be viewed as an *online* domain adaptation approach which leverages the synergy between human and machine capabilities *during system operation*. An extensive empirical evaluation of and comparison with state-of-the-art DA, UDA and DG methods confirmed the effectiveness of the HITL approach, which turned out to be competitive with, or even superior to such approaches, even in the proposed implementation based on simple relevance feedback algorithms originally devised for content-based image retrieval, as well as with novel versions of such algorithms we devised specifically for Re-Id. The considered RF-based HITL solution is very general, as it is model-agnostic, and can therefore be implemented on top of *any* Re-Id model. Moreover, its potential scope is broader than the one considered in this work: (i) Besides being an alternative to DA and UDA, as well as to DT and DG, in cross-domain scenarios, it is also *complementary* to them, i.e., it can be used *together* with these techniques with the aim of further improving their performance; (ii) Similarly, it can also be used in application scenarios not involving domain shift (i.e., where the target and source domains coincide), to improve the model's effectiveness further. In this context, an interesting research direction for future work is to investigate the combination of RF-based HITL with UDA, by exploiting either pseudo-label-based UDA methods to increase the number of true matches used to perform re-ranking, or the user's feedback to adapt the source model.

We conclude our work by pointing out the issue of the *trustworthiness* of learning-based methods, which is becoming more and more relevant due to their increasing adoption in critical applications, such as health and security, and may in principle, be addressed also leveraging the HITL approach. Trustworthiness is crucial in determining the acceptance of learning-based systems by end-users in these kinds of applications, and is undermined by two main factors (Holzinger, 2021): the lack of robustness to perturbations of input data (Kamath, Deshpande, Kambhampati Venkata, & N Balasubramanian, 2021), including domain shift issues considered in this work, and the difficulty, even of the most powerful learning methods, in explaining their predictions (Dwivedi et al., 2023; Rawal et al., 2022). The latter issue affects Re-Id systems too, as one can expect (Chen et al., 2021; Goyal, Patel, Truong, & Yanushkevich, 2021; Liao et al., 2020; Zhao, Luo, Yang, & Song, 2018). Under this viewpoint, it has recently been suggested that the HITL approach may be exploited, in the context of a human-centred design process, to promote the explainability and robustness of learning systems, and consequently to improve their reliability and trustworthiness, ensuring that humans remain in control (Holzinger, 2021). Investigating the design of HITL methods with this purpose also in the specific case of Re-Id systems is another very interesting direction for future work.

CRediT authorship contribution statement

Rita Delussu: Software, Investigation, Writing – review & editing, Visualization. **Lorenzo Putzu:** Software, Conceptualization, Methodology, Investigation, Writing – review & editing, Visualization. **Giorgio Fumera:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rita Delussu, Lorenzo Putzu, Giorgio Fumera reports financial support was provided by ESA Group. Lorenzo Putzu, Giorgio Fumera reports financial support was provided by Europe Horizons. Rita Delussu, Lorenzo Putzu, Giorgio Fumera reports a relationship with University of Cagliari that includes: employment.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported by the projects "Law Enforcement agencies human factor methods and Toolkit for the Security and protection of CROWDs in mass gatherings" (LETSCROWD), EU Horizon 2020 programme, grant agreement No. 740466, and "IMaging MAnagement Guidelines and Informatics Network for law enforcement Agencies" (IMMAGINA), European Space Agency, France, ARTES Integrated Applications Promotion Programme, contract No. 4000133110/20/NL/AF.

References

Ainam, J., et al. (2021). Unsupervised domain adaptation for person re-identification with iterative soft clustering. *Knowledge-Based Systems*, *212*, Article 106644. http://dx.doi.org/10.1016/j.knosys.2020.106644.

Ali, S., et al. (2010). Interactive retrieval of targets for wide area surveillance. In *ICM* (pp. 895–898). ACM, http://dx.doi.org/10.1145/1873951.1874106.

Chen, X., et al. (2021). Explainable person re-identification with attribute-guided metric distillation. In *ICCV* (pp. 11793–11802). CVF / IEEE, http://dx.doi.org/10.1109/ICCV48922.2021.01160.

Chen, F., et al. (2023a). Unsupervised person re-identification via multi-domain joint learning. *Pattern Recognition*, *138*, Article 109369.

Chen, S., et al. (2023b). MTNet: Mutual tri-training network for unsupervised domain adaptation on person re-identification. *Journal of Visual Communication and Image Representation*, *90*, Article 103749. http://dx.doi.org/10.1016/j.jvcir.2022.103749.

Chong, Y., et al. (2021). Style transfer for unsupervised domain-adaptive person re-identification. *Neurocomputing*, *422*, 314–321.

Delussu, R., et al. (2020). Online domain adaptation for person re-identification with a human in the loop. In *ICPR* (pp. 3829–3836). http://dx.doi.org/10.1109/ICPR48806.2021.9412485.

Deng, J., et al. (2016). Leveraging the wisdom of the crowd for fine-grained recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(4), 666–676. http://dx.doi.org/10.1109/TPAMI.2015.2439285.

Dwivedi, R., et al. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, *55*(9), http://dx.doi.org/10.1145/3561048.

Farenzena, M., et al. (2010). Person re-identification by symmetry-driven accumulation of local features. In *CVPR* (pp. 2360–2367). CVF / IEEE, http://dx.doi.org/10.1109/CVPR.2010.5539926.

Feng, H., et al. (2021). Complementary pseudo labels for unsupervised domain adaptation on person re-identification. *IEEE Transactions on Image Processing*, *30*, 2898–2907. http://dx.doi.org/10.1109/TIP.2021.3056212.

Ge, Y., et al. (2020). Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In *ICLR* (pp. 1–15).

Genç, A., et al. (2019). Cross-dataset person re-identification using deep convolutional neural networks: effects of context and domain adaptation. *Multimedia Tools and Applications*, *78*(5), 5843–5861. http://dx.doi.org/10.1007/s11042-018-6409-3.

Giacinto, G. (2007). A nearest-neighbor approach to relevance feedback in content based image retrieval. In *CIVR* (pp. 456–463). ACM, http://dx.doi.org/10.1145/1282280.1282347.

Gou, M., et al. (2017). DukeMTMC4ReID: A large-scale multi-camera person re-identification dataset. In *CVPR workshops* (pp. 1425–1434). http://dx.doi.org/10.1109/CVPRW.2017.185.

Goyal, D., Patel, N., Truong, T., & Yanushkevich, S. (2021). Towards explainable person re- identification. In *IEEE SSCI* (pp. 1–8). http://dx.doi.org/10.1109/SSCI50451.2021.9660071.

Hirzer, M., et al. (2011). Person re-identification by descriptive and discriminative classification. In *SCIA* (pp. 91–102). http://dx.doi.org/10.1007/978-3-642-21227-7_9.

Holzinger, A. (2021). The next frontier: AI we can really trust. *Communications in Computer and Information Science*, *1524 CCIS*, 427–440. http://dx.doi.org/10.1007/978-3-030-93736-2_33.

Kamath, S., Deshpande, A., Kambhampati Venkata, S., & N Balasubramanian, V. (2021). Can we have it all? On the trade-off between spatial and adversarial robustness of neural networks. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, & J. W. Vaughan (Eds.), *Advances in neural information processing systems, Vol. 34* (pp. 27462–27474). Curran Associates, Inc..

Khosla, A., et al. (2012). Undoing the damage of dataset bias. In *ECCV* (pp. 158–171).

Leng, Q., et al. (2020). A survey of open-world person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(4), 1092–1108. http://dx.doi.org/10.1109/TCSVT.2019.2898940.

Li, H., et al. (2021). Triple adversarial learning and multi-view imaginative reasoning for unsupervised domain adaptation person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, *1*, 1. http://dx.doi.org/10.1109/TCSVT.2021.3099943.

Liao, S., et al. (2020). Interpretable and generalizable person re-identification with query-adaptive convolution and temporal lifting. In *ECCV* (pp. 456–474). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-58621-8_27.

Lin, W., et al. (2015). The effect of low-level image features on pseudo relevance feedback. *Neurocomputing*, *166*, 26–37. http://dx.doi.org/10.1016/j.neucom.2015.04.037.

Lin, S., et al. (2021). Multi-domain adversarial feature generalization for person re-identification. *IEEE Transactions on Image Processing*, *30*, 1596–1607. http://dx.doi.org/10.1109/TIP.2020.3046864.

Liu, Y., Ge, H., Sun, L., & Hou, Y. (2022). Complementary attention-driven contrastive learning with hard-sample exploring for unsupervised domain adaptive person re-ID. *IEEE Transactions on Circuits and Systems for Video Technology*, *33*(1), 326–341. http://dx.doi.org/10.1109/TCSVT.2022.3200671.

Liu, C., et al. (2013). POP: Person re-identification post-rank optimisation. In *ICCV* (pp. 441–448). CVF / IEEE, http://dx.doi.org/10.1109/ICCV.2013.62.

Liu, Z., et al. (2019). Deep reinforcement active learning for human-in-the-loop person re-identification. In *ICCV* (pp. 6121–6130). CVF / IEEE, http://dx.doi.org/10.1109/ICCV.2019.00622.

Mekhazni, D., et al. (2020). Unsupervised domain adaptation in the dissimilarity space for person re-identification. In *ECCV, Vol. 12372* (pp. 159–174). http://dx.doi.org/10.1007/978-3-030-58583-9_10.

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2022). Human-in-the-loop machine learning: A state of the art. *Artificial Intelligence Review*, *56*(4), 3005–3054. http://dx.doi.org/10.1007/s10462-022-10246-w.

Navaneet, K. L., et al. (2020). Operator-in-the-loop deep sequential multi-camera feature fusion for person re-identification. *IEEE Transactions on Information Forensics and Security*, *15*, 2375–2385. http://dx.doi.org/10.1109/TIFS.2019.2957701.

Piras, L., et al. (2013). Passive-aggressive online learning for relevance feedback in content based image retrieval. In *ICPRAM* (pp. 182–187).

Qi, L., et al. (2019). A novel unsupervised camera-aware domain adaptation framework for person re-identification. In *ICCV* (pp. 8079–8088). CVF / IEEE, http://dx.doi.org/10.1109/ICCV.2019.00817.

Rawal, A., et al. (2022). Recent advances in trustworthy explainable artificial intelligence: Status, challenges, and perspectives. *IEEE Transactions on Artificial Intelligence*, *3*(6), 852–866. http://dx.doi.org/10.1109/TAI.2021.3133846.

Royer, A., et al. (2015). Classifier adaptation at prediction time. In *CVPR* (pp. 1401–1409). CVF / IEEE, http://dx.doi.org/10.1109/CVPR.2015.7298746.

Song, X., Liu, J., & Jin, Z. (2022). Dual prototype contrastive learning with Fourier generalization for domain adaptive person re-identification. *Knowledge-Based Systems*, *256*, Article 109851. http://dx.doi.org/10.1016/j.knosys.2022.109851.

Song, L., et al. (2020). Unsupervised domain adaptive re-identification: Theory and practice. *Pattern Recognition*, *102*, Article 107173. http://dx.doi.org/10.1016/j.patcog.2019.107173.

Tang, C., Xue, D., & Chen, D. (2022). Multi-level mutual supervision for cross-domain person re-identification. *Journal of Visual Communication and Image Representation*, *89*, Article 103674. http://dx.doi.org/10.1016/j.jvcir.2022.103674.

Verma, A., et al. (2023). Unsupervised domain adaptation for person re-identification via individual-preserving and environmental-switching cyclic generation. *IEEE Transactions on Multimedia*, *25*, 364–377. http://dx.doi.org/10.1109/TMM.2021.3126404.

Vezzani, R., Baltieri, D., & Cucchiara, R. (2013). People reidentification in surveillance and forensics: A survey. *ACM Computing Surveys*, *46*(2), 29:1–29:37. http://dx.doi.org/10.1145/2543581.2543596.

Wang, W., Liao, S., Zhao, F., Kang, C., & Shao, L. (2021). DomainMix: Learning generalizable person re-identification without human annotations. In *32nd British machine vision conference 2021, BMVC 2021,* (p. 355). BMVA Press.

Wang, H., et al. (2016). Human-in-the-loop person re-identification. In *ECCV* (pp. 405–422). http://dx.doi.org/10.1007/978-3-319-46493-0_25.

Wang, H., et al. (2018). Person re-identification in identity regression space. *International Journal of Computer Vision*, *126*(12), 1288–1310. http://dx.doi.org/10.1007/s11263-018-1105-3.

Wei, L., et al. (2018). Person transfer GAN to bridge domain gap for person re-identification. In *CVPR* (pp. 79–88). CVF / IEEE, http://dx.doi.org/10.1109/CVPR.2018.00016.

Wu, A., et al. (2019). Unsupervised person re-identification by camera-aware similarity consistency learning. In *ICCV* (pp. 6921–6930). CVF / IEEE, http://dx.doi.org/10.1109/ICCV.2019.00702.

Xu, Y., et al. (2021). Dual attention-based method for occluded person re-identification. *Knowledge-Based Systems*, *212*, Article 106554. http://dx.doi.org/10.1016/j.knosys.2020.106554.

Ye, M., et al. (2022). Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(06), 2872–2893. http://dx.doi.org/10.1109/TPAMI.2021.3054775.

Yuan, Y., et al. (2020). Calibrated domain-invariant learning for highly generalizable large scale re-identification. In *WACV* (pp. 3578–3587). http://dx.doi.org/10.1109/WACV45572.2020.9093521.

Zhang, X., et al. (2019). Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In *ICCV* (pp. 8221–8230). CVF / IEEE, http://dx.doi.org/10.1109/ICCV.2019.00831.

Zhang, C., et al. (2020a). Improving domain-adaptive person re-identification by dual-alignment learning with camera-aware image generation. *IEEE Transactions on Circuits and Systems for Video Technology*, *1*, 4334–4346. http://dx.doi.org/10.1109/TCSVT.2020.3047095.

Zhang, W., et al. (2020b). Adaptive attention-aware network for unsupervised person re-identification. *Neurocomputing*, *411*, 20–31. http://dx.doi.org/10.1016/j.neucom.2020.05.094.

Zhao, Y., Luo, S., Yang, Y., & Song, M. (2018). DeepSSH: Deep semantic structured hashing for explainable person re-identification. In *ICIP* (pp. 1653–1657). http://dx.doi.org/10.1109/ICIP.2018.8451107.

Zheng, L., et al. (2015). Scalable person re-identification: A benchmark. In *ICCV* (pp. 1116–1124). CVF / IEEE, http://dx.doi.org/10.1109/ICCV.2015.133.

Zhong, Z., et al. (2019). Invariance matters: Exemplar memory for domain adaptive person re-identification. In *CVPR* (pp. 598–607). CVF / IEEE, http://dx.doi.org/10.1109/CVPR.2019.00069.

Zhou, K., Liu, Z., Qiao, Y., Xiang, T., & Loy, C. C. (2023). Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(4), 4396–4415. http://dx.doi.org/10.1109/TPAMI.2022.3195549.

Zhou, K., et al. (2019). Learning generalisable omni-scale representations for person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*, 1. http://dx.doi.org/10.1109/TPAMI.2021.3069237.

Zhou, S., et al. (2021). Cross-view similarity exploration for unsupervised cross-domain person re-identification. *Neural Computing and Applications*, 1–11. http://dx.doi.org/10.1007/s00521-020-05566-3.

Zou, Y., et al. (2020). Joint disentangling and adaptation for cross-domain person re-identification. In *ECCV, Vol. 12347* (pp. 87–104). http://dx.doi.org/10.1007/978-3-030-58536-5_6.