Deep Reinforcement Active Learning for Human-In-The-Loop Person Re-Identification

Zimo Liu[†⋆], Jingya Wang[‡⋆], Shaogang Gong[§], Huchuan Lu[†]*, Dacheng Tao[‡]
† Dalian University of Technology, ‡ UBTECH Sydney AI Center, The University of Sydney, § Queen Mary University of London
lzm920316@gmail.com, jingya.wang@sydney.edu.au, s.gong@qmul.ac.uk, lhchuan@dlut.edu.cn, dacheng.tao@sydney.edu.au

Abstract

Most existing person re-identification(Re-ID) approaches achieve superior results based on the assumption that a large amount of pre-labelled data is usually available and can be put into training phrase all at once. However, this assumption is not applicable to most real-world deployment of the Re-ID task. In this work, we propose an alternative reinforcement learning based human-in-the-loop model which releases the restriction of pre-labelling and keeps model upgrading with progressively collected data. The goal is to minimize human annotation efforts while maximizing Re-ID performance. It works in an iteratively updating framework by refining the RL policy and CNN parameters alternately. In particular, we formulate a Deep Reinforcement Active Learning (DRAL) method to guide an agent (a model in a reinforcement learning process) in selecting training samples on-the-fly by a human user/annotator. The reinforcement learning reward is the uncertainty value of each human selected sample. A binary feedback (positive or negative) labelled by the human annotator is used to select the samples of which are used to fine-tune a pre-trained CNN Re-ID model. Extensive experiments demonstrate the superiority of our DRAL method for deep reinforcement learning based human-in-the-loop person Re-ID when compared to existing unsupervised and transfer learning models as well as active learning models.

1. Introduction

Person re-identification (Re-ID) is the problem of matching people across non-overlapping camera views distributed at distinct locations. Most existing supervised person Re-ID approaches employ a train-once-and-deploy scheme, that is, pairwise training data are collected and annotated manually for every pair of cameras before learning a model. Based on this assumption, supervised Re-ID methods have progressed on several benchmarks in recent years [21, 56, 35, 52, 25].

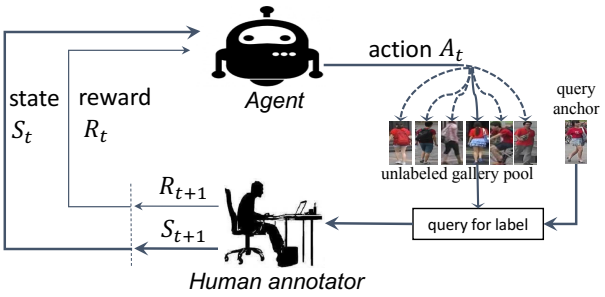


Figure 1: An illustration of Deep Reinforcement Active learning (DRAL). For each query anchor (probe), an agent (reinforcement active learner) will select sequential instances from gallery pool for human annotation with binary feedback (positive/negative) in an active learning process.

However, in practice this assumption is not easy to adapt due to a few reasons: Firstly, pairwise pedestrian data is prohibitive to be collected since it is unlikely that a large amount of pedestrian may reappear in other camera views. Secondly, the increasing number of camera views amplifies the difficulties in searching the same person among multiple camera views. To address these difficulties, one solution is to design unsupervised learning algorithms. A few works begin to focus on transfer learning or domain adaption technique for unsupervised Re-ID [11, 44, 28]. However, unsupervised learning based Re-ID models are inherently weaker compared to supervised learning based models, compromising Re-ID effectiveness in any practical deployment.

Another possible solution is following the semi-supervised learning scheme that decreases the requirement of data annotations. Successful researches have been done on either dictionary learning [27] or self-paced learning [14] based methods. These models are still based on a strong assumption that parts of the identities (e.g. one third of the

* Corresponding Author
⋆ Equal Contribution

2380-7504/19/$31.00 ©2019 IEEE
DOI 10.1109/ICCV.2019.00622

training set) are fully labelled for every camera view. This remains impractical for a Re-ID task with hundreds of cameras and 24/7 operations, typical in urban applications.

To achieve effective Re-ID given a limited budget cost on annotation, we focus on human-in-the-loop person Re-ID with selective labelling by human feedback on-the-fly [43]. This approach differs from the common once-and-done model learning approach. Instead, a step-by-step sequential active learning process is adopted by exploring human selective annotations on a much smaller pool of samples for model learning. These cumulatively labelled data by human binary verification are used to update model training for improving Re-ID performance. Such an approach to model learning is naturally suited for reinforcement learning together with active learning, the focus of this work.

Active learning is a technique for on-the-fly human data annotation that aims to sample actively the more informative training data for optimising model learning without exhaustive data labelling. Formally, some instance from an unlabelled set are selected and then annotated by a human oracle, and the label information can be employed for model training. These operations will repeat many times until it satisfies the termination criterion, e.g. the annotation budget is exhausted. The most critical in this process is the sample selection strategy. The more informative samples from less human annotation cost can greatly benefit the performance. Rather than a hand-design strategy, we propose to a reinforcement learning-based criterion. Fig 1 illustrates our design for a Deep Reinforcement Active Learning (DRAL) model. Specifically, we develop a model which introduces both active learning (AL) and reinforcement learning (RL) in a single human-in-the-loop model learning framework. By representing the AL part of our model as a sequence making process, since each action affects the sample correlations among unlabelled data pool (with similarity recomputed at each step), it will influence the decision of next step. By treating the uncertainty brought by the selected samples as the objective goal, the RL part of our model aims to learn a powerful sample selection strategy given human feedback annotations. Therefore, the informative samples selected from the RL policy can significantly boost the performance of Re-ID which in return enhance the ability of sample choosing strategy. The iterative training scheme will lead to a strong Re-ID model.

The main contributions of this work are: (1) We introduce a Deep Reinforcement Active Learning (DRAL) model, formulated to explore jointly both reinforcement learning and active learning principles in a single CNN deep learning framework. (2) We design an effective DRAL model for human-in-the-loop person Re-ID so that a deep reinforcement active learner (agent) can facilitate human-in-the-loop active learning strategy directly on a CNN deep network. Extensive comparative experiments show clearly the proposed DRAL model has advantages over existing supervised and transfer learning methods on scalability and annotation costs, over existing semi-supervised, unsupervised and active learning methods with significant performance gain whilst using much less annotations.

2. Related Work

Person Re-ID. Person Re-ID task aims to search the same people among multiple camera views. Recently, most person Re-ID approaches [50, 45, 8, 10, 33, 38, 7, 53, 19, 5, 51, 9, 39, 36] try to solve this problem under the supervised learning framework, where the training data is fully annotated. Despite the high performance these methods achieved, their large annotation cost cannot be ignored. To address the high labelling cost problem, some researchers propose to learn the model with only a few labelled samples or without any label information. Representative algorithms [32, 48, 2, 55, 23, 44, 28, 46] include domain transfer scheme, group association approaches, and some label estimation methods.

In addition to the above-mentioned approaches, some researchers aim to reduce the annotation cost in a human-in-the-loop (HITL) model learning process. When there is only a few annotated image samples, HITL model learning can be expected to improve the model performance by directly involving human interaction in the circle of model training, tuning or testing. With the human population correct the inaccuracies happen in machine learning predictions, the model could be efficiently corrected thereby leading to higher results. This circumstance sounds similar to the situation of person Re-ID task whose pre-labelling information is hard to be obtained with the gallery candidate size far beyond that of the query anchor. Motivated by this, Wang *et al.* [43] formulates a Human Verification Incremental Learning (HVIL) model which aims to optimize the distance metric with flexible human feedback continuously in real-time. The flexible human feedback (true, false, false but similar) employed in this model enables to involve more information and boost the performance in a progressive manner.

AL and RL. Active Learning has drawn many attention in the last few decades and been exploited in Natural Language Processing (NLP), data annotation and image classification task [41, 6, 4, 31]. Its procedure can be thought as human-in-the-loop setting, which allows the algorithm to interactively query the human annotator with the instances recognized as the most informative samples among the entire unlabelled data pool. This work is usually done by using some heuristic selection methods with limited effectiveness. Therefore, some researchers aim to address the shortcomings of the heuristic selection approaches by framing the active learning as a reinforcement learning problem to explicitly optimize a selection policy. In [15], rather than adopt-

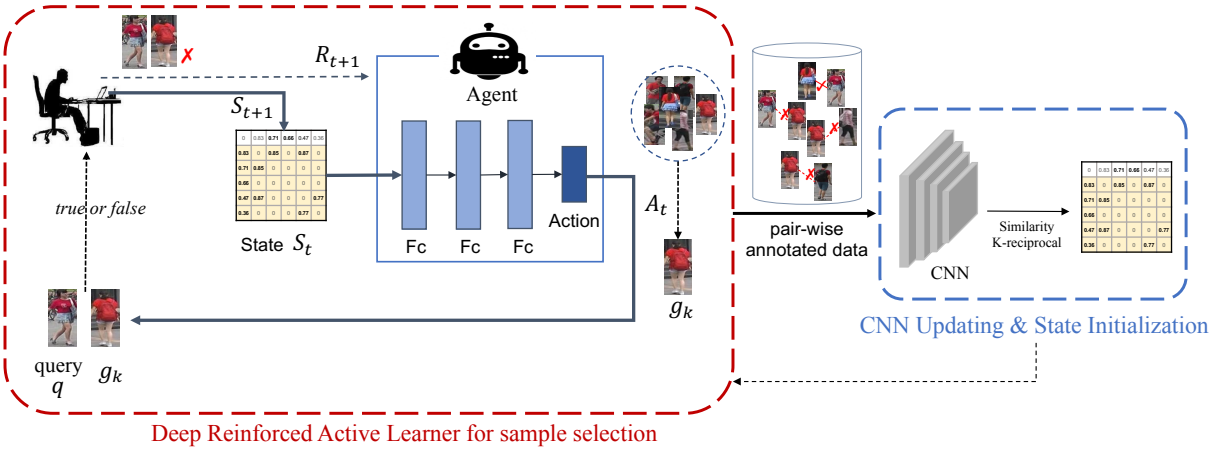


Figure 2: The Deep Reinforcement Active Learning (DRAL) framework: State measures the similarity relations among all instance. Action determines which gallery candidate will be sent for human annotator for querying. Reward is computed with different human feedback. A CNN is adopted for state initialization and being updated by the pairwise data annotated via a human annotator in-the-loop on-the-fly when the model is deployed. This iterative process stops when it reaches the annotation budget.

ing a fixed heuristic selection strategy, Fang *et al.* performs to learn a deep Q-network as an adaptive policy to select the data instances for labelling. Woodward *et al.* [47] try to solve the one-shot classification task by formulating an active learning approach which incorporates meta-learning with deep reinforcement learning. An agent learned via this approach enables to decide how and when to request label. Those successful applications indicate that reinforcement learning is a natural fit for active learning.

3. Methodology

3.1. Base CNN Network

We employ the Resnet-50 [20] architecture as the base net with ImageNet pre-train. To effectively learn the ID discriminative feature embedding, we adopt both cross entropy loss for classification and triplet loss for similarity learning synchronously.

The softmax Cross Entropy loss function defined as:

$$L_{\text{cross}} = -\frac{1}{n_{\text{b}}}\sum_{i=1}^{n_{\text{b}}}\log(p_{\text{i}}(y)) \tag{1}$$

where n_{b} denotes the batch size and $p_{\text{i}}(y)$ is the predicted probability on the groundtruth class y of input image.

Given triplet samples x_a, x_p, x_n, x_a is an anchor point. x_p is hardest positive sample in the same class of x_a , and x_n is a hardest negative sample of a different class of x_a. Finally we define the triplet loss as following:

$$L_{\text{tri}} = \sum_{x_a,x_p,x_n}^{n_{\text{b}}}[D_{x_a,x_p} - D_{x_a,x_n} + m] \tag{2}$$

where m is a margin parameter for the positive and negative pairs.

Finally, the total loss for can be calculated by:

$$L_{\text{total}} = L_{\text{cross}} + L_{\text{tri}} \tag{3}$$

3.2. A Deep Reinforced Active Learner - An Agent

The framework of the proposed DRAL is presented in Fig 2, of which "an agent" (model) is designed to dynamically select instances that are most informative to the query instance. As each query instance arrives, we perceive its n_s-nearest neighbors as the unlabelled gallery pool. At each discrete time step t, the environment provides an observation state S_t which reveals the instances' relationship, and receives a response from the agent by selecting an action A_t. For the action $A_t = g_k$, it requests the k-th instance among the unlabelled gallery pool being annotated by human oracle, who replies with binary feedback true or false against the query. This operation repeats until the maximum annotation amount for each query is exhausted. When plentiful enough pair-wise labelled data are obtained, the CNN parameters enable to be updated via triplet loss function, which in return generates a new initial state for incoming data. Through iteratively executing the sample selection and CNN network refreshing, the proposed algorithm could quickly escalate. This progress terminates with all query instances have been browsed once. More details about the proposed active learner are revealed in the following. To clarify on our formulation of the model, Table 1 and Algorithm 1 give the definitions of the notations and the entire process of the approach, respectively.

Table 1: Definitions of notations.

Notations	Description
$\mathcal{A}_t, S_t, R_t$	action, state and reward at time t
$\mathcal{T}_r, n$	train set and its size
$\mathcal{T}_p$	pairwise annotated data set
$Sim(i,j)$	similarity between samples i, j
d_i^j	Mahalanobis distance of i, j
q, g_k	query, the k-th gallery candidate
y_k^t	binary feedback of g_k at time t
X_p^t, X_n^t	positive/negative sample batch until time t
K_{max}	annotating sample number for each query
n_s	action size
κ	parameter of reciprocal operation
$thred$	threshold parameter

Algorithm 1 DRAL

Input: agent π, CNN weights w, $\mathcal{T}_r$ (size n), $\mathcal{T}_p = \emptyset$
  for $i = 1 : n$ **do**
    Sample query q and gallery pool g from $\mathcal{T}_r$
    while $t < K_{max}$ **do**
      $S_t \leftarrow (Sim, R(n_i, k))$ via Eq. 4-8
      $A_t : g_k \leftarrow \pi(S_t)$, requests label for pair (q, g_k)
      $\mathcal{T}_p \leftarrow \mathcal{T}_p \cup (q, g_k)$
      $(R_t, Sim) \leftarrow (S_t, A_t)$ via Eq. 9
    end while
    optimize $\pi^* \leftarrow \arg\max_{\pi} \mathbb{E}[R_t + \gamma R_{t+1} + ...]$
    optimize w by $\mathcal{T}_p$ after several steps
  end for

3.2.1 Action

The action set defines to select an instance from the unlabelled gallery pool, hence its size is the same as the pool. At each time step t, when encountered with the current state S_t, the agent decides the action to take based on its policy $\pi(A_t|S_t)$. Therefore the A_t instance of the unlabelled gallery pool will be selected querying by human oracle. Once $A_t = g_k$ is performed, the agent is unable to choose it again in the subsequent steps. The termination criterion of this process depends on a pre-defined K_{max} which restricts the maximal annotation amount for each query anchor.

3.2.2 State

Graph similarity has been widely employed for data selecting in active learning framework [16, 30] by digging the structural relationships among data points. Typically, a sparse graph is adopted which only connects data point to a few of its most similar neighbors to exploit their contextual information. In this work, we also construct a sparse similarity graph among query and gallery samples and take it as the state value. With a queried anchor q and its corresponding gallery candidate set $g = \{g_1, g_2, ..., g_{n_s}\}$, there Re-ID features could be extracted via the CNN network, where n_s is a pre-defined number of the gallery candidates. The similarity value $Sim(i,j)$ between every two samples $i, j (i \neq j)$ are then calculated as

$$Sim(i,j) = 1 - \frac{d_i^j}{\max\limits_{i,j \in q,g} d_i^j} \tag{4}$$

where d_i^j is the Mahalanobis distance of i, j, else set as 0. A k-reciprocal operation [57] is executed to build the sparse similarity matrix. For any node $n_i \in (q, g)$ of the similarity matrix Sim, its top κ-nearest neighbors are defined as $N(n_i, \kappa)$. Then the κ-reciprocal neighbors $R(n_i, \kappa)$ of n_i is obtained through

$$R(n_i, \kappa) = \{x_j | (n_i \in N(x_j, \kappa)) \wedge (x_j \in N(n_i, \kappa))\} \tag{5}$$

Compared to the previous description, the κ-reciprocal nearest neighbors are more related to the node n_i, of which the similarity value is remained otherwise be assigned with zero. This sparse similarity matrix is then taken as the initial state and imported into the policy network for action selection. Once the action is employed, the state value will be adjusted accordingly to better reveal the sample relations.

To better understand the update of state value, we illustrate an example in Fig 3. For a state S_t at time t, the optimal action $A_t = g_k$ is selected via the policy network, which indicates the gallery candidate g_k will be selected for querying by the human annotator. A binary feedback is the given as $y_k^t = \{1, -1\}$, which indicates g_k to be the positive pair or negative of the query instance. Therefore the similarity $Sim(q, g_k)$ between q and g_k will be set as

$$Sim(q, g_k) = \begin{cases} 1, & y_k^t = 1 \\ 0, & y_k^t = -1 \end{cases} \tag{6}$$

The similarities of the remaining gallery samples $g_i, i \neq k$ and query sample will also be re-computed, which aiming to zoom in the distance among positives and push out the distance among negatives. Therefore, with positive feedback, the similarity $Sim(q, g_i)$ is the average score between g_i with (q, g_k), where

$$Sim(q, g_i) = \frac{Sim(q, g_i) + Sim(q, g_k)}{2} \tag{7}$$

Otherwise, the similarity $Sim(q, g_i)$ will only be updated when the similarity among g_k and g_i is larger than a threshold $thred$, where

$$Sim(q, g_i) = \max(Sim(q, g_i) - Sim(g_k, g_i), 0) \tag{8}$$

The k-reciprocal operation will also be adopt afterwards, and a renewed state S_{t+1} is then obtained.

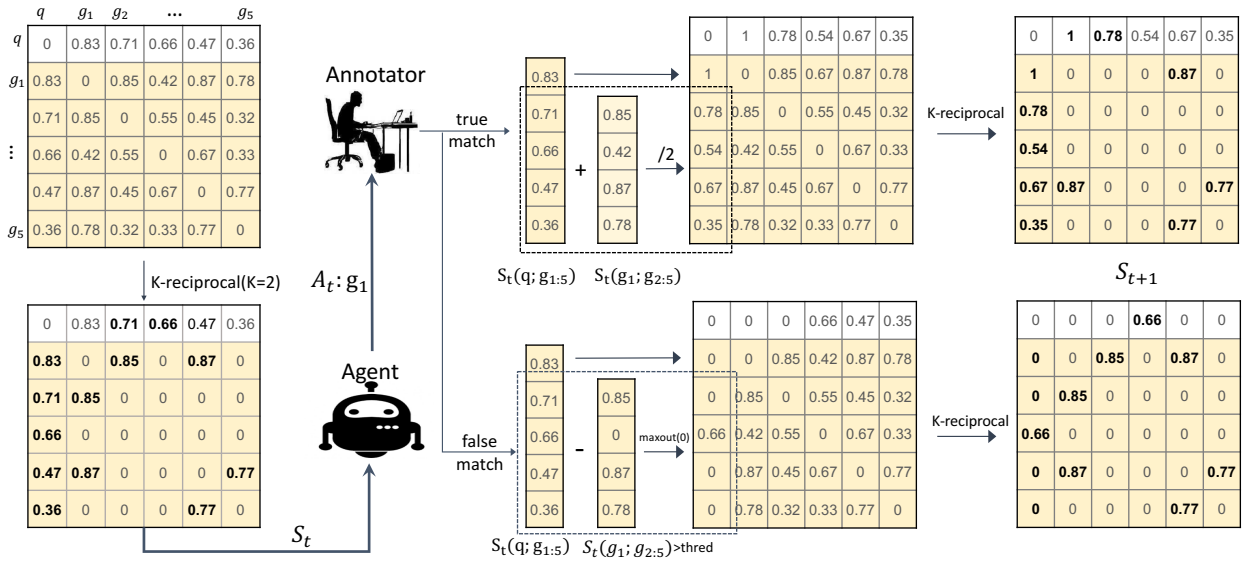


Figure 3: An example of state updating with different human feedback, which aims to narrow the similarities among instances sharing high correlations with negative samples, and enlarge the similarities among instances which are highly similar to the positive samples. The values with yellow background are the state imported into the agent.

3.2.3 Reward

Standard active learning methods adopt an uncertainty measurement, hypotheses disagreement or information density as the selection function for classification [4, 18, 58, 49] and retrieval task [17, 3]. Here, we use data uncertainty as the objective function of the reinforcement learning policy.

For data uncertainty measurement, higher uncertainty indicates that the sample is harder to be distinguished. Following the same principle of [42] which extends a triplet loss formulation to model heteroscedastic uncertainty in a retrieval task, we perform a similar hard triplet loss [21] to measure the uncertainty of data. Let the X_p^t, X_n^t indicate the positive and negative sample batch obtained until time t, $d_{g_k}^x$ be a metric function measuring Mahalanobis distances between any two samples g_k and x. Then the reward is computed as

$$R_t = [m + y_k^t(\max_{x_i \in X_p^t} d_{g_k}^{x_i} - \min_{x_j \in X_n^t} d_{g_k}^{x_j})]_+ \quad (9)$$

where $[\bullet]_+$ is the soft margin function by at least a margin m. Therefore all the future rewards$(R_{t+1}, R_{t+2}, ...)$ discounted by a factor γ at time t can be calculated as

$$Q^* = \max_{\pi} \mathbb{E}[R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} \cdots | \pi, S_t, A_t] \quad (10)$$

Once Q^* is learned, the optimal policy π^* can be directly inferred by selecting the action with the maximum Q value.

3.3. CNN Network Updating

For each query anchor, several samples are actively selected via the proposed DRAL agent and are manually annotated by the human oracle, and these pairwise data will be added to a updated training data pool. The CNN network is then updated gradually using fine-tuning. We use the triplet loss as the objective function, and when more labelled data is involved, the model becomes more robust and smarter. The renewed network is employed for Re-ID feature extraction, which in return helps the upgrade of the state initialization. We stop this iterative training scheme with a fixed annotation budget when each image in the training data pool has been browsed once by our DRAL agent.

4. Experiments

4.1. Dataset and Settings

Datasets For experimental evaluations, we report results on both large-scale and small-scale person re-identification benchmarks for robust analysis:

(1) The Market-1501 [54] is widely adapt large-scale re-id dataset that contains 1,501 identities obtained by Deformable Part Model pedestrian detector. It includes 32,668 images obtain from 6 non-overlapping camera views in the campus with 12936 images of 751 identities used for training. In testing stage, 3368 queries are used as the query set to search the true match among the remained candidates.

(2) CUHK01 [24] is one of the remarkable small-scale re-id dataset, which consists of 971 identities from two camera views, each identity has two images per camera view and thus totally including 3884 images which are manually cropped. The entire dataset is split into two parts: 485 identities for training and 486 for testing.

(3) DukeMTMC-ReID(Duke) [34] is one of the most popular large scale re-id dataset which consists 36411 pedestrian images captured from 8 different camera views. Among them, 16522 images (702 identities) are adopted for training, 2228 (702 identities) images are taken as query to be retrieved from the remaining 17661 images.

Evaluation Protocols Two evaluation metrics are adopted in this approach to evaluate the Re-ID performance. The first one is the Cumulated Matching Characteristics(CMC), and the second is the mean average precision(mAP) which taking person Re-ID task as an object retrieval problem.

Implementation Details. We implemented the proposed DRAL method in the Pytorch framework. We pre-train a resnet-50 multi-class identity discrimination network with the combination of triplet loss and cross entropy loss by 60 epochs(pre-train on Duke for Market1501 and CUHK01, pre-train on Market1501 for Duke), at a learning rate of 5E-4 by using the Adam optimizer. The final FC layer output feature vector (2,048-D) is extracted as the re-id feature vector in our model by resizing all the training images as 256×128. The policy network in the proposed method consists of three FC layers setting as 256. The proposed DRAL model is randomly initialized and then optimized with the learning rate at 2E-2, and (K_{max}, n_s, κ) are set as (10, 30, 15) by default. The balanced parameter $thred$ and m are set as 0.4 and 0.2, respectively. With every 25% of the training quires have been reviewed by the human annotator, we start to fine-tune the CNN network with learning rate at 5E-6.

4.2. Comparison With Unsupervised/Transfer Learning/Semi-Supervised Approaches

Human-in-the-loop person re-identification does not require the pre-labelling data, but receive user feedback for the input query little by little. It is feasible to label many of the gallery instances, but to cut down the human annotation cost, we perform to use the active learning technique for sample selecting. Therefore, we compare the proposed DRAL method with some active learning based approach and unsupervised/transfer/semi-supervised based methods, in the table we use 'uns/trans/semi', 'active' to indicate the training style. Moreover, the baseline results reported is computed by directly employing a pre-trained CNN model, and the upper bound result indicates that the model is fine-tuned on the dataset with fully supervised training data.

For unsupervised/transfer learning and semi-supervised setting, sixteen state-of-the-arts approaches are selected for comparing including UMDL [32], PUL [14], SPGAN [11], Tfusion [28], TL-AIDL [44], ARN [26], TAUDL [23], CAMEL [48], SSDAL [40], SPACO [29], One-Exampler [13] and DML [52]. In table 2, 3 and 4, we illustrate the rank-1, 5, 10 matching accuracy and mAP(%) performance on the Market1501 [54], Duke [34] and CUHK01 [24] dataset, of which the results of our approach are in bold. The proposed method achieves 84.2% and 66.26% at rank-1 and mAP, which outperforms the second best unsupervised/transfer/semi-supervised approaches

Table 2: Rank-1, 5, 10 accuracy and mAP (%) with some unsupervised, semi-supervised and adaption approaches on the Market1501 dataset.

style	Methods	Market1501			
		mAP	R-1	R-5	R-10
uns/trans/semi	UMDL [32]	22.4	34.5	52.6	59.6
	PUL [14]	20.7	45.5	60.7	66.7
	SPGAN [11]	26.9	58.1	76.0	82.7
	TFusion [28]	-	60.75	74.4	79.25
	TL-AIDL [44]	26.5	58.2	74.8	81.1
	ARN [26]	39.4	70.3	80.4	86.3
	TAUDL [23]	41.2	63.7	77.7	82.8
	CAMEL [48]	26.3	54.5	-	-
	SSDAL [40]	19.6	36.4	-	-
	SPACO [29]	-	68.3	-	-
	One-Exampler [13]	26.2	55.8	72.3	78.4
	DML [52]	46.57	-	-	-
active	Random	35.15	58.02	79.07	85.78
	QIU [22]	44.99	67.84	85.69	91.12
	QBC [1]	46.32	68.35	86.07	91.15
	GD [12]	49.3	71.44	87.05	91.42
	HVIL [43]	-	78.0	-	-
Ours	Baseline	20.04	42.79	62.32	70.04
	UpperBound	73.25	87.95	95.25	96.79
	DRAL	**66.26**	**84.2**	**94.27**	**96.59**

Table 3: Rank-1, 5, 10 accuracy and mAP (%) with some unsupervised, semi-supervised and adaption approaches on the Duke dataset.

style	Methods	Duke			
		mAP	R-1	R-5	R-10
uns/trans/semi	UMDL [32]	7.3	17.1	28.8	34.9
	PUL [14]	16.4	30.0	43.4	48.5
	SPGAN [11]	26.2	46.4	62.3	68.0
	TL-AIDL [44]	23.0	44.3	-	-
	ARN [26]	33.4	60.2	73.9	79.5
	TAUDL [23]	43.5	61.7	-	-
	CAMEL [48]	-	57.3	-	-
	One-Exampler [13]	28.5	48.8	63.4	68.4
active	Random	25.68	44.7	63.64	70.65
	QIU [22]	36.78	56.78	74.15	79.31
	QBC [1]	40.77	61.13	77.42	82.36
	GD [12]	33.58	53.5	69.97	75.81
Ours	Baseline	14.87	28.32	43.27	50.94
	UpperBound	60.93	77.96	88.69	91.61
	DRAL	**56**	**74.28**	**84.83**	**88.42**

Table 4: Rank-1, 5, 10 accuracy and mAP (%) with some unsupervised and adaption approaches on the CUHK01 dataset.

style	Methods	CUHK01			
		mAP	R-1	R-5	R-10
uns/trans	TSR [37]	-	22.4	35.9	47.9
	UCDTL [32]	-	32.1	-	-
	CAMEL [48]	61.9	57.3	-	-
	TRSTP [28]	-	60.75	74.44	79.25
active	Random	52.46	51.03	71.09	81.28
	QIU [22]	56.95	54.84	76.85	85.29
	QBC [1]	58.88	57.1	80.04	86.83
	GD [12]	54.79	52.37	75.21	83.44
Ours	Baseline	45.55	43.21	65.74	73.46
	UpperBound	79.96	79.22	93.00	95.37
	DRAL	**71.52**	**74.07**	**88.99**	**93.93**

by 13.9% and 19.69% on Market1501 [54] benchmark. For Duke [34] and CUHK01 [24] datasets, DRAL also achieves fairly good performance with rank-1 matching rate at 74.28% and 74.07%. These results demonstrate clearly the effectiveness of our active sample selection strategy implemented by the DRAL method, and shows that without annotating large quantities of training data, a good re-identification model can be built effectively by DRAL.

4.3. Comparisons with Active Learning

Beyond the approaches as mentioned above, we further compare with some active learning based approaches which involve human-machine interaction during training. We choose four active learning strategy as comparisons of which the model is trained through the same framework as our method, of which an iterative procedure of these active sample selection strategy and CNN parameter updating is executed until the annotation budget is achieved. Here 20% of the entire training samples(around 4% pairs) are selected via the reported active learning approaches, which indicates 388, 2588, 3304 are set as the annotation budget for termination on the CUHK01 [24], Market1501 [54], and Duke [34] dataset, respectively Beside these active learning methods, we also compare the performance with another active learning approach HVIL [43], which runs experiments under human-in-the-loop setting. The details of these approaches are described as follows: (1) Random, as a baseline active learning approach, we randomly pick some samples for querying; (2) Query Instance Uncertainty [22] (QIU), QIU strategy selects the samples with the highest uncertainty for querying; (3) Query By Committee [1] (QBC), QBC is a very effective active learning approach which learns an ensemble of hypotheses and queries the instances that cause maximum disagreement among the committee; (4) Graph Density [12] (GD), active learning by GD is an algorithm which constructs graph structure to identify highly connected nodes and determine the most representative data for querying. (5) Human Verification Incremental Learning [12] (HVIL), HVIL is trained with the human-in-the-loop setting which receives soft user feedback (true, false, false but similar) during model training, requiring the annotator to label the top-50 candidates of each query instance.

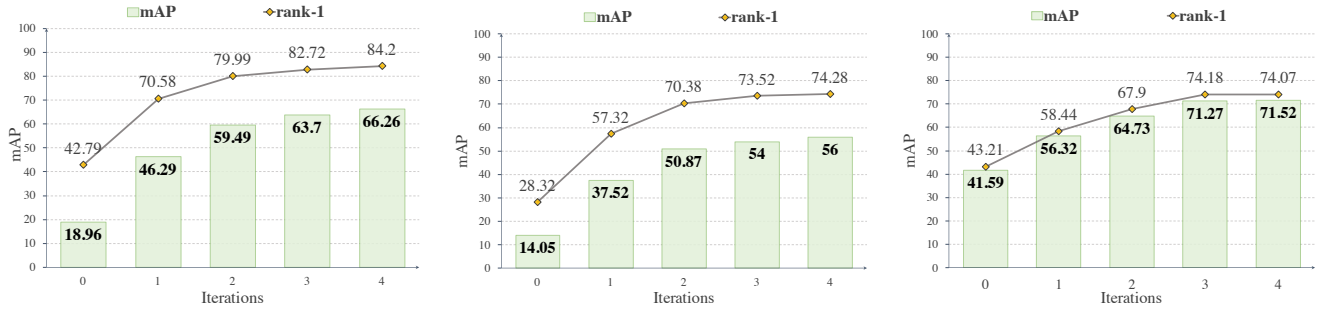
Table 2, 3, 4 compares the rank-1, 5, 10 and mAP rate from the active learning models against DRAL, where the baseline model result is from directly employing the pre-trained CNN model. We can observe from these results that (1) all the active learning methods perform better than the random picking strategy, which validates that active sample selection does benefit person Re-ID performance. 2) DRAL outperforms all the other active learning methods, with rank-1 matching rate exceeds the second best models QBC, HVIL and GD by 16.97%, 6.2% and 13.15% on the CUHK01 [24], Market1501 [54] and Duke [34] dataset, respectively, with a much lower annotation cost. This suggests that DRAL is more effective than other active learning methods for person Re-ID by introducing the policy as sample selection strategy.

4.4. Comparison at Different Annotation Cost

In this work, cost is measured via the annotation number between image pairs. With training set size n, the cost for the fully supervised setting will be $n*(n-1)/2$, and $10*n$ for the reported DRAL result. Therefore our DRAL annotates about 0.12%(Duke [34]), 0.15%(Market1501 [54]) and 1%(CUHK01 [24]) pairs. We further compare the performance of the proposed DRAL approach in a varying amount of labelled data (indicate by K_{max}) with fully supervised learning(UpperBound) on the three reported datasets. With the enlarge of training data size, the cost of annotating all data shows exponential increasement. Among the results, the baseline is obtained by directly employing the pre-trained CNN for testing. For the fully supervised setting, with all the training data annotated, it enables to fine-tune the CNN parameters with both the triplet loss and the cross-entropy loss to looking for better performance. For DRAL method, we present the performance with K_{max} setting as 3, 5 and 10 in Table 5. As can be observed, 1) with more annotated data, the model becomes stronger with increasing annotation cost. With the annotation number for each query increases from 3 to 10, the rank-1 matching rate improves 13.37%, 8.72% and 15.43% on the Duke [34], Market1501 [54] and CUHK01 [24] benchmarks. 2) compared to the fully supervised setting, the proposed active learning approach shows only around 4% rank-1 accuracy falling on each dataset. However, the annotation cost of DRAL is far below the supervised one.

Table 5: Rank-1 accuracy and mAP (%) result by directly employing(Baseline), fully supervised learning(UpperBound), and DRAL with varied K_{max} on the three reported dataset, where n indicates the training instance number for each benchmark. The annotation cost is calculated through the times of labelling behavior for every two samples.

Methods	Duke				Market1501				CUHK01				cost
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	
Baseline	14.05	28.32	43.27	50.94	18.96	42.79	62.32	70.04	41.59	43.21	65.74	73.46	0
DRAL	43.82	64.77	78.19	82.81	52.5	75.48	89.9	93.26	55.71	58.64	77.78	85.6	$n*3$
	51.48	70.51	84.16	87.43	59.71	81.06	93.21	95.1	66.98	70.06	87.14	91.87	$n*5$
	56	74.28	83.43	88.42	66.26	84.2	94.27	96.59	71.52	74.07	88.99	93.93	$n*10$
UpperBound	60.93	77.96	88.69	91.61	73.25	87.95	95.25	96.79	79.96	79.22	93.00	95.37	$n*(n-1)/2$



(a) Market1501 (b) Duke (c) CUHK01

Figure 4: Rank-1 accuracy and mAP(%) improvement with respect to the iterations on the (a) Market1501, (b) Duke and (c) CUHK01 dataset. The gray line and green bar(bold number) indicates the rank-1 accuracy and mAP respectively.

4.5. Effects with Number of Iterations

The promise of active learning is that, through iteratively increasing the size of labelled data, the model performance is enhanced gradually. For each input query, we only associate the label to the gallery candidates derived from the DRAL, and adopted these pairwise labelled data for CNN parameter updating. We set the iteration as a fixed number 4 in our experiments on all the datasets. Fig 4 shows the rank-1 accuracy and mAP improvement with respect to the iterations on the three datasets. From these results, we can observe that the performance of the proposed DRAL active learner improves quickly, with rank-1 accuracy increases around 20%∼40% over the first two iterations on all three benchmarks, and the improvement in model performance starts to flat out after five iterations. This suggests that for person Re-ID, fully supervising may not be essential. Once the informative samples/information have been obtained, a sufficiently good Re-ID model can be derived at the cost of a much smaller annotation workload by exploring a sample selection strategy on-the-fly.

5. Conclusion

In this work, we addressed the problem of how to reduce human labelling effort in conventional data pre-labelling for person re-identification model training. With limited annotation cost or inaccessible large quantity of pre-labelled training data, our model design aims to maximise the effectiveness of Re-ID model learning with a small number of selective sample labelling. The key task for the model design becomes how to select more informative samples at a fixed annotation cost. Specifically, we formulated a Deep Reinforcement Active Learning (DRAL) method with a flexible reinforcement learning policy to select informative samples (ranked list) for a given input query. Those samples are then fed into a human annotator so that the model can receive binary feedback (true or false) as reinforcement learning reward for DRAL model updating. Moreover, an iterative scheme is executed for the update of DRAL and Re-ID model. Extensive comparative evaluations were conducted on both large-scale and small-scale Re-ID benchmarks to demonstrate our model robustness.

Acknowledgement

This work is supported by National Natural Science Foundation of China No.61725202, 61829102, 61751212; Fundamental Research Funds for the Central Universities under Grant No.DUT19GJ201; Vision Semantics Limited; the China Scholarship Council; Alan Turing Institute; Innovate UK Industrial Challenge Project on Developing and Commercialising Intelligent Video Analytics Solutions for Public Safety (98111-571149); and the Australian Research Council Projects: FL-170100117, DP-180103424.

References

[1] Naoki Abe and Hiroshi Mamitsuka. Query learning strategies using boosting and bagging. In *ICML*, 1998. 6, 7

[2] Slawomir Bak, Peter Carr, and Jean-Francois Lalonde. Domain adaptation through synthesis for unsupervised person re-identification. In *ECCV*, 2018. 2

[3] Björn Barz, Christoph Käding, and Joachim Denzler. Information-theoretic active learning for content-based image retrieval. In *PR*, 2018. 5

[4] William H. Beluch, Tim Genewein, Andreas Nürnberger, and Jan M. Köhler. The power of ensembles for active learning in image classification. In *CVPR*, 2018. 2, 5

[5] Xiaobin Chang, Timothy M. Hospedales, and Tao Xiang. Multi-level factorisation net for person re-identification. In *CVPR*, 2018. 2

[6] Moitoeya Chatterjee and Anton Leuski. An active learning based approach for effective video annotation and retrieval. In *NIPS*, 2015. 2

[7] Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang. Beyond triplet loss: A deep quadruplet network for person re-identification. In *CVPR*, 2017. 2

[8] Yilun Chen, Zhicheng Wang, Yuxiang Peng, Zhiqiang Zhang, Gang Yu, and Jian Sun. Cascaded pyramid network for multi-person pose estimation. In *CVPR*, 2018. 2

[9] De Cheng, Yihong Gong, Sanping Zhou, Jinjun Wang, and Nanning Zheng. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In *CVPR*, 2016. 2

[10] Dahjung Chung, Khalid Tahboub, and Edward J. Delp. A two stream siamese convolutional neural network for person re-identification. In *ICCV*, 2017. 2

[11] Weijian Deng, Liang Zheng, Guoliang Kang, Yi Yang, Qixiang Ye, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person reidentification. In *CVPR*, 2018. 1, 6

[12] Sandra Ebert, Mario Fritz, and Bernt Schiele. RALF: A reinforced active learning formulation for object class recognition. In *CVPR*, 2012. 6, 7

[13] Yu Wu *et al*. Progressive learning for person re-identification with one example. *TIP*, 2019. 6

[14] Hehe Fan, Liang Zheng, Chenggang Yan, and Yi Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM*, 2018. 1, 6

[15] Meng Fang, Yuan Li, and Trevor Cohn. Learning how to active learn: A deep reinforcement learning approach. In *EMNLP*, 2017. 2

[16] Eyal En Gad, Akshay Gadde, Amir Salman Avestimehr, and Antonio Ortega. Active learning on weighted graphs using adaptive and non-adaptive approaches. In *ICASSP*, 2016. 4

[17] Philippe Henri Gosselin and Matthieu Cord. Active learning methods for interactive image retrieval. *TIP*, 2008. 5

[18] Husheng Guo and Wenjian Wang. An active learning-based SVM multi-class classification model. *PR*, 2015. 5

[19] Yiluan Guo and Ngai-Man Cheung. Efficient and deep person re-identification using multi-level similarity. In *CVPR*, 2018. 2

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 3

[21] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *CoRR*, 2017. 1, 5

[22] David D. Gale Lewis and William A. Gale. Training text classifiers by uncertainty sampling. In *SIGIR*, 1994. 6, 7

[23] Minxian Li, Xiatian Zhu, and Shaogang Gong. Unsupervised person re-identification by deep learning tracklet association. In *ECCV*, 2018. 2, 6

[24] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transfered metric learning. In *ACCV*, 2012. 5, 6, 7

[25] Wei Li, Xiatian Zhu, and Shaogang Gong. Harmonious attention network for person re-identification. In *CVPR*, 2018. 1

[26] Yu-Jhe Li, Fu-En Yang, Yen-Cheng Liu, Yu-Ying Yeh, Xiaofei Du, and Yu-Chiang Frank Wang. Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification. In *CVPR*, 2018. 6

[27] Xiao Liu, Mingli Song, Dacheng Tao, Xingchen Zhou, Chun Chen, and Jiajun Bu. Semi-supervised coupled dictionary learning for person re-identification. In *CVPR*, 2014. 1

[28] Jianming Lv, Weihang Chen, Qing Li, and Can Yang. Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns. In *CVPR*, 2018. 1, 2, 6, 7

[29] Fan Ma, Deyu Meng, Qi Xie, Zina Li, and Xuanyi Dong. Self-paced co-training. In *ICML*, 2017. 6

[30] Yifei Ma, Tzu-Kuo Huang, and Jeff G. Schneider. Active search and bandits on graphs using sigma-optimality. In *UAI*, 2015. 4

[31] Sujoy Paul, Jawadul H. Bappy, and Amit K. Roy-Chowdhury. Non-uniform subset selection for active learning in structured data. In *CVPR*, 2017. 2

[32] Peixi Peng, Tao Xiang, Yaowei Wang, Massimiliano Pontil, Shaogang Gong, Tiejun Huang, and Yonghong Tian. Unsupervised cross-dataset transfer learning for person re-identification. In *CVPR*, 2016. 2, 6, 7

[33] Xuelin Qian, Yanwei Fu, Yu-Gang Jiang, Tao Xiang, and Xiangyang Xue. Multi-scale deep learning architectures for person re-identification. In *ICCV*, 2017. 2

[34] Ergys Ristani, Francesco Solera, Roger S. Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *ECCV Workshops*, 2016. 6, 7

[35] M. Saquib Sarfraz, Arne Schumann, Andreas Eberle, and Rainer Stiefelhagen. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. *CVPR*, 2018. 1

[36] Yantao Shen, Hongsheng Li, Shuai Yi, Dapeng Chen, and Xiaogang Wang. Person re-identification with deep similarity-guided graph neural network. In *ECCV*, 2018. 2

[37] Zhiyuan Shi, Timothy M. Hospedales, and Tao Xiang. Transferring a semantic representation for person re-identification and search. In *CVPR*, 2015. 7

[38] Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. Pose-driven deep convolutional model for person re-identification. In *ICCV*, 2017. 2

[39] Chi Su, Fan Yang, Shiliang Zhang, Qi Tian, Larry S. Davis, and Wen Gao. Multi-task learning with low rank attribute embedding for person re-identification. In *ICCV*, 2015. 2

[40] Chi Su, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. Deep attributes driven multi-camera person re-identification. In *ECCV*, 2016. 6

[41] Hang Su, Zhaozheng Yin, Takeo Kanade, and Seungil Huh. Active sample selection and correction propagation on a gradually-augmented graph. In *CVPR*, 2015. 2

[42] Ahmed Taha, Yi-Ting Chen, Teruhisa Misu, Abhinav Shrivastava, and Larry Davis. Unsupervised data uncertainty learning in visual retrieval systems. *CoRR*, 2019. 5

[43] Hanxiao Wang, Shaogang Gong, Xiatian Zhu, and Tao Xiang. Human-in-the-loop person re-identification. In *ECCV*, 2016. 2, 6, 7

[44] Jingya Wang, Xiatian Zhu, Shaogang Gong, and Wei Li. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In *CVPR*, 2018. 1, 2, 6

[45] Yicheng Wang, Zhenzhong Chen, Feng Wu, and Gang Wang. Person re-identification with cascaded pairwise convolutions. In *CVPR*, 2018. 2

[46] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *CVPR*, 2018. 2

[47] Mark Woodward and Chelsea Finn. Active one-shot learning. *CoRR*, 2017. 3

[48] Hong-Xing Yu, Ancong Wu, and Wei-Shi Zheng. Cross-view asymmetric metric learning for unsupervised person re-identification. In *ICCV*, 2017. 2, 6, 7

[49] Chicheng Zhang and Kamalika Chaudhuri. Beyond disagreement-based agnostic active learning. In *NIPS*, 2014. 5

[50] Li Zhang, Tao Xiang, and Shaogang Gong. Learning a discriminative null space for person re-identification. In *CVPR*, 2016. 2

[51] Ying Zhang, Baohua Li, Huchuan Lu, Atshushi Irie, and Xiang Ruan. Sample-specific svm learning for person re-identification. In *CVPR*, 2016. 2

[52] Ying Zhang, Tao Xiang, Timothy M Hospedales, and Huchuan Lu. Deep mutual learning. *CVPR*, 2018. 1, 6

[53] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *CVPR*, 2017. 2

[54] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *ICCV*, 2015. 5, 6, 7

[55] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *ICCV*, 2017. 2

[56] Zhedong Zheng, Liang Zheng, and Yi Yang. Pedestrian alignment network for large-scale person re-identification. *TCSVT*, 2018. 1

[57] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *CVPR*, 2017. 4

[58] Jingbo Zhu, Huizhen Wang, Benjamin K. Tsou, and Matthew Y. Ma. Active learning with sampling by uncertainty and density for data annotations. *TASLP*, 2010. 5