Highly Efficient Regression for Scalable Person Re-Identification

Hanxiao Wang
hanxiao.wang@qmul.ac.uk
Shaogang Gong
s.gong@qmul.ac.uk
Tao Xiang
t.xiang@qmul.ac.uk

Vision Group,
School of Electronic Engineering and Computer Science,
Queen Mary, University of London
London, E1 4NS, UK

Abstract

Existing person re-identification models are poor for scaling up to large data required in real-world applications due to: (1) Complexity: They employ complex models for optimal performance resulting in high computational cost for training at a large scale; (2) Inadaptability: Once trained, they are unsuitable for incremental update to incorporate any new data available. This work proposes a truly scalable solution to re-id by addressing both problems. Specifically, a Highly Efficient Regression (HER) model is formulated by embedding the Fisher's criterion to a ridge regression model for very fast re-id model learning with scalable memory/storage usage. Importantly, this new HER model supports faster than real-time incremental model updates therefore making real-time active learning feasible in re-id with human-in-the-loop. Extensive experiments show that such a simple and fast model not only outperforms notably the state-of-the-art re-id methods, but also is more scalable to large data with additional benefits to active learning for reducing human labelling effort in re-id deployment.

1 Introduction

Person re-identification (re-id) refers to the problem of matching people across disjoint camera views at different locations and times [13]. It is challenging since a person's appearance often undergoes dramatic changes in a multi-camera environment, affected by viewing angles, body poses, illuminations and background clutters. Re-id is further compounded by the homogeneous clothing styles among different individuals. As a result, the concentration of most existing state-of-the-art methods [1, 20, 23, 24, 25, 26, 37, 41, 42, 44, 47, 51, 52] has been anchored at reducing intra-person appearance disparity whilst enhancing inter-person appearance individuality to optimise person identity-discriminative information. By deploying such learning principles on exhaustively pre-labelled pairwise training data, these methods have reported ever-increasing accuracies on existing re-id benchmarks.

Then, should we expect from them a scalable re-id solution for deployment in a real-world environment? The answer is no. Specifically, a truly scalable re-id system [12] requires: (1) Low model complexity with scalable computational cost and memory usage in model training; and (2) High model adaptability supporting fast model update to incorporate

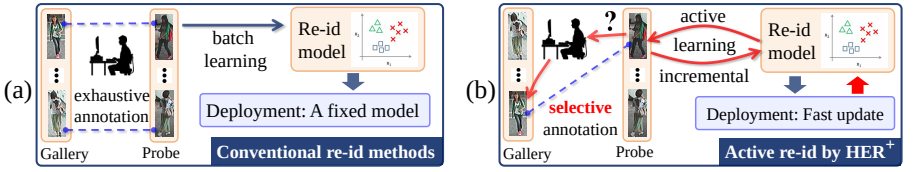


Figure 1: (a) Conventional re-id: A re-id model is trained on a fully labelled training set, then fixed for deployment; (b) Active re-id by HER$^+$: A training set is actively labelled incrementally on-the-fly as a re-id model is incrementally learned, and further updated without re-training during future deployment.

any new and increasingly larger data, e.g. unknown camera pairs. None of the aforementioned state-of-the-art satisfies these two requirements. By fixating on the re-id matching performance only, existing re-id models are typically solved by either slow iterative optimisation [1, 23, 24, 25, 51, 52] or costly generalised eigenproblems [26, 37, 44, 47], requiring hours or even days to train when the data size grows over 10^4. Moreover, most of the models are restricted to learning in a batch-mode, therefore failing to offer a solution to efficient model update without re-training from scratch when new labelled data become available. In a real-world application scenario, a human operator using a re-id system will generate new labels in the process of deployment. It is highly desirable that a re-id model can grow continuously its knowledge whilst being used. Given the existing re-id models, this can only be achieved by re-training from scratch, putting a huge burden on both model updating time and data storage[1], rendering existing models poorly suited for scaling up to cumulatively large data. Moreover, such re-training cannot provide immediate responses to the operator, which is a fundamental flaw for any systems with human-in-the-loop.

To make person re-id more scalable for real-world deployment, a model must be simple with very fast learning and inference algorithms and supports incremental learning. To that end, we propose an extremely simple but very fast re-id model by exploring multivariate ridge regression [15] to enable regression capable of performing re-id verification tasks. The new model, Highly Efficient Regression (HER) for re-id, has four advantages over existing methods: (1) Low model complexity with a very simple and fast closed-form solution, involved with only a set of linear equations. It is readily scalable to large data if more efficient algorithms like LSQR [32] are used. (2) By embedding Fisher's criterion [10, 11], HER does not sacrifice any discriminative power for speed, and outperforms the state-of-the-art re-id models with complex algorithms. (3) HER$^+$, an additional incremental formulation of HER, enables real-time incremental model update to incorporate new data of which the existing batch-based re-id models are incapable. (4) HER$^+$ makes *active learning re-id* with human-in-the-loop (Fig. 1) feasible to reduce human labelling costs. This is important as heavy human supervision is another bottleneck to scalable re-id.

Contributions: (1) A Highly Efficient Regression (HER) re-id model is formulated. The model is both very simple, can be implemented by one line of code, and scalable to large data. (2) An incremental extension HER$^+$ is further formulated for real-time model update to incorporate new labelled data. (3) HER$^+$ in active learning is explored with three novel joint exploration-exploitation criteria proposed for active re-id. Extensive experiments (Sec. 4) on three benchmark datasets VIPeR [14], CUHK01 [22] and CUHK03 [23] have shown the superiority of the proposed HER model over seven state-of-the-art models, including Mid-level Filter [51], Deep+ [1], kLFDA [44], XQDA [26], MLAPG [25], NFST [47], and Ensembles [35], on both re-id performance and efficiency. Our experiments also reveal the benefits of HER$^+$ for active learning in re-id with reduced human annotation effort.

[1] Need to keep all the data for training.

2 Related Work

Most existing state-of-the-art in re-id [6, 20, 23, 25, 26, 32, 35, 37, 41, 42, 44, 51, 52] focus only on learning the person identity-discriminative information for pursuing high performance. Existing approaches include exploring view-invariant visual features [9, 31, 45, 50], imposing pairwise or list-wise constraints for ranking [9, 30, 35, 42], discriminative subspace/distance metric learning [20, 25, 26, 37, 44, 47, 52], and deep learning [1, 6, 23, 40]. However, existing models have poor scalability due to two factors: High complexity to train; and low adaptability to update. For example, many methods rely on iterative optimisation[1, 23, 24, 25, 51, 52] which are extremely slow in training. The current fastest closed-form solutions are those solved by generalised eigenproblems [26, 37, 44, 47] which are still expensive to compute. Moreover, most contemporaries only assume a batch-mode learning scheme: To incorporate any new data, a system has to keep all the past training data and re-train a new model from scratch. This re-training approach makes them unscalable to large-scale deployment in the real-world.

Ridge regression [15, 16], as a regularised least squares model, is one of the most well-studied machine learning models. It has a simple closed-form solution solved by a linear system, and thus low model complexity. Furthermore, many well-optimised algorithms [34] can be readily applied to large data. Finally, its adaptable solution supports efficient model update for incremental learning [28]. The new HER model casts re-id into such a regression problem, benefiting from all of its advantages in scalability. To explore ridge regression for discriminative re-id verification tasks, the proposed HER model is further embedded with the criterion of Fisher Discriminant Analysis (FDA) [10, 11] to encode person identity-discriminative information. The relationship between FDA and linear regression has been studied for binary [7] and multi-class [15, 36] classification tasks. Recently, similar connections have been discovered for their regularised counterparts [21, 48, 49]. However, this work is the first to formulate it for a verification setting as in re-id. For its incremental extension, our model HER$^+$, differs significantly to [28] which only supports updates on a single sample without regularisation employed.

The proposed HER$^+$ enables active learning in re-id. Active learning [3, 8, 17, 19, 29, 33] argues that not all samples are equally important to model learning, and thus requires the training set to be actively selected on-the-fly as the model is updated with new labels. For active learning to be plausible in engaging human-in-the-loop, real-time model response with a very fast updating algorithm is essential. A recent attempt at active re-id was reported in [5]. Instead of learning a generalised cross-view matching function, it trains multi-class SVM person classifiers on known identities. The model cannot be deployed to re-identify a person without having abundant person-specific training images in advance. Thus, it is not applicable to most existing benchmarks [14, 22, 23] nor common re-id settings where each training person is often only captured by a single or a very few shot(s) and during testing the person to be re-identified does not assume to have any training data [12]. Moreover, their model updates still require expensive re-training. Compared to [5], the proposed HER$^+$ with active learning criteria learns a generalised re-id matching function, specially designed for the most common re-id conditions. Moreover, it achieves better human effort efficiency, since HER$^+$ can be incrementally updated whenever new labels arrive, and in turn immediately benefits further active data selection process in a loop.

3 Methodology

In the conventional re-id setting, a pre-labelled training set consists of n cross-view images and their identity labels are assumed to exist, denoted as $\{\mathbf{X} \in \mathbb{R}^{d\times n}, \mathbf{l} \in \mathbb{Z}^n\}$, where $\mathbf{X} =$

$[\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n]$, $\boldsymbol{x}_i \in \mathbb{R}^d$ is the i-th image's feature vector, $l_i \in \{1, \cdots, c\}$ is its identity label, and c is the total amount of person identities. After model training, the learned generalised re-id model is deployed to match person images of some independent probe and gallery population by measuring their inferred distances.

3.1 Highly Efficient Regression for Re-Id

Given labelled data $\{\mathbf{X} \in \mathbb{R}^{d \times n}, \mathbf{l} \in \mathbb{Z}^n\}$, we aim to learn a discriminative projection $\mathbf{P} \in \mathbb{R}^{d \times k}$. Specifically, we expect that in the projected k-dim subspace, the samples of the same person become closer, and those of different people are further apart.

Suppose there exists a set of *'ideal'* vectors $\boldsymbol{y}_i$ in the $\mathbb{R}^k$ subspace which perfectly satisfy the above discriminative relationship to represent each training sample $\boldsymbol{x}_i$, and together they construct a matrix $\mathbf{Y} \in \mathbb{R}^{n \times k}$ in a row-wise manner. Treating $\mathbf{Y}$ as regression outputs and $\mathbf{P}$ the regression parameter, a discriminative projection can thus be estimated with multivariate ridge regression, which minimises a least mean squared error between $\mathbf{Y}$ and the projected features of training samples, $\mathbf{X}^\top \mathbf{P}$, with the magnitude of $\mathbf{P}$ being l_2-regularised:

$$\mathbf{P} = \arg\min_{\mathbf{P}} \tfrac{1}{2} \|\mathbf{X}^\top \mathbf{P} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{P}\|_F^2, \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm, λ controls the regularisation strength. $\mathbf{Y}$ is known as an indicator matrix. The above formulation leads to a simple closed-form solution:

$$\mathbf{P} = \left(\mathbf{X}\mathbf{X}^\top + \lambda \mathbf{I}\right)^\dagger \mathbf{X}\mathbf{Y}, \tag{2}$$

where $(\cdot)^\dagger$ denotes a Moore-Penrose inverse, and $\mathbf{I}$ an identity matrix.

To frame the re-id task as a regression problem with Eq. (2), we need to find the optimal regression outputs, $\mathbf{Y}$, so that the person identity-discriminative information is encoded by the learned projection. To tackle this problem, we propose to explore the criterion of FDA [10, 11]. Specifically, in FDA the intra-person appearance disparity and inter-person appearance difference are measured by two scatter matrices:

$$\mathbf{S}_w = \tfrac{1}{n} \textstyle\sum_{j=1}^{c} \sum_{l_i=j} (\boldsymbol{x}_i - \boldsymbol{u}_j)(\boldsymbol{x}_i - \boldsymbol{u}_j)^\top, \qquad \mathbf{S}_b = \textstyle\sum_{j=1}^{c} n_j (\boldsymbol{u}_j - \boldsymbol{u})(\boldsymbol{u}_j - \boldsymbol{u})^\top, \tag{3}$$

where $\mathbf{S}_w$ and $\mathbf{S}_b$ are denoted as within-class and between-class scatter respectively, $\boldsymbol{u}_j$ and $\boldsymbol{u}$ being the class-wise means and the global mean, and n_j the sample size of class j. Then a discriminative assignment of $\mathbf{Y}$ for re-id can be obtained by optimising the FDA's criterion, with the earlier derived representation in Eq. (2) as a constraint:

$$\mathbf{Y} = \arg\max_{\mathbf{Y}} \mathit{trace}\Big(\left(\mathbf{P}^\top \mathbf{S}_w \mathbf{P}\right)^\dagger \left(\mathbf{P}^\top \mathbf{S}_b \mathbf{P}\right)\Big), \quad s.t.\ \mathbf{P} = \left(\mathbf{X}\mathbf{X}^\top + \lambda \mathbf{I}\right)^\dagger \mathbf{X}\mathbf{Y}. \tag{4}$$

There exists a simple solution [28, 46] $\mathbf{Y} \in \mathbb{R}^{n \times k}$ to Eq. (4) as follows:

$$k = c; \quad \mathbf{Y}_{ij} = \tfrac{1}{\sqrt{n_j}}, \ \text{if } l_i = j; \quad \text{and} \quad \mathbf{Y}_{ij} = 0, \ \text{if } l_i \neq j. \tag{5}$$

Equation (5) has three interesting properties: (1) Output matrix $\mathbf{Y}$ can be directly constructed with negligible cost before training. Once $\mathbf{Y}$ is set, our Highly Efficient Regression (HER) solution for re-id can be directly acquired by Eq. (2); (2) Unlike existing re-id approaches [26, 37, 44, 47] which also exploit Fisher's criterion, the calculations of $\mathbf{S}_w$ and $\mathbf{S}_b$ are no longer required by HER's solution in Eq. (2). (3) By setting $\mathbf{Y}$ as Eq. (5), all samples of the same identity tend to be projected onto one single point, thus the learned subspace becomes maximally person identity-discriminative. A similar idea was considered by a recent

study [47] which learns a discriminative null space for projection. But since no regularisation is used in [47], the learned null space on training data is often not optimal for testing due to over-fitting. In contrast, the HER model is less prone to over-fitting because of regularisation (Eq. (1)), and achieves better re-id performance (Sec. 4).

On computational efficiency, HER is much more efficient than iterative optimisation re-id approaches, e.g. [1, 23, 24, 25, 51, 52], since our closed-form solution is significantly simpler and faster in computation; Compared to those models based on solving eigenproblems [26, 37, 44, 47], HER's solution is also more efficient. For a feature dimension d, sample size n, and $m = \min(d,n)$, performing an eigen-decomposition to solve a generalised eigenproblem or a null space requires $\frac{3}{2}dnm + \frac{9}{2}m^3$ floating point addition/multiplications [38], whereas solving the linear system in Eq. (2) takes $\frac{1}{2}dnm + \frac{1}{6}m^3$ [9]. This gives maximally 9 times speed-up by HER (see Sec. 4 for empirical results).

As a linear model, HER can be further kernalised to capture non-linear transforms. Assume the kernel matrix is $\mathbf{K} \in \mathbb{R}^{n\times n}$, then kernelised projection vectors $\mathbf{Q} \in \mathbb{R}^{n\times c}$ are:

$$\mathbf{Q} = (\mathbf{K}\mathbf{K}^\top + \lambda\mathbf{K})^\dagger \mathbf{K}\mathbf{Y}. \tag{6}$$

3.2 Incremental HER$^+$

Most conventional re-id models, and the proposed HER, are still limited in their scalability, since they only consider a batch learning scheme: To update a model with newly labelled data, they have to add the new data to the overall training pool and re-train from scratch. In a real-world where data cumulation can increase significantly, such a re-training strategy will become extremely expensive, e.g. take hours or days to perform even one update.

To further improve the scalability of HER, we introduce an incremental learning formulation HER$^+$, enabling fast model updates without the need for re-training from scratch. Suppose at time t, $\mathbf{X}_t \in \mathbb{R}^{d\times n_t}$ is the features of n_t previously labelled images of c_t person identities, $\mathbf{Y}_t \in \mathbb{R}^{n_t\times c_t}$ is their indicator matrix defined by Eq. (5); $\mathbf{X}' \in \mathbb{R}^{d\times n'}$ the features of n' newly labelled images of c' new person classes. and $\mathbf{Y}' \in \mathbb{R}^{n'\times(c_t+c')}$ is its corresponding indicator matrix defined by Eq. (5). Let $\mathbf{T}_t = \mathbf{X}_t\mathbf{X}_t^\top + \lambda\mathbf{I}$, then HER's projection $\mathbf{P}_t \in \mathbb{R}^{d\times c_t}$ at time t can then be written as $\mathbf{P}_t = \mathbf{T}_t^\dagger \mathbf{X}_t\mathbf{Y}_t$ (Eq. (2)). Next, we derive an incremental update version of HER with a two-step updating scheme.

Updating $\mathbf{T}^\dagger$ - The first step is to update matrix $\mathbf{T}^\dagger$. After incorporating the new data at t, the updated data matrix and indicator matrix can be represented as:

$$\mathbf{X}_{t+1} = [\mathbf{X}_t,\, \mathbf{X}'], \quad \mathbf{Y}_{t+1} = \begin{bmatrix} \mathbf{Y}_t \oplus \mathbf{O} \\ \mathbf{Y}' \end{bmatrix}, \tag{7}$$

where we define operator $(\cdot)\oplus\mathbf{O}$ as augmenting a matrix by padding appropriate numbers of zero columns on the right. Since $\mathbf{T}_t = \mathbf{X}_t\mathbf{X}_t^\top + \lambda\mathbf{I}$, we write $\mathbf{T}_{t+1}$ as:

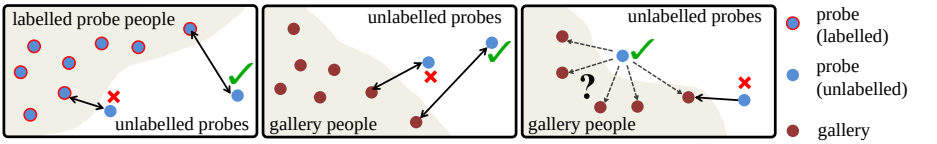
$$\mathbf{T}_{t+1} = \mathbf{T}_t + \mathbf{X}'\mathbf{X}'^\top. \tag{8}$$

Applying the Sherman-Morrison-Woodbury formula [43] to Eq. (8), the update of $\mathbf{T}^\dagger$ is:

$$\mathbf{T}_{t+1}^\dagger = \mathbf{T}_t^\dagger - \mathbf{T}_t^\dagger\mathbf{X}'\big(\mathbf{I} + \mathbf{X}'^\top\mathbf{T}_t^\dagger\mathbf{X}'\big)^\dagger\mathbf{X}'^\top\mathbf{T}_t^\dagger. \tag{9}$$

Updating P - Next we update projection matrix $\mathbf{P}$. Eq. (2) and Eq. (7) together give:

$$\mathbf{P}_{t+1} = \mathbf{T}_{t+1}^\dagger\mathbf{X}_{t+1}\mathbf{Y}_{t+1} = (\mathbf{T}_{t+1}^\dagger\mathbf{X}_t\mathbf{Y}_t)\oplus\mathbf{O} + \mathbf{T}_{t+1}^\dagger\mathbf{X}'\mathbf{Y}'. \tag{10}$$



(a) appearance diversity (b) matching uncertainty (c) ranking uncertainty

Figure 2: Joint exploration-exploitation criteria for active re-id.

Expanding $\mathbf{T}_{t+1}^{\dagger}$ in the first term with Eq. (9), and consider $\mathbf{P}_t = \mathbf{T}_t^{\dagger}\mathbf{X}_t\mathbf{Y}_t$, the update of $\mathbf{P}$ is:

$$\mathbf{P}_{t+1} = \left(\mathbf{P}_t - \mathbf{T}_t^{\dagger}\mathbf{X}'\left(\mathbf{I} + \mathbf{X}'^{\top}\mathbf{T}_t^{\dagger}\mathbf{X}'\right)^{\dagger}\mathbf{X}'^{\top}\mathbf{P}_t\right) \oplus \mathbf{O} + \mathbf{T}_{t+1}^{\dagger}\mathbf{X}'\mathbf{Y}'. \tag{11}$$

The above updating scheme (Eq. (9) and Eq. (11)) forms our HER$^+$ algorithm. It shows that the previous data $\{\mathbf{X}_t, \mathbf{Y}_t\}$ is not needed for model updates. Analogous to Eq. (6), kernelisation can also be applied as a pre-processing step.

Implementation consideration - The HER$^+$ algorithm supports updates on both a single and/or a small chunk of data with $n' \geqslant 1$. If the data chunk size $n' \ll d$, where d is the feature dimension, it is faster to perform n' separate updates on each new sample instead of by chunk. The reason is that in such a way the Moore-Penrose matrix inverse in Eq. (9) and Eq. (11) can be reduced to n' separate scaler inverses whose computation is much cheaper.

3.3 Active Re-Id by Joint Exploration-Exploitation

The efficient model updates achieved by HER$^+$ makes active learning plausible in a re-id system in order to reduce human labelling effort. To differentiate from the aforementioned conventional setting, we define a more scalable *active re-id* setting as follows:

Active re-id - An *unlabelled* probe set $\widetilde{\mathcal{P}}$ and gallery set $\widetilde{\mathcal{G}}$ are the only available data before training. Assume at time step $t \in \{1, \cdots, \tau\}$ where τ is a limited human labelling budget, m_t denotes the currently learned re-id model, $\widetilde{\mathcal{P}}_t$ and $\widetilde{\mathcal{G}}_t$ denote the remaining unlabelled data at time t. *Active re-id* describes the following procedure: (1) Image(s) $\mathcal{I}_t^p \in \widetilde{\mathcal{P}}_t$ of a new training probe identity l_t is actively selected by m_t, according to its usefulness/importance measured by certain active sampling criteria; (2) A ranking list of unlabelled gallery images $\widetilde{\mathcal{G}}_t$ against the selected probe is then generated by m_t; (3) Human annotators verify $\mathcal{I}_t^p$'s cross-view matching image(s) $\mathcal{I}_t^g \in \widetilde{\mathcal{G}}_t$ in the ranking list; (4) Model m_{t+1} is updated from new annotation $(\mathcal{I}_t^p, \mathcal{I}_t^g, l_t)$, and repeat from step (1).

To select the samples that would maximise model discrimination capacity, we propose a joint exploration-exploitation active sampling strategy, consists of three criteria (Fig. 2):

Appearance diversity exploration - The diversity of training persons' appearances is critical for a re-id model to generalise well, thus the preferred next probe image to annotate should lie in the most unexplored part within the population. Assume at time t, the distance between any two samples $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ obtained by the current re-id model $m_t = \mathbf{P}_t$ is:

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2 | m_t) = (\boldsymbol{x}_1 - \boldsymbol{x}_2)^{\top}\mathbf{P}_t\mathbf{P}_t^{\top}(\boldsymbol{x}_1 - \boldsymbol{x}_2). \tag{12}$$

Let $\widetilde{\mathcal{P}}_t$ and $\mathcal{P}_t$ denote the unlabelled and labelled part of probe set $\widetilde{\mathcal{P}}$ at time t respectively ($\widetilde{\mathcal{P}}_t \bigcup \mathcal{P}_t = \widetilde{\mathcal{P}}$), we thus measure the diversity degree of an unlabelled probe sample $\boldsymbol{x}_i^p \in \widetilde{\mathcal{P}}_t$ by the distance to its *within-view nearest neighbour* in $\mathcal{P}_t$ (Fig. 2 (a)):

$$\varepsilon_1(\boldsymbol{x}_i^p) = \min\, d(\boldsymbol{x}_i^p, \boldsymbol{x}_j^p | m_t), \quad s.t.\ \boldsymbol{x}_i^p \in \widetilde{\mathcal{P}}_t,\ \boldsymbol{x}_j^p \in \mathcal{P}_t. \tag{13}$$

Matching uncertainty exploitation - Uncertainty-based exploitative sampling schemes has been widely researched for classification tasks [8, 18, 39], querying the least certain sample for human to annotate. Tailor-made for re-id tasks, our second criterion here prefers the probe samples staying far away from the gallery after projection at time t, i.e. the re-id model m_t remains unclear on what are the corresponding cross-view appearances of these 'poorly-matched' probe images. We measure the matching uncertainty of an unlabelled probe sample $\boldsymbol{x}_i^p \in \widetilde{\mathcal{P}}_t$ by the distance to its *cross-view nearest neighbour* in $\widetilde{\mathcal{G}}$ (Fig. 2 (b)):

$$\varepsilon_2(\boldsymbol{x}_i^p) = \min\, d(\boldsymbol{x}_i^p, \boldsymbol{x}_j^g | m_t), \quad s.t.\ \boldsymbol{x}_i^p \in \widetilde{\mathcal{P}}_t,\ \boldsymbol{x}_j^g \in \widetilde{\mathcal{G}}. \tag{14}$$

Ranking uncertainty exploitation - Due to similar appearances among different identities, a weak re-id model could generate close ranking scores for those visually-ambiguous gallery identities to a given probe. It would thus be useful to ask human to label such a probe sample to enhance a re-id model's discrimination power (Fig. 2 (c)). We define a distribution over all gallery samples $\boldsymbol{x}_j^g \in \widetilde{\mathcal{G}}$, conditioned on a given probe $\boldsymbol{x}_i^p$ according to their ranking scores:

$$p_{m_t}(\boldsymbol{x}_j^g | \boldsymbol{x}_i^p) = \frac{1}{Z_i^t} e^{-d(\boldsymbol{x}_i^p, \boldsymbol{x}_j^g | m_t)}, \quad \text{where } Z_i^t = \sum_k e^{-d(\boldsymbol{x}_i^p, \boldsymbol{x}_k^g | m_t)},\ \boldsymbol{x}_k^g \in \widetilde{\mathcal{G}}. \tag{15}$$

Large entropy of the above distribution means that the ranking scores are close to each other, indicating m_t's uncertainty on its returned ranking list. Thus, our third criterion is:

$$\varepsilon_3(\boldsymbol{x}_i^p) = -\sum_j p_{m_t}(\boldsymbol{x}_j^g | \boldsymbol{x}_i^p) \log p_{m_t}(\boldsymbol{x}_j^g | \boldsymbol{x}_i^p), \quad s.t.\ \boldsymbol{x}_i^p \in \widetilde{\mathcal{P}}_t,\ \boldsymbol{x}_j^g \in \widetilde{\mathcal{G}}. \tag{16}$$

Joint exploration-exploitation - Similar to [4, 8], we combine exploitation and exploration into our final active selection strategy. Specifically, our final selection criterion is a sum of $\varepsilon_1, \varepsilon_2, \varepsilon_3$, where each score is normalised to $(0,1)$:

$$\varepsilon(\boldsymbol{x}_i^p) = \varepsilon_1(\boldsymbol{x}_i^p) + \varepsilon_2(\boldsymbol{x}_i^p) + \varepsilon_3(\boldsymbol{x}_i^p). \tag{17}$$

The unlabelled probe samples can thus be sorted according to this selection criterion, and the one with highest score is then selected for human annotation. Finally, whenever such newly labelled data is obtained, our HER$^+$ model can perform an immediate incremental update. Deployed together with the above active sampling criteria, our HER$^+$ model helps to focus human effort only on labelling those most useful samples, thus maximises the cost-effectiveness in learning a scalable re-id model.

4 Experiments

We conducted experiments under both the conventional supervised re-id setting and the new active re-id setting to evaluate our HER (Sec. 3.1) and HER$^+$ (Sec. 3.2) models respectively. We also evaluated the proposed active sampling criteria (Sec. 3.3).

Datasets and features - Three benchmarks, VIPeR [14], CUHK01 [22], and CUHK03 (manual version) [23] were considered for our evaluation. VIPeR contains a total of 632 people with one image per person per view, CUHK01 contains 971 people with two images per person per view, whilst CUHK03 contains 13,164 images of 1,360 people with a maximum of five images per person per view. By default the Local Maximal Occurrence (LOMO) features [26] (26,960 dimensions) are adopted for image representation.

dataset	VIPeR [12]				CUHK01 [23]				CUHK03 [23]			
rank	R1	R5	R10	R20	R1	R5	R10	R20	R1	R5	R10	R20
l_2 norm	15.6	27.7	36.2	49.2	20.5	37.1	45.3	55.3	11.3	27.3	39.8	55.8
Mid-level [51]	29.1	52.3	66.0	79.9	34.3	55.1	65.0	74.9	-	-	-	-
Deep+ [1]	34.8	63.6	75.6	84.5	47.5	71.6	80.3	87.5	54.7	86.5	93.9	98.1
kLFDA [44]	38.6	69.2	80.4	89.2	54.6	80.5	86.9	92.0	45.8	77.1	86.8	93.1
XQDA [26]	40.0	68.1	80.5	91.1	63.2	83.9	90.0	94.2	52.2	82.2	92.1	96.3
MLAPG [25]	40.7	69.9	82.3	92.4	64.2	85.4	90.8	94.9	58.0	**87.1**	94.7	**98.0**
NFST [47]	42.3	71.5	82.9	92.1	65.0	85.0	89.9	94.4	58.9	85.6	92.5	96.3
HER (LOMO)	**45.1**	**74.6**	**85.1**	**93.3**	**68.3**	**86.7**	**92.6**	**96.2**	**60.8**	87.0	**95.2**	97.7
Ensembles [65]	45.9	77.5	88.9	**95.8**	53.4	76.4	84.4	90.5	62.1	89.1	94.3	97.8
HER (fusion)	**53.0**	**79.8**	**89.6**	95.5	**71.2**	**90.0**	**94.4**	**97.3**	**65.2**	**92.2**	**96.8**	**99.1**

Table 1: Re-id comparisons under the conventional setting (100% labelled training set).

4.1 Conventional Re-Id Evaluation

Settings and Comparisons - We first considered the standard fully-supervised re-id setting to evaluate the proposed HER model (Sec. 3.1). Seven state-of-the-art baselines were compared: Mid-level Filter [51], Deep+ [1], kLFDA [44], XQDA [26], MLAPG [25], NFST [47], and Ensembles [65]. Whenever code is available and features can be replaced, we compared them using the same LOMO features. In addition, we also compared with the Ensembles fusion model in which four types of features with two different matching metrics are fused [65]. For comparison, we fused two types of features for HER: LOMO [26] and that of [27]. For each feature type we trained one HER model independently and performed a final score-level fusion (HER (fusion)). For data partitions, VIPeR and CUHK01 were randomly split into two equal halves for training/testing as in [26, 47], and repeated by 10 times for averaging. On CUHK03, we adopted the standard 20 training/testing splits [23] and the single-shot testing protocol [26, 47]. In implementation, we applied the same RBF kernel settings of kLFDA [44] and NFST [47]. The free parameter λ in Eq. (1) and the parameters of other models were all determined by cross-validation.

Results - The Cumulated Matching Characteristics (CMC) curve is adopted as the evaluation metric. The results are shown in Table 1. It is evident that HER significantly outperforms all existing competitors on all three datasets. Taking the Rank-1 recognition rate as a comparison, HER (LOMO) has notably improved the current state-of-the-art NFST [47] from 42.3% to 45.1% on VIPeR, from 65.0% to 68.3% on CUHK01, and from 58.9% to 60.8% on CUHK03 when the same LOMO feature is used. This shows the superiority of HER's regularised discriminative projection over the null-space model NFST. Although both models encourage the same class samples to be projected onto a single point, the NFST model does not employ regularisation therefore is sensitive to over-fitting. Through fusing different feature representations, HER model's performance can be further boosted. Specifically, HER (fusion) outperforms the Ensembles fusion model [65], with Rank-1 rates of 53.0%, 71.2% and 65.2% for the three benchmarks respectively, 7.1%, 17.8%, and 3.1% better than the Ensembles; despite that the Ensembles model fuses four feature types and two re-id matching models, whilst a single HER model fuses only two feature types.

To evaluate the computational efficiency of HER on model training time, critical for scaling up to large data, we chose three representative existing models with code available:

time(sec)	**HER**	kLFDA	XQDA	MLAPG
VIPeR	**1.2**	5.0	4.1	50.9
CUHK01	**4.2**	45.9	51.9	746.6
CUHK03	**248.8**	2203.2	3416.0	4.0×10^4

Table 2: Training time for conventional setting.

time$_U$(sec)	VIPeR	CUHK01	CUHK03
HER$^+$	**0.03**	**0.2**	**0.6**
HER	0.4	2.3	53.8

Table 3: Model updating time (50% labelled).

rank	dataset labelled (%)	VIPeR [ ] 10	20	30	40	50	100	CUHK01 [ ] 10	20	30	40	50	100
R1(%)	Random	15.8	20.8	27.2	30.1	35.2	**45.2**	21.4	34.4	44.8	48.4	53.7	**64.7**
	Density [8]	15.7	23.3	27.6	30.4	34.2	**45.2**	23.5	36.4	46.3	49.1	53.9	**64.7**
	JointE^2	**18.2**	**25.7**	**30.7**	**34.6**	**37.1**	**45.2**	**29.9**	**41.0**	**47.2**	**51.6**	**55.8**	**64.7**
R5(%)	Random	35.8	45.2	51.8	57.1	61.4	**74.5**	44.1	58.6	68.5	73.5	77.1	**85.2**
	Density [8]	38.1	50.7	57.1	59.4	64.2	**74.5**	47.4	63.4	70.4	**75.2**	78.2	**85.2**
	JointE^2	**41.0**	**51.0**	**57.2**	**61.7**	**66.8**	**74.5**	**54.9**	**64.8**	**71.1**	75.1	**79.8**	**85.2**

Table 4: Re-id performances under active re-id scheme (10% to 50% training data labelled). We also include incrementally training with 100% (fully) labelled training data for comparison to Table 1.

kLFDA, XQDA, and MLAPG. Whilst kLFDA and XQDA are solved by generalised eigen-problems (closed-form), MLAPG is computed by iterative optimisation using line-searching. We recorded each model's training time in each trial and averaged over 10 trials. Our experiments were performed on a Linux server@2.6GHz CPU with 384GB memory. Table 2 reports the averaged results: HER takes 1.2 seconds to train on a small dataset VIPeR (over 10^2 images). This is respectively $\times$4.2 times, $\times$3.4 times, and $\times$42.4 times faster than kLFDA, XQDA, and MLAPG. On a much larger dataset CUHK03 (over 10^4 images), comparing to kLFDA (2,203.2 sec), XQDA (3,416.0 sec) and MLAPG (40,000.0 sec), HER only takes 248.8 seconds to train. That is respectively $\times$8.8 times, $\times$13.7 times, and $\times$160.8 times faster. This shows significant advantage of HER over other models for scaling up to larger data. Although we could not evalute the training time of Deep+ model [1] due to unavailable code, typical deep model training time is expected much longer than those in Table 2.

4.2 Active Re-Id Evaluation

Settings - Next we consider the active re-id setting (Sec. 3.3) to evaluate our HER$^+$ algorithm and active sampling criteria. Our experiments were conducted on VIPeR and CUHK01. For each dataset, we train HER$^+$ model incrementally, with the next probe sample randomly sampled (Random), or actively selected by the new joint exploration-exploitation criteria (JointE^2). We also compared with another active sampling strategy [8] which finds the densest region of the probe sample space (Density). The groundtruth matched-images of selected probes were given as simulated human-in-the-loop labelling. As in a realistic re-id system, an operator can only afford to actively label a small proportion of the vast amount of unlabelled data, we varied the labelling budget from 10% to 50% of the overall training set. Our experiments were conducted in 10 trials with the averaged results reported.

Results - Table 4 shows the Rank-1 and Rank-5 recognition rates on both datasets. It is evident that: (1) On VIPeR, the Rank-1 rate of JointE^2 is averagely 3.4% higher than Random, and it also compares favourably against the alternative Density criteria [8]. (2) Active re-id achieves better performances with less human labelling efforts. For example, on VIPeR active labelling only 30% of the data achieves a 30.7% Rank-1 rate, already higher than that of randomly labelling 40% of the data. Also, when applied with JointE^2, only trained with 50% labelled data on CUHK01, the HER$^+$ model achieved 55.8% for Rank-1, already higher than the Mid-level, Deep+, and kLFDA models trained on 100% fully labelled training set (Table 1). (3) When incrementally trained with 100% labelled training set, HER$^+$ achieved comparable results with the batch-based HER (Table 1), suggesting that incremental updates of HER$^+$ does not sacrifice re-id performances.

Finally we evaluated computational costs of incremental model update versus re-training. A further experiment was conducted: We first trained a HER model by using 50% labelled training data on each dataset. We then provided additional one pair of matched images as

newly labelled data from online confirmation, and updated the trained HER model as follows: (1) Incremental update by HER$^+$ on the two newly arrived images; (2) Re-train HER using an enlarged training data pool consisting of the initial 50% labelled data plus the new pair. The model updating times are presented in Table 3 (averaged results over 10 trials). It is evident that HER$^+$ only takes 0.03 second, 0.2 second and 0.6 second respectively to perform real-time model update on the three benchmarks. This is $\times$13.3 times, $\times$11.5 times and $\times$89.7 times faster than re-training HER in batch mode. In particular, on a larger dataset CUHK03, the advantage of incremental model update by HER$^+$ over re-training is more significant. Given that HER is already the fastest batch learning model over others by 1-2 orders of magnitude (Table 2), the incremental model update by HER$^+$ is thus over $\times 10^4$ times faster than re-training a conventional existing model such as kLFDA, XQDA and MLAPG. This offers significant advantage in scaling up to large data in real-world deployments.

5 Conclusion

A highly scalable re-id model HER is proposed by formulating re-id as a ridge regression problem. Compared to existing re-id models, HER is much faster and cheaper to compute, whilst yields superior re-id performance. An incremental HER$^+$ model is introduced to enable sustainable incremental cumulative model update and to facilitate active learning re-id for reducing human supervision in re-id deployment. Extensive experiments show the scalability of HER and HER$^+$ and their potential for large size data re-id deployments.

References

[1] Ejaz Ahmed, Michael J. Jones, and Tim K. Marks. An improved deep learning architecture for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3908–3916. IEEE, 2015. ISBN 978-1-4673-6964-0.

[2] Deng Cai, Xiaofei He, and Jiawei Han. Srda: An efficient algorithm for large-scale discriminant analysis. *Knowledge and Data Engineering, IEEE Transactions on*, 20 (1):1–12, 2008.

[3] Nicolas Cebron and Michael R Berthold. Active learning for object classification: from exploration to exploitation. *Data Mining and Knowledge Discovery*, 18(2):283–299, 2009.

[4] Jiaxin Chen, Zhaoxiang Zhang, and Yunhong Wang. Relevance metric learning for person re-identification by exploiting listwise similarities. *Image Processing, IEEE Transactions on*, 24(12):4741–4755, 2015.

[5] Abir Das, Rameswar Panda, and Amit Roy-Chowdhury. Active image pair selection for continuous person re-identification. In *IEEE International Conference on Image Processing*, Quebec, Canada, September 2015.

[6] Shengyong Ding, Liang Lin, Guangrun Wang, and Hongyang Chao. Deep feature learning with relative distance comparison for person re-identification. *Pattern Recognition*, pages 2993–3003, 2015.

[7] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[8] Sandra Ebert, Mario Fritz, and Bernt Schiele. Ralf: A reinforced active learning formulation for object class recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3626–3633, Providence, Rhode Island, United States, June 2012.

[9] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. Person re-identification by symmetry-driven accumulation of local features. In *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, California, United States, June 2010.

[10] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

[11] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Academic press, 2013.

[12] Shaogang Gong, Marco Cristani, Change Loy Chen, and Timothy M. Hospedales. The re-identification challenge. In *Person Re-Identification*. Springer, 2014.

[13] Shaogang Gong, Marco Cristani, Shuicheng Yan, and Chen Change Loy. *Person re-identification*, volume 1. Springer, 2014.

[14] Douglas Gray, Shane Brennan, and Hai Tao. Evaluating appearance models for recognition, reacquisition and tracking. In *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, 2007.

[15] Trevor Hastie, Robert Tibshirani, Jerome Friedman, and James Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 2005.

[16] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[17] Timothy M Hospedales, Shaogang Gong, and Tao Xiang. A unifying theory of active discovery and learning. In *European Conference on Computer Vision*, pages 453–466. Florence, Italy, October 2012.

[18] Ajay J Joshi, Fatih Porikli, and Nikolaos Papanikolopoulos. Multi-class active learning for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2372–2379, Miami, Florida, United States, June 2009.

[19] Jaeho Kang, Kwang Ryel Ryu, and Hyuk-Chul Kwon. Using cluster-based sampling to select initial training set for active learning in text classification. In *Advances in knowledge discovery and data mining*, pages 384–388. Sydney, Australia, May 2004.

[20] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M. Roth, and Horst Bischof. Large scale metric learning from equivalence constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, Rhode Island, United States, June 2012.

[21] Kibok Lee and Junmo Kim. On the equivalence of linear discriminant analysis and least squares. In *AAAI Conference on Artificial Intelligence*, 2015.

[22] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[23] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, Ohio, United States, June 2014.

[24] Zhen Li, Shiyu Chang, Feng Liang, Thomas Huang, Liangliang Cao, and John Smith. Learning locally-adaptive decision functions for person verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3610–3617, 2013.

[25] Shengcai Liao and Stan Z. Li. Efficient psd constrained asymmetric metric learning for person re-identification. In *IEEE International Conference on Computer Vision*, December 2015.

[26] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li. Person re-identification by local maximal occurrence representation and metric learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2197–2206, Boston, Massachusetts, United States, June 2015.

[27] Giuseppe Lisanti, Iacopo Masi, and Alberto Del Bimbo. Matching people across camera views using kernel canonical correlation analysis. In *ACM International Conference on Distributed Smart Cameras*, Venice, Italy, November 2014.

[28] Li-Ping Liu, Yuan Jiang, and Zhi-Hua Zhou. Least square incremental linear discriminant analysis. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*, pages 298–306. IEEE, 2009.

[29] Chen Change Loy, Timothy M Hospedales, Tao Xiang, and Shaogang Gong. Stream-based joint exploration-exploitation active learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1560–1567, Providence, Rhode Island, United States, June 2012.

[30] Chen Change Loy, Chunxiao Liu, and Shaogang Gong. Person re-identification by manifold ranking. In *IEEE International Conference on Image Processing*, volume 1. Citeseer, 2013.

[31] Bingpeng Ma, Yu Su, and Frédéric Jurie. Local descriptors encoded by fisher vectors for person re-identification. In *ECCV workshop*, 2012.

[32] Alexis Mignon and Frédéric Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, Rhode Island, United States, June 2012.

[33] Thomas Osugi, Deng Kim, and Stephen Scott. Balancing exploration and exploitation: A new algorithm for active machine learning. In *IEEE International Conference on Data Mining*, Houston, Texas, United States, November 2005.

[34] Christopher C Paige and Michael A Saunders. Lsqr: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software (TOMS)*, 8(1):43–71, 1982.

[35] Sakrapee Paisitkriangkrai, Chunhua Shen, and Anton van den Hengel. Learning to rank in person re-identification with metric ensembles. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1846–1855, 2015.

[36] Cheong Hee Park and Haesun Park. A relationship between linear discriminant analysis and the generalized minimum squared error solution. *SIAM Journal on Matrix Analysis and Applications*, 27(2):474–492, 2005.

[37] Sateesh Pedagadi, James Orwell, Sergio A. Velastin, and Boghos A. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, Portland, Oregon, United States, June 2013.

[38] Roger Penrose. A generalized inverse for matrices. In *Proc. Cambridge Philos. Soc*, volume 51, pages 406–413. Cambridge Univ Press, 1955.

[39] Burr Settles and Mark Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1070–1079. Association for Computational Linguistics, 2008.

[40] Hailin Shi, Xiangyu Zhu, Shengcai Liao, Zhen Lei, Yang Yang, and Stan Z Li. Constrained deep metric learning for person re-identification. *arXiv preprint arXiv:1511.07545*, 2015.

[41] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. Person re-identification by video ranking. In *European Conference on Computer Vision*, Zurich, Switzerland, September 2014.

[42] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. Person re-identification by discriminative selection in video ranking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, January 2016.

[43] Max A Woodbury. Inverting modified matrices. *Memorandum report*, 42:106, 1950.

[44] Fei Xiong, Mengran Gou, Octavia Camps, and Mario Sznaier. Person re-identification using kernel-based metric learning methods. In *European Conference on Computer Vision*. Zurich, Switzerland, September 2014.

[45] Yang Yang, Jimei Yang, Junjie Yan, Shengcai Liao, Dong Yi, and Stan Z. Li. Salient color names for person re-identification. In *European Conference on Computer Vision*, 2014.

[46] Jieping Ye. Least squares linear discriminant analysis. In *Proceedings of the 24th international conference on Machine learning*, pages 1087–1093. ACM, 2007.

[47] Li Zhang, Tao Xiang, and Shaogang Gong. Learning a discriminative null space for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[48] Zhihua Zhang, Guang Dai, and Michael I Jordan. A flexible and efficient algorithm for regularized fisher discriminant analysis. In *Machine Learning and Knowledge Discovery in Databases*, pages 632–647. Springer, 2009.

[49] Zhihua Zhang, Guang Dai, Congfu Xu, and Michael I Jordan. Regularized discriminant analysis, ridge regression and beyond. *The Journal of Machine Learning Research*, 11: 2199–2228, 2010.

[50] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Unsupervised salience learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, Portland, Oregon, United States, June 2013.

[51] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Learning mid-level filters for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, Ohio, United States, June 2014.

[52] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Re-identification by relative distance comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 653–668, March 2013.