Fast Open-World Person Re-Identification

Xiatian Zhu, Botong Wu, Dongcheng Huang, Wei-Shi Zheng

***Abstract*—Existing person re-identification (re-id) methods typically assume that (1) any probe person is guaranteed to appear in the gallery target population during deployment (i.e. closed-world), and (2) the probe set contains only a limited number of people (i.e. small search scale). Both assumptions are artificial and breached in real-world applications, since the probe population in target people search can be extremely vast in practice due to the ambiguity of probe search space boundary. Therefore, it is *unrealistic* that any probe person is assumed as one target people, and a large-scale search in person images is inherently demanded. In this work, we introduce a new person re-id search setting, called Large Scale Open-World (LSOW) re-id, characterised by huge size probe images and open person population in search thus more close to practical deployments. Under LSOW, the under-studied problem of person re-id efficiency is essential in addition to that of commonly-studied re-id accuracy. We therefore develop a novel fast person re-id method, called Cross-view Identity Correlation and vErification (X-ICE) hashing, for joint learning of cross-view identity representation *binarisation* and *discrimination* in a unified manner. Extensive comparative experiments on three large scale benchmarks have been conducted to validate the superiority and advantages of the proposed X-ICE method over a wide range of the state-of-the-art hashing models, person re-id methods, and their combinations.**

***Index Terms*—Person re-identification, large probe population, open search space, fast search, efficient matching, hashing.**

I. Introduction

THE aim of person re-identification (re-id) is to match people across non-overlapping cameras distributed over wide physical areas [1]. Person re-id is inherently challenging due to the large unknown variations across camera views in human pose, illumination condition, view angle, occlusion and background clutter. Re-id is usually performed by matching visual appearance features of person images [2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18]. Two unscalable assumptions are often made by existing methods: **(1)** *Closed-world matching* where every probe person guarantees to exist in the gallery set, which however is largely invalid for real-world applications owning to no such prior knowledge available in deployments; and **(2)** *Small search space* in contrast to enormous search space in practice. Except the former closed-world assumption, this is mainly due to the neglect of open-world matching nature with no precise search space boundary available. Hence, it is inevitable to consider a sufficiently large number of probe people where an unknown high fraction are non-target persons.

Fig. 1: Large scale open-world person re-identification aims to re-identify a gallery target person (right box) among an *inherently immense* probe search population (left box). This is due to no prior knowledge about the search space boundary and many non-target people are inevitable in deployments. Target identity is indicated with coloured bounding box.

Open-world person re-id has been recently investigated by a few studies [19,20]. Specifically, Liao et al. [19] only discussed a generic open-world re-id evaluation metrics. Zheng et al. [20] presented a watch-list based open-world re-id setting along with a transfer learning algorithm for overcoming label scarcity. However, both works consider only a *small scale* re-id matching scenario (i.e. the search space consists of a limited number of probe people), while the inherent *large scale* search scalability problem is still overlooked.

In this work, we propose a more realistic re-id setting, called **Large Scale Open-World** (LSOW) person re-id. LSOW has four important features: **(I)** *Vast probe search population* - The probe image set captured by many cameras in open world contains inevitably a large number of non-target people (also known as imposters), th search space is therefore inherently "large". **(II)** *Fast disjoint-view search* - Fast search in large data pools has been extensively investigated in image retrieval [21]. However, the re-id problem is not a conventional image retrieval problem, as it is particularly constrained by searching person images across disjoint views. Cross-view search can be largely influenced by significant appearance variation of a person due to view transformation, pose and lighting condition change, occlusion and etc. Thus, any fast re-id search models should intrinsically address these challenges while achieving

This work was supported partially by the National Key Research and Development Program of China (2016YFB1001002), NSFC (61522115, 61472456, 61573387, 61661130157), the Royal Society Newton Advanced Fellowship (NA150459), Guangdong Province Science and Technology Innovation Leading Talents (2016TX03X157). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiaochun Cao. (Corresponding author: Wei-Shi Zheng)

Xiatian Zhu is with School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510275, China; and also with School of Electronic Engineering and Computer Science, Queen Mary University of London, United Kingdom. *E-mail*: xiatian.zhu@qmul.ac.uk.

Botong Wu is with School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510275, China; and also with School of Electronic Engineering and Computer Science, Peking University, Beijing, 100871, China. *E-mail*: botongwu@pku.edu.cn.

Dongcheng Huang is with School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510275, China, and is also with Collaborative Innovation Center of High Performance Computing, NUDT, China. *E-mail*: orangehdc@outlook.com.

Wei-Shi Zheng is with School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510275, China, and is also with Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China. *E-mail*: wszheng@ieee.org / zhwshi@mail.sysu.edu.cn.

rapid matching. **(III)** *Sparse training person identities* - In practice only a limited number of persons with cross-view pairwise labelled data are available for building discriminative person re-id models. **(IV)** *Zero-shot transfer learning* - In contrast to conventional fast search methods that consider the matching of training classes in deployment, re-id requires the model transfer knowledge induced from seen training person classes to unseen test person classes in cross-view sense. In a nutshell, the LSOW re-id we investigate here can be regarded as a hybrid of "open-world" re-id and fast search across disjoint views. It challenges existing re-id models in search efficiency and fast search models in re-id efficacy. Under LSOW, re-id efficiency becomes substantially critical: Without the capability of "fast search" over a huge probe population, performing person re-id is not practically applicable and usable even with satisfied recognition accuracy.

To overcome the LSOW challenges, we propose a fast person re-identification matching approach. It enables to not only learn jointly cross-view identity correlation and discrimination, but also perform efficient cross-view matching in deployment. This is realised by exploiting the hashing strategy commonly used in large scale nearest neighbour search [21]. Specifically, we formulate a novel cross-view identity discriminative hashing approach to simultaneously binarising identity representation and learning person discrimination in a unified formulation. In deployment, person images can be represented by short hash codes. Fast re-id search is then achieved by efficient hamming distance. Note, the proposed model is unique to conventional fast search approaches [22,23,24,25,26,27,28,29,30,31] due to the capability of addressing the significant matching challenges inherent to large viewing condition variations across cameras under the "open-world" setting. It is also unique to conventional re-id approaches [4,6,12,32,33,34,35,36,37] due to the capability of learning compact binary representation for effective fast matching.

We extend significantly our preliminary work [38] by making three **contributions** in this manuscript: **(1)** We propose a more realistic Large Scale Open-World (LSOW) person re-id problem. The LSOW eliminates two artificial assumptions made by existing re-id models that fundamentally prevent them from being scalable and applicable to real-world deployments. **(2)** We develop a new person re-id method, called *Cross-view Identity Correlation and vErification* (X-ICE), for efficiently and effectively addressing LSOW re-id matching. In particular, X-ICE learns a re-id discriminative binary coding space by simultaneous cross-view identity correlation hashing and person class discrimination verification. **(3)** We extensively evaluate a wide range (14) of state-of-the-art hashing methods for LSOW person re-id. To our knowledge, this is the first attempt to investigate fast search solutions for large scale re-id in open world. We validated the effectiveness and advantages of X-ICE by extensively comparing both state-of-the-art hashing and person re-id models on three large benchmarking datasets, CUHK03 [39], SYSU [36], and Market-1501 [40].

II. Related Work

Person re-identification. Existing person re-id methods focus on either extracting discriminative view-invariant features [3,6,16,18,32,33,39,41,42,43,44,45] or learning matching distance metrics [4,7,8,12,14,34,35,36,37,46,47,48,49,50,51,52,53]. They typically assume an impractical closed-world person re-id scenario – probe and gallery people are completely overlapping in model deployment. This is not true considering the complex camera network topology in common video surveillance sites, unknown probe search scope, and inevitable occurrence of imposters. In practice, re-id of gallery target people is carried out against a large probe search population, i.e. large scale open-world person matching. The recent person search work [54] considers jointly detection and re-id in a closed-world scenario. While a few works consider open-world person re-id [19,20], their setting is still not practical for real-world deployments, due to the small search scale assumption. In contrast to these existing works, the proposed Large Scale Open-World (LSOW) re-id setting eliminates both closed-world and small probe search space assumptions. This opens a more meaningful research topic for developing scalable person re-id methods. Under the LSOW setting, we further investigate particularly the under-studied but critical re-id matching efficiency issue. This is done by jointly exploring learning to hash and discriminative person re-id matching in a principled formulation.

Hashing. Hashing is commonly adopted in large scale similarity search, due to its low time and space complexity [21]. From data modality view point, existing hashing methods can be broadly grouped into two classes: (1) single-modality based, and (2) multi-modality based. Algorithmically, cross-view methods [22,23,24] should be regarded as a special case of the latter if treating a camera view as an individual modality. In the literature, single-modal hashing methods have been extensively investigated. Representative unsupervised and supervised models include Locality Sensitive Hashing [25,26], Spectral Hashing [27], PCA Hashing [28], Anchor Graphs Hashing [29], Kernel-based Supervised Hashing [30], Supervised Discrete Hashing [31], and so forth. Multi-modal hashing methods can be summarised as some joint learning of individual data modalities for establishing a shared cross-modal coding space. In this space, semantically similar cross-modal samples are enforced to be close, otherwise distant. As such, cross-modal search and matching can be similarly realised as the single-modal case. Notable multi-modal hashing models are Predictable Dual-view Hashing [55], Cross-View Hashing [56], Cross-Modality Similarity Sensitive Hashing [57], Deep Hashing [58], to name a few.

All these existing hashing methods are designed for generic classification tasks given a large search database, e.g. matching the category semantics (among the seen ones in model training) of a query sample. They are less suitable (see evaluations in Section IV-C) for the more challenging person re-id problem characterised by disjoint training and test person classes, more subtle difference between classes, and complex appearance change of the same class across camera views. The proposed LSOW problem is even more difficult due to: (1) A limited amount of target person class training data, (2) A large number of (potentially infinite) person classes in deployment, and (3) Many different person classes may share

visually similar appearance. All these issues pose additional modelling challenges to existing hashing methods. In this study, we jointly cope with these challenges by formulating a new cross-camera hashing based person re-id method. The proposed method combines the advantages of both supervised hashing and person re-id models in a principled manner in order to favourably solve the LSOW problem.

III. Fast Open-World Person Re-Identification

A. Problem Statement

Suppose we need to identify a small set of n_g gallery (target) people $\tilde{G} = \{\tilde{\boldsymbol{I}}_i^g\}_{i=1}^{n_g}$ (see Figure 1) captured from m_g cameras $\{\text{Cam}_i^g\}_{i=1}^{m_g}$ in deployment. To automate the person re-id process, we extract the visual features $\tilde{\boldsymbol{x}}_i^g \in \mathbb{R}^{1\times d}$ (with d the feature dimension) to characterise the appearance pattern of corresponding person images. The feature matrix for all gallery people is denoted as $\tilde{\boldsymbol{X}}_g \in \mathbb{R}^{n_g\times d}$ where each row represents a person image. The search space is formed by n_p probe person images $\tilde{P} = \{\tilde{\boldsymbol{I}}_i^p\}_{i=1}^{n_p}$ captured by different m_p cameras $\{\text{Cam}_i^p\}_{i=1}^{m_p}$ with disjoint field of view against any of n_g gallery cameras. The visual features of probe images $\tilde{P}$ are denoted as $\tilde{\boldsymbol{X}}_p \in \mathbb{R}^{n_p\times d}$. For both $\tilde{P}$ and $\tilde{G}$ image sets, one person may be associated with multiple images from the same camera view, i.e. multi-shot re-id setting [59]. For brevity and clarity, in the remainder, we may use feature vectors to stand for the corresponding images. Test data are indicated with mathematics mode accent $\tilde{*}$ (e.g. $\tilde{\boldsymbol{x}}$ as a test image feature vector) for clear differentiation from training data (e.g. $\boldsymbol{x}$).

In real-world applications, the probe set $\tilde{P}$ can be rather vast, e.g. $n_p >> n_g$. Also, we have no knowledge whether a probe image $\tilde{\boldsymbol{x}}$ describes one gallery target person in prior to re-identification, i.e. open-world. There can be a large quantity of non-target people (*imposters*) in the probe set $\tilde{P}$. For building a discriminative re-id model, a reasonable amount of human labelling budget is often allocated to collect a set of pairwise training data $D_{\text{tr}} = \{\boldsymbol{x}_i^p, \boldsymbol{x}_j^g, \boldsymbol{S}_{ij}\}_{i=1}^{n}$ for one or multiple camera pairs. The label $\boldsymbol{S}_{ij} \in \{0, 1\}$ indicates whether a image pair describes the same person (1) or not (0). To make the labelled data effective for re-id the gallery people $\tilde{G}$, human annotators are likely to form camera pairs by selecting one from gallery views $\{\text{Cam}_i^g\}_{i=1}^{m_g}$ and the one from probe views $\{\text{Cam}_i^p\}_{i=1}^{m_p}$ in constructing the cross-camera training data. Due to limited budget and likely large number of gallery/probe cameras involved, it shall be impossible to exhaustively enumerate all such camera pairs. We call the problem above *Large Scale Open-World* (LSOW) person re-identification. The proposed LSOW is more realistic to practical deployments, different significantly from existing settings with *closed-world* and *small search scale* assumptions [59].

B. Approach Overview

Under LSOW, it is desired and necessary to resolve the person re-id efficiency issue. To this end, we propose exploring the hashing scheme, a well-known fast approximate nearest neighbourhood search approach by learning short binary codes [21]. However, traditional hashing methods are typically developed for the generic classification problem, rather than person re-id requiring the challenging knowledge transfer from seen training classes to unseen test classes. Hence, they are potentially suboptimal. In this work, we formulate a new model for LSOW. Formally, we assume n cross-view true matching training image pairs $\{(\boldsymbol{x}_i^p, \boldsymbol{x}_i^g)\}_{i=1}^{n}$ from n_{id} different persons, with their corresponding feature matrices: $\boldsymbol{X}_p \in \mathbb{R}^{n\times d}$ (of training probe images) and $\boldsymbol{X}_g \in \mathbb{R}^{n\times d}$ (of training gallery images), where $\boldsymbol{x}_i^p$ and $\boldsymbol{x}_i^g$ are the i-th row of $\boldsymbol{X}_p$ and $\boldsymbol{X}_g$, respectively. Each training image $\boldsymbol{x}_i$ from any camera is associated with a one-hot identity label vector $\boldsymbol{y}_i \in \mathbb{R}^{n_{\text{id}}\times 1}$ with the corresponding element as "1" and all others as "0". The feature data are preprocessed to be zero-centered [28,60], i.e. $\sum_{i=1}^{n} \boldsymbol{x}_i^p = \boldsymbol{0}$ and $\sum_{i=1}^{n} \boldsymbol{x}_i^g = \boldsymbol{0}$ where $\boldsymbol{0}$ is d-dimensional zero vector. The identity label matrix $\boldsymbol{S}$ is defined on cross-view image pairs, with elements as:

$$\boldsymbol{S}_{ij} = \begin{cases} 1 & \text{if } \boldsymbol{x}_i^p \text{ and } \boldsymbol{x}_j^g \text{ are of the same person,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We want to learn two hashing functions in training:

$$f_p(\boldsymbol{x}_i^p) = \boldsymbol{x}_i^p \boldsymbol{W}_p, \qquad f_g(\boldsymbol{x}_j^g) = \boldsymbol{x}_j^g \boldsymbol{W}_g, \tag{2}$$

where $\boldsymbol{W}_p \in \mathbb{R}^{d\times c}$ (for probe views) and $\boldsymbol{W}_g \in \mathbb{R}^{d\times c}$ (for gallery views) denote the to-be-learned function parameters (data projection matrices). The hash codes of length c can be obtained by thresholding as

$$\begin{aligned} \boldsymbol{B}_p &= \text{sign}(\boldsymbol{X}_p \boldsymbol{W}_p) \in \{-1, 1\}^{n_p\times c}, \\ \boldsymbol{B}_g &= \text{sign}(\boldsymbol{X}_g \boldsymbol{W}_g) \in \{-1, 1\}^{n_g\times c}, \end{aligned} \tag{3}$$

where the element-wise function $\text{sign}(\cdot)$ returns "1" for positive numbers and "-1" for non-positive numbers. For performing LSOW re-id matching, two steps are included: (1) Encoding person images into compact hash codes; (2) Matching the identity of a given gallery image against a large open probe population by the efficient hamming distance in hash coding space. For achieving re-id discrimination, we shall require that the hash codes are similar for intra-identity images and dissimilar for inter-identity ones in cross-camera sense. This is realised by formulating a novel *Cross-view Identity Correlation and vErification* (X-ICE) model.

C. Joint Correlation Hashing and Discriminative Verification

X-ICE has two parts: (I) cross-view identity correlation hashing, and (II) cross-view identity verification regularisation.

(I) Cross-view Identity Correlation Hashing. For extracting person-sensitive appearance information, we exploit positive pairwise training data since they encode most discriminative knowledge for person re-id. To characterise the underlying identity correlation across camera views, we adopt the cosine similarity between cross-camera person images:

$$\begin{aligned} \text{cosine}\big(f_p(\boldsymbol{x}_i^p), f_g(\boldsymbol{x}_j^g)\big) &= \frac{f_p(\boldsymbol{x}_i^p)\big(f_g(\boldsymbol{x}_j^g)\big)^\top}{\|f_p(\boldsymbol{x}_i^p)\|_2 \|f_g(\boldsymbol{x}_j^g)\|_2} \\ &= \frac{\boldsymbol{x}_i^p \boldsymbol{W}_p \boldsymbol{W}_g^\top \boldsymbol{x}_j^{g\top}}{\sqrt{\boldsymbol{x}_i^p \boldsymbol{W}_p \boldsymbol{W}_p^\top \boldsymbol{x}_i^{p\top}} \sqrt{\boldsymbol{x}_j^g \boldsymbol{W}_g \boldsymbol{W}_g^\top \boldsymbol{x}_j^{g\top}}}. \end{aligned} \tag{4}$$

In spirit of maximum margin [61,62], we further replace the ratio relation with subtraction in Eqn. (4) as:

$$\begin{aligned}&\text{cosine}\big(f_p(\boldsymbol{x}_i^p), f_g(\boldsymbol{x}_j^g)\big) \approx \\ &\boldsymbol{x}_i^p \boldsymbol{W}_p \boldsymbol{W}_g^\top \boldsymbol{x}_j^{g\top} - \sqrt{\boldsymbol{x}_i^p \boldsymbol{W}_p \boldsymbol{W}_p^\top \boldsymbol{x}_i^{p\top}} \sqrt{\boldsymbol{x}_j^g \boldsymbol{W}_g \boldsymbol{W}_g^\top \boldsymbol{x}_j^{g\top}}.\end{aligned} \tag{5}$$

Then, we define the hashing quantisation loss:

$$l_{\text{quan}} = \sum_{s\in\{p,g\}} \|\boldsymbol{B}_s - \boldsymbol{X}_s \boldsymbol{W}_s\|_F^2, \tag{6}$$

with $\|\cdot\|_F$ denoting the Frobenius/Euclidean norm. After combining the quantisation loss into Eqn. (5), we have the following hashing optimisation (minimisation) problem:

$$\begin{aligned}O_{\text{ic}} = &\underbrace{\Big(\|\boldsymbol{B}_p - \boldsymbol{X}_p \boldsymbol{W}_p\|_F^2 + \|\boldsymbol{B}_g - \boldsymbol{X}_g \boldsymbol{W}_g\|_F^2\Big)}_{\text{Quantisation loss}} - \alpha \sum_{(i,j)} \boldsymbol{S}_{ij} \\ &\times \underbrace{\left(\boldsymbol{x}_i^p \boldsymbol{W}_p \boldsymbol{W}_g^\top \boldsymbol{x}_j^{g\top} - \sqrt{\boldsymbol{x}_i^p \boldsymbol{W}_p \boldsymbol{W}_p^\top \boldsymbol{x}_i^{p\top}} \sqrt{\boldsymbol{x}_j^g \boldsymbol{W}_g \boldsymbol{W}_g^\top \boldsymbol{x}_j^{g\top}}\right)}_{\text{Approximated cross-view positive correlation}} \\ \text{s.t.} \quad &\boldsymbol{W}_p^\top \boldsymbol{W}_p = \boldsymbol{I}_{c\times c}, \quad \boldsymbol{W}_g^\top \boldsymbol{W}_g = \boldsymbol{I}_{c\times c},\end{aligned} \tag{7}$$

where α is a trade-off parameter. The two constraints underneath enforce $\boldsymbol{W}_p$ and $\boldsymbol{W}_g$ to be orthogonal projections. Note that, we use a single model to characterise either all probe cameras ($\boldsymbol{W_p}$) or all gallery cameras ($\boldsymbol{W_g}$). This not only simplifies the model learning and deployment task, but also mitigates significantly the tedious requirement of collecting per camera-pair training data (prohibitive in real-world since there are a quadratic number of camera pairs).

View Context Discrepancy Regularisation. Visual context has proven important in various vision problems [63,64,65]. In person re-id, visual context refers to the similarity/dissimilarity relation of different camera views in terms of imaging characteristics and environmental factors, e.g. viewpoint, background and illumination conditions. Intuitively, similar imaging condition between probe and gallery cameras should mean small discrepancy between hashing models $\boldsymbol{W}_p$ (probe) and $\boldsymbol{W}_g$ (gallery), and vice verse. Motivated by [36], we enforce a View Context Discrepancy (VCD) regularisation into our cross-view identity correlation hashing algorithm. The purpose is to globally and contextually regularise the identity coding procedure by *explicitly* imposing the viewing condition correlation constraint in model optimisation.

Formally, we model the discrepancy between hashing models $\boldsymbol{W}_p$ and $\boldsymbol{W}_g$ by the Bregman divergence [66,67]:

$$d_{\text{breg}}^h = h(\boldsymbol{W}_p) - h(\boldsymbol{W}_g) - \Delta h(\boldsymbol{W}_g)^\top (\boldsymbol{W}_p - \boldsymbol{W}_g), \tag{8}$$

where h denotes a strictly convex function: $h : \mathbb{R}^{d\times c} \to \mathbb{R}$, with its derivative defined as $\Delta h(\cdot)$. We adopt the Frobenious norm due to its formulation consistency with the proposed identity correlation modelling (Eqn. (7)) and therefore facilitating model optimisation. Specifically, by setting $h(\boldsymbol{W}_*) = \|\boldsymbol{W}_*\|_F^2$, we have

$$R_{\text{vcd}} = \|\boldsymbol{W}_p - \boldsymbol{W}_g\|_F^2. \tag{9}$$

We then extend our objective function (Eqn. (7)) as

$$O_{\text{ic}} + \underbrace{\lambda_{\text{vcd}} \|\boldsymbol{W}_p - \boldsymbol{W}_g\|_F^2}_{\text{VCD}}, \tag{10}$$

where $\lambda_{\text{vcd}} > 0$ is the trade-off parameter for balancing hashing loss and VCD regularisation.

Upper Bound Approximation. It is difficult to exactly optimise O_{ic}. We therefore derive an upper bound. Specifically, by the Jensen's inequality [68], we have

$$\begin{aligned}&\sqrt{\boldsymbol{x}_i^p \boldsymbol{W}_p \boldsymbol{W}_p^\top \boldsymbol{x}_i^{p\top}} \sqrt{\boldsymbol{x}_j^g \boldsymbol{W}_g \boldsymbol{W}_g^\top \boldsymbol{x}_j^{g\top}} \\ \leq\ &\frac{1}{2}\Big(\boldsymbol{x}_i^p \boldsymbol{W}_p \boldsymbol{W}_p^\top \boldsymbol{x}_i^{p\top} + \boldsymbol{x}_j^g \boldsymbol{W}_g \boldsymbol{W}_g^\top \boldsymbol{x}_j^{g\top}\Big),\end{aligned} \tag{11}$$

which can be used to simplify the approximated cross-view positive correlation in Eqn. (7). The matrix form is:

$$\begin{aligned}&\text{tr}(\boldsymbol{W}_p^\top \boldsymbol{X}_p^\top \boldsymbol{S} \boldsymbol{X}_g \boldsymbol{W}_g) - \\ &\frac{1}{2}\Big(\text{tr}(\boldsymbol{W}_p^\top \boldsymbol{X}_p^\top \boldsymbol{L}_r \boldsymbol{X}_p \boldsymbol{W}_p) + \text{tr}(\boldsymbol{W}_g^\top \boldsymbol{X}_g^\top \boldsymbol{L}_c \boldsymbol{X}_g \boldsymbol{W}_g)\Big).\end{aligned} \tag{12}$$

As such, we build an upper bound $O_{\text{ic}}^{\text{ub}}$ as

$$\begin{aligned}&O_{\text{ic}} \leq O_{\text{ic}}^{\text{ub}} = \\ &\Big(\|\boldsymbol{B}_p - \boldsymbol{X}_p \boldsymbol{W}_p\|_F^2 + \|\boldsymbol{B}_g - \boldsymbol{X}_g \boldsymbol{W}_g\|_F^2\Big) - \\ &\alpha\Big(\text{tr}(\boldsymbol{W}_p^\top \boldsymbol{X}_p^\top \boldsymbol{S} \boldsymbol{X}_g \boldsymbol{W}_g) - \\ &\frac{1}{2}\text{tr}(\boldsymbol{W}_p^\top \boldsymbol{X}_p^\top \boldsymbol{L}_r \boldsymbol{X}_p \boldsymbol{W}_g) - \frac{1}{2}\text{tr}(\boldsymbol{W}_g^\top \boldsymbol{X}_g^\top \boldsymbol{L}_c \boldsymbol{X}_g \boldsymbol{W}_g)\Big),\end{aligned} \tag{13}$$

where $\boldsymbol{L}_r$ and $\boldsymbol{L}_c$ are diagonal matrices with elements as the row and column summation of identity label matrix $\boldsymbol{S}$. We then assemble parameters from different camera views as

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{X}_p & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_g \end{bmatrix}, \tilde{\boldsymbol{S}} = \alpha \begin{bmatrix} -\boldsymbol{L}_r & \boldsymbol{S} \\ \boldsymbol{S}^\top & -\boldsymbol{L}_c \end{bmatrix} + \lambda_{\text{vcd}} \begin{bmatrix} -\boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{I} & -\boldsymbol{I} \end{bmatrix} \tag{14}$$

Hence we obtain

$$\begin{aligned}&O_{\text{ic}}^{\text{ub}} + \lambda_{\text{vcd}} R_{\text{vcd}} \\ &= \|\boldsymbol{B} - \boldsymbol{Z}\boldsymbol{W}\|_F^2 - \text{tr}(\boldsymbol{W}^\top \boldsymbol{Z}^\top \tilde{\boldsymbol{S}} \boldsymbol{Z} \boldsymbol{W}) \\ &= \|\boldsymbol{B}\|_F^2 + \|\boldsymbol{Z}\boldsymbol{W}\|_F^2 - 2\text{tr}(\boldsymbol{B}\boldsymbol{W}^\top \boldsymbol{Z}^\top) - \text{tr}(\boldsymbol{W}^\top \boldsymbol{Z}^\top \tilde{\boldsymbol{S}} \boldsymbol{Z} \boldsymbol{W}) \\ &= nc + \text{tr}(\boldsymbol{Z}\boldsymbol{W}\boldsymbol{W}^\top \boldsymbol{Z}^\top) - 2\text{tr}(\boldsymbol{B}\boldsymbol{W}^\top \boldsymbol{Z}^\top) - \text{tr}(\boldsymbol{W}^\top \boldsymbol{Z}^\top \tilde{\boldsymbol{S}} \boldsymbol{Z} \boldsymbol{W}) \\ &\text{s.t.} \quad \boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}, \quad \text{with} \quad \boldsymbol{W} = \Big[\boldsymbol{W}_p^\top; \boldsymbol{W}_g^\top\Big]^\top \in \mathbb{R}^{2d\times c}\end{aligned} \tag{15}$$

where $\boldsymbol{I}$ is identity matrix. We impose the orthogonality constraint on $\boldsymbol{W}$ for facilitating optimisation. We call this above formulation *Cross-view Identity Correlation Hashing*.

(II) Cross-view Identity Verification Regularisation. We leverage the available person class labels by *Cross-view Identity Verification* regularisation for further benefiting re-id discriminative hashing. Specifically, we introduce a linear transformation $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n_{\text{id}}}] \in \mathbb{R}^{c\times n_{\text{id}}}$ to model the relation between binary hash codes and identity class labels.

This connection is quantified by a loss function $l_{\text{ie}}(\boldsymbol{y}_i, \boldsymbol{U}^\top \boldsymbol{b}_i)$ under cross-view identification, e.g. hinge loss[1]:

$$l_{\text{ie}}(\boldsymbol{y}_i, \boldsymbol{U}^\top \boldsymbol{b}_i) = \|\boldsymbol{U}\|_F^2 + \eta_{\text{hinge}} \sum_{i=1}^{n} \varepsilon_i \tag{16}$$
$$\text{s.t.} \quad \forall i, j \quad \boldsymbol{u}_{k_i}^\top \boldsymbol{b}_i - \boldsymbol{u}_j^\top \boldsymbol{b}_i + \boldsymbol{y}_{i,j} \geq 1 - \varepsilon_i, \quad \varepsilon_i \geq 0$$

where k_i denotes the person identity class of sample $\boldsymbol{x}_i$, ε_i refers to the non-negative slack variable, η_{hinge} is a balance parameter, $\boldsymbol{y}_i$ is identity one-hot label vector of image $\boldsymbol{x}_i$ with the element $\boldsymbol{y}_{i,k_i} = 1$ and all others 0. In essence, identity verification regularisation by hinge loss enforces a one-vs-all optimisation *constraint* through posing discriminative margins between different person classes in the coding space. This is because $\boldsymbol{U}$ is no longer useful in deployment. This design not only increases the identity discrimination of learned hash functions, but also helps open-world re-id matching due to the inherent person class verification regularisation given by the loss function.

By incorporating the identity verification loss (Eqn. (16)), we extend our model objective (Eqn. (15)) as follow:

$$O_{\text{ice}} = O_{\text{ic}} + \lambda_{\text{vcd}} R_{\text{vcd}} + \underbrace{\lambda_{\text{ie}} \sum_{i=1}^{n} l_{\text{ie}}(\boldsymbol{y}_i, \boldsymbol{U}^\top \boldsymbol{b}_i)}_{\text{Identity Verification}} \tag{17}$$
$$= O_{\text{ic}} + \lambda_{\text{ie}} O_{\text{ie}} + \lambda_{\text{vcd}} R_{\text{vcd}},$$
$$\text{s.t.} \quad \boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}, \tag{18}$$

where

$$O_{\text{ie}} = \sum_{i=1}^{n} l_{\text{ie}}(\boldsymbol{y}_i, \boldsymbol{U}^\top \boldsymbol{b}_i) = l_{\text{ie}}(\boldsymbol{Y}, \boldsymbol{U}^\top \boldsymbol{B}^\top) \tag{19}$$

where $\boldsymbol{Y} \in \{0,1\}^{n_{\text{id}} \times (n_p + n_g)}$ represents all training identity labels $\boldsymbol{y}_i$ (n_{id} the training identity size, n_p/n_g the probe/gallery image size), $\boldsymbol{B} = [\boldsymbol{B}_p; \boldsymbol{B}_g]$ is a row-wise aggregation of the probe ($\boldsymbol{B}_p$) and gallery ($\boldsymbol{B}_g$) hash code matrices (Eqn. (3)); λ_{ie} is the weight of O_{ie}. We call our model "*Cross-view Identity Correlation and vErification*" (X-ICE) hashing.

D. Model Optimisation

To learn the proposed X-ICE model, we develop an alternating optimisation algorithm to infer model parameters, i.e. $\boldsymbol{W}$, $\boldsymbol{B}$, and $\boldsymbol{U}$. Specifically, we start by randomly initialising $\boldsymbol{W}^{(0)}$ and computing $\boldsymbol{B}^{(0)}$ with Eqn. (3). We then perform iteratively the following three routines until the model converges or the pre-defined maximal iteration number n_{it} reaches. Algorithm 1 summarises the optimisation of X-ICE.

(i) Fix $\boldsymbol{W}^{(t)}$ and $\boldsymbol{B}^{(t)}$ to optimise $\boldsymbol{U}^{(t)}$. $O_{\text{ic}}^{\text{ub}}$ and R_{vcd} are constant and we only need to optimise O_{ie}. The loss function l_{ie} in Eqn. (16) is a standard multi-class SVM formulation [69,70]. It can be solved with any off-the-shelf solvers [71].

(ii) Fix $\boldsymbol{B}^{(t)}$ and $\boldsymbol{U}^{(t)}$ to optimise $\boldsymbol{W}^{(t+1)}$. O_{ie} is fixed and we need to optimise $O_{\text{ic}}^{\text{ub}} + \lambda_{\text{vcd}} R_{\text{vcd}}$. Through introducing a Lagrangian multiplier $\boldsymbol{\Lambda}$ with respect to the constraint $\boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}$, we can rewrite Eqn. (15) as:

$$L = O_{\text{ic}}^{\text{ub}}(\boldsymbol{W}) + \lambda_{\text{vcd}} R_{\text{vcd}} - \frac{1}{2}\text{tr}\Big(\boldsymbol{\Lambda}(\boldsymbol{W}^\top \boldsymbol{W} - \boldsymbol{I})\Big). \tag{20}$$

As $\boldsymbol{W}^\top \boldsymbol{W}$ is symmetric, so is this Lagrangian multiplier $\boldsymbol{\Lambda}$. By setting the gradient of Eqn. (20) w.r.t $\boldsymbol{W}$ to zero, we have

$$\frac{\partial L(\boldsymbol{W}, \boldsymbol{\Lambda})}{\partial \boldsymbol{W}} = \frac{\partial O_{\text{ic}}^{\text{ub}}(\boldsymbol{W}) + \lambda_{\text{vcd}} R_{\text{vcd}}}{\partial \boldsymbol{W}} - \boldsymbol{W}\boldsymbol{\Lambda} = \boldsymbol{0}. \tag{21}$$

For expression simplicity, we define

$$\boldsymbol{G} = \frac{\partial O_{\text{ic}}^{\text{ub}}(\boldsymbol{W}) + \lambda_{\text{vcd}} R_{\text{vcd}}}{\partial \boldsymbol{W}} = 2(\boldsymbol{Z}^\top \boldsymbol{Z}\boldsymbol{W} - \boldsymbol{Z}^\top \tilde{\boldsymbol{S}} \boldsymbol{Z}\boldsymbol{W} - \boldsymbol{Z}^\top \boldsymbol{B}). \tag{22}$$

After multiplying both sides of Eqn. (21) by $\boldsymbol{W}^\top$, applying $\boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}$ and the symmetric property of $\boldsymbol{\Lambda}$, we have

$$\boldsymbol{\Lambda} = \boldsymbol{W}^\top \boldsymbol{G} = \boldsymbol{G}^\top \boldsymbol{W}. \tag{23}$$

From Eqns. (21) (22) (23), we obtain

$$\begin{aligned} \frac{\partial L(\boldsymbol{W}, \boldsymbol{\Lambda})}{\partial \boldsymbol{W}} &= \boldsymbol{G} - \boldsymbol{W}\boldsymbol{G}^\top \boldsymbol{W} \\ &= \boldsymbol{G}\boldsymbol{W}^\top \boldsymbol{W} - \boldsymbol{W}\boldsymbol{G}^\top \boldsymbol{W} \\ &= (\boldsymbol{G}\boldsymbol{W}^\top - \boldsymbol{W}\boldsymbol{G}^\top)\boldsymbol{W}. \end{aligned} \tag{24}$$

By further introducing a skew-symmetric matrix [72]:

$$\boldsymbol{A} = \boldsymbol{G}\boldsymbol{W}^\top - \boldsymbol{G}^\top \boldsymbol{W}, \tag{25}$$

we can subsequently update iteratively $\boldsymbol{W}$ by the Crank-Nicolson-like scheme [73]:

$$\boldsymbol{W}_{(\nu+1)}^t = \boldsymbol{W}_{(\nu)}^t - \frac{\delta}{2}\boldsymbol{A}(\boldsymbol{W}_{(\nu)}^t + \boldsymbol{W}_{(\nu+1)}^t), \tag{26}$$

where δ is the step size. By solving Eqn. (26), we obtain

$$\boldsymbol{W}_{(\nu+1)}^t = \boldsymbol{Q}\boldsymbol{W}_{(\nu)}^t, \tag{27}$$

with

$$\boldsymbol{Q} = (\boldsymbol{I} + \frac{\delta}{2}\boldsymbol{A})^{-1}(\boldsymbol{I} - \frac{\delta}{2}\boldsymbol{A}).$$

Hereafter, we iteratively update $\boldsymbol{W}_{(\nu)}^t$ with Eqn. (27) using the Barzilai-Borwein method [72]. In particular, we start from $\boldsymbol{W}_{(0)}^t = \boldsymbol{W}^{(t)}$ and stop optimising $\boldsymbol{W}_{(\nu)}^t$ until it converges or the maximum iteration number n_{wit} reaches. We set $\boldsymbol{W}^{(t+1)}$ with the final $\boldsymbol{W}_{(\nu)}^t$. Note that solving $\boldsymbol{W}$ alone is by a separate inner iterative procedure, different from the outer iteration among $\boldsymbol{U}$, $\boldsymbol{B}$ and $\boldsymbol{W}$.

(iii) Fix $\boldsymbol{W}^{(t+1)}$ and $\boldsymbol{U}^{(t)}$ to optimise $\boldsymbol{B}^{(t+1)}$. In this case, R_{vcd} is constant, and we need to optimise

$$\min_{\boldsymbol{B}} \quad \lambda_{\text{ie}} l_{\text{ie}}(\boldsymbol{Y}, \boldsymbol{U}^\top \boldsymbol{B}^\top) - 2\text{tr}(\boldsymbol{B}\boldsymbol{W}^\top \boldsymbol{Z}^\top), \tag{28}$$

which is a mixed-integer NP-hard optimisation problem. Note that, in model optimisation, $\boldsymbol{B}$ depends on not only $\boldsymbol{W}$ but also $\boldsymbol{U}$ for imposing identity class discrimination constraint. A typical practice is by continuous relaxation: first obtaining a continuous solution to $\boldsymbol{B}$ which is subsequently thresholded to generate the binary codes [27,28,29]. However, such approximation may be sub-optimal. In this study, we seek for

[1]The loss function for cross-view identity verification regularisation can be alternative forms, and we further discuss the regression (reg) loss later.

Algorithm 1: Learning the proposed X-ICE model

Input: Training data: $\boldsymbol{X}_p$, $\boldsymbol{X}_g$; identity label matrix: $\boldsymbol{S}$; hash code length: c; iteration numbers: n_{it}, n_{wit}; hyper-parameters: α, λ_{vcd}, λ_{ie};

Output: Cross-view hashing function parameter: $\boldsymbol{W}$;

(I) Parameter initialisation
Randomly initialise $\boldsymbol{W}^{(0)}$;
Compute $\boldsymbol{B}^{(0)}$ with Eqn. (3);
(II) Model optimisation
for $t = 0 : n_{it} - 1$ **do**
  (i) Optimise $\boldsymbol{U}^{(t)}$ when fixing $\boldsymbol{W}^{(t)}$ and $\boldsymbol{B}^{(t)}$;
    – For the hinge loss, by Eqn. (16);
    – For the regression loss, by Eqn. (36);
  (ii) Optimise $\boldsymbol{W}^{(t+1)}$ when fixing $\boldsymbol{B}^{(t)}$ and $\boldsymbol{U}^{(t)}$;
    – For either hinge or regression loss, by Eqn. (27);
  (iii) Optimise $\boldsymbol{B}^{(t+1)}$ when fixing $\boldsymbol{U}^{(t)}$ and $\boldsymbol{W}^{(t+1)}$;
    – For the hinge loss, by Eqn. (34);
    – For the regression loss, by Eqn. (41);
end
Return $\boldsymbol{W} = \boldsymbol{W}^{(n_{\text{it}})}$.

the *exact* optimal solution in spirit of discrete hashing [31]. In particular, we perform sample-wise optimisation as

$$\min_{\boldsymbol{B}} \quad \|\boldsymbol{b}_i^s - \boldsymbol{x}_i^s \boldsymbol{W}_s\|_F^2, \quad s \in \{p, g\} \tag{29}$$

$$\text{s.t.} \quad \forall j \quad \boldsymbol{u}_{k_i}^\top \boldsymbol{b}_i^s - \boldsymbol{u}_j^\top \boldsymbol{b}_i^s + \boldsymbol{y}_{i,j} \geq 1 - \varepsilon_i, \tag{30}$$

where k_i is the identity class of image $\boldsymbol{x}_i^s$. By transforming the constraints Eqn. (30) as

$$\forall j \quad C_j = (\boldsymbol{u}_{k_i} - \boldsymbol{u}_j)^\top \boldsymbol{b}_i^s + (\boldsymbol{y}_{i,j} - 1 + \varepsilon_i) \geq 0, \tag{31}$$

and incorporating them with Eqn. (29), we have

$$\min_{\boldsymbol{B}} \quad \|\boldsymbol{b}_i^s - \boldsymbol{x}_i^s \boldsymbol{W}_s\|_F^2 - \lambda_b \sum_{j=1}^{n_{\text{id}}} C_j \tag{32}$$

$$\equiv \max_{\boldsymbol{B}} \quad \boldsymbol{b}_i^{s\top} \big(\boldsymbol{x}_i^s \boldsymbol{W}_s + \frac{\lambda_b}{2} \sum_{j=1}^{n_{\text{id}}} (\boldsymbol{u}_{k_i} - \boldsymbol{u}_j)\big), \tag{33}$$

where λ_b is a balancing parameter (we set $\lambda_b = 1$ in our experiments). In other words, Eqn. (33) is a transformed optimisation formulation of Eqn. (29) subject to the constraints Eqn. (30). The optimal solution of Eqn. (33) is

$$\boldsymbol{b}_i^s = \text{sign}\big(\boldsymbol{x}_i^s \boldsymbol{W}_s + \frac{\lambda_b}{2} \sum_{j=1}^{n_{\text{id}}} (\boldsymbol{u}_{k_i} - \boldsymbol{u}_j)\big). \tag{34}$$

In this way, we can obtain the optimal binary codes for all training samples.

E. Alternative Identity Verification Loss Function

Apart from the hinge loss for l_{ie} (Eqn. (16)), other function forms can be flexibly adopted in our X-ICE model. We additionally consider the Euclidean regression (reg) loss. As such, we also need to modify the optimisation as Section III-D. Specifically, in step **(i)**, instead of Eqn. (16) we solve

$$\min_{\boldsymbol{U}} \quad \|\boldsymbol{Y} - \boldsymbol{U}^\top \boldsymbol{B}^\top\|_F^2, \tag{35}$$

which has a closed-formed solution:

$$\boldsymbol{U} = \boldsymbol{B}^{-1} \boldsymbol{Y}^\top. \tag{36}$$

In step **(iii)**, rather than Eqn. (29) we need to minimise

$$\begin{aligned} &\lambda_{\text{ie}} \|\boldsymbol{Y} - \boldsymbol{U}^\top \boldsymbol{B}^\top\|_F^2 - 2\text{tr}(\boldsymbol{B}\boldsymbol{W}^\top \boldsymbol{Z}^\top) \\ =&\lambda_{\text{ie}} \Big(\|\boldsymbol{Y}\|_F^2 - 2\text{tr}(\boldsymbol{B}\boldsymbol{U}\boldsymbol{Y}) + \|\boldsymbol{B}\boldsymbol{U}\|_F^2\Big) - 2\text{tr}(\boldsymbol{B}\boldsymbol{W}^\top \boldsymbol{Z}^\top) \\ =&\underbrace{\lambda_{\text{ie}} \|\boldsymbol{Y}\|_F^2}_{\text{const}} + \lambda_{\text{ie}} \|\boldsymbol{B}\boldsymbol{U}\|_F^2 - 2\text{tr}\Big(\boldsymbol{B}(\lambda_{\text{ie}} \boldsymbol{U}\boldsymbol{Y} + \boldsymbol{W}^\top \boldsymbol{Z}^\top)\Big), \end{aligned} \tag{37}$$

where $\|\boldsymbol{Y}\|_F^2 = n_p + n_g$ is constant so ignorable. For notation brevity, we denote $\boldsymbol{R} = \lambda_{\text{ie}} \boldsymbol{U}\boldsymbol{Y} + \boldsymbol{W}^\top \boldsymbol{Z}^\top$. Rather than learning the discrete $\boldsymbol{B}$ in one time, we alternatively optimise it in a bitwise manner. Formally, at one time we optimise only the i-th column $\ddot{\boldsymbol{b}}_i \in \{-1, 1\}^{(n_p+n_g)\times 1}$ of $\boldsymbol{B}$ (i.e. the i-th bit of all training images) whilst all other columns are fixed. We denote $\boldsymbol{B}_{-i} = \boldsymbol{B} \setminus \ddot{\boldsymbol{b}}_i$. Similarly, we define $\ddot{\boldsymbol{u}}_i$ as the i-th row of $\boldsymbol{U}$ and $\boldsymbol{U}_{-i} = \boldsymbol{U} \setminus \ddot{\boldsymbol{u}}_i$. As such, we have

$$\begin{aligned} \|\boldsymbol{B}\boldsymbol{U}\|_F^2 &= \text{tr}(\boldsymbol{U}^\top \boldsymbol{B}^\top \boldsymbol{B}\boldsymbol{U}) \\ &= \underbrace{\text{tr}(\ddot{\boldsymbol{u}}_i \ddot{\boldsymbol{b}}_i^\top \ddot{\boldsymbol{b}}_i \ddot{\boldsymbol{u}}_i^\top)}_{\text{const}} + 2\ddot{\boldsymbol{u}}_i^\top \boldsymbol{U}_{-i}^\top \boldsymbol{B}_{-i}^\top \ddot{\boldsymbol{b}}_i + \text{const}, \end{aligned} \tag{38}$$

where $\text{tr}(\ddot{\boldsymbol{u}}_i \ddot{\boldsymbol{b}}_i^\top \ddot{\boldsymbol{b}}_i \ddot{\boldsymbol{u}}_i^\top) = (n_p + n_g)\ddot{\boldsymbol{u}}_i^\top \ddot{\boldsymbol{u}}_i$ is constant. Then, we define $\ddot{\boldsymbol{r}}_i$ as the i-th row of $\boldsymbol{R}$ and $\boldsymbol{R}_{-i} = \boldsymbol{R} \setminus \ddot{\boldsymbol{r}}_i$. We analogously obtain

$$\text{tr}(\boldsymbol{B}\boldsymbol{R}) = \ddot{\boldsymbol{r}}_i^\top \ddot{\boldsymbol{b}}_i + \text{const}. \tag{39}$$

After integrating Eqns. (38) and (39) into Eqn. (37), our optimisation problem becomes minimising

$$(\lambda_{\text{ie}} \ddot{\boldsymbol{u}}_i^\top \boldsymbol{U}_{-i}^\top \boldsymbol{B}_{-i}^\top - \ddot{\boldsymbol{r}}_i^\top) \ddot{\boldsymbol{b}}_i, \tag{40}$$

which can be directly solved as

$$\ddot{\boldsymbol{b}}_i = \ddot{\boldsymbol{r}}_i - \lambda_{\text{ie}} \boldsymbol{B}_{-i} \boldsymbol{U}_{-i} \ddot{\boldsymbol{r}}_i. \tag{41}$$

As such, we compute all $\ddot{\boldsymbol{b}}_i$ of $\boldsymbol{B}$ iteratively and stop until there is no change in each binary bit. Typically, we perform only ≤ 5 times of optimisation for each bit in our experiments.

IV. Experiments

A. Datasets and Evaluation Settings

Datasets: Three large-scale contemporary person re-id datasets were utilised in our evaluations: **(I)** The *CUHK03* dataset [39] has $13,164$ images from $1,360$ people captured by 6 surveillance cameras in a university. Each person identity is observed by two different cameras, with an average of 4.8 images per view (Figure 2(a)). **(II)** The *SYSU* dataset [36] contains totally $24,446$ images from 502 people captured by 2 cameras on a university campus (Figure 2(b)). **(III)** The *Market-1501* dataset [40] was collected from 6 surveillance cameras near a university supermarket, including $32,668$ bounding boxes of $1,501$ identities (Figure 2(c)). All three re-id datasets are challenging due to the significant unknown covariates across different camera views, random inter-object occlusion and distracting background clutters.

Baseline methods: We extensively considered a wide range of state-of-the-art hashing methods, including **(I)** unsupervised

(a) CUHK03 [39] (b) SYSU [36] (c) Market-1501 [40]

Fig. 2: Example person images, with two images in each column corresponding to the same person for every dataset.

Locality Sensitive Hashing (LSH) [25], *Spectral Hashing* (SH) [27], *Scalable Graph Hashing* (SGH) [74], *Iterative Quantisation* (ITQ) [75]; **(II)** supervised *Kernel-based Supervised Hashing* (KSH) [30], Canonical Correlation Analysis [76] + Iterative quantisation [75] (CCA+ITQ) where CCA is utilised for supervised projection learning, *Fast Hashing* (FH) [77], *Supervised Discrete Hashing* (SDH) [31], *Column Sampling based Discrete Supervised Hashing* (COSDISH) [78]; **(III)** multi-modal *Semantic Preserving Hashing* (SePH) [79], *Semantic Correlation Maximization* (SCM) [60], *Cross-View Hashing* (CVH) [56], *Collective Matrix Factorization Hashing* (CMFH) [80], *Cross-Modality Similarity Sensitive Hashing* (CMSSH) [57]. For evaluating how competitive the hashing based re-id models are against conventional person re-id approach under the LSOW re-id scenario, we further evaluated **(IV)** five state-of-the-art re-id methods, KISSME [34], CVDCA [36], XQDA [6], MLAPG [11], and DNS [12].

Evaluation protocol: For simulating the practical large scale open-world person re-id scenario, we created specifically the following data partitions. We first split the whole person identity population randomly into two disjoint parts: one for training ($360/202/501$) and one for test ($1000/300/1000$) on CUHK03, SYSU and Market-1501, respectively. In the test data, we selected randomly 10 target people for re-identification. As a result, there are $990/290/990$ probe imposters for CUHK03, SYSU and Market-1501. To achieve statistically reliable evaluations, we repeated 10 folds of training/testing data splits, on each of which we further performed 10 times of target people random selection. We utilised the averaged results over all $100 (= 10 \times 10)$ trials for performance comparison among all methods.

We set the probe and gallery camera view(s) as followings. For CUHK03, person image data are provided in the form of camera pair. We therefore used one camera of each pair as gallery view and the other as probe view for ensuring the cross-view matching property. For SYSU which has two cameras, we similarly used one camera as gallery view and the other as probe view. For Market-1501 which provides the camera label for each person image, we utilised a 2(gallery)/4(probe) camera split. The purpose is to simulate the open-world person re-id scenario as well as possible: (1) Large search space, i.e. more probe views to be searched against with a large number of persons; and (2) Multiple gallery camera views.

We considered two performance evaluation criteria [20]: (I) *Set Verification* (SV): Verifying whether a given probe person belongs to any gallery target person; and (II) *Individual Verification* (IV): verifying whether a given probe image is of one specific target person. IV is a special case of SV when there is only *one* target person in the gallery set.

Evaluation metrics: Given the existence of imposters (i.e. non-target people) in the probe population, we need to measure how well the probe images of target people are correctly verified and how well the probe images of imposters are successfully filtered out. We used two metrics [20]: (1) True Target Rate (TTR) and (2) False Target Rate (FTR):

$$\text{TTR} = \frac{N_{t2t}}{N_t}, \quad \text{FTR} = \frac{N_{nt2t}}{N_{nt}} \tag{42}$$

where N_t and N_{nt} denote the numbers of probe images from target and non-target people; N_{t2t} is the number of correctly verified probe images of target people; N_{nt2t} denotes the number of probe images of non-targets but verified as target people. TTR and FTR can be applied for both set and individual verification criteria due to their intrinsic connection.

Specifically, for *set verification* we compute TTR and FTR as below: (1) We first compute the matching distance $\{d(\tilde{\boldsymbol{x}}^p, \tilde{\boldsymbol{x}}_i^g)\}$ between a given probe $\tilde{\boldsymbol{x}}^p$ and all gallery $\{\tilde{\boldsymbol{x}}_i^g\}$; We denote $i^* = \arg\min_i \{d(\tilde{\boldsymbol{x}}^p, \tilde{\boldsymbol{x}}_i^g)\}$ and consider the i^*-th gallery image as the most matched person. (2) Given a threshold θ_m, we verify $\tilde{\boldsymbol{x}}^p$ as the target person if $d(\tilde{\boldsymbol{x}}^p, \tilde{\boldsymbol{x}}_{i^*}^g) < \theta_m$, otherwise as an imposter. (3) We count a correct target match only when $d(\tilde{\boldsymbol{x}}^p, \tilde{\boldsymbol{x}}_{i^*}^g) < \theta_m$, also $\tilde{\boldsymbol{x}}_{i^*}^g$ and $\tilde{\boldsymbol{x}}^p$ are from the same target person. In contrast, we mark a false target match when $d(\tilde{\boldsymbol{x}}^p, \tilde{\boldsymbol{x}}_{i^*}^g) < \theta_m$ but $\tilde{\boldsymbol{x}}^p$ is actually from an imposter. (4) We repeat these steps for every probe image. (5) We compute TTR and FTR scores for all probes by Eqn. (42). For *individual verification*, we compute TTR and FTR for each target person with the same steps as set verification. In this case, the gallery set contains the images of this target person alone. We average over all target people to obtain the final TTR and FTR scores w.r.t. a given θ_m. We can obtain different (TTR, FTR) pairs and form a Receiver Operating Characteristic (ROC) curve by varying θ_m. When evaluating different methods, we compare their TTR measures against a series of FTR so that model performance can be measured under different verification standards.

Additionally, we utilised mean Average Precision (mAP) to evaluate the holistic ranking performance. First, we compute

TABLE I: Evaluating model components. (Metric: TTR (%) at varying FTRs (%). IV: Individual Verification, SV: Set Verification. Both SV and IV utilise the same TTR/FTR metric.)

Loss	Metric	Dataset	CUHK03 [39]					SYSU [36]					Market-1501 [40]				
		FTR	1%	5%	10%	20%	30%	1%	5%	10%	20%	30%	1%	5%	10%	20%	30%
hinge	IV	X-IC	42.45	72.59	84.52	93.29	96.61	54.79	80.04	88.75	94.93	97.59	55.52	81.10	89.07	95.42	97.83
		X-ICE\VCD	42.78	73.19	84.23	93.19	96.76	54.50	81.02	89.75	95.65	97.89	57.85	84.44	91.93	96.31	97.99
		ICE	47.09	74.18	85.73	93.68	96.83	60.48	83.36	90.66	95.97	97.92	63.79	87.10	93.18	97.14	98.67
		X-ICE	**49.67**	**79.60**	**89.50**	**96.09**	**98.48**	**61.86**	**84.10**	**91.47**	**96.35**	**98.26**	**66.52**	**88.03**	**93.66**	**97.15**	**98.55**
	SV	X-IC	12.02	30.17	43.58	60.59	72.32	17.25	38.18	52.45	68.59	78.18	21.73	43.62	56.69	71.37	79.88
		X-ICE\VCD	13.31	31.98	45.45	62.27	72.85	19.36	41.86	55.32	70.91	80.08	18.53	43.11	58.41	74.09	82.62
		ICE	15.40	35.34	48.50	64.48	73.75	**24.43**	**47.42**	60.09	73.36	81.36	20.69	46.80	62.91	77.69	85.03
		X-ICE	**16.41**	**37.50**	**50.14**	**66.56**	**77.30**	23.32	46.84	**60.48**	**74.20**	**82.37**	**26.81**	**52.73**	**66.47**	**79.66**	**86.16**
regression	IV	X-IC	42.45	72.59	84.52	93.29	96.61	54.79	80.04	88.75	94.93	97.59	55.52	81.10	89.07	95.42	97.83
		X-ICE\VCD	41.74	70.61	83.87	92.99	96.78	55.62	81.45	89.73	95.68	97.85	56.75	83.87	91.6	96.61	98.3
		ICE	44.06	73.36	85.29	94.21	96.91	59.38	81.74	89.40	95.33	97.42	61.16	85.22	91.93	96.73	98.39
		X-ICE	**49.96**	**78.18**	**88.96**	**95.88**	**97.98**	**63.13**	**84.86**	**91.52**	**96.17**	**98.08**	**64.18**	**86.98**	**92.91**	**97.09**	**98.59**
	SV	X-IC	12.02	30.17	43.58	60.59	72.32	17.25	38.18	52.45	68.59	78.18	21.73	43.62	56.69	71.37	79.88
		X-ICE\VCD	12.54	32.36	45.47	61.15	71.09	20.67	43.26	56.63	71.36	80.49	17.24	41.84	56.82	73.62	82.58
		ICE	14.77	33.83	46.77	63.17	73.14	23.85	46.65	59.26	72.30	80.40	19.64	43.63	59.33	75.80	83.38
		X-ICE	**16.37**	**37.36**	**49.71**	**65.49**	**76.03**	**25.94**	**49.94**	**62.59**	**75.91**	**83.30**	**22.27**	**48.12**	**62.87**	**77.13**	**84.55**

TABLE II: Evaluating different loss functions for cross-view identity verification. (Metric: TTR (%) at varying FTRs (%)).

Criterion	Dataset	CUHK03 [39]					SYSU [36]					Market-1501 [40]				
	FTR	1%	5%	10%	20%	30%	1%	5%	10%	20%	30%	1%	5%	10%	20%	30%
IV	hinge	49.67	**79.60**	**89.50**	**96.09**	**98.48**	61.86	84.10	91.47	**96.35**	**98.26**	**66.52**	**88.03**	**93.66**	**97.15**	98.55
	regression	**49.96**	78.18	88.96	95.88	97.98	**63.13**	**84.86**	**91.52**	96.17	98.08	64.18	86.98	92.91	97.09	**98.59**
SV	hinge	**16.41**	**37.50**	**50.14**	**66.56**	**77.30**	23.32	46.84	60.48	74.20	82.37	**26.81**	**52.73**	**66.47**	**79.66**	**86.16**
	regression	16.37	37.36	49.71	65.49	76.03	**25.94**	**49.94**	**62.59**	**75.91**	**83.30**	22.27	48.12	62.87	77.13	84.55

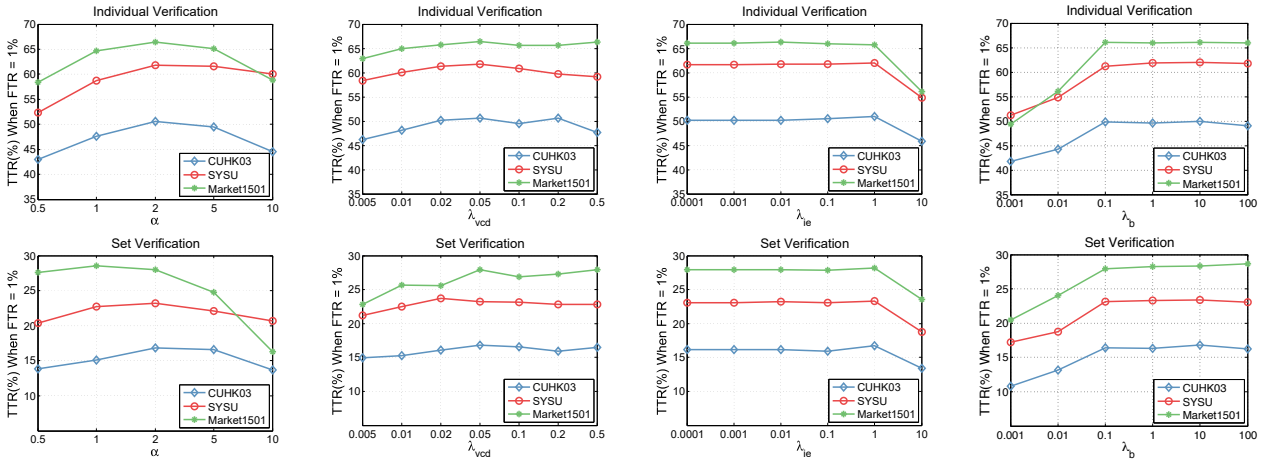


Fig. 3: Evaluating parameter sensitivity.

the Average Precision (AP) for each probe, i.e. the area under the Precision-Recall curve; mAP is calculated as the mean of APs over all probes. Hence, mAP provides a comprehensive metric by considering the quality of all rank lists.

Visual features: We adopted a state-of-the-art re-id feature LOMO [6] for person image representation. To remove redundancy and noise, we performed principal component analysis on the raw features and used the top-1000 dominant components as the final features. We also evaluated the deep feature and discussed the effect of different representations.

Implementation details: For fair comparison, we used the same evaluation protocol for all compared models. We utilised the codes released by the original authors if available with their recommended parameter settings. The default parameter settings in our evaluations are: $c = 256$ (Eqn. (3)); $\alpha = 2$ (Eqn. (7)); $\lambda_{\text{vcd}} = 0.05$ (Eqn. (10)); $\lambda_{\text{ie}} = 0.01$ (Eqn. (17)); $\eta_{\text{hinge}} = 0.01$ (Eqn (16)); $n_{\text{it}} = 20$, $n_{\text{wit}} = 10$ (Eqn. (27)), and $\lambda_b = 1$ (Eqn. (33)).

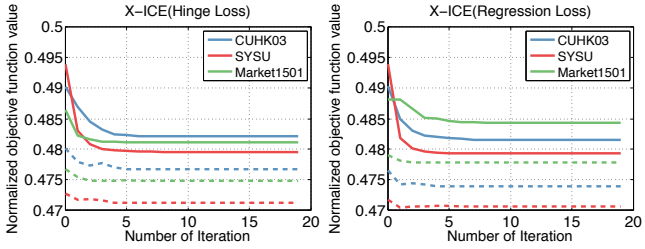


Fig. 4: Model convergence analysis by tracking the hashing objective function Eqn. (7) (dashed curves) and its upper bound Eqn. (13) (solid curves) over training iterations.

B. Evaluating Our Proposed Method

We evaluated the X-ICE method in the following aspects: (1) Effect of model components; (2) Effect of different loss functions in cross-view identity verification regularisation;

TABLE III: Comparing state-of-the-art hashing methods. (Metrics: TTR (%) at varying FTRs (%), and mAP (%).)

Dataset	CUHK03 [39]						SYSU [36]						Market-1501 [40]					
Metric	Individual Verification					mAP	Individual Verification					mAP	Individual Verification					mAP
	1%	5%	10%	20%	30%	(%)	1%	5%	10%	20%	30%	(%)	1%	5%	10%	20%	30%	(%)
LSH [25]	15.03	34.87	48.12	64.66	75.20	1.91	21.21	43.73	57.42	72.17	81.17	5.48	37.00	61.56	73.43	85.12	90.92	8.28
SH [27]	11.97	27.8	39.99	54.73	65.35	1.49	17.60	34.21	45.63	60.14	70.27	4.18	38.54	58.69	69.47	80.71	86.75	9.29
SGH [74]	16.95	37.37	50.71	66.45	76.76	2.36	27.18	49.21	61.74	75.34	82.92	8.03	37.75	63.16	75.05	86.13	91.76	8.69
ITQ [75]	17.31	39.29	53.06	69.12	80.24	2.70	26.51	49.55	63.18	77.04	84.87	7.47	40.72	67.17	78.67	88.51	93.41	10.84
CCA+ITQ [75]	28.11	51.15	65.05	78.95	86.37	4.28	50.50	73.75	83.16	91.08	94.67	18.12	57.75	80.30	87.53	93.47	96.17	15.02
KSH [30]	32.29	57.54	69.78	81.73	88.96	5.49	53.23	77.28	85.88	92.62	95.62	22.29	59.03	81.83	89.01	94.26	96.41	17.34
FH [77]	20.01	40.07	52.32	67.35	77.39	1.03	29.48	50.56	62.42	75.58	83.65	8.07	28.24	48.88	60.59	73.92	81.61	5.07
SDH [31]	38.80	66.82	78.83	88.15	93.03	7.31	46.09	72.34	82.76	90.75	94.51	17.99	58.03	81.26	88.22	94.16	96.29	15.57
COSDISH [78]	13.19	29.18	40.33	56.88	68.23	1.55	38.04	61.43	72.73	83.95	89.38	11.51	39.29	62.26	73.49	83.68	89.04	8.44
CMSSH [57]	10.46	32.46	49.80	68.67	80.45	1.25	11.06	33.76	50.51	70.55	82.29	3.18	8.88	29.25	46.49	67.02	79.56	1.55
CVH [56]	2.83	10.09	17.81	31.05	42.51	0.39	5.76	19.67	31.77	49.33	62.21	1.30	3.51	13.38	22.62	37.31	50.55	0.53
CMFH [80]	11.85	31.23	46.40	64.63	75.56	1.27	25.73	54.24	68.73	82.09	89.14	6.32	24.96	52.96	67.52	81.62	89.11	4.07
SCM [60]	5.43	17.84	28.77	44.72	58.70	0.59	14.83	32.93	45.22	60.35	70.95	3.92	13.41	31.49	43.44	58.75	69.09	2.04
SePH [79]	26.98	52.69	65.88	79.29	86.24	4.18	37.15	64.01	75.75	86.09	91.56	13.56	41.88	70.39	80.72	88.89	93.09	8.80
X-ICE(hinge)	49.67	**79.60**	**89.50**	**96.09**	**98.48**	**11.66**	61.86	84.10	91.47	**96.35**	**98.26**	**29.93**	**66.52**	**88.03**	**93.66**	**97.15**	98.55	**21.47**
X-ICE(reg)	**49.96**	78.18	88.96	95.88	97.98	11.23	**63.13**	**84.86**	**91.52**	96.17	98.08	29.44	64.18	86.98	92.91	97.09	**98.59**	20.68
Metric	Set Verification					mAP	Set Verification					mAP	Set Verification					mAP
LSH [25]	4.81	13.97	21.83	35.01	45.88	1.91	7.25	18.13	27.35	41.18	52.31	5.48	17.17	33.22	43.85	57.75	67.36	8.28
SH [27]	3.91	12.00	19.93	31.72	43.41	1.49	7.22	16.36	24.39	36.62	46.96	4.18	21.52	36.77	46.41	58.33	68.00	9.29
SGH [74]	5.42	14.34	22.87	36.84	48.89	2.36	10.20	22.06	31.91	46.08	56.88	8.03	16.59	34.04	45.33	59.12	69.60	8.69
ITQ [75]	5.45	14.9	23.96	37.95	49.51	2.70	8.81	21.47	31.42	45.82	56.73	7.47	17.39	35.66	47.25	60.90	70.46	10.84
CCA+ITQ [75]	10.52	23.86	34.05	49.95	59.58	4.28	21.08	42.10	53.63	67.05	75.76	18.12	18.30	45.05	60.16	73.75	81.38	15.02
KSH [30]	11.65	27.43	38.19	52.68	63.62	5.49	22.07	43.13	55.44	69.26	77.25	22.29	21.74	47.36	61.28	74.78	82.22	17.34
FH [77]	7.66	17.92	26.82	40.15	52.08	1.03	12.82	26.19	35.86	48.86	59.28	8.07	13.32	27.17	36.78	50.13	60.84	5.07
SDH [31]	13.59	31.37	43.59	60.08	70.34	7.31	15.46	36.48	49.00	63.30	72.79	17.99	20.95	47.01	61.17	75.19	82.16	15.57
COSDISH [78]	5.07	13.37	21.71	34.27	43.74	1.55	16.86	33.10	43.36	57.20	66.38	11.51	15.80	35.00	46.81	60.49	69.66	8.44
CMSSH [57]	2.32	10.32	19.42	32.44	45.06	1.25	2.66	11.13	19.29	33.87	46.11	3.18	2.20	8.93	17.42	31.57	44.29	1.55
CVH [56]	1.31	5.62	12.06	23.29	32.45	0.39	1.75	7.65	14.44	27.18	38.12	1.30	1.57	6.83	12.83	24.39	35.74	0.53
CMFH [80]	3.49	11.91	20.06	34.46	45.35	1.27	7.25	21.42	32.99	48.97	60.98	6.32	7.95	22.60	33.55	48.81	60.49	4.07
SCM [60]	1.91	7.64	14.74	27.22	37.74	0.59	5.75	15.55	23.82	37.04	48.23	3.92	5.29	15.23	23.20	36.53	47.93	2.04
SePH [79]	9.64	23.22	33.51	49.79	60.94	4.18	12.40	30.40	42.85	57.81	67.73	13.56	11.78	32.98	46.91	63.28	73.64	8.80
X-ICE(hinge)	**16.41**	**37.50**	**50.14**	**66.56**	**77.30**	**11.66**	23.32	46.84	60.48	74.20	82.37	**29.93**	**26.81**	**52.73**	**66.47**	**79.66**	**86.16**	**21.47**
X-ICE(reg)	16.37	37.36	49.71	65.49	76.03	11.23	**25.94**	**49.94**	**62.59**	**75.91**	**83.30**	29.44	22.27	48.12	62.87	77.13	84.55	20.68

(a) CUHK03 [39] (b) SYSU [36] (c) Market-1501 [40]

Fig. 5: Visualising person re-id performance by four top methods X-ICE (1st row), KSH (2nd row), CCA+ITQ (3rd row) and SDH (4th row). For each dataset, the left-most image is the probe person image, followed by top 10 most matched gallery images by respective methods with red boxes indicating true matches.

(3) Sensitivity of model parameters; (4) Analysis of model convergence.

Effect of model components. We introduce three stripped-down variants of our full X-ICE for component analysis: (1) X-IC: With Eqn. (10) as the model objective therefore lacking the cross-view identity verification regularisation component. This allows evaluating the efficacy of both identity correlation hashing and identity verification. (2) X-ICE\VCD: Removing the View Context Discrepancy (VCD) regularisation (Eqn. (9)) from X-ICE for evaluating view correlation modelling. (3) ICE: Learning a uni-view re-id model from the assembled training data of all cameras, so that camera view information is discarded. This allows evaluating cross-view modelling. Table I. shows that with only cross-view identity correlation hashing, our X-IC model is already able to effectively perform LSOW person re-id. By accommodating identity verification regularisation, re-id performance can be consistently boosted across all datasets. This is in alignment with the previous finding that discriminative learning on training person classes is generalisable for recognising unseen test classes [81,82]. This also suggests the great complementary benefits between cross-view identity correlation hashing and class discriminative verification regularisation in our formulation. Without the cross-view modelling, ICE performs clearly poorer than X-

ICE. This implies the importance of camera view domain information in person matching. Camera view correlation regularisation is also critical, as indicated by the performance drop with X-ICE\VCD.

Effect of identity verification loss function. We evaluated the influence of different loss functions (e.g. hinge and regression) in cross-view identity verification regularisation. Table II shows that the two loss functions produce similar re-id accuracies, with hinge loss slightly better than regression (reg) loss on CUHK03 and Market-1501 but worse on SYSU. This suggests the flexibility of X-ICE in choosing loss function.

Sensitivity of model parameters. We analysed the impact of model parameters α (Eqn. (7)), λ_{vcd} (Eqn. (10)), λ_{ie} (Eqn. (17)), and λ_b (Eqn. (33)). Figure 3 reveals four observations: (1) "α" is most sensitive among the four, with the best values lying in $[2, 5]$. In X-ICE, the essence of α is about Identity Correlation (IC) learning. When $\alpha = 2$, we obtained 75.61%/85.20%/73.98% IC gain in training on CUHK03/SYSU/Market-1501, respectively. This justifies the effectiveness of our model design and optimisation. (2) "λ_{vcd}" is less sensitive and the optimal value may depend on specific camera viewing conditions. For example, higher values should be used when viewing condition is similar among different cameras such as on Market-1501, whilst lower ones for opposite cases like on CUHK03 and SYSU. (3) "λ_{ie}" has a wide satisfactory range and small values are typically preferred. The plausible reason is that, fitting overwhelmingly the training person classes may render the final model less generalisable to unseen test person classes. (4) "λ_b" also has a wide satisfactory range and large values (>0.1) are required. This suggests the positive effects of identity verification regularisation and the necessity of sufficiently ensuring discriminative margins among different identity classes during model optimisation.

Model convergence analysis. We adopt the upper bound minimisation strategy for approximately optimising our hashing function parameter $\boldsymbol{W}$ (Section III-C). This upper bound $O_{\text{ic}}^{\text{ub}}$ (Eqn. (13)) is supposed to decline gradually in training time. To validate this, we tracked normalised $O_{\text{ic}}^{\text{ub}}$ and O_{ic} (Eqn. (7)) in parallel over optimisation iterations. As shown in Figure 4, we indeed observed the expected trend. Additionally, it is shown that the X-ICE can converge within a small number of iterations on all three datasets. This validates empirically our optimisation algorithm design and derivation.

C. Comparing State-of-the-Art Hashing Methods

We evaluated a wide range (14) of state-of-the-art hashing models for LSOW re-id. Table III shows that the X-ICE model surpasses all the hashing competitors on all three datasets by a large margin in both mAP and set/individual verification rates. This demonstrates the efficacy and advantages of X-ICE over existing hashing methods for LSOW re-id matching. This superiority is due to a collective effect of identity correlation hashing, inter-camera contextual regularisation, and person class discrimination in a jointly optimised cross-view model (Table I). Conceptually, the proposed X-ICE model joins the merits of both supervised hashing (i.e. cross-camera hashing by identity correlation) and person re-id (i.e. inter-camera context modelling by VCD and person class discrimination learning by Identity Verification Regularisation) models in a principled manner, therefore yielding favourable performance.

Among existing hashing methods, best performers are KSH, SDH, CCA+ITQ, COSDISH and SePH, with the former four taking single-modality modelling and the last one multi-modality modelling. This implies that existing multi-modal hashing methods do not necessarily have advantages over single-modal counterparts in LSOW re-id matching. It is observed in surprise that the worst performers are supervised methods FH, CVH and CMSSH, rather than unsupervised models LSH, SH, SGH, and ITQ. The plausible reasons are: (1) The state-of-the-art LOMO feature possesses good re-id discrimination and cross-view invariance property, which makes unsupervised methods fairly effective. (2) The neglect of cross-view identity correlation modelling by existing supervised hashing methods may lead to model overfit in discriminative learning. In all unsupervised methods, there is no clear winner: ITQ generates the best results on CUHK03 and Market-1501; SGH and ITQ are top-2 on SYSU; and LSH is very competitive to other alternatives on all three datasets. This seems reasonable because all these unsupervised hashing models do not exploit labelled data for discriminative model learning. A qualitative evaluation is presented in Figure 5.

Hash code length. We evaluated the hash code length effect using top-5 hashing methods (KSH, SDH, CCA+ITQ, COSDISH, SePH). We used four code lengths $\{32, 64, 128, 256\}$. Table IV shows that longer hash codes generally yield better re-id performances across all methods. This is consistent with existing findings in the hashing literature. The X-ICE surpasses all competitors in every case. This validates the advantages of our method over alternatives across various code lengths.

D. Comparing State-of-the-Art Person Re-Id Methods

We compared the proposed X-ICE method with five state-of-the-art supervised person re-id models (e.g. KISSME [34], CVDCA [36], XQDA [6], MLAPG [11], DNS [12]). Table V shows that the X-ICE hashing method is very competitive in re-id accuracy as compared to these strong non-hashing person re-id models in many cases, although sometimes outperformed by $6{\sim}12\%$ in TTR across different FTRs in individual verification and by $\leq 15\%$ in set verification, and by $4{\sim}10\%$ in mAP. This performance gap is partially attributed to information loss in converting long float-valued feature representations into short binary-valued hash codes. This form change typically leads to degraded representation capability as experienced in existing fast similarity search models [21]. Importantly, X-ICE demonstrates the critical efficient search advantage over all these re-id competitors. For example, X-ICE is at least over two orders of magnitude faster than non-hashing person re-id methods in searching a gallery person against the large probe population. We measured the search time on a workstation of Intel CPU @ 2.66 GHz, 4.0GB RAM. This suggest a favourable trade-off between person search efficiency (due to feature binarisation) and effectiveness (due to feature discrimination) for LSOW person re-id by the proposed X-ICE model.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2017.2740564, IEEE Transactions on Image Processing

TABLE IV: Evaluating the effect of hash code length. (Metric: TTR (%) when FTR = 1%; IV: Individual Verification, SV: Set Verification.)

Criterion	Dataset	CUHK03 [39]				SYSU [36]				Market-1501 [40]			
	Code length (bits)	32	64	128	256	32	64	128	256	32	64	128	256
IV	CCA+ITQ [75]	29.09	30.35	**28.81**	28.11	39.72	44.44	45.87	**50.50**	43.90	50.84	53.69	**57.75**
	KSH [30]	20.74	23.18	25.38	**32.29**	30.44	37.39	46.31	**53.23**	38.27	45.70	51.44	**59.03**
	SDH [31]	11.46	19.87	29.72	**38.80**	15.00	27.11	37.65	**46.09**	19.05	33.01	47.18	**58.03**
	COSDISH [78]	1.86	4.05	7.41	**13.19**	6.66	13.82	25.59	**38.04**	6.85	15.11	27.49	**39.29**
	SePH [79]	4.53	9.34	16.93	**26.98**	9.17	16.55	26.78	**37.15**	11.24	19.78	30.22	**41.88**
	X-ICE(hinge)	35.22	42.01	47.50	**49.67**	43.24	52.26	58.96	**61.86**	45.32	56.08	62.00	**66.52**
	X-ICE(reg)	37.12	42.29	46.24	**49.96**	42.32	53.07	57.34	**63.13**	45.96	53.18	59.26	**64.18**
SV	CCA+ITQ [75]	9.19	10.87	**10.63**	10.52	12.38	16.08	18.97	**21.49**	11.31	13.58	15.80	**18.72**
	KSH [30]	5.91	8.70	9.40	**11.65**	11.25	14.61	18.96	**22.07**	11.46	16.07	19.38	**21.74**
	SDH [31]	2.90	6.36	10.67	**13.59**	4.08	8.83	12.68	**15.46**	6.38	12.01	18.24	**20.95**
	COSDISH [78]	1.53	1.36	2.88	**5.07**	2.83	5.10	10.86	**16.86**	2.02	5.85	11.49	**15.80**
	SePH [79]	1.44	2.82	5.55	**9.64**	2.45	5.62	8.68	**12.40**	4.53	7.30	9.68	**11.78**
	X-ICE(hinge)	9.46	12.46	15.67	**16.41**	13.00	18.43	22.53	**23.32**	12.91	18.93	23.82	**26.81**
	X-ICE(reg)	10.62	12.80	14.39	**16.37**	11.57	18.39	21.60	**25.94**	11.33	15.73	19.59	**22.27**

TABLE V: Comparing state-of-the-art non-hashing person re-id methods. (Metrics: TTR (%) at varying FTRs (%), and mAP (%); ST: *Search Time* (smaller is better), with unit set as the search time of X-ICE.)

Dataset		CUHK03 [39]							SYSU [36]							Market-1501 [40]						
Metric		Individual Verification					mAP	ST	Individual Verification					mAP	ST	Individual Verification					mAP	ST
		1%	5%	10%	20%	30%	(%)	-	1%	5%	10%	20%	30%	(%)	-	1%	5%	10%	20%	30%	(%)	-
X-ICE(hinge)		49.67	**79.60**	**89.50**	**96.09**	**98.48**	**11.66**	**1**	61.86	84.10	91.47	**96.35**	**98.26**	**29.93**	**1**	**66.52**	**88.03**	**93.66**	**97.15**	98.55	**21.47**	**1**
X-ICE(reg)		**49.96**	78.18	88.96	95.88	97.98	11.23	**1**	**63.13**	**84.86**	**91.52**	96.17	98.08	29.44	**1**	64.18	86.98	92.91	97.09	**98.59**	20.68	**1**
non-hash	KISSME [34]	33.66	61.67	74.69	86.14	91.29	7.97	954	33.67	58.30	70.79	83.57	90.08	19.31	2056	59.40	81.55	89.19	94.87	96.84	21.64	1447
	CVDCA [36]	41.73	68.65	80.63	89.91	94.52	8.50	367	58.39	81.89	89.75	94.94	97.09	23.88	486	32.87	52.90	63.15	74.81	81.45	11.25	456
	XQDA [6]	56.96	82.67	91.71	97.04	98.29	15.93	1775	**59.93**	**83.31**	**91.04**	**96.32**	**98.09**	**34.28**	4311	71.40	90.00	94.70	97.98	99.04	28.95	3185
	MLAPG [11]	53.97	83.05	**92.35**	**97.56**	**99.02**	12.79	**128**	55.61	79.55	88.17	94.77	97.54	29.13	**159**	66.71	87.93	94.04	97.73	98.95	22.30	**154**
	DNS [12]	**59.68**	**84.68**	92.25	97.05	98.50	**17.52**	316	59.62	82.12	89.52	95.10	97.37	29.29	164	**75.33**	**91.55**	**95.83**	**98.26**	**99.13**	**31.48**	262
Metric		Set Verification					mAP	ST	Set Verification					mAP	ST	Set Verification					mAP	ST
X-ICE(hinge)		**16.41**	**37.50**	**50.14**	**66.56**	**77.30**	**11.66**	**1**	23.32	46.84	60.48	74.20	82.37	**29.93**	**1**	**26.81**	**52.73**	**66.47**	**79.66**	**86.16**	**21.47**	**1**
X-ICE(reg)		16.37	37.36	49.71	65.49	76.03	11.23	**1**	**25.94**	**49.94**	**62.59**	**75.91**	**83.30**	29.44	**1**	22.27	48.12	62.87	77.13	84.55	20.68	**1**
non-hash	KISSME [34]	9.00	23.77	34.78	50.64	62.29	7.97	954	12.35	25.26	35.66	50.95	62.16	19.31	2056	31.42	50.39	61.30	73.88	81.37	21.64	1447
	CVDCA [36]	15.27	33.04	44.94	60.61	70.89	8.50	367	21.55	**46.22**	**59.37**	**73.29**	81.31	23.88	486	15.55	31.37	41.00	53.67	63.02	11.25	456
	XQDA [6]	15.43	38.83	54.55	71.38	80.87	15.93	1775	**23.33**	45.07	58.14	73.04	**82.23**	**34.28**	4311	34.18	58.71	70.53	82.21	88.50	28.95	3185
	MLAPG [11]	13.86	38.19	54.15	71.54	82.00	12.79	**128**	22.00	43.00	56.29	71.31	80.06	29.13	**159**	30.79	55.14	68.69	81.78	88.62	22.30	**154**
	DNS [12]	**18.91**	**41.09**	**54.86**	**71.88**	**81.65**	**17.52**	316	20.42	39.01	50.95	65.82	75.57	29.29	164	**41.34**	**62.74**	**73.12**	**83.63**	**89.42**	**31.48**	262

TABLE VI: Evaluating model training time (in seconds) on CUHK03.

LSH	SH	SGH	ITQ	CCA+ITQ	KSH	FH
0	8.08	552.75	5.43	6.17	**3290.22**	469.21
SDH	COSDISH	CMSSH	CVH	CMFH	SCM	SePH
18.37	2040.76	2591.98	5.19	47.69	1109.21	2997.52
CVDCA	XQDA	MLAPG	DNS	KISSME	X-ICE(hinge)	X-ICE(reg)
42.25	45.22	332.53	127.51	16.32	243.67	540.38

E. Further Analysis

Model training time. We evaluated the model training time on CUHK03. Table VI shows that LSH (no model learning) and KSH (expensive kernel based learning) is the fastest and slowest, while the X-ICE is moderately fast. Since model learning is conducted off-line, a high training cost does not pose stringent constraint on model deployment.

Joining re-id & hashing. We evaluated the re-id+hashing joining approach using five state-of-the-art re-id (KISSME, MLAPG, CVDCA, XQDA and DNS) and two top hashing (KSH and SDH) models, resulting in totally 10 LSOW solutions. In each solution, we first learn a re-id model for projecting the visual features of person images into a discriminative subspace; We then train a hashing model in the subspace for allowing fast search. Table VII shows that re-id+hashing is competitive. For example, MLAPG performs well when integrated with KSH or SDH. However, the X-ICE model still yields the best overall mAP performances on all three datasets. On the other hand, the X-ICE requires no feature subspace projection and therefore not only giving more efficient deployment, but also eliminating the need for tuning the subspace dimension. This validates the superiority of our joint learning scheme over the re-id+hashing approach.

Larger search pool. We evaluated competitive hashing models on larger search pools by using 34,574 person images from an auxiliary dataset [84] (independent of CUHK03, SYSU, and Market-1501) as additional imposters. In this larger scale evaluation, we considered only hashing methods due to their unique fast search capability as compared to conventional re-id models. Table VIII shows that all these methods suffer lower mAP performances given more open search spaces, but the X-ICE model remains the best. This validates the clear scalability and superiority of the proposed model in larger scale deployments.

Moreover, we enlarged the Market-1501 dataset by adding 237,256 person bounding box images from its video based sibling dataset MARS [82]. We call this dataset "ExMarket". We conducted an experimental evaluation with comparisons to top-2 hashing competitors KSH [30] and SDH [31] on ExMarket. Table IX suggests the consistent performance advantages of the proposed X-ICE method over top-performing alternatives by a clear margin.

Effect of visual features. We evaluated the effect of visual features by additionally examining data-driven deep features. This also allows to examine the interaction between features

TABLE VII: Evaluating the re-id+hashing joining approach. (Metrics: TTR (%) at varying FTRs (%), and mAP (%).)

Dataset	CUHK03 [39]						SYSU [36]						Market-1501 [40]					
Metric	Individual Verification					mAP	Individual Verification					mAP	Individual Verification					mAP
	1%	5%	10%	20%	30%	(%)	1%	5%	10%	20%	30%	(%)	1%	5%	10%	20%	30%	(%)
KISSME [34] & KSH [30]	29.38	54.83	67.19	79.46	86.89	5.09	49.03	73.24	82.68	90.46	94.14	18.92	58.99	81.83	89.09	94.04	96.55	16.04
MLAPG [11] & KSH [30]	**50.79**	76.87	86.18	92.38	95.34	11.62	61.32	83.21	89.48	94.31	97.98	27.88	63.47	87.90	**93.86**	97.06	98.33	19.03
CVDCA [36] & KSH [30]	15.83	37.27	53.38	71.09	82.10	3.10	44.82	71.41	82.04	90.44	94.27	15.85	27.20	52.90	65.84	79.60	86.93	5.63
XQDA [6] & KSH [30]	41.96	67.57	77.89	86.77	90.92	8.35	58.28	80.05	87.72	93.50	96.05	24.98	62.53	84.90	91.37	95.62	97.54	18.03
DNS [12] & KSH [30]	47.86	76.26	86.02	93.19	95.89	10.57	55.73	81.94	89.79	95.27	97.40	25.25	**66.94**	88.00	93.07	96.64	98.15	20.25
KISSME [34] & SDH [31]	32.25	57.86	71.42	83.77	90.10	5.83	44.72	69.39	79.45	88.60	93.14	16.39	59.17	82.59	90.04	94.91	97.00	16.46
MLAPG [11] & SDH [31]	49.10	77.33	85.92	91.51	94.27	9.82	60.78	83.28	89.74	94.59	96.51	26.94	62.57	87.19	92.32	95.39	96.93	17.23
CVDCA [36] & SDH [31]	47.02	76.78	86.98	94.25	96.98	9.83	53.16	79.21	88.43	94.56	97.20	23.52	44.97	78.13	88.31	95.13	97.55	11.25
XQDA [6] & SDH [31]	36.85	64.13	76.79	87.47	92.94	7.24	49.28	72.92	82.10	89.45	93.32	19.39	63.01	84.81	91.22	96.09	97.84	18.61
DNS [12] & SDH [31]	49.82	79.00	88.81	95.41	97.82	11.40	52.56	79.46	88.55	94.99	97.44	25.33	60.31	86.17	93.29	**97.16**	**98.66**	18.83
X-ICE(hinge)	49.67	**79.60**	**89.50**	**96.09**	**98.48**	**11.66**	61.86	84.10	91.47	**96.35**	**98.26**	**29.93**	66.52	**88.03**	93.66	97.15	98.55	**21.47**
X-ICE(reg)	49.96	78.18	88.96	95.88	97.98	11.23	**63.13**	**84.86**	**91.52**	96.17	98.08	29.44	64.18	86.98	92.91	97.09	98.59	20.68
Metric	Set Verification					mAP	Set Verification					mAP	Set Verification					mAP
KISSME [34] & KSH [30]	9.61	24.25	35.07	50.58	61.27	5.09	20.47	40.10	51.87	65.46	74.18	18.92	23.14	47.99	62.18	75.28	82.86	16.04
MLAPG [11] & KSH [30]	**19.82**	**42.35**	**55.56**	69.76	78.81	11.62	23.89	49.94	62.95	76.08	83.39	27.88	20.92	50.20	65.69	80.37	87.66	19.03
CVDCA [36] & KSH [30]	2.02	9.48	22.07	39.77	53.12	3.10	16.42	36.13	48.95	64.02	73.61	15.85	8.52	23.96	35.99	50.62	62.09	5.63
XQDA [6] & KSH [30]	14.78	34.24	46.90	62.54	72.07	8.35	25.58	48.47	60.00	72.64	80.28	24.98	21.89	48.67	63.98	78.13	85.28	18.03
DNS [12] & KSH [30]	16.52	37.58	51.27	66.52	76.67	10.57	18.10	42.44	56.93	72.55	81.20	25.25	20.36	50.41	**67.46**	80.74	87.45	20.25
KISSME [34] & SDH [31]	11.18	26.50	38.00	53.37	63.00	5.83	19.83	36.99	48.70	62.55	72.10	16.39	20.34	45.88	60.92	75.43	82.95	16.46
MLAPG [11] & SDH [31]	13.75	38.41	54.33	**70.54**	**80.47**	9.82	24.48	**50.33**	**63.30**	**76.31**	**83.78**	26.94	18.83	49.67	65.85	**80.80**	**88.01**	17.23
CVDCA [36] & SDH [31]	14.85	35.91	50.29	67.41	78.20	9.83	18.41	39.57	52.94	68.08	77.42	23.52	11.97	31.80	47.63	65.58	76.68	11.25
XQDA [6] & SDH [31]	12.15	29.77	41.93	57.84	68.73	7.24	22.05	41.36	52.05	65.44	74.12	19.39	24.48	49.42	64.02	77.28	84.71	18.61
DNS [12] & SDH [31]	14.28	36.24	50.69	67.99	78.47	11.40	16.26	37.46	51.20	67.04	77.03	25.33	14.63	36.35	53.49	72.91	82.17	18.83
X-ICE(hinge)	16.41	37.50	50.14	66.56	77.30	**11.66**	23.32	46.84	60.48	74.20	82.37	**29.93**	**26.81**	**52.73**	66.47	79.66	86.16	**21.47**
X-ICE(reg)	16.37	37.36	49.71	65.49	76.03	11.23	**25.94**	49.94	62.59	75.91	83.30	29.44	22.27	48.12	62.87	77.13	84.55	20.68

TABLE VIII: Evaluating larger search pools with 34,574 imposters. (Metrics: TTR (%) at varying FTRs (%), and mAP (%).)

Dataset	CUHK03 [39]						SYSU [36]						Market-1501 [40]					
Metric	Individual Verification					mAP	Individual Verification					mAP	Individual Verification					mAP
	1%	5%	10%	20%	30%	(%)	1%	5%	10%	20%	30%	(%)	1%	5%	10%	20%	30%	(%)
CCA+ITQ [75]	11.38	30.57	41.38	58.39	68.58	0.46	52.23	73.90	82.90	91.11	94.82	9.36	58.80	78.75	86.10	93.03	96.09	8.78
KSH [30]	12.88	29.84	45.29	61.17	72.55	0.61	49.54	73.07	82.67	90.20	93.84	9.91	61.75	80.26	87.30	92.91	95.89	9.76
SDH [31]	17.50	38.86	52.90	66.89	77.44	0.64	46.50	69.65	79.78	88.93	93.49	9.08	63.64	82.74	87.76	93.06	95.87	8.75
COSDISH [78]	10.24	27.87	41.18	60.10	70.75	0.32	36.70	59.61	71.14	82.30	88.50	5.24	38.61	61.00	71.50	82.72	88.67	4.11
SePH [79]	6.91	23.42	36.71	54.03	66.41	0.23	30.62	54.42	67.24	80.48	87.30	4.28	32.71	55.94	67.64	80.18	86.58	2.26
X-ICE(hinge)	28.31	56.05	68.47	82.47	89.98	**1.48**	66.40	**86.88**	**93.07**	**96.87**	**98.28**	**17.97**	**75.66**	**91.50**	**95.79**	**98.38**	99.01	**12.57**
X-ICE(reg)	**29.47**	**56.51**	**70.17**	**82.81**	**90.43**	1.43	**67.20**	86.79	92.70	96.37	98.10	17.90	72.85	90.59	95.66	97.97	**99.06**	12.30
Metric	Set Verification					mAP	Set Verification					mAP	Set Verification					mAP
CCA+ITQ [75]	3.32	10.98	18.65	33.45	45.43	0.46	24.90	45.08	56.37	69.13	77.16	9.36	31.17	53.66	63.42	74.51	80.19	8.78
KSH [30]	4.41	11.24	18.51	30.81	40.89	0.61	23.47	41.78	53.78	66.29	74.93	9.91	35.25	54.75	65.50	76.21	82.19	9.76
SDH [31]	5.47	15.04	24.09	39.59	51.16	0.64	19.59	38.69	50.83	64.40	72.97	9.08	34.87	56.43	66.22	76.53	83.73	8.75
COSDISH [78]	6.52	21.86	35.57	56.37	71.08	0.32	17.41	32.90	42.78	57.16	67.76	5.24	22.78	38.74	48.15	60.15	68.28	4.11
SePH [79]	0.96	5.80	11.27	23.96	35.84	0.23	12.43	26.98	38.05	51.95	63.23	4.28	17.86	33.28	43.45	56.41	65.02	2.26
X-ICE(hinge)	6.28	20.30	33.13	51.57	**64.50**	**1.48**	27.61	52.02	64.67	78.61	85.92	**17.97**	**38.73**	**64.16**	**75.18**	**85.08**	**90.76**	**12.57**
X-ICE(reg)	**7.28**	**22.84**	**36.04**	**52.86**	64.39	1.43	**31.66**	**55.27**	**66.83**	**79.65**	**86.49**	17.90	33.98	60.63	73.71	84.94	90.46	12.30

TABLE IX: Evaluating larger search performance on ExMarket. (Metrics: TTR (%) at varying FTRs (%), and mAP (%).)

Metric	Individual Verification					Set Verification					mAP
	1%	5%	10%	20%	30%	1%	5%	10%	20%	30%	(%)
KSH [30]	52.59	73.31	81.72	89.00	92.39	26.83	47.56	57.39	68.52	77.20	15.10
SDH [31]	51.98	73.56	80.05	86.58	90.90	23.87	47.96	58.23	69.71	76.83	14.55
X-ICE(hinge)	**59.93**	**82.31**	**88.63**	**93.84**	**96.36**	28.91	48.87	**61.34**	**73.54**	**81.39**	**17.96**
X-ICE(reg)	59.38	81.68	88.57	93.74	95.70	**29.88**	**50.04**	59.53	72.35	81.01	17.62

TABLE X: Evaluating the effect of different visual features. (Metrics: TTR (%) at FTR = 1%, and mAP (%). IV: Individual Verification, SV: Set Verification.)

Method	Feature	CUHK03 [39]			SYSU [36]			Market-1501 [40]		
		IV	SV	mAP	IV	SV	mAP	IV	SV	mAP
DCNN [83]	Deep	47.87	14.38	13.62	58.77	19.64	29.69	78.37	31.58	33.65
KSH [30]	LOMO	32.29	11.65	5.49	53.23	22.07	22.29	59.03	21.74	17.34
	Deep	51.08	**19.00**	14.77	60.38	21.55	31.54	79.50	34.39	34.96
SDH [31]	LOMO	38.80	13.59	7.31	46.09	15.46	17.99	58.03	20.95	15.57
	Deep	36.88	15.93	9.11	52.93	16.60	22.23	71.00	35.61	26.73
X-ICE(hinge)	LOMO	49.67	16.41	11.66	61.86	23.32	29.93	66.52	26.81	21.47
	Deep	51.79	18.29	**15.33**	63.08	23.35	32.95	**80.90**	**41.52**	**37.34**
X-ICE(reg)	LOMO	49.96	16.37	11.23	63.13	25.94	29.44	64.18	22.27	20.68
	Deep	**52.62**	17.63	15.21	**64.05**	**24.57**	**33.07**	80.37	41.01	37.23

and hashing functions on the LSOW re-id. Specifically, we adopted the Deep Convolutional Neural Network (DCNN) to learn visual features in the hashing context [83] and used the 4096-dimensional activation of the 7-th fully-connection layer (previous to the latent hashing layer) as image representation. As such, all compared hashing methods can benefit equally from the deep learning advantages. Table X shows that the X-ICE model outperforms again all alternatives. This suggests the consistent superiority of our method over existing models under different representations, thanks to the joint learning of cross-view identity correlation hashing and person class verification regularisation. Two additional observations are: (1)

Non-deep hashing methods (e.g. KSH, SDH, and X-ICE) can be well integrated with deep features. (2) While DCNN enjoys the merit of jointly learning feature and hashing functions, it is still inferior to KSH and our X-ICE. This suggests that feature learning and hashing function learning are two important and complementary aspects of a LSOW re-id method.

V. Conclusion

We presented a more realistic Large Scale Open-World (LSOW) person re-id problem setting. LSOW is uniquely characterised by vast probe search population with a large number of imposters, without the unrealistic closed-world and small search scale assumptions as made in existing re-id methods. Importantly, LSOW raises the re-id matching efficiency requirement and moves the re-id research a step further towards practical deployments. To address LSOW re-id, we proposed a new Cross-view Identity Correlation and vErification (X-ICE) hashing re-id model. This is achieved by a joint learning of cross-view identity correlation hashing and discriminative person class verification regularisation. The X-ICE model is learned by a principled alternating optimisation algorithm. Extensive comparative evaluations have demonstrated the superiority and advantages of the proposed X-ICE method over a wide range of hashing and re-id competitors on three large re-id benchmarks.

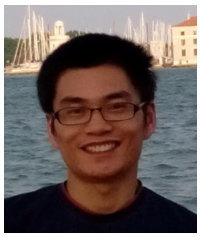
Acknowledgment

Xiatian Zhu and Botong Wu equally contributed to this work.

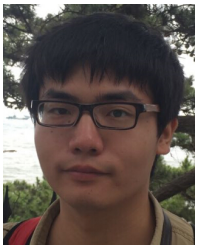
References

[1] S. Gong, M. Cristani, C. L. Chen, and T. M. Hospedales, "The re-identification challenge," in *Person Re-Identification*. Springer, 2014.
[2] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in *ECCV*, 2012.
[3] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *CVPR*, 2010.
[4] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *BMVC*, 2010.
[5] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *CVPR*, 2013.
[6] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *CVPR*, 2015.
[7] Y. Li, Z. Wu, S. Karanam, and R. J. Radke, "Multi-shot human re-identification using adaptive fisher discriminant analysis," in *BMVC*, 2015.
[8] X. Li, W.-S. Zheng, X. Wang, T. Xiang, and S. Gong, "Multi-scale learning for low-resolution person re-identification," in *ICCV*, 2015.
[9] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by video ranking," in *ECCV*, 2014.
[10] ——, "Person re-identification by discriminative selection in video ranking," *IEEE TPAMI*, vol. 38, no. 12, pp. 2501–2514, Dec 2016.
[11] S. Liao and S. Z. Li, "Efficient psd constrained asymmetric metric learning for person re-identification," in *ICCV*, 2015.
[12] L. Zhang, T. Xiang, and S. Gong, "Learning a discriminative null space for person re-identification," in *CVPR*, 2016.
[13] Y. Zhang, B. Li, H. Lu, A. Irie, and X. Ruan, "Sample-specific svm learning for person re-identification," in *CVPR*, 2016.
[14] D. Chen, Z. Yuan, B. Chen, and N. Zheng, "Similarity learning with spatial constraints for person re-identification," in *CVPR*, 2016.
[15] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *CVPR*, 2014.
[16] E. Ahmed, M. J. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification." in *CVPR*, 2015.
[17] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognit.*, vol. 48, no. 10, pp. 2993–3003, 2015.
[18] T. Xiao, H. Li, W. Ouyang, and X. Wang, "Learning deep feature representations with domain guided dropout for person re-identification," in *CVPR*, 2016.
[19] S. Liao, Z. Mo, Y. Hu, and S. Z. Li, "Open-set person re-identification," *arXiv*, 2014.
[20] W. S. Zheng, S. Gong, and T. Xiang, "Towards open-world person re-identification by one-shot group-based verification," *IEEE TPAMI*, vol. 38, no. 3, pp. 591–606, March 2016.
[21] J. Wang, H. T. Shen, J. Song, and J. Ji, "Hashing for similarity search: A survey," *arXiv*, 2014.
[22] L. Liu, M. Yu, and L. Shao, "Multiview alignment hashing for efficient image search," *IEEE TIP*, vol. 24, no. 3, pp. 956–966, 2015.
[23] C. Xu, D. Tao, and C. Xu, "Large-margin multi-view information bottleneck," *IEEE TPAMI*, vol. 36, no. 8, pp. 1559–1572, 2014.
[24] ——, "A survey on multi-view learning," *arXiv*, 2013.
[25] A. Gionis, P. Indyk, R. Motwani *et al.*, "Similarity search in high dimensions via hashing," in *VLDB*, vol. 99, no. 6, 1999, pp. 518–529.
[26] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Proceedings of ACM annual symposium on Computational geometry*, 2004.
[27] Y. Weiss, A. Torralba, and R. Fergus, "Spectral hashing," in *NIPS*, 2009.
[28] J. Wang, S. Kumar, and S.-F. Chang, "Semi-supervised hashing for scalable image retrieval," in *CVPR*, 2010, pp. 3424–3431.
[29] W. Liu, J. Wang, S. Kumar, and S.-F. Chang, "Hashing with graphs," in *ICML*, 2011.
[30] W. Liu, J. Wang, R. Ji, Y.-G. Jiang, and S.-F. Chang, "Supervised hashing with kernels," in *CVPR*, 2012.
[31] F. Shen, C. Shen, W. Liu, and H. Tao Shen, "Supervised discrete hashing," in *CVPR*, 2015.
[32] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *CVPR*, 2014.
[33] N. McLaughlin, J. Martinez del Rincon, and P. Miller, "Recurrent convolutional network for video-based person re-identification," in *CVPR*, 2016.
[34] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *CVPR*, 2012.
[35] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE TPAMI*, vol. 35, no. 3, pp. 653–668, 2013.
[36] Y.-C. Chen, W.-S. Zheng, P. C. Yuen, and J. Lai, "An asymmetric distance model for cross-view feature mapping in person re-identification," in *IEEE TCSVT*, 2015.
[37] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *CVPR*, 2015.
[38] B. Wu, Q. Yang, W.-S. Zheng, Y. Wang, and J. Wang, "Quantized correlation hashing for fast cross-modal search," in *IJCAI*, 2015.
[39] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *CVPR*, 2014.
[40] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *ICCV*, 2015.
[41] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *ECCV*, 2008.
[42] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," *IEEE TPAMI*, vol. 35, no. 7, pp. 1622–1634, 2013.
[43] X. Ma, X. Zhu, S. Gong, X. Xie, J. Hu, K.-M. Lam, and Y. Zhong, "Person re-identification by unsupervised video matching," *Pattern Recognit.*, vol. 65, pp. 197–210, 2017.
[44] W. Li, X. Zhu, and S. Gong, "Person re-identification by deep joint learning of multi-loss classification," in *IJCAI*, 2017.
[45] A. Wu, W.-S. Zheng, H. Yu, S. Gong, and J. Lai, "Rgb-infrared cross-modality person re-identification," in *ICCV*, 2017.
[46] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *ECCV*, 2014.
[47] R. Zhang, L. Lin, R. Zhang, W. Zuo, and L. Zhang, "Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification," *IEEE TIP*, vol. 24, no. 12, pp. 4766–4779, 2015.
[48] W.-S. Zheng, X. Li, T. Xiang, S. Liao, J. Lai, and S. Gong, "Partial person re-identification," in *ICCV*, 2015.
[49] H. Wang, X. Zhu, T. Xiang, and S. Gong, "Towards unsupervised open-set person re-identification," in *ICIP*, 2016.
[50] H. Wang, S. Gong, X. Zhu, and T. Xiang, "Human-in-the-loop person re-identification," in *ECCV*, 2016.
[51] J. You, A. Wu, X. Li, and W.-S. Zheng, "Top-push video-based person re-identification," in *CVPR*, 2016.

[52] Y.-C. Chen, X. Zhu, W.-S. Zheng, and J.-H. Lai, "Person re-identification by camera correlation aware feature augmentation," *IEEE TPAMI (DOI: 10.1109/TPAMI.2017.2666805)*, 2017.
[53] H. Yu, A. Wu, and W.-S. Zheng, "Cross-view asymmetric metric learning for unsupervised person re-identification," in *ICCV*, 2017.
[54] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, "End-to-end deep learning for person search," in *arXiv*, 2016.
[55] M. Rastegari, J. Choi, S. Fakhraei, D. Hal, and L. Davis, "Predictable dual-view hashing," in *ICML*, 2013.
[56] S. Kumar and R. Udupa, "Learning hash functions for cross-view similarity search," in *IJCAI*, 2011.
[57] M. M. Bronstein, A. M. Bronstein, F. Michel, and N. Paragios, "Data fusion through cross-modality metric learning using similarity-sensitive hashing," in *CVPR*, 2010.
[58] D. Wang, P. Cui, M. Ou, and W. Zhu, "Deep multimodal hashing with orthogonal regularization," in *ICAI*, 2015.
[59] S. Gong, M. Cristani, C. C. Loy, and T. M. Hospedales, "The re-identification challenge," in *Person Re-Identification*. Springer, January 2014, pp. 1–20.
[60] D. Zhang and W.-J. Li, "Large-scale supervised multimodal hashing with semantic correlation maximization," in *AAAI*, 2014.
[61] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *CVPR*, 2012.
[62] H. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," in *NIPS*, 2003.
[63] P. Carbonetto, N. De Freitas, and K. Barnard, "A statistical model for general contextual object recognition," in *ECCV*, 2004.
[64] A. Oliva and A. Torralba, "The role of context in object recognition," *Trends in Cognitive Sciences*, vol. 11, no. 12, pp. 520–527, 2007.
[65] W.-S. Zheng, S. Gong, and T. Xiang, "Quantifying and transferring contextual information in object detection," *IEEE TPAMI*, vol. 34, no. 4, pp. 762–777, 2012.
[66] S. Si, D. Tao, and B. Geng, "Bregman divergence-based regularization for transfer subspace learning," *IEEE TKDE*, vol. 22, no. 7, pp. 929–942, 2010.
[67] H. H. Bauschke and J. M. Borwein, "Joint and separate convexity of the bregman distance," *Studies in Computational Mathematics*, vol. 8, pp. 23–36, 2001.
[68] M. D. Perlman, "Jensen's inequality for a convex vector-valued function on an infinite-dimensional space," *Journal of Multivariate Analysis*, vol. 4, no. 1, pp. 52–65, 1974.
[69] J. Weston and C. Watkins, "Multi-class support vector machines," Tech. Rep., 1998.
[70] K. Crammer and Y. Singer, "On the algorithmic implementation of multiclass kernel-based vector machines," *JMLR*, vol. 2, no. December, pp. 265–292, 2001.
[71] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *JMLR*, vol. 9, no. Aug, pp. 1871–1874, 2008.
[72] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," *Mathematical Programming*, vol. 142, no. 1-2, pp. 397–434, 2013.
[73] G. D. Smith, *Numerical solution of partial differential equations: finite difference methods*. Oxford university press, 1985.
[74] Q.-Y. Jiang and W.-J. Li, "Scalable graph hashing with feature transformation," in *IJCAI*, 2015.
[75] Y. Gong and S. Lazebnik, "Iterative quantization: A procrustean approach to learning binary codes," in *CVPR*, 2011.
[76] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.
[77] G. Lin, C. Shen, Q. Shi, A. Hengel, and D. Suter, "Fast supervised hashing with decision trees for high-dimensional data," in *CVPR*, 2014.
[78] W.-C. Kang, W.-J. Li, and Z.-H. Zhou, "Column sampling based discrete supervised hashing," in *AAAI*, 2016.
[79] Z. Lin, G. Ding, M. Hu, and J. Wang, "Semantics-preserving hashing for cross-view retrieval," in *CVPR*, 2015.
[80] G. Ding, Y. Guo, and J. Zhou, "Collective matrix factorization hashing for multimodal data," in *CVPR*, 2014.
[81] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *CVPR*, 2014.
[82] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "Mars: A video benchmark for large-scale person re-identification." in *ECCV*, 2016.
[83] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen, "Deep learning of binary hash codes for fast image retrieval," in *CVPRW*, 2015.
[84] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, "Joint detection and identification feature learning for person search," in *CVPR*, 2017.



Xiatian Zhu received his B.Eng. and M.Eng. from University of Electronic Science and Technology of China, and his Ph.D. (2015) from Queen Mary University of London. He won The Sullivan Doctoral Thesis Prize (2016), an annual award representing the best doctoral thesis submitted to a UK University in the field of computer or natural vision. His research interests include computer vision and machine (deep) learning. Homepage: http://www.eecs.qmul.ac.uk/~xiatian/.



Botong Wu received his B.S. degree (2015) from Sun Yat-sen University. He now is a PhD student at Peking University. He is majored in Computer Vision and Digital Arts. His research interests include machine learning and pattern recognition.

Dongcheng Huang received his B.Eng. (2014) and M.Sc. (2017) from Sun Yat-sen University. His research interests are in computer vision and machine learning.

Wei-Shi Zheng is now a Full Professor at Sun Yat-sen University. He has now published more than 90 papers, including more than 60 publications in main journals (TPAMI, TIP, PR) and top conferences (ICCV, CVPR, IJCAI). His research interests include person/object association and activity understanding in visual surveillance. He has joined Microsoft Research Asia Young Faculty Visiting Programme. He is a recipient of Excellent Young Scientists Fund of the NSFC, and a recipient of Royal Society-Newton Advanced Fellowship of United Kingdom. Homepage: http://isee.sysu.edu.cn/%7ezhwshi/.