Neural Computing and Applications
https://doi.org/10.1007/s00521-019-04424-1

REVIEW ARTICLE

A half-precision compressive sensing framework for end-to-end person re-identification

Longlong Liao[1,2] · Zhibang Yang[3] · Qing Liao[4] · Kenli Li[5] · Keqin Li[6] · Jie Liu[1] · Qi Tian[7]

Received: 31 December 2018 / Accepted: 7 August 2019

Abstract
Compressive sensing (CS) approaches are useful for end-to-end person re-identification (Re-ID) in reducing the overheads of transmitting and storing video frames in distributed multi-camera systems. However, the reconstruction quality degrades appreciably as the measurement rate decreases for existing CS methods. To address this problem, we propose a half-precision CS framework for end-to-end person Re-ID named HCS4ReID, which efficiently recoveries detailed features of the person-of-interest regions in video frames. HCS4ReID supports half-precision CS sampling, transmitting and storing CS measurements with half-precision floats, and CS reconstruction with two measurement rates. Extensive experiments implemented on the PRW dataset indicate that the proposed HCS4ReID achieves 1.55 × speedups over the single-precision counterpart on average for the CS sampling on an Intel HD Graphics 530, and only half-network bandwidth and storage space are needed to transmit and store the generated CS measurements. Comprehensive evaluations demonstrate that the proposed HCS4ReID is a scalable and portable CS framework with two measurement rates, and suitable for end-to-end person Re-ID. Especially, it achieves the comparable performance on the reconstructed PRW dataset against CS reconstruction with single-precision floats and a single measurement rate.

Keywords Compressive sensing · Half-precision float · Pedestrian detection · Person re-identification

1 Introduction

Person re-identification (Re-ID) is a challenging and fundamental task to retrieve a given person image in the gallery of all pedestrian images captured across non-overlapping camera views. It increasingly receives attention as a key component of the large-scale intelligent surveillance [3, 40] and is deployed in smart cities, multi-camera forensic search, public transportation and object-level video advertising [46, 47]. Although a variety of powerful person Re-ID algorithms have been proposed over the past few years, most of them usually focus on designing

✉ Longlong Liao
liaolonglong16@nudt.edu.cn

Zhibang Yang
yangzb@ccsu.edu.cn

Qing Liao
liaoqing@hit.edu.cn

Kenli Li
lkl@hnu.edu.cn

Keqin Li
lik@newpaltz.edu

Jie Liu
liujie@nudt.edu.cn

Qi Tian
Qi.Tian@utsa.edu

[1] College of Computer, National University of Defense Technology, Changsha 410073, China

[2] State Key Laboratory of High Performance Computing, Changsha 410073, China

[3] College of Computer Engineering and Applied Mathematics, Changsha University, Changsha 410022, China

[4] Department of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China

[5] College of Information Science and Engineering, Hunan University, Changsha 410082, China

[6] Department of Computer Science, State University of New York, New Paltz, NY 12561, USA

[7] Department of Computer Science, University of Texas at San Antonio, San Antonio, TX 78249, USA

Published online: 14 August 2019

discriminative features [39, 55] and constructing more robust metric learning algorithms [18] either individually or sequentially. Recently, end-to-end person Re-ID approaches are also exploited in [44]. It performs re-identification without pre-cropped pedestrian images, and aims at recognizing a queried person image in a gallery of pedestrians detected from captured video frames with distributed multi-camera systems. It can simplify and facilitate the applications of person Re-ID in real-world scenarios. However, there are few research works paying attention to reduce the storage and transmission of captured pedestrian images for end-to-end person Re-ID.

Newly created distributed multi-camera surveillance systems adapt high-resolution cameras, and they generate an increasing number of high-resolution surveillance images and videos. The data collected over these distributed multi-camera networks make it impractical to transport all captured raw data to remote servers, and then to analyse and store them. Even if the storage space on servers is large enough, consideration must be given to the cost of the transmission and the availability of the high-speed network. Thus, it becomes one of the stumbling blocks on the way to wide implementation of person Re-ID in real large-scale surveillance scenarios.

Compressive sensing (CS) methods demonstrate that images and videos can be reconstructed with high fidelity [14]. This acquisition strategy enables image and video sampling with a sub-Nyquist measurement rate [21, 31]. Meanwhile, CS conducts data sampling and compression at the same time, which is efficient for limited transmission bandwidth and storage space, and enables low-cost video capturing on a range of edge devices. However, existing CS algorithms focus on improving the performance of CS reconstruction for an entire image or video frame. As a result, the reconstruction quality degrades appreciably as the measurement rate decreases. For end-to-end person Re-ID, the person-of-interest regions are more critical than the rest regions, and therefore CS construction quality of image blocks in these regions significantly affects the performance of person Re-ID and its usability in real surveillance systems.

Meanwhile, low precision data formats are sufficient for deep learning algorithms and signal processing applications, since the final results suffer no noticeable loss [9]. For instance, IEEE half-precision floats (FP16) are sufficient not just for the inference of deep learning models but also for training them. Arithmetic operations with half-precision floats are faster than the corresponding operations with single-precision floats when hardware devices natively support it, e.g. the integrated GPUs in 6th generation Intel processors or newer ones. OpenCL [16, 26, 36] is a portable interface for parallel programming on compute devices, and allows the same code to be executed across a variety of processors and accelerators. It is used to accelerate computationally intensive applications across different devices and architectures by writing portable code. Part of industry-leading hardware vendors (e.g. Intel, Altera and XILINX) have provided OpenCL implementations for half-precision routines on their hardware devices. Thus, half-precision applications can be ported to several hardware platforms with increasing compute devices supporting half-precision routines. On the other hand, FP16 has limited numerical range. The data presented with FP16 takes less storage space and transmission bandwidth than the same data with the 32-bit single-precision floats (FP32).

It is crucial to sample as few CS measurements as possible for reconstructing images/videos in person Re-ID, and still to retain enough local features in the person-of-interest regions for re-identifying a specific person. To address this problem, we propose a half-precision CS framework for end-to-end person Re-ID, which is named HCS4ReID. The proposed framework supports half-precision CS with two measurement rates, where the higher one is used to sample and reconstruct the person-of-interest regions while the lower one is used on the rest regions. That is, measurement matrices and CS measurements are represented with half-precision floats, and the CS sampling is accelerated with OpenCL [4, 15, 45] when the hardware devices natively support half-precision arithmetic operations.

The key contributions in this paper are as follows.

- We propose a half-precision CS framework with two measurement rates for end-to-end person Re-ID named HCS4ReID. HCS4ReID utilizes two measurement rates on different image blocks to sample and reconstruct RGB video frames. It utilizes a higher one on image blocks belonging to the person-of-interest regions, and a lower one on the rest image blocks, which aims to reduce the cost of transmitting and storing captured videos in a large-scale distributed multi-camera system.
- We propose a half-precision CS sampling method that represents the data of measurement matrices and CS measurements with half-precision floats instead of the default single-precision floats. This not just reduces the required transmission bandwidth and storage space, but also accelerates the process of CS sampling with OpenCL.
- We present an end-to-end person Re-ID method for searching the queried person in a gallery of CS reconstructed images/videos, which are recovered with the proposed HCS4ReID. The difference from the existing researches about person Re-ID, the presented end-to-end person Re-ID method, aims to study the efficiency of the proposed half-precision CS method in

the real surveillance scenarios, instead of improving the performance of end-to-end person Re-ID methods.

The remainder of this paper is organized as follows. Section 2 describes a brief overview of related references on the end-to-end person Re-ID, CS sampling and reconstruction, and half-precision floats used in deep neural networks. Then, Sect. 3 explores the detailed architecture of the proposed HCS4ReID that is a half-precision CS framework with two measurement rates for end-to-end person Re-ID. In Sect. 4, the implementation approaches of the proposed HCS4ReID framework are presented. Meanwhile, a set of extensive experiments for evaluating the performance of the proposed HCS4ReID also executed, and the results are also discussed in detail. Finally, Sect. 5 concludes the paper including the limitations of this paper and new points for future investigations.

2 Related work

2.1 End-to-end person Re-ID

Deformable Part Model (DPM) [17] and Aggregated Channel Features (ACF) [12] are the most commonly used off-the-shelf pedestrian detectors, but they rely on hand-crafted features and linear classifiers to detect pedestrians. Driven by the surge of various deep neural networks (DNNs), a range of DNN-based pedestrian detectors has been proposed in recent years. For instance, Faster R-CNN [35] introduces a region proposal network that shares full-image convolutional features with the detection network, to enable nearly cost-free region proposals. It is improved and adapted to detect pedestrians in [24, 50]. On the other hand, many state-of-the-art object detection networks (e.g. YOLOv3 [22]) can also be employed as pedestrian detectors. Despite the impressive recent progress in pedestrian detection, it has been rarely considered with person re-ID as a whole procedure.

Person Re-ID is usually viewed as an image retrieval problem, i.e. searching the queried person in a gallery of pedestrian images. It is fundamental for various surveillance applications, such as finding criminals, cross-camera person tracking, and person activity analysis. Various deep learning-based person Re-ID methods have been proposed in recent years [1, 27]. Ding et al. [10] and Cheng et al. [8] exploited triplet samples for training person Re-ID models to minimize the feature distance between the same person and maximize the distance between different persons. Xiao et al. [43] proposed to learn features by classifying identities.

Most existing researches about person Re-ID mainly focus on learning features and metric learning approaches. Supervised [19] or unsupervised [2] approaches are also proposed to extract relevant features and to combine them into a single similarity function. Several distance learning approaches are available, a new relative distance comparison was proposed in [59] for maximizing the probability of a pair of true matches. Li et al. [27] and Ahmed et al. [1] utilized a pair of cropped pedestrian images as input, and employed a binary verification loss function to train DNNs for person Re-ID. Other research works reformulate the person Re-ID as a ranking problem, where the potential true match is assigned with the highest rank rather than a distance metric learning, in this way, the re-identification problem is cast into a relative ranking problem [6, 7].

Although numerous person Re-ID methods and related datasets have been proposed, they mainly focus on matching pre-defined pedestrian images between queries and candidates. In real-world scenarios, the annotations of pedestrian bounding boxes are unavailable, and the target person is needed to be detected and searched from the gallery of the whole scene video frames [57]. Real person re-ID applications adopt pedestrian detectors to automatically obtain cropped pedestrian images from captured video frames, which lead to end-to-end person re-ID systems.

Thus, there is still a big gap between the ideal problem setting and real-world applications since most of the existing person Re-ID methods assume perfect pedestrian detections. Nevertheless, the manually cropped bounding boxes are unavailable in practical applications. Pedestrian detectors inevitably produce false alarms, misdetections, and misalignments, which could harm the final searching performance significantly. To close the gap, Zhang et al. [51] detected persons in photo albums and recognition-specific person using face and global signatures. However, the settings in this research are not typical for person re-ID where pedestrians are observed by surveillance cameras and faces are not clear enough. Xiao et al. [44] proposed a new deep learning method for person search to joint pedestrian detection and person Re-ID. The difference from these existing works, we focus on study whether the proposed half-precision CS framework is efficient in terms of end-to-end person Re-ID, and how to use it in real end-to-end person Re-ID applications.

2.2 CS sampling and reconstruction

Existing CS reconstruction approaches are usually classified into two categories: iterative optimization-based CS methods (e.g. D-AMP [29], IWR [11]) and deep network-based CS methods (e.g. SDA [32], DeepInverse [31], ReconNet [23], CSNet [37], ISTA-Net [48]).

D-AMP [29] is extended from the approximate message passing (AMP) framework by integrating multiple types of

denoisers within its iterations. To improve the CS reconstruction quality of the block-based CS of video through a weighting process, Dinh et al. [11] designed a weighting process to limit the solution space of the recovered signal, and combined the weighting process with simplified Landweber iterations to form an iterative weighted recovery (IWR) algorithm. Although there are theoretical convergence guarantees for these iterative optimized-based CS methods, potentially available training data are not fully utilized by these methods.

Inspired by the powerful learning ability of DNNs, several DNN-based CS reconstruction methods have been proposed to learn the inverse mapping from the CS measurement domain to the original signal domain. SDA [32] is a stacked denoising auto-encoder that recovers images from CS measurements. It consists of fully connected layers, which means a larger network when the signal size grows. This imposes a large computational complexity and leads to overfitting. DeepInverse [31] uses fully convolutional layers to build the DNN model for reconstructing CS images. ReconNet [23] uses fully connected layers and convolutional layers to create the DNN model for regressing an image block from its CS measurement. CSNet [37] applies a neural network to train a sampling matrix rather than uses a manual-designed one. Then, it uses a DNN-based method to reconstruct images, and can get the improved quality by achieving an optimal signal recovery. ISTA-Net [48] casts the iterative shrinkage-thresholding algorithm (ISTA) into DNN form, and solves the proximal mapping associated with the sparsity-inducing regularizer. These DNN-based methods are the data-driven methods that use no hand-designed models. The training dataset and test dataset are provided for learning the structure within the data, such that they can compete with state-of-the-art methods. Especially, all the parameters in ISTA-Net are learned end-to-end, rather than being hand-crafted. ISTA-Net can reduce the network complexity and the training time while ensures a good reconstruction quality. Unfortunately, they need to be trained for specific random measurement matrices and noise level, which are not well designed for CS reconstruction.

2.3 Half-precision floats

Half-precision floats are introduced in the IEEE 754-2008 standard. They have a smaller range and lower precision than 32-bit single-precision floats and consist of 1 sign bit, 5 bits of exponent, and 10 fractional bits. They are intended for storing floating-point values in applications where lower precision is sufficient for performing arithmetic computations. The arithmetic with half-precision floats is faster than corresponding one with single-precision floats, if the hardware devices natively support half-precision floats routines.

Although most of the scientific calculation requires 32-bit single or 64-bit double precision floats, artificial intelligence approaches can perform with 16-bit half-precision floats. To decrease the consumption of memory and reduce the time taken by the training and inference, several research works explore half-precision representation for parameters of DNN models. Their results show that both the training and inference of DNN models can be efficiently performed with lower precision, using 16-bit multipliers for training and inference with minimal even no loss in accuracy. For instance, Micikevicius et al. [30] propose a mixed precision training approach to train DNNs using half-precision floats and get a significant computation speedup while their accuracy has no significant loss.

Half-precision allows significantly more programs data to reside in the same caches, and the data can be moved faster through the memory hierarchy to maximize compute resources. Thus, half-precision floats are suitable for better usage of cache and reduction in bandwidth bottlenecks for operations like matrix multiplications. Therefore, half-precision floats are especially suitable to perform CS sampling and store its results.

3 Architecture of HCS4ReID

A half-precision CS framework with two measurement rates (MRs) named HCS4ReID is proposed for end-to-end person Re-ID. It aims at reducing the required network bandwidth and storage space in the case of both accelerating the process of CS sampling and preserving the accuracy of end-to-end person Re-ID. It starts from raw video frames captured by cameras or drones, then the raw video frames are processed using half-precision CS sampling and reconstruction with two MRs to recovery video frames. Finally, a gallery of pedestrian images is automatically created with YOLOv3 [22]. Given a specific person-of-interest image, person Re-ID algorithms are used to match it with person images in the generated gallery. Figure 1 shows the architecture of HCS4ReID, which consists of three parts: (1) half-precision video frames CS sampling with two MRs, (2) CS reconstruction of video frames, and (3) end-to-end person Re-ID, which are illustrated in detail as follows.

3.1 Half-precision CS sampling

Since the block-based CS reconstruction scheme is capable to facilitate the low-cost sampling and recovery of images and videos, the proposed framework HCS4ReID applies block-based CS sampling and reconstruction methods. In

real-world applications related to person Re-ID, the video frames acquired by various cameras are natural images, which consist of C channels such as red (R), green (G) and blue (B) channels. The CS sampling of a natural image consists of three independent CS processes on each channel, i.e.

$$\mathbf{y}_i = \Phi \mathbf{x}_i, i \in \{R, G, B\},$$

where $\mathbf{y}_i \in \mathbb{R}^M$ denotes the CS measurements on the channel i, $\Phi \in \mathbb{R}^{M \times N}$ is the measurement matrix for each channel, $\mathbf{x}_i \in \mathbb{R}^N$ denotes the vectorized version of an image block in the channel i which is normalized to the range 0–1. As $M \leq N$, the measurement rate is defined as $\frac{M}{N}$. For a given measurement rate, the corresponding measurement matrix Φ is constructed by generating a random Gaussian matrix and then orthogonalizing its rows, i.e. $\Phi\Phi^T = I$, where I is the identity matrix.

3.1.1 Identify person-of-interest image blocks

First, a pedestrian detector (e.g. YOLOv3 [22]) is used to get the bounding boxes of pedestrians in raw video frames, and zero-padding is utilized to keep the CS sampled image blocks constant in each channel. Second, each channel $i \in \{R, G, B\}$ of captured video frames is divided into 33×33 image blocks with no overlap. Given a $W \times H$ video frame, bounding boxes of n pedestrians in the frame are denoted as the lists of $b_j = [x_j, y_j, w_j, h_j], j = 0, 1, \ldots, n$, where x_j and y_j denote the coordinates of the top-left corner of the jth predicted bounding box in the frame, w_j and h_j denote the width and height of the jth predicted bounding box. To denote the specific image blocks overlapping with the predicted pedestrian bounding boxes, a matrix $\mathbf{P}$ is introduced, i.e.

$$\mathbf{P} = \{p_{st}\}, s \in [0, (W-1)/33], t \in [0, (H-1)/33], \quad (1)$$

where p_{st} denote the identifiers of 33×33 image blocks for the given frame. $p_{st} = 1$ denotes the image block p_{st} and one of pedestrian bounding boxes b_j contains the same pixels, and $p_{st} = 0$ denotes the opposite case. Thus, it generates two matrices $\mathbf{x}_{ip} \in \mathbb{R}^{1089 \times M_p}$ and $\mathbf{x}_{ib} \in \mathbb{R}^{1089 \times M_b}$ of image regions in the i-th channel by concatenating vectorized image block size of 1089×1, where M_p and M_b denote the number of image blocks in the predicted pedestrian region and the rest regions of the given video frame, respectively. Thus, the CS sampling on each channel performs as two large matrix multiplications instead of several small matrix multiplications, and this can accelerate the speed of CS sampling of the proposed HCS4ReID.

3.1.2 Half-precision CS measurement matrices

Half-precision is a useful data format for storing floating-point numbers since it requires half of the storage space and the memory bandwidth. On the other hand, hardware devices may enable higher operations per second at half-precision since these arithmetic operations require less silicon area and power than single-precision ones.

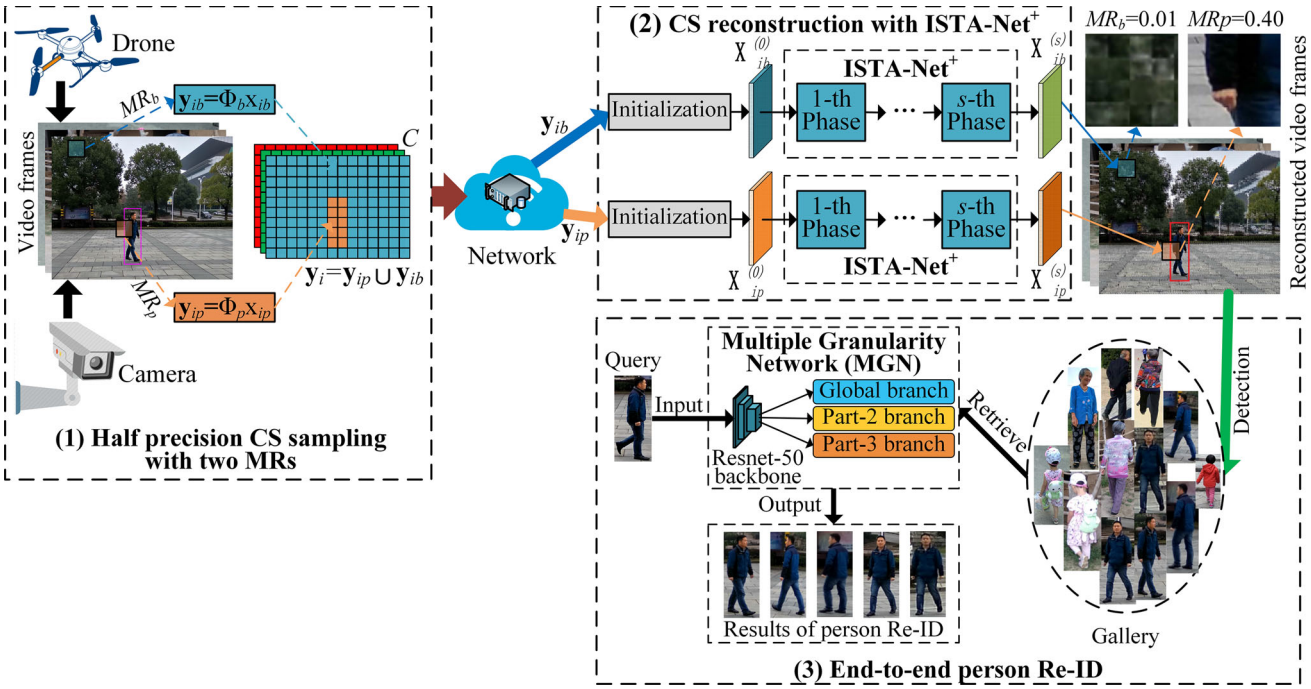

Fig. 1 Half-precision CS framework with two measurement rates for end-to-end person Re-ID

Nevertheless, for existing CS methods, values of normalized image vectors and measurement matrices are presented with 32-bit single-precision floats or 64-bit double precision floats. Meanwhile, for the block-based CS sampling $\mathbf{y}_i = \Phi\mathbf{x}_i$, the main computation is the element-wise product of the measurement matrix Φ and the normalized image vector $\mathbf{x}_i$. It is the memory-bandwidth limited and computationally intensive arithmetic operations. Therefore, the proposed framework HCS4ReID utilizes half-precision floats to represent and store the values of measurement matrices and normalized image vectors. This is capable to accelerate related computations like addition and multiplication by reducing memory reads and writes which usually consume a lot of runtime. Thus, CS sampling is performed with half-precision floats and is accelerated when the given hardware devices that natively support the half-precision arithmetic operations. Meanwhile, the output measurement values are half-precision floats, which are suitable for reducing the required network bandwidth and storage space.

3.1.3 CS sampling with two half-precision MRs

For existing CS methods, the benefits of CS sampling disappear in terms of reducing the cost of data transmission over networks and data storage on servers when the measurement rate grows. Besides, for person Re-ID, the reconstruction quality of the person-of-interest regions is more critical than the rest regions in video frames. Therefore, HCS4ReID employs two different MRs to CS sampling of raw RGB video frames. As shown in Step (1) of Fig. 1, for each channel of video frames, it uses the higher measurement rate MR_p to sample the image blocks in the person-of-interest regions that are identified with $p_{st} = 1$, while it uses a lower one MR_b to sample the rest image blocks in the given video frames that are identified with $p_{st} = 0$.

Thus, HCS4ReID generates two half-precision CS measurements $\mathbf{y}_{ip}$ and $\mathbf{y}_{ib}$ in the i-th channel, i.e.

$$\begin{cases} \mathbf{y}_{ip} = \Phi_p\mathbf{x}_{ip}, & \text{if } p_{st} = 1, \\ \mathbf{y}_{ib} = \Phi_b\mathbf{x}_{ib}, & \text{if } p_{st} = 0, \end{cases} \tag{2}$$

where $\Phi_p \in R^{M_p \times 1089}$ denotes the half-precision measurement matrix corresponding to the measurement ratios MR_p, which is used in image blocks belong to the person-of-interest regions. $\Phi_b \in R^{M_b \times 1089}$ denotes the half-precision measurement matrix corresponding to the measurement ratios MR_b, which is used in rest image blocks. The size of $\mathbf{y}_{ip}$ is $M_p \times num_p$ while the size of $\mathbf{y}_{ib}$ is $M_b \times num_b$, where num_p and num_b are the number of image blocks in the person-of-interest regions and the number of rest image blocks in a given video frame, respectively. Combining two matrices $\mathbf{y}_{ip}$ and $\mathbf{y}_{ib}$, HCS4ReID gets the results of half-precision CS measurement denoted by $\mathbf{y}_i = \mathbf{y}_{ip} \cup \mathbf{y}_{ib}$ for each channel (1), which is transmitted to remote servers for reconstructing CS images/videos.

3.2 CS reconstruction with two MRs

As shown in Step (2) of Fig. 1, the CS reconstruction network can be one of block-based image CS reconstruction networks like ISTA-Net$^+$ [48], ReconNet [23] and CSNet [37]. ISTA-Net is a state-of-the-art one that makes full use of the merits of both optimized-based and deep network-based CS methods, its enhanced version ISTA-Net$^+$ is composed of s phases, and each phase corresponds to one iteration in ISTA. Thus, HCS4ReID adopts its enhanced version ISTA-Net$^+$ to reconstruct CS video frames.

When the servers received CS measurements $\mathbf{y}_i = \mathbf{y}_{ip} \cup \mathbf{y}_{ib}, i \in \{R, G, B\}$ represented by half-precision floats from the CS sampling terminals, the CS measurements are first converted into single-precision floats since most high-performance compute devices do not support half-precision arithmetic at present.

Second, two CS reconstruction networks ISTA-Net$^+$ take $\mathbf{y}_{ip}$ and $\mathbf{y}_{ib}$ as their inputs respectively, and compute the corresponding initializations $\mathbf{x}^{(p0)}$ and $\mathbf{x}^{(b0)}$ with Eq. 3.

$$\begin{cases} \mathbf{x}^{(p0)} = \mathbf{X}_{ip}\mathbf{Y}_{ip}^T(\mathbf{Y}_{ip}\mathbf{Y}_{ip})^{-1}\mathbf{y}_{ip}, \\ \mathbf{x}^{(b0)} = \mathbf{X}_{ib}\mathbf{Y}_{ib}^T(\mathbf{Y}_{ib}\mathbf{Y}_{ib})^{-1}\mathbf{y}_{ib}, \end{cases} \tag{3}$$

where $\mathbf{x}^{(p0)}$ is the initialization for the input CS measurement $\mathbf{y}_{ip}$, and $\mathbf{x}^{(b0)}$ is the initialization for the input CS measurement $\mathbf{y}_{ib}$. $\mathbf{X}_{ip} = [\mathbf{x}_{ip}^1, \ldots, \mathbf{x}_{ip}^s]$ denotes input image blocks in the regions of detected pedestrians, and $\mathbf{Y}_{ip} = [\mathbf{y}_{ip}^1, \ldots, \mathbf{y}_{ip}^s]$ denotes their corresponding CS measurements. By contrast, $\mathbf{X}_{ib} = [\mathbf{x}_{ib}^1, \ldots, \mathbf{x}_{ib}^s]$ indicates input image blocks in the rest regions, and $\mathbf{Y}_{ib} = [\mathbf{y}_{ib}^1, \ldots, \mathbf{y}_{ib}^s]$ indicates their corresponding CS measurements.

Third, the initializations $\mathbf{x}^{(p0)}$ and $\mathbf{x}^{(b0)}$ are inferred with the s phases of iterations, which are shown in Step (3) of Fig. 1. Its kth ISTA iteration is cast into two separate modules, named $\mathbf{r}^{(k)}$ and $\mathbf{x}^{(k)}$ separately. The former module aims to generate the immediate reconstruction results $\mathbf{r}^{(k)}$ of the k-th phase with the input $\mathbf{x}^{(k-1)}$, while the latter one aims to compute $\mathbf{x}^{(k)}$ according to the input $\mathbf{r}^{(k)}$. When two MRs are used in the proposed CS4ReID, there are double $\mathbf{r}^{(k)}$ and double $\mathbf{x}^{(k)}$ in the k-th phases, i.e. $\mathbf{r}_{ip}^{(k)}$, $\mathbf{x}_{ip}^{(k)}$, $\mathbf{r}_{ib}^{(k)}$ and $\mathbf{x}_{ib}^{(k)}$. Their values are updated according to Eqs. 4 and 5, respectively.

$$\begin{cases} \mathbf{r}_{ip}^{(k)} = \mathbf{x}_{ip}^{(k-1)} - \rho \Phi_p^T (\Phi_p \mathbf{x}_{ip}^{(k-1)} - \mathbf{y}_{ip}), \\ \mathbf{x}_{ip}^{(k)} = \text{argmin}_{\mathbf{x}_{ip}} \frac{1}{2} ||\mathbf{x}_{ip} - \mathbf{r}_{ip}^{(k)}||_2^2 + \lambda ||F(\mathbf{x}_{ip})||_1, \end{cases} \tag{4}$$

$$\begin{cases} \mathbf{r}_{ib}^{(k)} = \mathbf{x}_{ib}^{(k-1)} - \rho \Phi_b^T (\Phi_b \mathbf{x}_{ib}^{(k-1)} - \mathbf{y}_{ib}), \\ \mathbf{x}_{ib}^{(k)} = \text{argmin}_{\mathbf{x}_{ib}} \frac{1}{2} ||\mathbf{x}_{ib} - \mathbf{r}_{ib}^{(k)}||_2^2 + \lambda ||F(\mathbf{x}_{ib})||_1, \end{cases} \tag{5}$$

where k is the ISTA iteration index, ρ denotes the pre-defined step size, and λ is the pre-defined regularization parameter in the ISTA. ρ and λ do not change with k, but they can be tuned. $F(\cdot)$ denotes a combination of two linear convolution operators (without bias terms) separated by a rectified linear unit (ReLU). It is formulated in matrix form as $F(\mathbf{x}) = \mathbf{B}ReLU(\mathbf{A}\mathbf{x})$, where $\mathbf{A}$ and $\mathbf{B}$ correspond to the two convolution operators [48]. By performing the CS construction with the above s phases, Step (2) in Fig. 1 outputs the reconstructed blocks $\mathbf{x}_{ip}^{(s)}$ and $\mathbf{x}_{ib}^{(s)}$, corresponding to the input CS measurements $\mathbf{y}_{ip}$ and $\mathbf{y}_{ib}$.

Finally, the reconstructed blocks compose an intermediate reconstructed image for each channel, and three channels are merged into a natural video frame as the final output of CS reconstruction with ISTA-Net$^+$ in Step (2) of the proposed CS4ReID.

3.3 End-to-end person Re-ID

End-to-end person Re-ID in Step (3) of CS4ReID contains two phrases: generation of a gallery of person images and re-identification of the given queried person in the generated gallery.

3.3.1 Generation of gallery

Different pedestrian detectors produce galleries of different sizes, and a good detector is more likely to recall all person-of-interest images and get high precision. The number of detected pedestrians per image also affects the accuracy of end-to-end person re-ID. When too few pedestrians are detected in a video frame, it is highly possible that the person-of-interest bounding boxes are not detected, so the performance of end-to-end person Re-ID is compromised. By contrast, distractors may have a negative influence on end-to-end person Re-ID when there are too many false positive detected pedestrians are detected, thus the accuracy of overall end-to-end person Re-ID slowly decreases as the number of pedestrians per video frame increases.

YOLOv3 [22] is a faster detector than other detectors with relatively high APs performance by using the new multi-scale predictions. There are significant benefits over other detection methods in terms of accuracy and speed of the inference. Thus, to detect pedestrians in reconstructed video frames, YOLOv3 is utilized to automatically get the locations of pedestrians. Then, a set of pedestrian images can be obtained by cropping the regions occupied by detected pedestrians, and a gallery of automatically detected pedestrian images can be generated for person Re-ID.

3.3.2 Person Re-ID with MGN

Multiple granularities network (MGN) [41] is a multi-branch deep network for person Re-ID. Its backbone is ResNet-50 [20] for achieving competitive performances in existing person Re-ID algorithms [38]. It divides the subsequent part after *res_conv4_1* block into three independent branches, i.e. one Global Branch for learning global feature representations without any partition information, and two local branches (Part-2 Branch and Part-3 Branch) for learning local feature representations. To obtain the most powerful pedestrian discrimination, global and multi-granularity local features are concatenated as the output feature representation of a pedestrian image. Thus, MGN is more efficient and robust to scenarios with large variances than other person Re-ID methods. It utilizes softmax loss for classification, and triplet loss for metric learning as loss functions in the training phase. During the testing phase, the generated gallery and the given queried person image are inputs of MGN, the output results are the re-identified pedestrian images as shown in Step (3) of Fig. 1.

3.4 Algorithmic description of CS4ReID

In summary, the proposed framework performs as described in Fig. 2. Its three steps correspond to the three parts in the proposed half-precision CS framework CS4ReID discussed in above, respectively.

4 Experimental results

4.1 Dataset

To evaluate the performance of the proposed framework HCS4ReID, a large-scale person Re-ID dataset is needed. Most of person Re-ID datasets (e.g. CUHK03 [27], Market1501 [57] and MARS [56]) only contain pre-cropped pedestrian bounding boxes images without the complete video frames, but our experimental evaluation needs a person Re-ID dataset which contains whole scene images related to pedestrians. The person re-identification in the wild (PRW) dataset [58] is a large-scale dataset with comprehensive baselines for pedestrian detection and person Re-ID in raw video frames, acquiring with six cameras

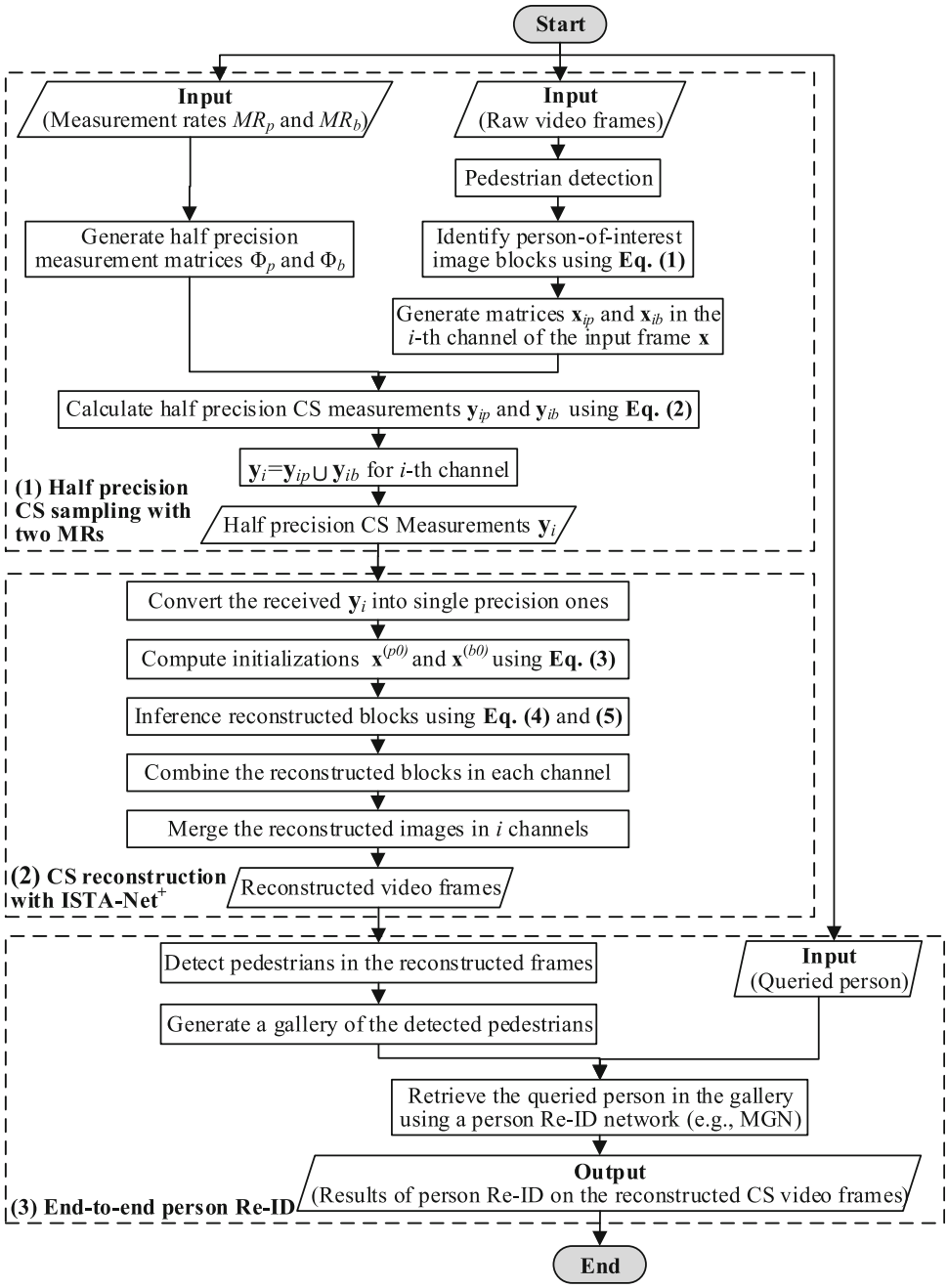

Fig. 2 Algorithmic diagram for the proposed CS4ReID

in the summer of 2014. It contains 11,816 raw video frames in which pedestrians are annotated with their bounding box positions and identities for evaluating both pedestrian detection and person re-ID, and annotated 932 identities belong to 34,304 pedestrian bounding boxes. It is divided into a training set with 5704 raw video frames and 482 pedestrian IDs, and a testing set with 6112 raw video frames and 450 pedestrian IDs. Thus, the PRW dataset is suitable for the performance evaluation of end-to-end person Re-ID methods.

4.2 Implementation approaches

The implementation of the proposed framework HCS4ReID contains half-precision CS sampling with two MRs, CS reconstruction with two MRs, and end-to-end person Re-ID, which respectively corresponds to the three steps shown in Fig. 1.

First, the pedestrian detector for CS sampling with two MRs and end-to-end person Re-ID is implemented with the state-of-the-art object detector YOLOv3 [22], which is named YOLOv3-PRW. It uses the 416×416 sized

YOLOv3 model and is retrained on the PRW dataset with the OpenCL-based deep neural network framework UHCL-Darknet [25, 28, 49].

Second, half-precision CS sampling is implemented with OpenCL [5, 45, 52] and C to compute CS measurements on an OpenCL device, i.e. Intel HD Graphics 530. The computations of two CS measurements in each channel shown in Eq. 2 are mainly general matrix multiply (GEMM) routines. They are implemented with half-precision GEMM kernels, which are tuned on the OpenCL device according to the size of measurement matrices (i.e. Φ_p and Φ_b) and normalized image matrices (i.e. $\mathbf{x}_{ip}$ and $\mathbf{x}_{ib}$) [13, 33, 53]. On the other hand, the corresponding CS reconstruction of video frames is implemented based on ISTA-Net$^+$ with Tensorflow. ISTA-Net$^+$ with $s = 9$ phases and a batch size of 64 is trained for a range of measurement rates $\{0.01, 0.04, 0.10, 0.25, 0.40, 0.50\}$ respectively, using the same set of 91 images in [23] as the training dataset.

Third, the MGN for person Re-ID is implemented with PyTorch, which is trained on the PRW dataset. To capture more detailed information from pedestrian images, input images are resized to 384×128. The weights of ResNet-50 are used to initialize the backbone and branches of MGN. The training of MGN uses the same settings and strategies in [41], e.g. each mini-batch is sampled with randomly selected 16 identities and randomly sampled four images for each identity to cooperate the requirement of triples loss. For the testing, it extracts the features corresponding to original pedestrian images and the horizontally flipped versions, and uses the average of these extracted features as the final features.

The experiments are performed on a workstation that consists of an Intel Core i7-6700 CPU (integrating an Intel HD Graphics 530), a NVIDIA GTX 1080Ti GPU and a 8 GB DDR4 memory, and runs Ubuntu 16.04.5 with the GCC compiler 5.3.2 and Intel OpenCL driver 5.0.

4.3 Evaluation protocols

First, average precision (AP) under intersection over union (IoU) > 0.5 is usually used to evaluate pedestrian detection [34, 42, 54]. However, Zheng et al. [58] found that IoU > 0.7 is more effective than IoU > 0.5 for measuring influences of detected pedestrians on the accuracy of person Re-ID, since the localization ability of pedestrian detectors affect the performance of end-to-end person Re-ID. Thus, AP under IoU > 0.7 is utilized to evaluate the performance of trained pedestrian detector YOLOv3-PRW.

Second, the speedup over the corresponding single-precision CS sampling and mean compression ratio (mCR) on reconstructed video frames are used to measure the performance of the proposed half-precision CS sampling method. They aim to evaluate the influence on accelerating CS sampling and reducing the overheads of required network transmission and storage space. The speedups are measured with the time taken by half-precision CS sampling over the time taken by corresponding single-precision CS sampling on the same hardware device (e.g. Intel HD Graphics 530 in the Intel Core i7-6700 CPU). The mCRs are calculated by the size of RGB video frames against the size of the corresponding CS measurements generated by HCS4ReID.

Third, mean peak signal-to-noise ratio (mPSNR) is used to measure the reconstruction quality of the proposed framework HCS4ReID. It is calculated with the average of PSNR in terms of the reconstructed video frames against the corresponding original video frames coming from the gallery of the PRW dataset.

Finally, as PSNR mainly indicates the quality of CS reconstruction in terms of a whole image using the same measurement rate, and is limited in evaluating the quality of CS reconstruction when two different measurement rates are concurrently used in single image reconstruction. Therefore, following [57], mean average precision (mAP) and the rank-1, 3, 5, 10 accuracies are used to evaluate the performance of person Re-ID, which indicate the applicability and efficiency of the proposed HCS4ReID in terms of end-to-end person Re-ID. The mAP is the mean across all queries' average precision to determine the correctness of detected bounding boxes, where the average precision is calculated for each query based on the precision-recall curve. The rank-1, 3, 5, 10 accuracies denote the possibility to locate at least one true positive retrieve result in the top-1, 3, 5, 10 ranks respectively, where a true positive retrieve result is a matching, the queried person bounding box overlaps with the annotated person ground truths in test dataset with IoU which is greater or equal to 0.5.

4.4 Evaluation results

The performance of half-precision CS sampling with two MRs, the quality of corresponding CS reconstruction and APs of pedestrian detection on the reconstructed testing set of the PRW dataset are shown in Table 1, where $AP^{0.7}$ denotes AP under IoU > 0.7 of pedestrian detection on reconstructed video frames. Apart from mPSNR and APs used to evaluate the quality of reconstructed video frames, the performance of person Re-ID on the CS reconstructed video frames is shown in Table 2. As the person-of-interest regions are more critical than the rest regions for video frames in real applications like end-to-end person Re-ID, $MR_p \geq MR_b$ is used in this experiments. $MR_p = MR_b$ is a special case that HCS4ReID is used as the traditional CS

methods, which suggests that HCS4ReID only uses single MR in CS sampling and reconstruction.

4.4.1 Performance of half-precision CS sampling

As shown in Tables 1 and 2, the proposed half-precision CS framework HCS4ReID gets the nearly same performance with the traditional single-precision CS, e.g. they get the relatively high mPSNR for CS reconstruction, similar AP for pedestrian detection, comparable mAP and Rank-1, 3, 5, 10 accuracies for person Re-ID on reconstructed video frames of the PRW dataset. This suggests that half-precision CS sampling is sufficient for CS reconstruction and end-to-end person Re-ID on the reconstructed video frames.

Especially, the half-precision CS sampling achieves $1.37\times$ $-1.96\times$ speedups against the corresponding single-precision CS sampling, e.g. the time of CS sampling with half-precision floats reduces from 11.26 ms/frame to 7.87 ms/frame when $\mathrm{MR}_p/\mathrm{MR}_b = 0.40/0.01$. Thus, the average speedup of $1.55\times$ is provided by the half-precision CS sampling over the corresponding single-precision one. This denotes that the proposed half-precision CS sampling not just provides the comparable reconstruction quality with the corresponding single-precision CS sampling, but also accelerates the CS sampling compared with the single-precision one. The reason of this is that half-precision floats have inherent advantages over 32-bit single-precision floats: (1) they are half the size and fit into a lower level of cache with lower latency for accessing memory. (2) They take up half the cache space, which frees up cache space for other related data in a running program. (3) They require half the memory bandwidth, which frees up the bandwidth for other operations in the given program.

On the other hand, HCS4ReID saves approximately a half of network bandwidth and storage space against the corresponding single-precision CS sampling. For instance, the mCR of CS measurements sampled with the half-precision floats is 12.31 while the counterpart sampled with the single-precision floats is 6.16 when $\mathrm{MR}_p/\mathrm{MR}_b = 0.10/0.10$. The reason for this is that half-precision floats only require half the storage space and memory bandwidth of the single-precision floats.

4.4.2 Performance of CS reconstruction with two MRs

The construction quality of video frames is improved as the used measurement rates increase, and the size of CS measurement data also significantly increases, i.e. the corresponding mCR significantly decreases. For instance, the size of CS measurements increases 54.32 times on average when MR_p and MR_b grow from 0.01 to 0.50.

Table 1 Performance of half-precision CS sampling and reconstruction

MR_p/MR_b	Half-precision CS			Single-precision CS			Speedup
	mCR	mPSNR (dB)	$\mathrm{AP}^{0.7}(\%)$	mCR	mPSNR (dB)	$\mathrm{AP}^{0.7}$ (%)	
0.50/0.50	0.97	41.10	71.45	0.48	41.10	71.45	1.90 ×
0.40/0.40	1.21	38.68	71.40	0.60	36.68	71.35	1.96 ×
0.25/0.25	1.94	34.79	70.57	0.97	34.79	70.50	1.66 ×
0.10/0.10	4.86	28.64	67.06	2.43	28.64	67.02	1.44 ×
0.04/0.04	12.31	23.93	52.53	6.16	23.93	52.35	1.58 ×
0.01/0.01	52.70	20.53	14.44	26.43	20.53	14.37	1.43 ×
0.50/0.40	1.19	38.90	71.46	0.60	38.90	71.42	1.88 ×
0.50/0.25	1.80	35.20	71.16	0.91	35.20	71.18	1.40 ×
0.50/0.10	3.79	29.13	70.17	1.90	29.13	70.20	1.40 ×
0.50/0.04	7.12	24.46	68.38	3.57	24.46	68.22	1.61 ×
0.50/0.01	13.92	21.13	65.62	7.03	21.13	65.62	1.56 ×
0.40/0.25	1.86	35.09	71.19	0.93	35.10	71.16	1.47 ×
0.40/0.10	3.99	29.11	70.14	2.00	29.11	70.14	1.55 ×
0.40/0.04	7.75	24.45	68.30	3.89	24.45	68.30	1.69 ×
0.40/0.01	16.07	21.13	65.62	8.11	21.13	65.62	1.43 ×
0.25/0.10	4.36	29.02	69.76	2.19	29.02	69.78	1.50 ×
0.25/0.04	9.01	24.42	67.84	4.52	24.42	67.83	1.46 ×
0.25/0.01	21.14	21.12	65.46	10.67	21.11	65.49	1.38 ×
0.10/0.04	10.99	24.28	64.88	5.52	24.28	64.93	1.37 ×
0.10/0.01	31.97	21.05	62.68	16.16	21.05	62.68	1.37 ×
0.04/0.01	41.93	20.88	50.14	21.23	20.88	50.00	1.42 ×

Table 2 Performance of HCS4ReID for end-to-end person Re-ID

MR_p/MR_b	Half-precision CS					Single-precision CS				
	mAP (%)	Rank (R) (%)				mAP (%)	Rank (R) (%)			
		R-1	R-3	R-5	R-10		R-1	R-3	R-5	R-10
0.50/0.50	69.27	84.54	90.67	92.76	95.28	69.24	84.88	90.86	92.85	95.08
0.40/0.40	69.07	84.92	90.56	92.70	95.23	69.05	84.34	90.90	92.51	95.23
0.25/0.25	68.42	84.68	90.42	92.90	95.09	68.41	84.68	90.32	92.56	94.94
0.10/0.10	64.34	83.74	90.36	92.65	95.08	64.27	83.84	90.41	92.65	94.89
0.04/0.04	54.51	80.90	87.67	90.40	93.57	54.46	80.70	87.27	90.55	93.57
0.01/0.01	38.10	52.11	63.77	69.72	76.74	38.17	52.27	64.73	68.58	77.59
0.50/0.40	69.31	84.49	90.67	92.71	95.14	69.29	84.69	90.71	92.95	94.99
0.50/0.25	69.30	84.34	90.76	92.70	95.04	69.28	84.49	90.91	92.76	95.04
0.50/0.10	69.52	84.39	90.71	92.81	94.99	69.48	84.39	90.76	92.76	95.09
0.50/0.04	69.47	84.48	90.71	92.90	94.99	69.48	84.18	90.80	93.09	94.94
0.50/0.01	69.87	84.67	90.75	92.99	95.18	69.54	84.39	90.76	92.76	95.09
0.40/0.25	69.13	84.29	90.56	92.61	94.94	69.13	84.34	90.47	92.80	95.04
0.40/0.10	69.29	84.14	90.52	92.70	95.04	69.34	84.39	90.52	92.76	95.19
0.40/0.04	69.32	84.53	90.41	92.75	94.99	69.27	84.14	90.61	92.99	95.09
0.40/0.01	69.72	84.67	90.61	92.94	95.43	69.71	84.78	90.71	93.00	95.18
0.25/0.10	68.61	84.81	90.51	93.14	95.03	68.60	84.76	90.31	92.89	95.03
0.25/0.04	68.50	84.81	90.75	93.18	95.08	68.54	84.86	90.65	92.99	94.94
0.25/0.01	68.97	84.62	90.85	92.99	95.13	69.00	94.87	90.95	93.04	95.18
0.10/0.04	64.56	84.13	90.80	92.50	95.03	64.52	84.03	90.56	92.50	95.13
0.10/0.01	64.90	84.47	90.56	92.16	95.08	64.87	83.98	90.51	92.60	95.18
0.04/0.01	55.07	79.82	87.04	89.81	93.71	55.06	79.68	87.43	90.01	93.47

However, when MR_p and MR_b decrease to 0.01, the reconstruction quality is too low compared with the original video frames, and thus the reconstructed video frames are less useful for applications in real scenarios (e.g. end-to-end person Re-ID). The AP is only 14.37 on the reconstructed video frames, and the AP only reaches 20.07% of the AP obtained on the corresponding raw video frames. Meanwhile, the mAP and Rank-1, 3, 5, 10 accuracies are only 54.95%, 61.56%, 70.44%, 75.16%, 80.74% of the corresponding ones obtained on the raw video frames for end-to-end person Re-ID.

The proposed HCS4ReID supports CS sampling and reconstruction with two different MRs, and it uses higher measurement rate in the person-of-interest regions than the rest regions in the same video frame, i.e. $MR_p > MR_b$, to sample and reconstruct video frames. Thus, for the same MR_b in HCS4ReID, APs significantly increase with the growth of MR_p, while the corresponding mPSNRs slightly increase. For example, the mPSNR and AP are 20.53 dB and 14.44% when $MR_p = 0.01$ and $MR_b = 0.01$, while they increase to 21.13 dB and 65.62% when $MR_p = 0.40$ and $MR_b = 0.01$. On the other hand, for the same MR_p, the mPSNR increases more significantly than the AP as MR_b grows. For instance, when $MR_p = 0.40$, the mPSNR increases from 21.13 to 35.09 dB while the AP only increases from 65.62 to 71.19% as MR_b increase from 0.01 to 0.25. This means that the increase of MR_p is more efficient than the increase of MR_b for improving the performance of pedestrian detection when $MR_p > MR_b$, since the construction quality of the person-of-interest regions has more influences than the rest regions on the accuracy of pedestrian detection in end-to-end person Re-ID.

As shown in Table 2, MR_p nearly determines the performance of end-to-end person Re-ID, while MR_b has few influences on the performance of end-to-end person Re-ID. Especially, the results obtained with $MR_p > MR_b$ approximately equal to or even better than the corresponding results obtained when MR_b increases to MR_p. For instance, the mAP and Rank-1, 3, 5, 10 accuracies are 68.97%, 84.62%, 90.85%, 92.99%, 95.13% respectively when $MR_p = 0.25$ and $MR_b = 0.01$. Most of them are higher than the corresponding ones when $MR_p = 0.25$ and $MR_b = 0.25$. Besides, for the same MR_p, there is only a slight improvement on the performance of end-to-end person Re-ID when MR_b increases. This suggests that the minimum of MR_b (e.g. 0.01) is enough for end-to-end person Re-ID, while MR_p can be determined according to the size of available network bandwidth and storage space in real scenarios.

Therefore, the proposed HCS4ReID accelerates the CS sampling with half-precision floats and significantly improves mCR while mPSNR and AP decrease slightly. Besides, it is sufficient for CS sampling and reconstruction with the lowest MR_b and higher MR_p to achieve comparable performance, which is obtained with MR_b and MR_p equalling to the selected MR_p. That is, the construction quality of the person-of-interest regions is more important for determining the performance of end-to-end person Re-ID. Interestingly, when $MR_p = 0.25$ and $MR_p = 0.01$, HCS4ReID reaches the comparable performance of end-to-end person Re-ID on the reconstructed PRW dataset, while the mCR reaches 21.14.

4.4.3 Comparison to state-of-the-art CS methods

Table 3 compares the performance of HCS4ReID with three state-of-the-art DNN-based CS algorithms (e.g. DeepInverse [31], ReconNet [23] and CSNet [37]) for end-to-end person Re-ID. For a fair comparison, all compared methods are trained on the same dataset consisting of 91 images [23] with the default parameter settings. HCS4ReID uses 0.01 as the low measurement rate MR_b to obtain the highest mCR for various MR_p, while all compared methods use the same measurement rates $MR_p = MR_b$ since the conventional CS approaches only support a single measurement rate.

The mCR of HCS4ReID significantly improves by 4.11–30.26 times compared to the mCR of three compared CS approaches when they use the same MR_p. Especially, the mCR of HCS4ReID achieves 13.92 while the mCRs of the compared methods only achieves 0.46–0.47 when $MR_p = 0.5$. HCS4ReID also achieves the mCR of 41.93 when $MR_p = 0.04$ compared to the mCR of 4.49 obtained by the method DeepInverse.

For the average PSNR reconstruction performance of entire video frames, HCS4ReID obtains lower mPSNR than other state-of-the-art DNN-based CS approaches. For the pedestrian detection performance, HCS4ReID also obtains lower APs than the ones obtained by ReconNet and CSNet when MR_p is 0.50, 0.40, 0.25 and 0.10. Especially, CSNet nearly obtains the highest mPSNR and AP since it uses a convolution layer to implement CS sampling instead of performing CS sampling based on a random Gaussian matrix. The reason for this is that the compared CS algorithms use MR_p to sample the entire video frames, while

Table 3 Performance comparison of DNN-based CS algorithms for end-to-end person Re-ID

MR_p/MR_b	Algorithm	mCR	mPSNR (dB)	$AP^{0.7}$(%)	mAP (%)	Rank (R) (%)			
						R-1	R-3	R-5	R-10
0.50/0.50	DeepInverse [31]	0.47	25.86	64.08	56.82	81.12	87.79	91.00	94.01
0.50/0.50	ReconNet [23]	0.46	31.86	66.78	63.25	81.31	89.37	91.62	94.07
0.50/0.50	CSNet [37]	0.46	35.66	70.15	67.21	84.24	90.61	92.85	95.14
0.50/0.01	HCS4ReID	13.92	21.13	65.62	69.87	84.67	90.75	92.99	95.18
0.40/0.40	DeepInverse	0.59	24.76	59.57	53.46	79.20	86.99	89.72	93.08
0.40/0.40	ReconNet	0.60	31.25	63.19	60.14	81.41	88.65	91.01	93.62
0.40/0.40	CSNet	0.62	34.92	69.87	67.99	84.19	90.71	93.04	95.14
0.40/0.01	HCS4ReID	16.07	21.13	65.62	69.72	84.67	90.61	92.94	95.43
0.25/0.25	DeepInverse	0.92	23.19	59.29	52.32	77.99	85.25	89.00	92.45
0.25/0.25	ReconNet	0.97	29.85	69.68	65.53	84.29	90.81	92.51	94.70
0.25/0.25	CSNet	0.98	34.53	69.86	68.03	84.44	90.95	92.51	95.28
0.25/0.01	HCS4ReID	21.14	21.12	65.46	68.97	84.62	90.85	92.99	95.13
0.10/0.10	DeepInverse	2.12	23.01	45.38	47.98	73.27	82.15	86.15	91.22
0.10/0.10	ReconNet	2.43	25.87	63.17	57.64	81.51	88.66	91.53	93.92
0.10/0.10	CSNet	2.51	27.72	63.16	62.59	84.03	90.17	92.21	94.89
0.10/0.01	HCS4ReID	31.97	21.05	62.68	64.90	84.47	90.56	92.16	95.08
0.04/0.04	DeepInverse	4.49	21.09	30.87	42.47	64.71	75.33	80.13	95.27
0.04/0.04	ReconNet	6.46	23.06	45.75	48.46	74.13	82.82	86.24	90.09
0.04/0.04	CSNet	6.24	26.20	57.14	59.61	82.56	89.53	92.53	94.69
0.01/0.01	DeepInverse	10.20	19.87	12.67	36.78	49.85	63.15	69.45	76.25
0.01/0.01	ReconNet	26.51	20.51	13.33	38.93	53.13	65.66	69.82	76.85
0.01/0.01	CSNet	25.60	23.17	30.15	51.61	74.00	82.94	86.07	90.18
0.04/0.01	HCS4ReID	41.93	20.88	50.14	55.07	79.82	87.04	89.81	93.71

HCS4ReID uses MR_p to sample the person-of-interest regions and utilizes the lower measurement rate (i.e. $MR_b = 0.01$) to sample the other regions of a frame. However, HCS4ReID obtains the highest mAP when MR_p is 0.50, 0.40, 0.25 and 0.10, and also achieves comparable end-to-end person Re-ID performance on the reconstructed frames for the Rank-1, 3, 5, 10 accuracies. The mAP and Rank-1, 3, 5, 10 accuracies of HCS4ReID achieve 55.07%, 79.82%, 87.04%, 89.81% and 93.71% respectively when $MR_p = 0.04$ and $MR_b = 0.01$. They outperform the corresponding performance obtained by the compared CS methods when $MR_p = 0.01$ and $MR_b = 0.01$. Meanwhile, they are higher than the corresponding ones obtained by DeepInverse and ReconNet when $MR_p = 0.04$ and $MR_b = 0.04$, and there is no significant Rank accuracy loss compared to the performance obtained by CSNet when $MR_p = 0.04$ and $MR_b = 0.04$.

Therefore, compared to the state-of-the-art CS algorithms, the proposed HCS4ReID significantly improves mCR and then reduces the required transmission bandwidth and storage space of CS measurements, while preserving the end-to-end person Re-ID performance (e.g. mAp and Rank accuracies) on the reconstructed frames.

5 Conclusion

The proposed CS framework HCS4ReID supports half-precision CS sampling with two measurement rates, which accelerates the process of CS sampling and reduces the overheads of network transmission and storage space for the generated CS measurements. It achieves 1.55 × speedups over the corresponding single-precision CS sampling, and only need half-network bandwidth and storage space. It also supports CS reconstruction with two different measurement rates for end-to-end person Re-ID. Thus, it is more suitable than existing CS approaches for end-to-end person Re-ID on the reconstructed video frames. Besides, HCS4ReID is a scalable and portable CS framework supporting half-precision sampling and reconstruction with two measurement rates. Its mCR achieves 21.14, while its mAP and Rank-1, 3, 5, 10 accuracies achieve 68.97%, 84.62%, 90.85%, 92.99%, 95.13% for end-to-end person Re-ID on the reconstructed video frames, respectively. HCS4ReID can be easily extended to video CS sampling and reconstruction using other block-based CS algorithms and pedestrian detectors. It also can be ported to more OpenCL-accelerated platforms with an increasing number of compute devices with half-precision routines supported.

For the limitation, the proposed CS framework HCS-4ReID is more suitable for the image CS than the video CS. The reason for this is that the proposed CS framework does not take into account the fact that the spatial locations of moving objects are continuous, and these moving objects generally form a region of a surveillance video frame with a stationary background. Then, object tracking methods are more suitable than object detection ones to locate the continuous region of moving objects in video frames during video CS sampling and reconstruction with multiple measurement rates. Therefore, we will explore an approach to sampling and recovering CS surveillance video with multiple measurement rates based on object tracking in future papers. The approach will be utilized in large-scale distributed video surveillance scenarios, such as smart city, intelligent transportation, etc.

Acknowledgements The research was partially funded by the Program of National Natural Science Foundation of China (Grant No. 61751204), the National Outstanding Youth Science Program of National Natural Science Foundation of China (Grant No. 61625202), the International (Regional) Cooperation and Exchange Program of National Natural Science Foundation of China (Grant No. 61661146006), the National Key R&D Program of China (Grant Nos. 2016YFB0201303, 2016YFB0200201), the National Natural Science Foundation of China (Grant Nos. 61772182, 61802032), Science and Technology Plan of Changsha (K1705032). The authors would like to thank Tianming Jin for his help in improving the paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Ahmed E, Jones M, Marks TK (2015) An improved deep learning architecture for person re-identification. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), pp 3908–3916. https://doi.org/10.1109/CVPR.2015.7299016
2. Bashir K, Xiang T, Gong S (2008) Feature selection on gait energy image for human identification. In: 2008 IEEE international conference on acoustics, speech and signal processing, pp 985–988. https://doi.org/10.1109/ICASSP.2008.4517777
3. Chen C, Li K, Teo SG, Chen G, Zou X, Yang X, Vijay RC, Feng J, Zeng Z (2018) Exploiting spatio-temporal correlations with multiple 3d convolutional neural networks for citywide vehicle flow prediction. In: 2018 IEEE international conference on data mining (ICDM), pp 893–898. https://doi.org/10.1109/ICDM.2018.00107
4. Chen J, Fang J, Liu W, Tang T, Yang C (2018) CLMF: a fine-grained and portable alternating least squares algorithm for parallel matrix factorization. Future Gener Comput Syst. https://doi.org/10.1016/j.future.2018.04.071
5. Chen J, Li K, Tang Z, Bilal K, Yu S, Weng C, Li K (2017) A parallel random forest algorithm for big data in a spark cloud computing environment. IEEE Trans Parallel Distrib Syst 28(4):919–933. https://doi.org/10.1109/TPDS.2016.2603511
6. Chen S, Guo C, Lai J (2016) Deep ranking for person re-identification via joint representation learning. IEEE Trans Image Process 25(5):2353–2367. https://doi.org/10.1109/TIP.2016.2545929

7. Chen Y, Duffner S, Baskurt A, Stoian A, Dufour JY (2018) Similarity learning with listwise ranking for person re-identification. In: 2018 25th IEEE international conference on image processing (ICIP), pp 843–847. https://doi.org/10.1109/ICIP.2018.8451628
8. Cheng D, Gong Y, Zhou S, Wang J, Zheng N (2016) Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 1335–1344. https://doi.org/10.1109/CVPR.2016.149
9. Courbariaux M, Bengio Y, David JP (2015) Training deep neural networks with low precision multiplications. ArXiv preprint arXiv:1412.7024v5
10. Ding S, Lin L, Wang G, Chao H (2015) Deep feature learning with relative distance comparison for person re-identification. Pattern Recognit 48(10):2993–3003. https://doi.org/10.1016/j.patcog.2015.04.005
11. Dinh KQ, Jeon B (2017) Iterative weighted recovery for block-based compressive sensing of image/video at a low subrate. IEEE Trans Circuits Syst Video Technol 27(11):2294–2308. https://doi.org/10.1109/TCSVT.2016.2587398
12. Dollár P, Appel R, Belongie S, Perona P (2014) Fast feature pyramids for object detection. IEEE Trans Pattern Anal Mach Intell 36(8):1532–1545. https://doi.org/10.1109/TPAMI.2014.2300479
13. Duan M, Li K, Li K (2018) An ensemble cnn2elm for age estimation. IEEE Trans Inf Forensics Secur 13(3):758–772. https://doi.org/10.1109/TIFS.2017.2766583
14. Duarte MF, Davenport MA, Takhar D, Laska JN, Sun T, Kelly KF, Baraniuk RG (2008) Single-pixel imaging via compressive sampling. IEEE Signal Process Mag 25(2):83–91. https://doi.org/10.1109/MSP.2007.914730
15. Fang J, Varbanescu AL, Liao X, Sips H (2014) Evaluating vector data type usage in opencl kernels. Concurr Comput Pract Exp 27(17):4586–4602. https://doi.org/10.1002/cpe.3424
16. Fang J, Zhang P, Tang C, Huang T, Yang C (2017) Implementing and evaluating OpenCL on an ARMv8 multi-core CPU. In: IEEE international symposium on parallel and distributed processing with applications. IEEE Computer Society, Guangzhou, Guangdong, China, pp 860–867. https://doi.org/10.1109/ISPA/IUCC.2017.00131
17. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D (2010) Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645. https://doi.org/10.1109/TPAMI.2009.167
18. Ge Y, Gu X, Chen M, Wang H, Yang D (2018) Deep multi-metric learning for person re-identification. In: 2018 IEEE international conference on multimedia and expo (ICME), pp 1–6. https://doi.org/10.1109/ICME.2018.8486502
19. Gray D, Tao H (2008) Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Forsyth D, Torr P, Zisserman A (eds) Computer Vision: ECCV 2008. Springer, Berlin, pp 262–275
20. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778. https://doi.org/10.1109/CVPR.2016.90
21. Iliadis M, Spinoulas L, Katsaggelos AK (2018) Deep fully-connected networks for video compressive sensing. Dig Signal Process 72:9–18. https://doi.org/10.1016/j.dsp.2017.09.010
22. Joseph R, Ali F (2018) Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767
23. Kulkarni K, Lohit S, Turaga P, Kerviche R, Ashok A (2016) Reconnet: non-iterative reconstruction of images from compressively sensed measurements. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 449–458. https://doi.org/10.1109/CVPR.2016.55
24. Li J, Liang X, Shen S, Xu T, Feng J, Yan S (2018) Scale-aware fast R-CNN for pedestrian detection. IEEE Trans Multimed 20(4):985–996. https://doi.org/10.1109/TMM.2017.2759508
25. Li K, Tang X, Li K (2014) Energy-efficient stochastic task scheduling on heterogeneous computing systems. IEEE Trans Parallel Distrib Syst 25(11):2867–2876. https://doi.org/10.1109/TPDS.2013.270
26. Li K, Tang X, Veeravalli B, Li K (2015) Scheduling precedence constrained stochastic tasks on heterogeneous cluster systems. IEEE Trans Comput 64(1):191–204. https://doi.org/10.1109/TC.2013.205
27. Li W, Zhao R, Xiao T, Wang X (2014) Deepreid: deep filter pairing neural network for person re-identification. In: 2014 IEEE conference on computer vision and pattern recognition, pp 152–159. https://doi.org/10.1109/CVPR.2014.27
28. Liao L, Li K, Li K, Yang C, Tian Q (2018) UHCL-Darknet: an OpenCL-based deep neural network framework for heterogeneous multi-/many-core clusters. In: Proceedings of the 47th international conference on parallel processing, ICPP 2018. ACM, New York, NY, USA, pp 44:1–44:10. https://doi.org/10.1145/3225058.3225107
29. Metzler CA, Maleki A, Baraniuk RG (2016) From denoising to compressed sensing. IEEE Trans Inf Theory 62(9):5117–5144. https://doi.org/10.1109/TIT.2016.2556683
30. Micikevicius P, Narang S, Alben J, Diamos GF, Elsen E, Garca D, Ginsburg B, Houston M, Kuchaiev O, Venkatesh G, Wu H (2018) Mixed precision training. In: The 6th international conference on learning representations (ICLR 2018), pp 1–12
31. Mousavi A, Baraniuk RG (2017) Learning to invert: signal recovery via deep convolutional networks. In: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 2272–2276. https://doi.org/10.1109/ICASSP.2017.7952561
32. Mousavi A, Patel AB, Baraniuk RG (2015) A deep learning approach to structured signal recovery. In: 2015 53rd annual allerton conference on communication, control, and computing (Allerton), pp 1336–1343. https://doi.org/10.1109/ALLERTON.2015.7447163
33. Nugteren C (2018) Clblast: a tuned OpenCL BLAS library. In: Proceedings of the international workshop on OpenCL, IWOCL '18. ACM, New York, NY, USA, pp 5:1–5:10. https://doi.org/10.1145/3204919.3204924
34. Ouyang W, Wang X (2013) Joint deep learning for pedestrian detection. In: 2013 IEEE international conference on computer vision, pp 2056–2063. https://doi.org/10.1109/ICCV.2013.257
35. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
36. Shi L, Chen H, Sun J, Li K (2012) vCUDA: GPU-accelerated high-performance computing in virtual machines. IEEE Trans Comput 61(6):804–816. https://doi.org/10.1109/TC.2011.112
37. Shi W, Jiang F, Zhang S, Zhao D (2017) Deep networks for compressed image sensing. In: 2017 IEEE international conference on multimedia and expo (ICME). IEEE, pp 877–882
38. Sun Y, Zheng L, Yang Y, Tian Q, Wang S (2018) Beyond part models: person retrieval with refined part pooling. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (eds) European conference on computer vision (ECCV). Springer, Cham, pp 501–518
39. Tao D, Guo Y, Yu B, Pang J, Yu Z (2018) Deep multi-view feature learning for person re-identification. IEEE Trans Circuits Syst Video Technol 28(10):2657–2666. https://doi.org/10.1109/TCSVT.2017.2726580

40. Vezzani R, Baltieri D, Cucchiara R (2013) People reidentification in surveillance and forensics: a survey. ACM Comput Surv 46(2):29:1–29:37. https://doi.org/10.1145/2543581.2543596
41. Wang G, Yuan Y, Chen X, Li J, Zhou X (2018) Learning discriminative features with multiple granularities for person re-identification. In: Proceedings of the 26th ACM international conference on multimedia, MM '18. ACM, New York, NY, USA, pp 274–282. https://doi.org/10.1145/3240508.3240552
42. Wojek C, Dollar P, Schiele B, Perona P (2012) Pedestrian detection: an evaluation of the state of the art. IEEE Trans Pattern Anal Mach Intell 34:743–761. https://doi.org/10.1109/TPAMI.2011.155
43. Xiao T, Li H, Ouyang W, Wang X (2016) Learning deep feature representations with domain guided dropout for person re-identification. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR) pp 1249–1258. https://doi.org/10.1109/CVPR.2016.140
44. Xiao T, Li S, Wang B, Lin L, Wang X (2017) Joint detection and identification feature learning for person search. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 3376–3385. https://doi.org/10.1109/CVPR.2017.360
45. Xu Y, Li K, He L, Zhang L, Li K (2015) A hybrid chemical reaction optimization scheme for task scheduling on heterogeneous computing systems. IEEE Trans Parallel Distrib Syst 26(12):3208–3222. https://doi.org/10.1109/TPDS.2014.2385698
46. Zhang H, Cao X, Ho JKL, Chow TWS (2017) Object-level video advertising: an optimization framework. IEEE Trans Ind Inform 13(2):520–531. https://doi.org/10.1109/TII.2016.2605629
47. Zhang H, Ji Y, Huang W, Liu L (2018) Sitcom-star-based clothing retrieval for video advertising: a deep learning framework. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3579-x
48. Zhang J, Ghanem B (2018) ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1828–1837
49. Zhang L, Li K, Xu Y, Mei J, Zhang F, Li K (2015) Maximizing reliability with energy conservation for parallel task scheduling in a heterogeneous cluster. Inf Sci 319:113–131. https://doi.org/10.1016/j.ins.2015.02.023
50. Zhang L, Lin L, Liang X, He K (2016) Is faster R-CNN doing well for pedestrian detection? In: Leibe B, Matas J, Sebe N, Welling M (eds) ECCV 2016. Springe, Cham, pp 443–457
51. Zhang N, Paluri M, Taigman Y, Fergus R, Bourdev L (2015) Beyond frontal faces: improving person recognition using multiple cues. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), pp 4804–4813. https://doi.org/10.1109/CVPR.2015.7299113
52. Zhang P, Fang J, Tang T, Yang C, Wang Z (2018) Mocl: an efficient OpenCL implementation for the matrix-2000 architecture. In: ACM international conference on computing frontiers. ACM, Ischia, Italy. https://doi.org/10.1145/3203217.3203244
53. Zhang P, Fang J, Tang T, Yang C, Wang Z (2018) Tuning streamed applications on Intel Xeon Phi: a machine learning based approach. In: the 32nd IEEE international parallel and distributed processing symposium (IPDPS'18). Vancouver, British Columbia, Canada, pp 515–525
54. Zhang S, Benenson R, Omran M, Hosang J, Schiele B (2016) How far are we from solving pedestrian detection? In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 1259–1267. https://doi.org/10.1109/CVPR.2016.141
55. Zhao L, Li X, Zhuang Y, Wang J (2017) Deeply-learned part-aligned representations for person re-identification. In: 2017 IEEE international conference on computer vision (ICCV), pp 3239–3248. https://doi.org/10.1109/ICCV.2017.349
56. Zheng L, Bie Z, Sun Y, Wang J, Su C, Wang S, Tian Q (2016) Mars: a video benchmark for large-scale person re-identification. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer vision: ECCV 2016. Springer, Cham, pp 868–884
57. Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q (2015) Scalable person re-identification: a benchmark. In: 2015 IEEE international conference on computer vision (ICCV), pp 1116–1124. https://doi.org/10.1109/ICCV.2015.133
58. Zheng L, Zhang H, Sun S, Chandraker M, Yang Y, Tian Q (2017) Person re-identification in the wild. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 3346–3355. https://doi.org/10.1109/CVPR.2017.357
59. Zheng W, Gong S, Xiang T (2011) Person re-identification by probabilistic relative distance comparison. In: CVPR 2011, pp 649–656. https://doi.org/10.1109/CVPR.2011.5995598