Enabling Open-Set Person Re-Identification for Real-World Scenarios

Tunc Alkanat, Egor Bondarev, and Peter H. N. De With
Electrical Engineering Department, Eindhoven University of Technology, Eindhoven, The Netherlands
Email: {t.alkanat, e.bondarev, p.h.n.de.with}@tue.nl

***Abstract*—Person re-identification (re-ID) is a significant problem of computer vision with increasing scientific attention. To date, numerous studies have been conducted to improve the accuracy and robustness of person re-ID to meet the practical demands. However, most of the previous efforts concentrated on solving the closed-set variant of the problem, where a query is assumed to always have a correct match within the set of known people (the gallery set). However, this assumption is usually not valid for the industrial re-ID use cases. In this study, we focus on the open-set person re-ID problem, where, in addition to the similarity ranking, the solution is expected to detect the presence or absence of a given query identity within the gallery set. To determine good practices and to assess the practicality of the person re-ID in industrial applications, first, we convert popular closed-set person re-ID datasets into the open-set scenario. Second, we compare performance of eight state-of-the-art closed-set person re-ID methods under the open-set conditions. Third, we experimentally determine the efficiency of using different loss function combinations for the open-set problem. Finally, we investigate the impact of a statistics-driven gallery refinement approach on the open-set person re-ID performance in the low false-acceptance rate (FAR) region, while simultaneously reducing the computational demands of retrieval. Results show an average detection and identification rate increase of 8.38% and 3.39% on the DukeMTMC-reID and Market1501 datasets, respectively, for a FAR of 1%.**

***Index Terms*—person re-identification, open-set, image retrieval**

I. Introduction

Person re-identification (re-ID) is the problem of identity matching of pedestrians using images obtained from different, non-overlapping camera views. In other words, given a set of images of known people (gallery set), person re-ID aims to search and match the identity of a person within a given query image to the person identities within the existing gallery images taken from disjoint cameras. Person re-ID is a crucial tool for multi-camera tracking, since it enables the extraction of inter-camera trajectories. With the current progress in smart cameras allowing reliable and individual object detection, person re-ID is the next important step for advanced autonomous surveillance systems and higher levels of event understanding. Considering a public place with multiple non-overlapping cameras, obtaining an estimate of the real-world trajectories of people is beneficial for determining motion characteristics and possible anomalies in people behavior. Also, person re-ID is particularly useful for person-of-interest search. This application aims to reveal the whereabouts of a person whose visual appearance is known. Existence of such applications made person re-ID a prominent area of research with increasing scientific attention.

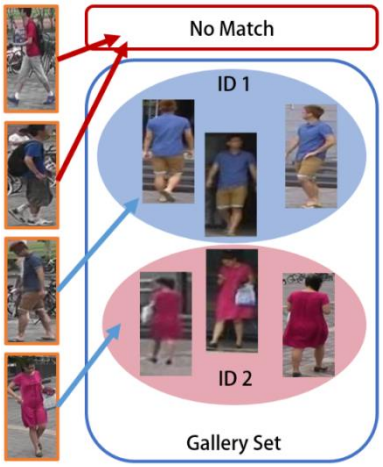

Figure 1. Depiction of an ideal open-set person re-ID method. Query images are marked with orange and gallery set includes two identities. Top two queries have no match third and fourth query images are of identities ID #1 and ID #2, respectively.

Although the problem of person re-ID has been studied heavily during the last decade, it is still far from being solved, mainly due to a large variety of challenges inherent to person re-ID itself. Occlusions, viewpoint variations, low-resolution images, clothing similarities and variable lighting/weather conditions, make this problem intrinsically complicated.

Another important yet overlooked difficulty associated with person re-ID is the so-called open-set problem (e.g. new persons entering the scene). Although some image retrieval tasks are suitable for closed-set scenarios, most of the practical applications operate in the open-set conditions. Given that person re-ID is primarily utilized to achieve Multi-target, Multi-camera (MTMC) tracking for surveillance applications in public places, an ideal person re-ID approach is expected to handle the case of newly entering pedestrians into the scene. To achieve this, an algorithm should be able to perform the so-called

Manuscript received November 30, 2019; revised March 27, 2020.

doi: 10.18178/joig.8.2.26-36

verification task [1]. In other words, an open-set algorithm must be able to determine whether a given query has a match in the gallery, as depicted in Fig. 1. There exist only a handful of previous studies in the literature that aim to perform person re-ID with the verification task, while most approaches consider the closed-set case only. Due to the strong correlation between the performance of closed-set and open-set scenarios, having a high-performance closed-set person re-ID algorithm is beneficial also for the open-set case. However, indirectly addressing the more practical case is sub-optimal.

In this study, we address the problem of open-set person re-ID directly. The contributions of this paper are as follows. (1) We convert two of the most commonly used person re-ID datasets into an open-set dataset. (2) We compare the performance of the most popular person re-ID methods in the open-set case. (3) We compare different combinations of popular loss functions and determine the effects of loss function on the open-set re-ID performance. (4) We explore the effects of statistical analysis and refinement of gallery embeddings on the open-set performance of existing algorithms.

The remainder of this paper is organized as follows. Section II presents an overview of existing open- and closed-set approaches in literature. Section III defines the problem of open-set person re-ID. Section IV describes the statistical gallery analysis and refinement approach. In Section V, we report on our experimental results for the open-set case. Finally, in Section VI, concluding remarks are given.

II. Related Work

The problem of person re-ID has been introduced in [2], where Zajdel *et al.* defined the problem in the context of MTMC. Then, scientific attention to person re-ID increased rapidly and researchers concentrated specifically on person re-ID, isolating it from the context of MTMC. The main focus of the person re-ID literature has been concentrating on the closed-set variant of the problem. Considering the correlation between open-set and closed-set performance, we present a brief overview of the closed-set person re-ID case in Section A, followed by a review of open-set studies.

A. Closed-Set Person Re-Identification

The existing person re-ID algorithms are mainly concentrated on two directions, namely, feature extraction and metric learning. In feature extraction, the aim is to design a transformation from image space to embedding space, where the resulting embeddings are discriminative in identity and robust to visual appearance changes. Early studies on person re-ID relied heavily on handcrafted features. In this approach, a set of features are extracted from each image and then, images are ranked by their similarity, according to the similarity of their corresponding feature vectors. In this class of approaches, feature extraction methods based on color histogram [3], Scale-Invariant Feature Transform (SIFT) [4], Speeded-up Robust Features (SURF) [5], Local Binary Patterns (LBP) [6], Maximally Stable Color Regions (MSCR) [7] and many more have been proposed. In some studies, multiple feature extraction strategies are combined to construct better feature vectors [8], [9]. Similarly, some studies employ features in various settings such as global [10], local [11] and a fusion of both feature types [12], [13]. Furthermore, part-based models have been proposed mainly to address the bounding-box alignment problem inherent to practical re-ID scenarios [14], [15].

The second important direction of the person re-ID research is finding approaches for metric learning. In these methods, the aim is to find a transformation of embedding space that maps points of the same identity closer to each other, while separating different identities farther apart from each other. For instance, in [16], Köstinger *et al.* derive a metric learning method called KISSME, by exploiting an equivalence constraint in an efficient formulation. In [17], Weinberger *et al.* formulate an approach to learn the Mahalanobis distance metric by enforcing k-nearest neighbors into the same identity, while placing embeddings of different identities farther away by a large margin. Another popular approach to metric learning is XQDA [18]. In XQDA, Liao *et al.* formulate to learn a discriminant low-dimensional subspace by cross-view quadratic discriminant analysis.

After the emergence of deep learning in 2012 [19], most computer vision research switched rapidly from handcrafted approaches to learning-based solutions and the field of person re-ID went in a similar direction [20]-[22]. Thanks to large datasets currently available, most top-performing methods in person re-ID depend on various deep learning approaches. Such approaches learn successful feature extraction and schemes for metric learning simultaneously from the labeled samples. Due to its superiority, numerous different deep learning approaches have been proposed for the problem of re-ID. Methods employing loss functions such as triplet loss [23], cross-entropy (CE) loss [24], spherical CE loss [25], center loss [26], quadruplet loss [27] and additional information such as pose [28], alignment [29], attention [30] and saliency [31] have been proposed and shown to perform better than the counterparts based on handcrafted feature extraction.

B. Open-Set Person Re-Identification

Compared to person re-ID, the open-set variant received only marginal attention. One of the first studies to address the open-set person re-ID in literature is [1], where Zheng *et al.* define the terms one- and multi-shot verification (the task of determining whether a query image is included in a gallery or not). Then, the authors formulate a transfer learning framework based on metric learning, to mine discriminative information about pedestrian images. They report results on ETHZ and i-LIDS datasets that include 119 and 146 identities, respectively. In [32], Cancela *et al.* propose a complete MTMC approach with both inter-camera and intra-camera tracking. For person re-ID, based on the first and last detection of each identity from each camera, the authors form a matrix where each entry is the pairwise

similarity score of the first and last appearance of two pedestrians. Then, they solve the similarity maximization problem by proposing an iterative algorithm. Authors report results on the SAIVT-Softbio dataset [33] that includes 150 identities. In [34], Chan *et al.* use a metric learning method, where they define a threshold for the minimum similarity between a pair of images to be considered as a match during training, by modifying the contrastive loss. The training aims at learning a transformation on embeddings that improves open-set performance. Authors report results on the iLIDS-VID [35] and the PRID2011 [36] datasets, by discarding some gallery identities that effectively reduce the gallery set to 100 and 60 identities, respectively. In [37], Zheng *et al.* address a similar problem of group-based verification and follow a metric learning approach, by simultaneously minimizing intraclass variation, maximizing interclass variation and imposing group separation. Authors report results on the i-LIDS, ETHZ, CAVIAR4REID [38] and VIPeR [39] datasets. The CAVIAR4REID dataset includes 72 identities and the VIPeR dataset includes two images of 632 identities.

In addition to the methods that directly address the problem of open-set person re-ID, there exist several closed-set person re-ID studies, reporting results on the open-set case. In [40], the authors report open-set person re-ID results on originally closed-set datasets like VIPeR and CUHK01 [41], by removing most of the identities from the gallery. Similarly in [42], Ma *et al.* report results for the open-set person re-ID case on modified versions of the iLIDS-VID and PRID2011 datasets.

Our research fully concentrates fully on the open-set case. We have found that the current algorithms are based on relatively small datasets for both training and evaluation, which leaves room for better learning and evaluation. Moreover, we also investigate the effects of a gallery refinement and various loss functions in order to reach higher performance with reduced computational demands.

III. Problem Definition

Most of the existing closed-set person re-ID approaches operate by extracting a feature vector for a given bounding-box of a pedestrian. Then, the problem of ranking is solved by calculating a distance metric between the query and gallery images and sorting the distances in ascending order. Mathematically, given an image I of HxW pixels, where $I \in \mathbb{R}^{HxWxC}$ with C channels, most closed-set person re-ID methods aim to find an identity-discriminating transformation ϕ such that, $\phi: I \to f \in \mathbb{R}^{f_S}$, where f is the embedding associated to image I and f_s is the dimensionality of the embedding space. Then, the similarity of any arbitrary image pair is measured by calculating a distance metric between the embeddings of each image, which is expressed by:

$$\boldsymbol{D}\big(\phi(I_1), \phi(I_2)\big) = \boldsymbol{D}(f_1, f_2) \tag{1}$$

where $\boldsymbol{D}$ is the embedding distance function, f_n is the embedding space point for image I_n for $n = 1, 2$.

Applying an embedding distance approach to achieve person re-ID is straightforward. Let a gallery set of N images be denoted as $G = \{I_n\}_{n=1}^{N}$. Also, assume that a mapping function $id(.)$ exists, that is defined as $id: \mathbb{R}^{HxWxC} \to \mathcal{Y} \in \mathbb{N}$, where $\mathcal{Y}$ is the identity label space. Then, the identity $\hat{y} \in \mathcal{Y}$ of the most likely match to a given query image $I_q \notin G$ can be found, by minimizing the distance in the embedding space, according to:

$$\begin{gathered} \hat{n} = \underset{n}{\operatorname{argmin}}\, \boldsymbol{D}(\phi(I_q), \phi(I_n)) \\ \hat{d} = \boldsymbol{D}\left(\phi(I_q), \phi(I_{\hat{n}})\right), \text{and} \\ \hat{y} = id(I_{\hat{n}}) \end{gathered} \tag{2}$$

here, $\hat{n}$ denotes the index of the most similar gallery image to the query and $\hat{d}$ is the distance between them in the embedding space. Similarly, identities of other most likely matches can be revealed by sorting the distances in ascending order.

As mentioned earlier, the closed-set person re-ID does not contain the verification task. The verification task in the context of person re-ID is to determine whether a given query identity is included in the gallery or not, whereas closed-set approaches aim to retrieve only the most similar gallery images to a given query image. To convert a closed-set method to the open-set case, one common approach to achieve verification is by utilizing query-to-gallery distances that are usually obtained as a side product of ranking. To do so, calculated distances are subject to a predefined threshold τ, which ensures that a query does not have a match in the gallery, when the closest gallery image has more than τ distance in embedding space. For this purpose, we introduce the piece-wise, thresholding function m: $\mathbb{R} \to \{0,1\}$ defined as:

$$m_\tau(\hat{d}) = \begin{cases} 0 & for\ \tau \leq \hat{d} \\ 1 & for\ \tau > \hat{d} \end{cases} \tag{3}$$

where $m_\tau(\hat{d}) = 1$ indicates a distance that is small enough to consider the corresponding image to be a match. Then, depending on whether the given query image has a match or not and whether verification returns positive or negative, we have four possible outcomes. Let the set of all identities included in the gallery be $F = \{id(x) | \forall x \in G\}$. Then, four possible outcomes of verification are specified by:

$$\begin{aligned} \text{True Positive} &= \{id(I_q) \in F \wedge m_\tau(\hat{d}) = 1\} \\ \text{True Negative} &= \{id(I_q) \notin F \wedge m_\tau(\hat{d}) = 0\} \\ \text{False Positive} &= \{id(I_q) \notin F \wedge m_\tau(\hat{d}) = 1\} \\ \text{False Negative} &= \{id(I_q) \in F \wedge m_\tau(\hat{d}) = 0\} \end{aligned} \tag{4}$$

The verification task in the context of person re-ID is depicted in Fig. 2. After verification, the problem is reduced to a closed-set person re-ID, considering only the query identities that are verified to exist in the gallery. To summarize, the aim of open-set person re-ID is to correctly identify the queries that have no match in the gallery and to retrieve correct matches for the queries that do have a match.

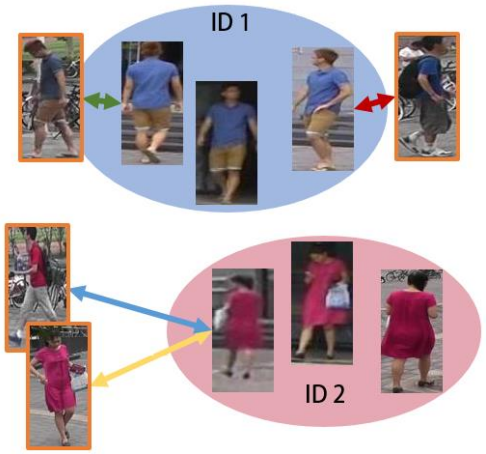

Figure 2. Visualization of all four possible outcomes of a verification task in the context of open-set person re-ID in a hypothetical 2D embedding space. A gallery has two identities ID1 and ID2, query images marked with orange. True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) outcomes are shown with green, blue, red and yellow arrows, respectively.

IV. Statistics-Driven Gallery Refinement

In this section, we give detailed information about the statistical analysis approach that we utilized in order to maximize the open-set performance. Analysis takes place in the gallery embedding space, and uses simple and fast computations that do not contradict with the practical demands. We use this statistical information to refine the gallery by discarding less informative and possibly misleading embeddings. By doing so, the number of false positives during the verification task is decreased and thus, the re-ID performance for low false acceptance rates is increased. This technique is applicable to all person re-ID approaches that are based on transforming images to an embedding space, as described in Section III.

Under realistic conditions, there may exist erroneous and misleading samples in a given gallery set. For example, consider a heavily occluded gallery entry, where a prominent portion of the target person is occluded by another person. For that specific gallery image, the corresponding embedding is expected to include parts resembling the embedding of the occluding person. Thus, in the embedding space, that occluded sample is expected to be located in the area that is far away from the mean of the same identity embeddings. In other words, transformation ϕ may fail to position all samples belonging to the same identity in close vicinity. Furthermore, there may exist a query, whose ID is not in the gallery and whose embedding is close to the occluded sample. In such a case, the verification task may fail producing a false positive. Similarly, such failures may happen due to other challenges associated with person re-ID. To alleviate the failures caused by misleading gallery samples, given that identity labels for the entire gallery are available, it is possible to detect and process unlikely samples belonging to each identity and improve the open-set results by discarding less trustworthy embeddings. Generalizing this idea, we hypothesize that, if a sample is located farther away from the statistical mean of the same identity embeddings in the gallery set, it becomes less reliable. This is reasonable because when the number of samples of a gallery identity grows to infinity, one can expect the mean of the same ID embeddings to be the most representative embedding for that specific identity-transformation (ϕ) pair.

Removing the outlier embeddings of an identity in the gallery is beneficial for both open- and closed-set cases of person re-ID. However, in the open-set case, the statistical refinement approach is not utilized only to clear outliers. Instead, this information introduces a way to modify the gallery content, such that it becomes possible to enforce the verification task to return a positive value only when the query has a very strong match in the gallery. Furthermore, a parameter is associated to this process, so that the trade-off between the open- and closed-set performance can be controlled, allowing the user to select an operating point that is most suitable for the application at hand. Now, we proceed by describing the steps of gallery refinement.

A. Dimensionality Reduction

The first step is to reduce the dimensionality of embeddings. A reduction of dimensionality is desired for two reasons: (a) removal of irrelevant dimensions with minimal variance and (b) reduced computational cost of distance calculation and embedding storage. To reduce dimensionality, Principal Component Analysis (PCA) is applied to embeddings. Following the notation in Section III, we denote the associated PCA transformation as $P: \mathbb{R}^{f_s} \rightarrow \mathbb{R}^{f_r}$. Then, the set of transformed gallery embeddings is $E^g = \{P(\phi(I)) | \forall I \in G\} \subset \mathbb{R}^{f_r}$.

B. Identity-Coupled Statistics Calculation

The next step is to find the sample mean and variances of each element of the feature vectors belonging to every identity in the reduced embedding space. Mathematically, let the set of reduced gallery embeddings belonging to the l^{th} identity be denoted as E_l^g and its cardinality be $N_l = |E_l^g|$. Then, we define the sample mean and variance by:

$$\begin{aligned} \mu_l[n] &= \frac{1}{N_l} \sum_{f \in E_l^g} f[n] \\ \sigma_l[n] &= \frac{1}{N_l} \sum_{f \in E_l^g} (f[n] - \mu_l[n])^2 \end{aligned} \tag{5}$$

where $n = 1,2, \ldots, f_r$.

C. Deviation Calculation of Samples

We proceed by calculating a single value for each reduced gallery embedding, such that it represents the deviation of that embedding from the statistical mean. This deviation is computed by

$$d_{l,m} = \sum_n \left(\frac{f_{l,m}[n] - \mu_l[n]}{\sigma_l[n]} \right)^2 \tag{6}$$

where $d_{l,m}$ is the deviation of m^{th} embedding from the mean of the l^{th} identity in the gallery embedding space.

D. Sorting & Elimination of Outliers

Lastly, gallery refinement concludes by sorting all embeddings belonging to the same identity in decreasing order with respect to their deviation from the mean, and

by discarding the gallery entries that comprise the highest α% of this metric.

V. Experimental Results

To demonstrate the effects of gallery refinement and to determine the good practices towards practical open-set person re-ID, we conducted extensive tests on popular public datasets. Here, we report on three sets of experiments. (1) Open-set evaluation of the existing state-of-the-art closed-set methods. (2) Evaluation of the effects of loss functions on open-set performance. (3) Impact of gallery refinement on performance.

A. Datasets

As summarized in Section II-B, previous studies on open-set person re-ID report results on datasets that lack the sufficient number of identities and images of training, query and/or gallery sets. This may lead to unrepresentative evaluation and be sub-optimal, especially for data-hungry deep learning-based algorithms. Therefore, we conduct our experiments on popular, large-scale datasets. Among those, we utilize the most popular datasets to conduct our experiments: Market1501 [43] and DukeMTMC-reID [44], [45].

Market1501. In this dataset, there are 1,501 identities, where 12,936 images corresponding to 751 identities are used for training. As for the testing, the remaining identities are used to construct the gallery and query sets that include 19,732 and 3,368 images, respectively. Market1501 is constructed using the Deformable Part Model (DPM) [46] for automatic detection and cropping of pedestrians in a given image and it includes images from 6 different cameras.

DukeMTMC-reID. This dataset includes 16,522 training images of 702 identities. As for the gallery and query sets, 2,228 and 17,661 images exist for another 702 identities. Images are obtained from 8 different cameras.

B. Methods

As in Section II, a wide variety of methods exist, addressing the challenge of person re-ID. To provide an up-to-date comparison of the performance, we have selected six recent and high-performing person re-ID methods.

HACNN. This approach combines soft and hard attention information by a novel Harmonious Attention module [47]. Li *et al.* utilize this module in a two-branch, multi-stream Inception [48], [49] architecture. Final features of each branch are constructed with global max pooling, followed by an FC layer of 512 elements. Concatenated features from branches are then used as embeddings. The network is trained with CE loss for person identification.

MLFN. Introduced by Chang *et al.*, Multi-Level Factorisation Net [50] is composed of multiple blocks containing multiple, identical sub-networks that specialize in extracting features from a different aspect of given images. Moreover, each block has a feature selection network that decides which sub-networks will be active during a forward pass. In MLFN, multiple blocks are combined sequentially such that higher layer blocks learn higher level features. The last step of the MLFN architecture is to fuse information from every block to construct the embeddings.

ResNet-50. This approach is based on the famous ResNet-50 architecture. The only modification to the original architecture lies in the size of the final Fully-Connected (FC) layer, which is set differently according to the number of identities in each dataset. This model is trained for CE loss of person identification. Then, embeddings are extracted after the global max pooling layer of the original architecture. Prior to training, the network is initialized with pre-trained weights from ImageNet [51].

MGN. Multiple granularity network [52] is introduced by Wang *et al.* in 2018. This approach uses a three-branch architecture based on ResNet-50, along with both triplet and CE loss. Each branch includes a global max pooling, followed by the CE and the triplet loss term. In addition, the second and third branch slice the feature volumes horizontally into 2 and 3 bins, respectively, prior to global max pooling. Then, max pooling is applied to each individual bin, resulting in sub-feature vectors to be trained for the person identification task using the CE loss. Note that, due to the lack of an official implementation or trained model for this method, all numerical results are computed using our implementation, where the closed set-performance is slightly less than reported in the original paper.

TriNet. Introduced in 2017 [23], this network is based on ResNet-50, pre-trained on ImageNet and is trained for triplet loss with a hard triplet mining strategy, known as batch-hard. The batch-hard strategy is found to be effective by a large number of succeeding studies. For each batch, this strategy randomly samples P identities and K images from each of those identities. Then, for each identity, network weights are only updated for the triplets that yield the highest loss within each batch. The ResNet-50 backbone is modified such that, the last FC layer is discarded in favor of two new FC layers. Embeddings are then constructed to be the output of the last layer, with a size of 128. Due to the relatively small embedding dimensionality, we do not apply PCA reduction to TriNet embeddings while applying GR.

SPReID. Human semantic parsing for person re-identification is introduced by Kalayeh *et al.* [24]. Authors use a two-branch network based on Inception V3 [53] backbone, one for re-ID feature extraction and the other for semantic parsing of five regions of the human body, namely, foreground, head, upper body, lower body and shoes. Then, embeddings are constructed by filtering the features for each semantic region and concatenating them.

BDB. Batch DropBlock Network is introduced by Dai *et al.* [54] in 2019. This network is a two-branch architecture with a ResNet-50-based backbone and is trained using triplet and CE loss. In the feature dropping branch of this architecture, the DropBlock layer [55] is incorporated in a batch wide consistent manner, where the same spatial block in feature tensors within a batch is

dropped. In the second branch, the global features are extracted.

BoT. Introduced in 2019 by Luo *et al.* [56], this architecture is a single branch ResNet-50-based approach trained by utilizing various strategies to enhance the performance, without significantly increasing the model complexity. In the paper, authors report on the performance increase of training strategies such as learning rate warm-up, random erasing, label smoothing and center loss along with architectural changes, such as setting the last stride of the backbone to one instead of two and utilizing batch-norm before the last FC layer. This model is trained using triplet, CE with label smoothing and center losses.

C. Evaluation Metrics

In this study, we use the Detection and Identification Rate (DIR) in combination with the False Accept Rate (FAR) [57], to assess the performance of a given method in open-set person re-ID. Let $\hat{g}$ be the most similar, same identity gallery entry index to a given query q that satisfies $\boldsymbol{D}\left(\phi(I_q), \phi(I_{\hat{g}})\right) = d_{\hat{g}}$. Furthermore, let the set of query entries that have a match in the gallery be P_G and let P_N be the queries that have no match. Then DIR and FAR are defined as:

$$DIR(\tau, k) = \frac{\left|\left\{q \middle| q \in P_G, rank(q) \leq k, m_\tau\left(d_{\hat{g}}\right)=1\right\}\right|}{|P_G|}$$
$$FAR(\tau) = \frac{|\{p|p \in P_N, m_\tau(\hat{d})=0\}|}{|P_N|} \quad (7)$$

where $rank(q)$ function gives the rank of $d_{\hat{g}}$ on the sorted sequence of distances between query embedding q and all gallery embeddings. Here, DIR is a function of τ and k, and represents the success of an algorithm to retrieve gallery images, such that the top-k include at least one correct match, if the gallery includes the given query identity. In contrast, the FAR measures the ability of an algorithm to correctly identify query images that do not have a match in the gallery.

D. Open-Set Evaluation of Existing Methods

Albeit more suitable to practical scenarios, to the best of our knowledge, there exist no prior studies that aim to benchmark the open-set performance of recent, deep learning-based methods. To alleviate this and contribute to open-set literature, we have conducted tests on the Market1501 and DukeMTMC-reID after converting them into open-set data. For the conversion, we have randomly discarded β% of all identities in gallery sets of these datasets. To reduce the effects of randomness, we have repeated the same procedure 100 times, for different sets of identities excluded from the gallery. The resulting DIR vs FAR curves are presented in Fig. 3 for 8 different methods on both datasets and for β values of 20 and 50.

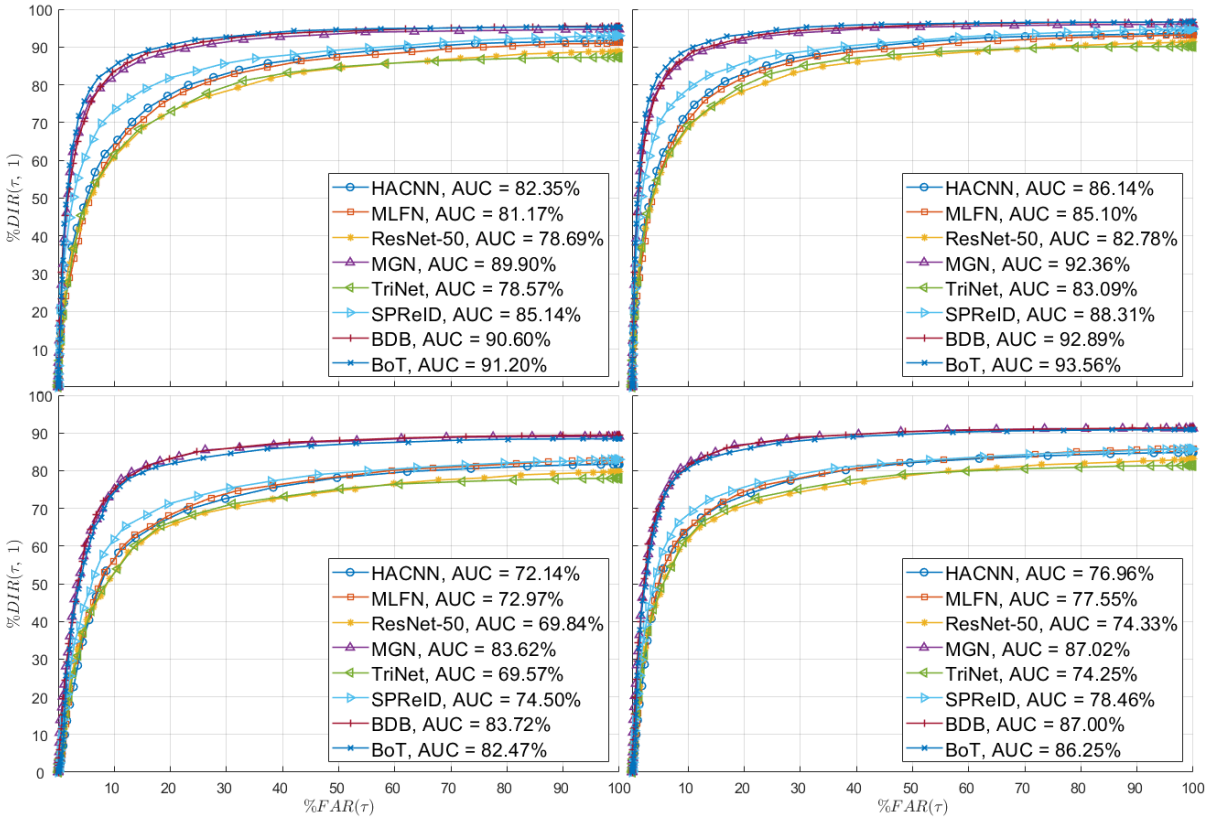

Figure 3. Open-set person re-ID performances for various methods shown in DIR vs FAR graphs, including area under curve percentages (AUC). The top graphs are for the Market1501 dataset and bottom are for the DukeMTMC-reID dataset. From left to right, β values are 20, 50, respectively. Note that, due to the modifications in the gallery set, the performances of the methods are not identical for the 100% FAR open-set case and the closed-set case, as reported individually in the corresponding papers. (Graphs are best viewed in color.)

E. Effects of the Loss on Open-Set Performance

As detailed in Section II-A, there exist numerous different loss functions to enhance the performance of person re-ID. Previous research has shown that, it is possible to improve the closed-set performance by incorporating additional information during training and by optimizing for multiple loss functions simultaneously. However, the contribution by each loss to the open-set performance in various FAR regions is unknown. To alleviate this and to determine an efficient loss combination that yield superior open-set performance, we have trained a simple ResNet-50 based architecture, with the combinations of triplet, softmax CE and center loss. The results are summarized in Table I. for FAR values of 0.1%, 1%, 5% and 10% and for six combinations of triplet, softmax CE and center loss functions.

TABLE I. OPEN-SET PERSON RE-IDENTIFICATION PERFORMANCE OF VARIOUS LOSS FUNCTIONS ON DUKEMTMC-REID DATASET FOR $\beta = 50$

	FAR			
LOSS	**0.1%**	**1%**	**5%**	**10%**
Triplet Only	4.22	27.63	59.21	70.75
Softmax Only	0.95	17.44	45.20	57.11
Triplet + Center	3.51	25.38	54.47	64.58
Softmax + Center	0.88	15.55	48.02	58.64
Softmax + Triplet	1.02	27.54	56.76	66.84
Softmax + Tr. + Center	2.68	29.53	60.14	70.27

Our tests are conducted on a ResNet-50-based architecture, pre-trained on ImageNet. A simple, yet realistic architecture has been chosen to isolate the effects of a loss function on the open-set performance. In the architecture, the output of the global mean pooling layer is used for testing and triplet/center loss training, if applicable. An additional FC layer is added to the network, for the cases where softmax CE loss is utilized. Training hyper-parameters are kept the same for every training instance and the models have been trained for 80 epochs. The initial learning rate is set to $2\text{x}10^{-4}$ and it is reduced to its one-tenth at epochs 40 and 60.

F. Performance of Gallery Refinement

The efficiency of statistical filtering is numerically evaluated for α (Percentage of deleted gallery entries) values of $\{0, 10, \ldots, 90\}$, for β values of 20 and 50 on both datasets and for six methods under consideration. The results are summarized in Table II. In this table, we present four rows for each of the eight methods. The first two rows are the open-set performance of each corresponding method with and without gallery refinement. Afterwards, we realize the performance boost by gallery refinement, Δ. Finally, the fourth row shows the optimum value of the parameter α. As the performance in the low FAR region is of our main interest, we report results for FAR values of 0.1%, 1%, 5% and 10%. The last row of Table II includes the average performance gain by GR with optimal α, at each suitable combination of β, dataset and FAR.

TABLE II. TOP-1 DIR VALUES OF ALL EIGHT METHODS, WITH AND WITHOUT GALLERY REFINEMENT (GR), FOR BOTH DATASETS, USING FARS OF 0.1%, 1%, 5%, 10% AND β VALUES OF 20%, 50%. EACH 4-TUPLE ROW DENOTES THE DIR VALUES WITH ON THE 1ST ROW, $\alpha = 0$, ON THE 2ND ROW, $\alpha = \alpha_{opt}$, ON THE THIRD ROW, THE IMPROVEMENT Δ, AND ON THE 4TH, THE OPTIMAL VALUE FOR α. THE BOTTOM ROW OF THE WHOLE TABLE IS THE AVERAGE PERFORMANCE GAIN FROM GR FOR EACH FAR AND β COMBINATION. BEST RESULTS FOR EACH COMBINED SETTING OF FAR, DATASET AND β ARE PRINTED IN BOLD

Dataset	**Market1501**								**DukeMTMC-reID**							
β	**20%**				**50%**				**20%**				**50%**			
FAR	**0.1%**	**1%**	**5%**	**10%**	**0.1%**	**1%**	**5%**	**10%**	**0.1%**	**1%**	**5%**	**10%**	**0.1%**	**1%**	**5%**	**10%**
HACNN	6.22	22.07	49.67	64.46	7.53	29.06	59.76	72.70	0.37	8.18	37.81	56.59	0.58	13.86	50.80	64.29
HACNN + GR	6.33	24.78	51.94	64.69	7.53	30.98	60.23	73.06	1.06	16.36	48.75	60.64	1.87	23.29	58.08	66.93
Δ	0.11	2.71	2.27	0.24	0.00	1.91	0.47	0.36	0.69	8.17	10.94	4.05	1.29	9.43	7.28	2.63
α_{opt}	20	50	40	20	0	40	30	20	90	90	60	50	80	90	50	50
MLFN	5.20	21.57	46.93	62.88	7.10	25.71	56.60	70.72	0.24	9.42	39.83	56.38	0.36	15.21	51.79	64.69
MLFN + GR	9.21	23.38	48.87	64.54	11.25	27.42	58.64	71.89	0.43	18.57	47.25	60.59	0.74	24.00	56.04	66.09
Δ	4.01	1.80	1.94	1.65	4.15	1.71	2.04	1.17	0.19	9.15	7.41	4.21	0.38	8.79	4.25	1.40
α_{opt}	50	40	60	40	50	50	40	40	90	80	70	40	90	80	60	30
ResNet50	3.69	22.35	46.91	60.87	5.69	29.37	55.99	68.50	0.85	10.53	39.90	52.96	1.39	16.67	47.23	61.81
ResNet50 + GR	7.06	26.21	48.11	61.64	9.61	30.90	57.46	69.31	3.66	18.08	44.12	57.30	4.73	24.98	52.60	63.86
Δ	3.37	3.86	1.20	0.77	3.92	1.54	1.47	0.81	2.81	7.55	4.22	4.34	3.34	8.32	5.37	2.04
α_{opt}	90	60	30	20	80	60	40	20	90	90	40	40	90	70	60	30
MGN	8.74	35.89	73.54	82.25	10.59	46.04	79.71	87.11	3.49	22.48	60.33	75.05	5.03	30.94	70.68	81.84
MGN + GR	11.64	41.82	74.89	83.12	15.26	51.47	80.63	87.33	**9.20**	29.05	65.10	**78.75**	**13.34**	38.16	**75.52**	83.13
Δ	2.90	5.93	1.35	0.87	4.67	5.43	0.92	0.22	5.71	6.57	4.77	3.70	8.30	7.23	4.84	1.29
α_{opt}	70	80	20	20	70	60	40	30	90	60	50	20	90	60	50	50
TriNet	2.19	19.04	48.31	61.53	3.39	26.02	56.22	69.27	0.11	11.60	39.06	52.79	1.15	16.35	48.11	62.33
TriNet + GR	4.47	23.08	50.70	63.96	7.11	29.55	58.79	70.82	3.99	17.48	44.92	59.32	5.83	23.98	55.03	65.42
Δ	2.29	4.03	2.39	2.43	3.73	3.53	2.57	1.55	3.88	5.88	5.86	6.53	4.67	7.63	6.92	3.08
α_{opt}	90	70	40	40	60	70	40	30	90	70	70	60	80	50	60	40

SPReID	10.07	32.59	61.66	73.31	12.64	39.77	70.40	78.91	1.76	14.85	45.69	61.90	2.34	20.70	56.61	68.36
SPReID + GR	**12.42**	34.10	62.81	73.85	**16.43**	42.01	70.98	79.46	3.01	19.97	51.46	63.60	3.96	26.75	59.72	69.74
Δ	2.36	1.50	1.14	0.53	3.79	2.24	0.58	0.55	1.24	5.12	5.78	1.71	1.62	6.06	3.11	1.38
α_{opt}	60	50	40	20	40	50	20	20	30	50	50	20	30	70	40	40
BDB	5.05	40.05	71.90	83.60	8.19	48.74	79.51	87.83	2.96	20.00	61.04	75.04	5.66	31.07	71.38	81.28
BDB + GR	11.91	**43.45**	75.61	84.64	15.43	51.90	81.53	88.42	7.68	**29.63**	**67.68**	78.62	9.93	37.49	75.05	**83.36**
Δ	6.86	3.40	3.71	1.04	7.24	3.16	2.02	0.59	4.72	9.63	6.64	3.57	4.27	6.42	3.66	2.08
α_{opt}	90	30	20	20	90	20	20	20	60	60	70	40	80	40	40	40
BoT	7.32	39.17	76.61	84.99	9.28	50.56	82.99	89.37	0.40	11.02	57.43	74.23	0.70	28.04	69.41	80.68
BoT + GR	9.81	43.02	**78.35**	**85.90**	12.78	**56.41**	**83.63**	**89.81**	0.62	26.00	65.45	77.49	2.10	**40.80**	74.50	82.68
Δ	2.49	3.85	1.75	0.91	3.51	5.86	0.63	0.44	0.22	14.98	8.02	3.26	1.41	12.76	5.09	1.99
α_{opt}	60	30	30	30	60	60	30	10	50	60	50	40	90	70	40	30
Δ_{avg}	3.05	3.39	1.97	1.06	3.88	3.17	1.34	0.71	2.43	8.38	6.71	3.92	3.16	8.33	5.07	1.99

G. Discussion

Methods in Open-Set Case. As Fig. 3 and Table II imply, in most of the cases, MGN, BoT and BDB outperform other methods by respectable margins. One interesting exception is the 0.1% FAR, Market1501 dataset case, where SPReID is found to be the best performing method, regardless of the value of β. Especially in Market1501 dataset, BoT outperforms all other methods in 5% and 10% FAR cases. Further, for both datasets, BoT is found to be the best-performing method for most experiments, while its performance is closely followed by BDB. BoT generally outperforms BDB in Market1501 tests, while BDB offers better performance in the DukeMTMC-ReID in general. On the other hand, MLFN and HACNN perform similarly in most of the experiments, as this is the case for ResNet-50 and TriNet. Interestingly, although the ResNet-50 method performs slightly better than TriNet in the 100% FAR case, in the low FAR region TriNet catches up with ResNet-50. Moreover, as it can be seen from Fig. 3, TriNet also performs better for the mediocre FAR region in all tests. Overall, the open-set evaluation of methods shows that it is possible to achieve a competitive low FAR performance by using a relatively light-weight architecture, such as BoT and BDB. However, the results also imply the necessity of determining the FAR operating region for the practical, open-set person re-ID problem, in order to make an informed decision about the choice of methods, considering computational demands and performance gains.

Efficiency of Loss Combinations. Results in Table I show that the triplet loss enhances the open-set person re-ID performance to a significant extent. All combinations that include triplet loss perform better than the counterparts that do not utilize it. This result is especially counter-intuitive, since many prior studies report similar closed-set performance for the triplet loss only and softmax CE loss only training. On the other hand, in the open-set case, the triplet loss trained network performs even better than the combination of the softmax CE and triplet losses. We conjecture that the success of triplet loss is a result of how it affects the gallery embedding space during training. Triplet loss increases with a smaller distance between the positive-negative pair and a higher distance between the positive-positive pair. As a result, the amount of push that a negative sample receives depends on the distance between the positive pair. This in turn pushes the negative sample more, if the distance between the positive pair is high, forcing the separation between identity clusters to be proportionate to the intra-identity variance. We argue that this property makes triplet loss better than other loss functions for an open-set operation with fixed-threshold verification. Another implication of Table I is that the center loss reduces the performance when it is used in combination with triplet loss only and softmax CE loss only. This result is consistent with our hypothesis on why the triplet loss works better, since the center loss tries to minimize the intra-identity variance in the embedding space as much as possible, which leads to the loss of intra-identity variance information.

Performance Boost of Gallery Refinement. Results in Table II show that, except for a single combination of methods, datasets, β and FARs, the gallery refinement consistently increases the open-set performance of all methods regardless of the β value and the dataset. Most remarkable increases are in the 1% FAR region, where 3.39% and 8.38% DIR increase is provided on the average for the Market1501 and DukeMTMC-reID datasets, respectively. The performance boost of GR is more prominent for the DukeMTMC-reID dataset averaging 5.00% DIR, while the average boost on the Market1501 dataset is 2.32%. We predict that this is because the DukeMTMC-reID data are challenging. In such a case, the intraclass (identity) variation in embedding space is expected to be higher and thus, GR is able to eliminate less informative and outlier embeddings efficiently. Another implication of results is that when FAR is increased by more than 1%, the improvement of GR generally drops. However, as the ideal operating point for any open-set person re-ID problem is to have 0% FAR and 100% DIR, we argue that the occurring drop in Δ is not important and does not contradict with practical demands.

Optimal Discard Ratio. The optimal α percentage for every setting combination is given in the respective cell of each method in Table II. For most of the cases, we observe that α_{opt} drops when FAR is increased. For very

low FAR cases such as 0.1%, α_{opt} may get as high as 90%. This corresponds to discarding most of the gallery entries and still improving open-set results. Another observation is that, unfortunately, α_{opt} shows a strong dependence on the FAR. This emphasizes how critical it is to determine the maximum acceptable FAR for a practical scenario. Further, Table II also reveals that α_{opt} values for the DukeMTMC-reID dataset are higher on the average, compared to the same FAR Market1501 results. We explain this again by stating that the DukeMTMC-reID dataset is more challenging than the Market1501.

Computational Cost. A thorough analysis of computational complexity is beyond the scope of this study. However, as the gallery refinement reduces the total number of entries within the gallery set by a significant ratio, and since the re-ID task is performed by calculating the distances between a given query and all gallery entries, it is clear that gallery refinement reduces the cost of retrieval. Assuming a linear computational complexity with respect to the number of gallery entries, the computational cost is expected to drop by the same percentage as the selected α. Furthermore, an additional computational cost decrease is also expected from the process of partial sorting of the query-to-gallery embedding-distance scores for ranking.

VI. Conclusions

In this study, we have investigated the problem of open-set person re-ID, which is rarely addressed in literature. Only a handful of papers are available that directly address the open-set variant as discussed in Section II-B. Moreover, up till now, the absence of a widely accepted open-set dataset made it also difficult to address the open-set person re-ID in detail. As the truly closed-set scenario is not practical for actual industrial applications, we conjecture that the work presented in this paper on open-set person re-ID is highly relevant for practical surveillance applications.

To alleviate the problem of open-set person re-ID dataset availability, we have conducted our tests on existing, popular closed-set datasets with modifications. These modifications involve the conversion of this data to the open-set case. This approach yields more representative evaluation, better training and it provides the possibility to compare the closed- and open-set, since the Market1501 and DukeMTMC-reID datasets have already been studied heavily for the closed-set scenario.

To the best of our knowledge, we have provided the first open-set performance benchmark in deep person re-ID. For an elaborate comparison, we have compared eight very recent and high-performing re-ID methods on popular, large-scale datasets. The operational conditions for these methods were detailed by choosing realistic parameter combinations that relate to surveillance in practice. A further contribution is the performance assessment of statistical analysis and the gallery refinement of gallery set embeddings. We have revealed that, the GR process could be efficiently exploited for realizing improvements in performance, which are particularly prominent for low FAR regions. The improvements in this region appear to be consistent for all investigated methods. Furthermore, we have identified the MGN, BoT and BDB algorithms with GR as the best practice for high performance.

Besides the deep learning architecture, another important design choice is the selection of a loss function. We evaluated different combinations of triplet, softmax and center loss functions in order to determine their impact on the open-set person re-ID performance. Our analysis has shown that the triplet loss is crucial for low FAR performance, while the contributions of center and softmax losses are circumstantial. In other words, we have found out that, unlike the closed-set case, utilizing a higher number of different losses does not necessarily improve performance in the low FAR, open-set case.

By simultaneously reducing computational cost of retrieval and increasing the open-set performance, we have the opinion that the post-processing of gallery embeddings helps in bridging the gap between industrial requirements and the use of re-ID systems. We predict that the next step for person re-ID lies in the challenging open-set case and a better cooperation with the industry under practical conditions.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Tunc Alkanat wrote the code and paper; Egor Bondarev and Peter H. N. De With reviewed the paper; all authors had approved the final version.

Acknowledgment

This research is funded by the European H2020 Interreg PASSAnT Project and Provincial Government of Noord-Brabant, The Netherlands.

References

[1] W. S. Zheng, S. Gong, and T. Xiang, "Transfer re-identification: From person to set-based verification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2650-2657.

[2] W. Zajdel, Z. Zivkovic, and B. Krose, "Keeping track of humans: Have I seen this person before?" in *Proc. IEEE International Conference on Robotics and Automation*, 2005, pp. 2081-2086.

[3] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2360-2367.

[4] K. Jüngling, C. Bodensteiner, and M. Arens, "Person re-identification in multi-camera networks," in *Proc. CVPR WORKSHOPS*, 2011, pp. 55-61.

[5] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, "Interest points harvesting in video sequences for efficient person identification," in *Proc. the Eighth International Workshop on Visual Surveillance-VS2008*, 2008.

[6] M. Hirzer, P. M. Roth, and H. Bischof, "Person re-identification by efficient impostor-based metric learning," in *Proc. IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*, 2012, pp. 203-208.

[7] B. Ma, Y. Su, and F. Jurie, "Local descriptors encoded by fisher vectors for person re-identification," in *Proc. European Conference on Computer Vision*, 2012, pp. 413-422.

[8] A. Mignon and F. Jurie, "Pcca: A new approach for distance learning from sparse pairwise constraints," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2666-2672.

[9] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. European Conference on Computer Vision*, 2008, pp. 262-275.

[10] H. Bouma, S. Borsboom, R. J. D. Hollander, S. H. Landsmeer, and M. Worring, "Re-identification of persons in multi-camera surveillance under varying viewpoints and illumination," in *Proc. SPIE-The International Society for Optical Engineering*, 2012, vol. 8359, p. 83590Q.

[11] M. Bäuml and R. Stiefelhagen, "Evaluation of local features for person re-identification in image sequences," in *Proc. 8th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2011, pp. 291-296.

[12] Y. Shen, W. Lin, J. Yan, M. Xu, J. Wu, and J. Wang, "Person re-identification with correspondence structure learning," in *Proc. IEEE International Conference on Computer Vision*, 2015, pp. 3200-3208.

[13] C. Liu, S. Gong, C. C. Loy, and X. Lin, "Person re-identification: What features are important?" in *Proc. European Conference on Computer Vision*, 2012, pp. 391-401.

[14] B. J. Prosser, W. S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person re-identification by support vector ranking," in *Proc. British Machine Vision Conference*, 2010, p. 6.

[15] E. Corvee, *et al.*, "Person re-identification using HAAR-based and DCD-based signature," in *Proc. the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2010, pp. 1-8.

[16] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2288-2295.

[17] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207-244, 2009.

[18] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2197-2206.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, 2012.

[20] X. Qian, Y. Fu, Y. G. Jiang, T. Xiang, and X. Xue, "Multi-scale deep learning architectures for person re-identification," in *Proc. IEEE International Conference on Computer Vision*, 2017, pp. 5399-5408.

[21] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 152-159.

[22] F. Wang, W. Zuo, L. Lin, D. Zhang, and L. Zhang, "Joint learning of single-image and cross-image representations for person re-identification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1288-1296.

[23] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," arXiv preprint arXiv:1703.07737, 2017.

[24] M. M. Kalayeh, E. Basaran, M. Gökmen, M. E. Kamasak, and M. Shah, "Human semantic parsing for person re-identification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1062-1071.

[25] X. Fan, W. Jiang, H. Luo, and M. Fei, "Spherereid: Deep hypersphere manifold embedding for person re-identification," *Journal of Visual Communication and Image Representation*, vol. 60, pp. 51-58, 2019.

[26] H. Jin, X. Wang, S. Liao, and S. Z. Li, "Deep person re-identification with improved embedding and efficient training," in *Proc. IEEE International Joint Conference on Biometrics*, 2017, pp. 261-267.

[27] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: A deep quadruplet network for person re-identification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 403-412.

[28] Y. J. Cho and K. J. Yoon, "Improving person re-identification via pose-aware multi-shot matching," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1354-1362.

[29] X. Zhang, *et al.*, "Alignedreid: Surpassing human-level performance in person re-identification," arXiv preprint arXiv:1711.08184, 2017.

[30] H. Liu, J. Feng, M. Qi, J. Jiang, and S. Yan, "End-to-end comparative attention networks for person re-identification," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3492-3506, 2017.

[31] R. Quispe and H. Pedrini, "Improved person re-identification based on saliency and semantic parsing with deep neural network models," arXiv preprint arXiv:1807.05618, 2018.

[32] B. Cancela, T. M. Hospedales, and S. Gong, "Open-world person re-identification by multi-label assignment inference," in *Proc. British Machine Vision Conference*, 2014.

[33] A. Bialkowski, S. Denman, S. Sridharan, C. Fookes, and P. Lucey, "A database for person re-identification in multi-camera surveillance networks," in *Proc. International Conference on Digital Image Computing Techniques and Applications*, 2012, pp. 1-8.

[34] S. Chan-Lang, Q. C. Pham, and C. Achard, "Closed and open-world person re-identification and verification," in *Proc. International Conference on Digital Image Computing: Techniques and Applications*, 2017, pp. 1-8.

[35] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by video ranking," in *Proc. European Conference on Computer Vision*, 2014, pp. 688-703.

[36] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. Scandinavian Conference on Image Analysis*, 2011, pp. 91-102.

[37] W. S. Zheng, S. Gong, and T. Xiang, "Towards open-world person re-identification by one-shot group-based verification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 3, pp. 591-606, 2016.

[38] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification," in *Proc. British Machine Vision Conference*, 2011, vol. 1, no. 2, p. 6.

[39] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, 2007, vol. 3, no. 5, pp. 1-7.

[40] S. Z. Chen, C. C. Guo, and J. H. Lai, "Deep ranking for person re-identification via joint representation learning," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2353-2367, 2016.

[41] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Proc. Asian Conference on Computer Vision*, 2012, pp. 31-44.

[42] X. Ma, X. Zhu, S. Gong, X. Xie, J. Hu, K. M. Lam, and Y. Zhong, "Person re-identification by unsupervised video matching," *Pattern Recognition*, vol. 65, pp. 197-210, 2017.

[43] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE International Conference on Computer Vision*, 2015, pp. 1116-1124.

[44] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *Proc. IEEE International Conference on Computer Vision*, 2017.

[45] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proc. European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking*, 2016.

[46] P. F. Felzenszwalb, *et al.*, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008, vol. 2, no. 6, p. 7.

[47] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2285-2294.

[48] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on

learning," in *Proc. Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
[49] T. Xiao, H. Li, W. Ouyang, and X. Wang, "Learning deep feature representations with domain guided dropout for person re-identification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1249-1258.
[50] X. Chang, T. M. Hospedales, and T. Xiang, "Multi-level factorisation net for person re-identification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2109-2118.
[51] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248-255.
[52] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *Proc. ACM Multimedia Conference on Multimedia Conference*, 2018, pp. 274-282.
[53] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818-2826.
[54] Z. Dai, M. Chen, X. Gu, S. Zhu, and P. Tan, "Batch dropblock network for person re-identification and beyond," arXiv preprint arXiv:1811.07130, 2018.
[55] G. Ghiasi, T. Y. Lin, and Q. V. Le, "Dropblock: A regularization method for convolutional networks," *Advances in Neural Information Processing Systems*, pp. 10727-10737, 2018.
[56] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, "Bag of tricks and a strong baseline for deep person re-identification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
[57] S. Liao, Z. Mo, J. Zhu, Y. Hu, and S. Z. Li, "Open-set person re-identification," arXiv preprint arXiv:1408.0872, 2014.

Tunc Alkanat was born in Izmir, Turkey. He received the B.S. and M.S. degrees in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 2013 and 2016, respectively.
He is currently working toward the Ph.D. degree at the Video Coding and Architectures Group, Eindhoven University of Technology, Eindhoven, The Netherlands. His research interests include image retrieval, anomaly detection, computer vision for surveillance and computational spectral imaging.

Egor Bondarev received his MSc degree in robotics and informatics at the State Polytechnic University, Belarus Republic, in 1997. In 2009 he has obtained his PhD degree in computer science at Eindhoven University of Technology (TU/e), The Netherlands in the research domain of performance predictions of real-time component-based systems on multiprocessor architectures.
Currently, he is an Assistant Professor at the Video Coding and Architectures group, TU/e, focusing on such research areas as multi-modal sensor fusion, smart surveillance with multi-camera systems and photorealistic 3D reconstruction of environments.
Dr. Bondarev has written and co-authored over 50 publications on real-time computer vision and image/3D processing algorithms. He is a leader of an internal research cluster on real-time data fusion from multi-modal sensors, such as thermal, depth, laser, RGB and acoustic. Egor Bondarev is involved in several European research projects and, currently, he is a TU/e project leader in the large international PANORAMA, PS-CRIMSON and APPS projects, all addressing challenges of multi-modal multi-camera smart surveillance.

Peter H. N. De With received his PhD degree (1992) from University of Technology Delft, The Netherlands. In 1984-1997, he worked for Philips Research Eindhoven on video compression and chaired a cluster for programmable TV architectures as senior TV Systems Architect. In 1997-2000, he was full professor at the University of Mannheim, Germany, Computer Engineering, and chair of Digital Circuitry and Simulation. In 2000-2007, he was with LogicaCMG in Eindhoven as a principal consultant and distinguished business consultant. He was also part-time professor at the TU/e, heading the chair on Video Coding and Architectures. In 2008-2010, he was VP Video (Analysis) Technology at Cyclomedia Technology. Since 2011, he has is full professor at Eindhoven University of Technology, faculty EE. From 2011-2018 he was scientific director of the Centre for Care and Cure Technologies.
Prof. Dr. Ir. De With has (co-)initiated an image analysis program with multiple regional and national hospitals for forms of oncology detection analysis, of which esophageal cancer detection projects became of leading international quality. He is also international expert in video surveillance for safety and security and has been involved in multiple EU projects on video analysis, object and behavior recognition with the industry. Dr. De With is Fellow of the IEEE and member of the Royal Holland Society of Sciences and Humanities. He is board member of the Gauss Foundation, DITSS and R&D advisor to multiple companies. He has (co-)authored over 400 papers on video coding, video analysis, architectures, 3D processing and their realization. He is (co-) recipient of multiple papers awards of the IEEE CES, VCIP and Transactions papers and a Eurasip Signal Processing award and holds some 30 patents.