Novelty Detection for Person Re-identification in an Open World

George Galanakis[1,2], Xenophon Zabulis[2] and Antonis A. Argyros[1,2]
[1]*Computer Science Department, University of Crete, Greece*
[2]*Institute of Computer Science, FORTH, Greece*

Keywords: Person Re-identification, Open World, Novelty Detection.

Abstract: A fundamental assumption in most contemporary person re-identification research, is that all query persons that need to be re-identified belong to a closed gallery of known persons, i.e., they have been observed and a representation of their appearance is available. For several real-world applications, this closed-world assumption does not hold, as image queries may contain people that the re-identification system has never observed before. In this work, we remove this constraining assumption. To do so, we introduce a novelty detection mechanism that decides whether a person in a query image exists in the gallery. The re-identification of persons existing in the gallery is easily achieved based on the persons representation employed by the novelty detection mechanism. The proposed method operates on a hybrid person descriptor that consists of both supervised (learnt) and unsupervised (hand-crafted) components. A series of experiments on public, state of the art datasets and in comparison with state of the art methods shows that the proposed approach is very accurate in identifying persons that have not been observed before and that this has a positive impact on re-identification accuracy.

1 INTRODUCTION

A key problem in vision-based person tracking is encountered when a person exits and then re-enters the field(s) of view of the cameras observing it. On that occasion, we wish to *re-identify* this person and associate it with its previous detection(s). Such a re-identification can support person tracking in large environments that are covered by multiple cameras, or improve single-view tracking against occlusions and viewing limitations.

Typically, re-identification methods assume that a *gallery set* contains person representations associated with person IDs. These representations may comprise of images as well as of global or local feature descriptors. A *query* or *probe image* of a person is compared against the gallery set, seeking a match with any of the persons therein. In most cases, the gallery is assumed to be a *closed-set*, i.e., it contains a representation for every person whose identity is going to be queried. In contrast, in *open-world person re-identification*, the gallery is assumed to be an *open-set*. As it is possible that this person has never been observed before, the query image may not match any of the persons in the gallery. In this context, prior to re-identification, it needs to be decided whether a person has been observed before or not. From a technical point of view, the open-world version of the problem is more challenging. At the same time, its solution can support a much more diverse set of application domains.

In this work, we treat the problem of open-world person re-identification. We do so by casting the problem of deciding whether a person belongs or not to the gallery as a *novelty detection* problem. A recent review on person re-identification (Zheng et al., 2016a) indicates novelty detection as an open issue. Novelty detection and the subsequent person re-identification operate on a newly proposed person descriptor that consists of a supervised (learnt) and an unsupervised (hand crafted) component.

Two major advantages of the proposed approach over existing ones are the following. First, it can be applied to a new setting (e.g., environment, set of cameras, gallery of known persons) without any additional training, as required by other learning approaches (Zheng et al., 2016b; Zhu et al., 2017). Second, it can be naturally adapted to operate in an online fashion. That is in contrast to methods as in (Wang et al., 2016) which require that the entire probe set is available, rather than gradually introduced. Therefore, we regard that our method constitutes a very attractive candidate for solving the person re-identification pro-

Galanakis, G., Zabulis, X. and Argyros, A.
Novelty Detection for Person Re-identification in an Open World.
DOI: 10.5220/0007368304010411
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 401-411
ISBN: 978-989-758-354-4

blem in open, real-world setups and scenarios. Such a scenario is considered in the CONNEXIONs Horizon 2020 project, which is funded by the European Commission, and will develop and demonstrate next-generation detection, prediction, prevention, and investigation services. In this context, the tracking and re-identification of persons in multiple heterogeneous cameras is of paramount importance.

2 RELATED WORK

Research topics that are relevant to person re-identification methods include the investigation of descriptors that can be used to represent persons as well as methods for comparing them. Re-identification methods themselves are categorized based on whether they operate under the assumption of an open or closed world. We focus in the second category, as this is the one treated in this paper by incorporating novelty detection mechanisms.

Features and Descriptors for Person Representation. Data-driven features have recently gained attention due to the proliferation of Convolutional Neural Networks (CNNs) and the availability of large-scale datasets for training. These factors gave rise to *learned* features, which are robust to illumination, scale and pose variations (Hermans et al., 2017; Chen et al., 2017; Zhou et al., 2017; Su et al., 2016; Wang et al., 2017; Qian et al., 2017; Li et al., 2018a; Song et al., 2018; Xu et al., 2018; Sarfraz et al., 2018), but also specific to the training data. *Hand-crafted* representations have also witnessed advances. Departing from conventional features such as color histograms, SIFT and HoG features, more sophisticated descriptors have been proposed (Farenzena et al., 2010; Liao et al., 2015; Matsukawa et al., 2016a; Gou et al., 2017). Context-specific knowledge has been also utilized, i.e., body-part segmentation (Su et al., 2017; Qian et al., 2017; Zheng et al., 2017a; Zhao et al., 2017) and attribute recognition (Shi et al., 2015; Su et al., 2016; Matsukawa and Suzuki, 2016; Qian et al., 2017; Chang et al., 2018). In greater relation to this work, learnt and hand-crafted features have been combined to support person re-identification (Wu et al., 2016).

Comparing Person Descriptors. Conventional measures such as Euclidean distance and cosine similarity have been adopted for this purpose (Farenzena et al., 2010; Wang et al., 2017). Learnt measures have been also used with better results (Zheng et al., 2016a). In such methods, discrepancies due to viewpoint and illumination variation between views are learned for specific view combinations (Yang et al., 2014; Liao et al., 2015; Martinel et al., 2015; Chen et al., 2015; Jose and Fleuret, 2016; Yu et al., 2017). CNNs have been also employed in this task (Hermans et al., 2017; Chen et al., 2017; Zhou et al., 2017; Wang et al., 2018; Xu et al., 2018). Pertinent methods are dataset-dependent, as they learn the change of person appearance across specific views.

Closed-world Person Re-identification. Closed-set re-identification methods mainly focus on a proper definition of the above components, i.e., discriminative features and/or distance metrics. Subsequently, most works perform pairwise comparison of query images against gallery ones. Pairwise comparisons result to a ranked list, on top of which lays the most probable gallery person (nearest neighbor). To improve results, some works exploit information about top-ranked persons to perform re-ranking (Zheng et al., 2015b; Lisanti et al., 2015; Zhong et al., 2017; Sarfraz et al., 2018). Other works, especially those that operate on videos, maintain images of a person in a set. Thus, comparison among query and gallery persons is formulated as set comparison (Wang et al., 2014) or graph matching (Ye et al., 2017).

During the evaluation of learning-based re-identification methods, datasets are split into *training* and *testing* subsets to avoid learning bias. Persons in these subsets are referred as *non-target* and *target*, respectively. Learning based on non-target persons is not practical in real-world applications, because it requires manual annotation of a large number of images.

Open-world Person Re-identification. Only a few works deal with open-world person re-identification. The decision on whether a person belongs to the gallery set has been approached as a novelty detection problem. Novelty detection was initially applied to object classification. In (Bodesheim et al., 2013; Liu et al., 2017) a discriminative null space is recommended, where images from the same, known, class are mapped to a single point in the null space. The most recent method (Liu et al., 2017) addresses the *incremental* case, where the gallery set is progressively expanded with new object representations. The novelty detection methods in (Bendale and Boult, 2016; Günther et al., 2017) employ normalization in class inclusion scores by applying Extreme Value Theory (EVT). In (Kliger and Fleishman, 2018), novel objects are detected using the Generative Adversarial Network (GAN) framework. Deep neural networks have been utilized by three recent works (Ruff et al., 2018; Masana et al., 2018; Perera and Patel, 2018) in an effort towards end-to-end feature extraction and novelty detection. The works in (Ruff et al., 2018; Perera and Patel, 2018) target one-class classification,

thus, in contrast to (Bodesheim et al., 2013), they do not expand naturally to multi-class novelty detection, but depend on some aggregation similar to One-Class SVMs (see Sec. 4). We also note that (Kliger and Fleishman, 2018; Ruff et al., 2018; Masana et al., 2018) are evaluated on datasets which are simpler and/or contain fewer classes (MNIST, CIFAR-10 etc). Moreover, if only few per class samples are available for training, application of DNNs is limited. Thereafter, we consider previous works as options for multi-class novelty detection.

A first example on how novelty detection is applied to open-world person re-identification appears in (Brun et al., 2011), where the kernel PCA algorithm is applied to graph representations of persons. An input representation is considered as novel if its squared distance from the first few of the principal components is above a certain threshold. This method is subject to careful parameter selection (number of principal components, kernel size), which is avoided in null space methods (Bodesheim et al., 2013). Three recent works (Wang et al., 2016; Zheng et al., 2016b; Zhu et al., 2017) deal with the open-world problem in very challenging settings. (Wang et al., 2016; Zheng et al., 2016b) use only one training image per person, while (Zhu et al., 2017) achieves computational efficiency in addition to accuracy. Although (Wang et al., 2016) depends only on target persons (unsupervised setting), its objective function requires the entire query set as input and can be applied only to a pair of views. This is particularly constraining as, in real-world settings, the entire query set is not available at once and more than two views are employed to achieve full coverage of the environment. The work in (Zheng et al., 2016b) capitalizes on a non-target dataset from all views, in order to project person representations to a new feature space. Intuitively, in this space, intra-identity representations are similar and inter-identity representations dissimilar. The work in (Zhu et al., 2017) proposes to learn two hashing functions, one for gallery and another one for probe images (obtained from disjoint views). Then, it compares hash codes using the Hamming distance. Both methods (Zheng et al., 2016b; Zhu et al., 2017) depend on a single dataset, containing images from specific viewpoints, captured at arbitrary illumination conditions and do not generalize to new conditions.

In the most recent work, (Li et al., 2018b), the GAN framework is utilized for learning how to discriminate between known and unknown persons. Their network architecture comprises of two types of discriminators; one between persons and other visual content, and the other between known and unknown persons. They consider a subset of the ids (persons) as known, while the rest of the ids are divided into training and testing. We argue that this method, although promising, is impractical in real-world scenarios. The reason is the following. Learning to discriminate between known versus unknown persons consists of utilization of both (a) known ids and (b) a set of few hundred unknown ids, during training. The later set essentially represents the negative samples, i.e. unknown persons. The method in (Li et al., 2018b) is demonstrated on training and testing datasets from same period of time and camera network ((Li et al., 2018b), Sec. 4.2). Therefore it depends on time-demanding acquisition and manual labeling of person images, apart from those of known ids. This assumption is fairly limiting. Moreover, the occurrence of an unknown id requires overly time-consuming training of the neural network that prohibits the real-time operation of the system. This time-consuming training is due to the incremental augmentation of the gallery of known persons.

Our Approach. In this work, we propose *OW-REID*, a novel method for open-world person re-identification. *OW-REID* operates on a person descriptor that consists of both hand-crafted and learnt features. It capitalizes on the Kernel Null Folley-Summon Transform (KNFST) (Bodesheim et al., 2013), a parameter-free novelty detection technique, to decide whether a person belongs to the gallery set or not. Novelty detection based on KNFST is possible without requiring an annotated dataset of non-target persons. This increases the exploitability of the method in real world conditions as it reduces dramatically its setup time and costs. The KNFST algorithm has been applied to re-identification (Zhang et al., 2016) but for addressing the closed-world version of the problem. In this context it is used as a feature mapping/reprojection technique and not as a novelty detection mechanism. More specifically, it is used to learn the transfer function normalizing person appearance across a number of views so that known people can be re-identified if observed by these cameras. In contrast, *OW-REID* learns a person-specific representation from a number of cameras, so that (a) these persons can be re-identified by these or other cameras and (b) unknown persons are identified as such.

Our Contribution. In summary, the contributions of this work are the following: (1) We propose *OW-REID*, a novel method for open-world person re-identification by treating it as a novelty detection problem; (2) We propose a hybrid person descriptor that consists of supervised (hand-crafted) and unsupervised (learnt) features. This contributes to better novelty detection accuracy and independence from the training data; (3) We provide an extensive, compa-

rative study and evaluation of the proposed approach against baseline methods and person descriptors on standard datasets.

3 OPEN-WORLD RE-IDENTIFICATION (*OW-REID*)

The proposed approach for novelty detection and open-world person re-identification (*OW-REID*) capitalizes on KNFST (Bodesheim et al., 2013), a method for detecting novel entities among general known object categories. In our framework, each category corresponds to a single person and contains different views of that person. KNFST is applied to person descriptors that represent a view of each person as a multidimensional feature vector.

3.1 Person Description

As person descriptors, we have considered both learnt and hand-crafted features that have been designed to optimize the accuracy of novelty detection and person re-identification.

As a first candidate, we considered the features proposed in (Hermans et al., 2017) where each image of a person is represented as an 128-D feature vector, called *TriNet*. TriNets are learned through training on two recent large-scale datasets, Market1501 (Zheng et al., 2015a) and MARS (zhe, 2016), achieving state-of-the-art re-identification accuracy in those particular datasets.

As a second candidate, we considered the recently proposed, hand-crafted GOG features (Matsukawa et al., 2016b). In (Matsukawa et al., 2016b), two variants for GOG are presented. The first, GOG_{rgb}, is obtained by applying the GOG descriptor on an RGB color representation of an image. In the second, $\text{GOG}_{\text{fusion}}$, the GOG descriptor is applied independently to the RGB, Lab, HSV and nRnG color encodings and the results are concatenated. In preliminary experiments, the GOG_{rgb} descriptor yielded better novelty detection accuracy and was adopted thereafter. In line with (Matsukawa et al., 2016b), we apply mean removal and $L2$ normalization. Also, using default settings, the GOG_{rgb} descriptor has $7,567$ dimensions. To retain only expressive components and improve accuracy, we retained 1000 components through PCA dimensionality reduction. Henceforth, we refer to the resultant descriptor simply as GOG.

In addition, we introduce *GOG+TriNet*, a new descriptor formed by the concatenation of the GOG and TriNet descriptors, in an effort to retain the advantages of both worlds.

3.2 Novelty Detection

Given a gallery set $\mathcal{G}$ and a query image q, we wish to deduce whether q depicts a person of known identity (i.e., whether $q \in \mathcal{G}$). Let c be the number of known persons. Query and gallery images are assumed to envelope persons, showing as less background as possible. Below we describe how KNFST (Bodesheim et al., 2013) is applied to solve this novelty detection problem.

Each image $\mathbf{x}$ (either query or gallery ones) are mapped to a descriptor/feature space $\mathcal{F}$ as discussed in Section 3.1. After this mapping has been performed, the original images are not utilized anymore. Let $i \in [1,c]$ enumerate persons in $\mathcal{G}$. Let also v_i be the number of images of person i, and $j \in [1,v_i]$ enumerate the images for person i. Then, $\mathbf{x}_{ij}$ denotes the person descriptor of the j^{th} image of the i^{th} person.

Let $\mathbf{X} \in \mathbb{R}^{n \times d}$ be a matrix containing a descriptor $\mathbf{x}_{ij}$ in each of its rows ($n = \sum_{i=1}^{c} v_i$ is the number of gallery images and d the dimensionality of $\mathbf{x}_{ij}$). The kernel $\mathbf{K} \in \mathbb{R}^{n \times n}$ is constructed to contain pairwise similarities for these n vectors. As in (Liu et al., 2017), pairwise similarities are measured utilizing the Radial Basis Function (RBF) kernel.

Using KNFST, a projection $P(\mathbf{x}) : \mathcal{F} \rightarrow \mathcal{N}$ is learned, where $\mathcal{N}$ denotes a $c-1$ null space. Each person i is represented by a point $\mathbf{t}_i \in \mathcal{N}$, $i \in [1,c]$, defined as $\mathbf{t}_i = \text{mean}_j(\mathbf{t}_{ij})$. The input parameters of KNFST are $\mathbf{K}$ and corresponding person labels. Details on how the projection is computed can be found in (Bodesheim et al., 2013). To provide some intuition, Fig. 1 provides a 2D visualization of how KNFST operates. Figure 1 left, shows 100 images of the RAiD dataset (10 images for each of 10 randomly chosen persons). Originally, these are 1128-D GOG+TriNet vectors in $\mathcal{F}$ that are projected on the 2D plane using t-SNE (Maaten and Hinton, 2008) to aid visualization. Images of the same person correspond to the same marker and color. We can see that images of the same person project to different points on the plane. Actually, for each person there are two clusters of points, each corresponding to images acquired by a different camera. Figure 1 right, shows a 2D projection of the representation of the same images in the $\mathcal{N}$ space. It can be verified that in this space, images of the same person are much more tightly clustered.

Novelty detection is then performed as follows. The feature vector $\mathbf{z}$ in $\mathcal{F}$ corresponding to a query image q is projected to point $\mathbf{t}$ in $\mathcal{N}$. A *novelty score* is computed, as the distance of $\mathbf{t}$ to its nearest neig-

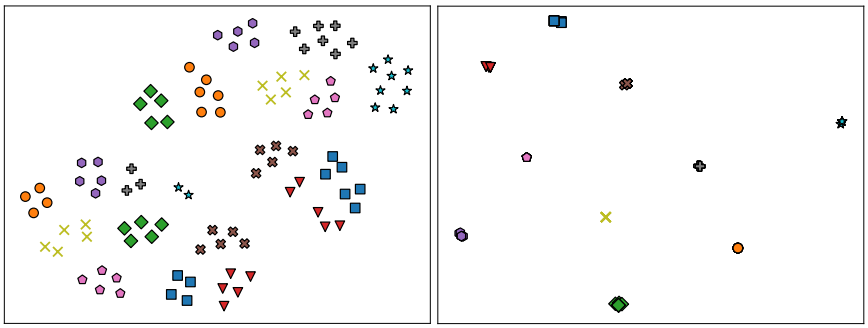

Figure 1: Visualization of feature vectors for 10 images per person and for 10 persons in the original feature space $\mathcal{F}$ (left) and in the KNFST's projection space $\mathcal{N}$ (right). Each person identity is represented by a unique marker and color. Both $\mathcal{F}$ and $\mathcal{N}$ spaces are projected to the 2D plane for visualization purposes (see text for details).

hbor $\mathbf{t}_i$ in $\mathcal{N}$ (ideally, the different persons in the gallery $\mathcal{G}$). If this distance is larger than a threshold θ, the query image is considered to be novel, i.e., depict a person that is not a member of the gallery. At the same time, if this distance is lower than θ, the identity of the nearest neighbor $\mathbf{t}_i$ of $\mathbf{t}$ in $\mathcal{N}$ solves the re-identification problem.

As soon as a novel person is detected, its views can be used to augment the gallery $\mathcal{G}$ and re-learn the KNFST-based mapping from $\mathcal{F}$ to $\mathcal{N}$. Thus, the proposed method can be turned into an online/incremental approach that extends an original gallery $\mathcal{G}$ of known persons, as views of previously unseen persons are encountered..

4 EXPERIMENTAL EVALUATION

To evaluate novelty detection, we adopt the recently proposed *cross-view scheme* (Zhu et al., 2017). In this scheme, $\mathcal{G}$ contains images from a subset of views (cameras). The query set $\mathcal{Q}$ contains images of both known and unknown persons from the remaining views. Considering the remaining views is particularly important because in a real-world setting, a person is typically first observed in a view and needs to be re-identified in another view, which usually is subject to different illumination conditions.

Evaluation datasets comprise of multiple images per person, per view. Persons in $\mathcal{G}$ and $\mathcal{Q}$ are imaged in a wide range of viewpoints and lighting conditions. Let $\mathcal{L}_g$ and $\mathcal{L}_q$ denote the disjoint sets of views in $\mathcal{G}$ and $\mathcal{Q}$, respectively; $\mathcal{L} = \mathcal{L}_g \cup \mathcal{L}_q$ denotes the set of all views. Let $\mathcal{Q}_k$ and $\mathcal{Q}_u$ denote the subsets of images of known and unknown persons, respectively; thus, $\mathcal{Q} = \mathcal{Q}_k \cup \mathcal{Q}_u$. In the experiments, we randomly choose the views in $\mathcal{L}_g$, to avoid bias to specific views. Similarly, to avoid bias over specific persons, we randomly chose the identities in $\mathcal{Q}_k$ and $\mathcal{Q}_u$. Finally, to avoid bias over a specific random selection of views or persons, we perform 300 random trials and report average accuracy, in each case. In all performed experiments, $\mathcal{G}$ is not incrementally updated, as it would be in a real-world system. This is to measure the accuracy of novelty detection over multiple trials and under the same conditions, regarding varying gallery sizes c and numbers of gallery images per person, v_i. Also, in line with (Bodesheim et al., 2013), relatively small values of c are considered (< 50).

Datasets. To evaluate *OW-REID*, three recent datasets have been employed, Market1501 (Zheng et al., 2015a), DukeMTMC-reID (Zheng et al., 2017b) and RAiD (Das et al., 2014). All datasets contain multiple persons and multiple views per person. For the first two datasets, images are split into training and testing sets. For these datasets, only the testing set is utilized in the evaluation, as the training set is reserved for learning methods (see Sec. 2). For RAiD, no such splitting is provided and the entire dataset is utilized.

The testing subsets of Market-1501 and DukeMTMC-reID datasets contain 750 and 1110 persons, observed from up to 6 and 8 cameras, respectively. For DukeMTMC-reID, we are interested in persons that are imaged in at least two views (702 in total), as required by the cross-view protocol. RAiD contains images for 43 walking persons from 4 views (2 indoor and 2 outdoor), with an exception of 2 persons that are imaged by only 3 views.

Novelty Detection Evaluation Metrics. Novelty detection accuracy is measured in (Bodesheim et al., 2013) using the area under the ROC curve, henceforth *auROC*. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR), for different values of θ. In our case, TPR denotes the ratio of persons correctly classified as unknown, while FPR denotes the ratio of known persons misclassified as unknown. The works in (Günther et al., 2017; Zheng et al., 2016b; Zhu et al., 2017) use the DIR

curve (Jain and Li, 2011), which correspondingly to the ROC curve, plots the Detection Identification Rate (DIR) over the False Alarm Rate (FAR), as a function of threshold θ. This curve has been also referred to as "True Target Rate (TTR) over False Target Rate (FTR)" in (Zheng et al., 2016b; Zhu et al., 2017). This measure is similar to ROC but, besides novelty detection, assesses identification accuracy as well. In this work, we borrow the idea from (Bodesheim et al., 2013) and use the area under the DIR curve, henceforth *auDIR*, which is determined based on FAR and DIR, where FAR is the ratio of unknown persons misclassified as known and DIR is the ratio of persons correctly classified as known and, also, correctly re-identified.

Experiment Types. We investigated two parameters that affect the performance of novelty detection: (a) the number of images v_i for each known person i, and (b) the number of known persons, c. A large value of v_i, implies that persons are imaged in a wide range of viewpoints and illumination conditions. In this case, novelty detection is provided with diverse gallery data for each person, which we anticipate as beneficial for novelty detection accuracy. In contrast, a large value of c implies greater ambiguity in novelty detection, as there is a greater chance for accidental similarity of an unknown person to a person in $\mathcal{G}$.

Given the above, novelty detection accuracy was assessed in a wide variety of settings, in two experiments, the **VPV** (varying person views) and the **VGS** (varying gallery size) experiments. More specifically, in the **VPV** experiment, the number of known persons was constant ($c = 20$) and the number of available images per person, v_i, was modulated as follows. We required v_i to be in range $[v_l, v_u]$ and systematically varied the limits of this range. The specific value of v_i depends on how many images are available for person i in the particular trial. If there are not enough gallery images ($v_i < v_l$), person i is considered as unknown. If more than enough images are available ($v_i > v_u$), we randomly choose a subset of v_u images from the available ones. As mentioned earlier, the datasets do not have the same number of images per person. For this reason, we evaluate these datasets in suitable ranges (see Section 4). In the **VGS** experiment, values $v_l = 20$, $v_u = 50$ were kept constant and c varied in the range $\{1,5,10,\ldots,50\}$, for all datasets.

Given the dataset sizes, in all experiments, we chose $|\mathcal{L}_g| = 5$ for the Market1501 and DukeMTMC-reID and $|\mathcal{L}_g| = 2$ for RAiD. In all experiments, the TriNet features used were as originally trained on the Market1501 dataset.

Comparison with Baselines. We compare the performance of *OW-REID* against two baseline methods: (a) distance thresholding in the original feature space and (b) Multiple One-Class SVMs. In (a), we compute pairwise distances of a query $\mathbf{z} \in \mathcal{F}$ against all $\mathbf{x}_{ij}$, as measured by the Euclidean (L2) distance. Query $\mathbf{z}$ is considered to belong to an unknown person, if the maximum of pairwise distances is larger than a threshold θ. For (b), we used the νSVM variant (Schölkopf et al., 2000), with $\nu = 0.5$ and $\gamma = 1/d$, where d is the dimensionality of $\mathcal{F}$. We utilized c One-Class SVM models, one for each person. Each model m_i is learned by utilizing all the v_i feature vectors $\mathbf{x}_{ij}$ for person i. Then, we compute the pairwise scores of $\mathbf{z}$ against every m_i, select the best score and compare it to threshold θ, as above. We note that we use the inverse distance as novelty score.

Tables 1 and 2 summarize mean auROC, of the compared methods for the **VPV** experiment (variable v_l) and **VGS** experiment (variable c), respectively. Tables 3 and 4 summarize auDIR quantities, for the same experiments. TriNet features dominate performance in the Market1501 dataset, independently of the novelty detection method. That is attributed to the fact that TriNet features were defined through training on exactly this dataset. For the other two datasets, we observe significant performance drop when TriNet features are utilized, independently of the novelty detection method. We conclude that learning features in a certain dataset and context, proves inadequate for supporting accurate novelty detection in other datasets, obtained under different conditions. We also stress that re-training features in a new setting is a cumbersome task. First, because it requires laborious manual annotation for the gallery images and, second, because in an open-world setting, examples of unseen persons are not available. In contrast, the proposed combination of TriNet with hand-crafted features yields the best performance, when an appropriate novelty detection method is utilized.

Indeed, the selection of the novelty detection method plays a significant role in the obtained performance. When hand-crafted (GOG) features are utilized, better results are obtained using *OW-REID*. Even more importantly, *OW-REID* combined with GOG+TriNet achieves a better overall performance against the other compared methods. *OW-REID* performs worse only for very small values of c, v_l. In these cases, the limited number of samples make it difficult for KNFST to learn an appropriate projection space $\mathcal{N}$. Nonetheless, *OW-REID* performs well when $v_l \geq 10$, an assumption that can easily be met in practice. *OW-REID* is computationally more efficient than the L2 baseline, because it compares a query with a single point in $\mathcal{N}$, instead of comparing it to all images in the gallery.

Table 1: Mean auROC for the **VPV** experiment (varying person views, v_l).

	method	features	1	5	10	15	20	25	30	35	40	45	50
Market1501	*OW-REID*	TriNet	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	-	-
		GOG	0.81	0.82	0.83	0.85	0.85	0.86	0.86	0.87	0.88	-	-
		GOG+TriNet	0.95	0.96	0.96	0.97	0.97	0.97	0.96	0.97	0.97	-	-
	OSVM	TriNet	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	-	-
		GOG	0.82	0.82	0.82	0.84	0.83	0.83	0.81	0.83	0.86	-	-
		GOG+TriNet	0.96	0.96	0.96	0.97	0.96	0.96	0.95	0.96	0.97	-	-
	L2	TriNet	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	-	-
		GOG	0.80	0.79	0.79	0.80	0.79	0.80	0.79	0.80	0.84	-	-
		GOG+TriNet	0.94	0.94	0.94	0.94	0.93	0.93	0.93	0.93	0.95	-	-

	method	features	1	5	10	15	20	25	30	35	40	45	50
DukeMTMC-reID	*OW-REID*	TriNet	0.71	0.71	0.72	0.72	0.74	0.74	0.76	0.79	0.79	0.85	**-**
		GOG	0.70	0.71	0.76	0.79	0.82	0.79	0.80	0.80	0.80	0.84	-
		GOG+TriNet	0.74	0.75	**0.79**	**0.82**	**0.84**	**0.82**	**0.83**	**0.83**	**0.83**	**0.86**	-
	OSVM	TriNet	0.70	0.70	0.72	0.71	0.71	0.73	0.75	0.78	0.77	0.79	-
		GOG	0.68	0.69	0.71	0.74	0.74	0.70	0.71	0.73	0.73	0.77	-
		GOG+TriNet	0.74	0.74	0.76	0.77	0.77	0.73	0.74	0.77	0.78	0.80	-
	L2	TriNet	**0.76**	**0.76**	0.76	0.75	0.76	0.77	0.78	0.80	0.79	0.82	-
		GOG	0.69	0.69	0.71	0.72	0.73	0.71	0.70	0.71	0.70	0.74	-
		GOG+TriNet	0.74	0.75	0.75	0.76	0.77	0.73	0.72	0.73	0.73	0.76	-

	method	features	2	5	10	15	20	30	40	50	60	70	80
RAiD	*OW-REID*	TriNet	0.65	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.70	0.71
		GOG	0.63	0.66	0.69	0.71	0.72	0.72	0.73	0.75	0.74	0.75	0.76
		GOG+TriNet	0.67	**0.73**	**0.76**	**0.78**	**0.79**	**0.78**	**0.79**	**0.80**	**0.76**	**0.78**	**0.79**
	OSVM	TriNet	0.64	0.67	0.68	0.69	0.70	0.69	0.69	0.70	0.69	0.69	0.70
		GOG	0.62	0.67	0.69	0.71	0.71	0.70	0.71	0.71	0.71	0.71	0.72
		GOG+TriNet	**0.69**	0.73	0.74	0.76	0.76	0.75	0.76	0.76	0.75	0.75	0.76
	L2	TriNet	0.69	0.72	0.72	0.74	0.74	0.73	0.73	0.74	0.74	0.74	0.75
		GOG	0.61	0.64	0.66	0.67	0.67	0.66	0.68	0.68	0.68	0.68	0.69
		GOG+TriNet	0.68	0.72	0.73	0.74	0.74	0.72	0.73	0.73	0.73	0.73	0.74

Table 2: Mean auROC for the **VGS** experiment (varying gallery size, c).

	method	features	1	5	10	15	20	25	30	35	40	45	50
Market1501	*OW-REID*	TriNet	0.72	0.99	0.99	0.98	0.98	0.98	0.98	0.97	0.97	0.97	0.97
		GOG	0.34	0.52	0.75	0.83	0.86	0.87	0.87	0.87	0.83	0.86	0.86
		GOG+TriNet	0.99	0.71	0.91	0.96	0.97	0.97	0.97	0.97	0.96	0.96	0.96
	OSVM	TriNet	0.55	0.99	0.99	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.97
		GOG	0.67	0.87	0.87	0.86	0.84	0.83	0.81	0.81	0.81	0.80	0.79
		GOG+TriNet	0.58	0.98	0.98	0.97	0.96	0.96	0.95	0.95	0.95	0.94	0.93
	L2	TriNet	**1.00**	**1.00**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.98**	**0.98**	**0.98**	**0.98**
		GOG	0.65	0.84	0.84	0.83	0.81	0.79	0.78	0.77	0.77	0.75	0.75
		GOG+TriNet	0.96	0.97	0.97	0.95	0.94	0.93	0.92	0.91	0.91	0.90	0.89

	method	features	1	5	10	15	20	25	30	35	40	45	50
DukeMTMC-reID	*OW-REID*	TriNet	0.79	0.77	0.76	0.74	0.74	0.73	0.72	0.72	0.70	0.69	0.69
		GOG	0.65	0.64	0.78	0.81	0.82	0.82	0.81	0.81	0.77	0.77	0.78
		GOG+TriNet	0.74	0.72	**0.82**	**0.84**	**0.84**	**0.84**	**0.83**	**0.83**	**0.78**	**0.77**	**0.78**
	OSVM	TriNet	0.78	0.77	0.76	0.72	0.70	0.71	0.69	0.68	0.66	0.65	0.65
		GOG	0.79	0.77	0.78	0.75	0.74	0.74	0.71	0.71	0.68	0.68	0.67
		GOG+TriNet	0.84	0.82	0.82	0.78	0.77	0.76	0.74	0.74	0.71	0.71	0.70
	L2	TriNet	**0.85**	**0.85**	0.81	0.77	0.76	0.75	0.73	0.72	0.71	0.70	0.70
		GOG	0.76	0.73	0.77	0.73	0.73	0.73	0.71	0.72	0.70	0.69	0.69
		GOG+TriNet	0.83	0.81	0.81	0.77	0.77	0.76	0.73	0.74	0.72	0.71	0.71

	method	features	1	5	10	15	20	25	30	35	40	45	50
RAiD	*OW-REID*	TriNet	0.62	0.76	0.73	0.71	0.69	0.68	0.67	0.67	0.64	**-**	**-**
		GOG	0.42	0.55	0.68	0.73	0.74	0.74	0.75	0.76	0.75	-	-
		GOG+TriNet	0.82	0.70	0.77	**0.79**	**0.79**	**0.76**	**0.76**	**0.77**	**0.75**	-	-
	OSVM	TriNet	0.63	0.78	0.73	0.71	0.69	0.68	0.67	0.67	0.65	-	-
		GOG	0.64	0.75	0.73	0.72	0.71	0.71	0.69	0.69	0.68	-	-
		GOG+TriNet	0.62	0.83	**0.79**	0.77	0.75	0.74	0.73	0.73	0.71	-	-
	L2	TriNet	**0.92**	**0.83**	0.79	0.75	0.73	0.72	0.71	0.70	0.68	-	-
		GOG	0.60	0.70	0.70	0.69	0.67	0.68	0.67	0.66	0.65	-	-
		GOG+TriNet	0.80	0.81	0.77	0.75	0.73	0.72	0.71	0.70	0.68	-	-

Comparison with the State of the Art. The method in (Wang et al., 2016) is not compatible to our evaluation protocol, because it only applies to a single view pair and requires all query data as a batch, rather than as a stream, as typically happens in real-world applications and in *OW-REID*. Also, none of the methods presented in (Zheng et al., 2016b; Zhu et al., 2017) is compatible to our experimental protocol, as their operation requires the availability of an annotated, non-target dataset. Aiming at a flexible and easy-to-setup method, *OW-REID* does not have this requirement. Additionally, there is no available implementation for (Zheng et al., 2016b; Zhu et al., 2017), so they cannot be tested according to our experimental protocol. However, in order to provide a basis of comparison of *OW-REID* to state of the art

Table 3: Mean auDIR for the **VPV** experiment (varying person views, v_l).

	method	features	1	5	10	15	20	25	30	35	40	45	50
Market1501	KNFST	TriNet	0.97	0.98	0.98	0.98	0.97	0.97	0.96	0.96	0.96	-	-
		GOG	0.70	0.71	0.72	0.75	0.74	0.76	0.74	0.74	0.74	-	-
		GOG+TriNet	0.93	0.94	0.94	0.95	0.94	0.95	0.94	0.93	0.93	-	-
	OSVM	TriNet	0.96	0.96	0.97	0.98	0.97	0.97	0.96	0.97	0.96	-	-
		GOG	0.71	0.71	0.71	0.74	0.73	0.74	0.71	0.72	0.72	-	-
		GOG+TriNet	0.93	0.94	0.94	0.95	0.94	0.94	0.93	0.93	0.93	-	-
	L2	TriNet	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.97**	**0.97**	**0.97**	-	-
		GOG	0.69	0.69	0.69	0.71	0.70	0.71	0.68	0.68	0.70	-	-
		GOG+TriNet	0.92	0.92	0.92	0.93	0.91	0.91	0.90	0.90	0.91	-	-

	method	features	1	5	10	15	20	25	30	35	40	45	50
DukeMTMC-reID	KNFST	TriNet	0.54	0.55	0.56	0.56	0.58	0.54	0.53	0.52	0.54	0.63	**-**
		GOG	0.47	0.49	0.54	0.60	0.64	0.59	0.59	0.56	0.57	0.65	-
		GOG+TriNet	**0.56**	**0.58**	**0.62**	**0.67**	**0.69**	**0.64**	**0.63**	**0.60**	**0.62**	**0.70**	-
	OSVM	TriNet	0.51	0.53	0.54	0.54	0.55	0.51	0.51	0.49	0.51	0.57	-
		GOG	0.46	0.48	0.51	0.56	0.59	0.52	0.52	0.52	0.53	0.60	-
		GOG+TriNet	0.56	0.57	0.59	0.63	0.64	0.58	0.58	0.56	0.59	0.65	-
	L2	TriNet	0.55	0.56	0.57	0.57	0.58	0.54	0.53	0.52	0.53	0.59	-
		GOG	0.46	0.48	0.51	0.55	0.59	0.54	0.53	0.51	0.52	0.59	-
		GOG+TriNet	0.56	0.58	0.59	0.62	0.65	0.59	0.57	0.55	0.57	0.64	-

	method	features	2	5	10	15	20	30	40	50	60	70	80
RAiD	KNFST	TriNet	0.37	0.46	0.48	0.49	0.49	0.48	0.48	0.48	0.49	0.50	0.51
		GOG	0.36	0.44	0.49	0.52	0.54	0.53	0.55	0.56	0.56	0.57	0.59
		GOG+TriNet	**0.46**	**0.56**	**0.60**	**0.63**	**0.64**	**0.63**	**0.64**	**0.65**	**0.63**	**0.64**	**0.66**
	OSVM	TriNet	0.37	0.43	0.46	0.48	0.48	0.47	0.47	0.48	0.48	0.49	0.50
		GOG	0.37	0.46	0.50	0.53	0.54	0.52	0.54	0.54	0.54	0.54	0.56
		GOG+TriNet	0.45	0.55	0.58	0.61	0.62	0.60	0.61	0.61	0.62	0.62	0.63
	L2	TriNet	0.40	0.47	0.49	0.52	0.53	0.51	0.51	0.51	0.52	0.53	0.54
		GOG	0.35	0.43	0.47	0.50	0.51	0.49	0.51	0.51	0.52	0.52	0.53
		GOG+TriNet	0.44	0.54	0.57	0.60	0.60	0.58	0.60	0.60	0.60	0.60	0.61

Table 4: Mean auDIR for the **VGS** experiment (varying gallery size, c).

	method	features	1	5	10	15	20	25	30	35	40	45	50
Market1501	KNFST	TriNet	0.72	0.98	0.98	0.98	0.97	0.97	0.97	0.96	0.96	0.96	0.95
		GOG	0.34	0.48	0.68	0.75	0.76	0.75	0.74	0.74	0.69	0.70	0.69
		GOG+TriNet	0.99	0.71	0.90	0.94	0.95	0.95	0.94	0.94	0.92	0.92	0.92
	OSVM	TriNet	0.55	0.98	0.98	0.98	0.97	0.97	0.96	0.96	0.96	0.96	0.95
		GOG	0.67	0.83	0.81	0.78	0.75	0.72	0.70	0.69	0.68	0.67	0.64
		GOG+TriNet	0.58	0.97	0.97	0.95	0.94	0.94	0.93	0.92	0.91	0.91	0.90
	L2	TriNet	**1.00**	**0.99**	**0.99**	**0.98**	**0.98**	**0.98**	**0.97**	**0.97**	**0.97**	**0.97**	**0.96**
		GOG	0.65	0.80	0.78	0.76	0.72	0.69	0.67	0.66	0.65	0.63	0.61
		GOG+TriNet	0.96	0.97	0.95	0.94	0.92	0.91	0.89	0.89	0.88	0.87	0.86

	method	features	1	5	10	15	20	25	30	35	40	45	50
DUKEMTMC-reID	KNFST	TriNet	0.72	0.70	0.66	0.61	0.58	0.56	0.54	0.53	0.51	0.50	0.49
		GOG	0.57	0.55	0.66	0.65	0.65	0.63	0.60	0.60	0.56	0.56	0.55
		GOG+TriNet	0.67	0.66	**0.73**	**0.72**	**0.70**	**0.69**	**0.66**	**0.65**	**0.59**	**0.59**	**0.59**
	OSVM	TriNet	0.71	0.69	0.65	0.58	0.54	0.53	0.50	0.48	0.45	0.44	0.43
		GOG	0.71	0.68	0.67	0.61	0.59	0.57	0.53	0.53	0.50	0.49	0.48
		GOG+TriNet	**0.77**	0.75	0.73	0.67	0.65	0.63	0.59	0.59	0.55	0.55	0.53
	L2	TriNet	0.77	**0.76**	0.69	0.61	0.58	0.56	0.52	0.51	0.49	0.48	0.46
		GOG	0.68	0.65	0.66	0.60	0.59	0.58	0.54	0.54	0.53	0.51	0.50
		GOG+TriNet	0.76	0.74	0.73	0.67	0.65	0.64	0.60	0.60	0.58	0.57	0.56

	method	features	1	5	10	15	20	25	30	35	40	45	50
RAiD	KNFST	TriNet	0.62	0.66	0.58	0.51	0.48	0.45	0.44	0.42	0.39	**-**	**-**
		GOG	0.42	0.48	0.56	0.57	0.56	0.54	0.53	0.52	0.51	-	-
		GOG+TriNet	0.82	0.65	0.68	**0.67**	**0.65**	**0.60**	**0.59**	**0.58**	**0.55**	-	-
	OSVM	TriNet	0.63	0.68	0.58	0.51	0.48	0.44	0.43	0.41	0.39	-	-
		GOG	0.64	0.67	0.62	0.57	0.53	0.51	0.49	0.48	0.45	-	-
		GOG+TriNet	0.62	**0.77**	**0.70**	0.64	0.61	0.58	0.56	0.55	0.52	-	-
	L2	TriNet	**0.92**	0.71	0.62	0.55	0.51	0.48	0.47	0.44	0.42	-	-
		GOG	0.60	0.63	0.58	0.55	0.51	0.49	0.47	0.45	0.43	-	-
		GOG+TriNet	0.80	0.75	0.68	0.63	0.60	0.57	0.55	0.53	0.50	-	-

methods, we present its performance comparatively to (Zhu et al., 2017), on the Market1501 dataset on which (Zhu et al., 2017) has been trained and evaluated.

Table 5 provides the results of this comparison. (Zhu et al., 2017) utilizes 500 random persons for training, which is in contrast to the training set provided by the authors of the Market1501 dataset (751 training persons). TriNet features are learned using the standard training set, thus, to avoid bias, and inline with our previous experiments, we utilized only the testing part of the dataset. To compensate for randomness, we repeat for 300 trials. The rest parameters were as in (Zhu et al., 2017), i.e., $|\mathcal{L}_g| = 2$ and $c = 10$. Under these settings, v_l is small; as discussed earlier, this is unfavorable to *OW-REID* which performs

Figure 2: Qualitative results on the RAiD dataset. Rows 1-2: gallery images. Rows 3-6: correct (green) and incorrect (red) novelty detection results (see text for details).

Table 5: True Target Rate (TTR, %) at varying False Target Rates (FTRs, %).

Method	1%	5%	10%	20%	30%
(Zhu et al., 2017)	**26.81**	**52.73**	66.47	79.66	86.16
OW-REID with GOG+TriNet (proposed)	13.53	46.20	**66.66**	**82.10**	**87.96**

much better when a larger number of views per person is available ($v_l \geq 10$). Nevertheless, we observe that *OW-REID* performs better than (Zhu et al., 2017) for larger FTR. Moreover, we note that *OW-REID* is directly applicable to new datasets, without the need for learning features or metrics under the new conditions.

Qualitative Results. Figure 2 shows qualitative results from the application of *OW-REID* on the RAiD dataset. The first two rows show representative images in a gallery of 10 persons. Two images per person are shown, each from a different view. Subsequent rows show indicative results of (a) known persons correctly identified as such (true negative), (b) known persons incorrectly identified as unknown (false positive), (c) unknown persons correctly identified as such (true positive) and (d) unknown persons incorrectly identified as known (false negative).

Overall, we observe that our method was successful in identifying known persons as such, even under different pose and/or lighting settings (third row). Nevertheless, some false positives occur as depicted in the fourth row. Such false positives never occurred in cases where the person's clothing color is prominent, e.g. orange and striped cyan t-shirts (4th and 8th column of gallery images, respectively), due to

less ambiguity. A common cause of false positives is the intervention of another person (cases 5 and 6). Similarly, we observe that a variety of unknown persons were identified as such (fifth row), while most of false negatives (sixth row) share similar appearance to gallery persons. Some obvious mistakes, such as the fourth case in sixth row, are considered as shortcomings of the method and require further investigation. Finally, we should note that there exist corner cases, where a person is identified as known due to some images, while as unknown due to others. Therefore, if multiple images are available (e.g. video sequence), we should consider aggregation of multiple novelty identifications to a single dominant result.

5 SUMMARY AND FUTURE WORK

We presented a new approach to the problem of open-world person re-identification. The method is based on a novelty detection technique appied to person descriptors that do not require re-training when the system is required to operate in a new setting (new cameras, illumination conditions, etc). Extensive experiments indicated the improved performance of the proposed approach over baseline and state of the art methods. Ongoing work includes testing of the approach when it operates in an online fashion, i.e., in a setup that involves the incremental learning (Liu et al., 2017) and incorporation of novel persons in the gallery of known persons. Future work aims at the utilization of multiple query images per person (MvsM scenario (Lisanti et al., 2015)) and further person description models, to further increase novelty detection accuracy.

ACKNOWLEDGMENTS

This work is partially funded by the H2020 projects CONNEXIONs (GA 786731) and Co4Robots (GA 731869).

REFERENCES

(2016). *MARS: A Video Benchmark for Large-Scale Person Re-identification*.

Bendale, A. and Boult, T. E. (2016). Towards Open Set Deep Networks. In *CVPR*.

Bodesheim, P., Freytag, A., Rodner, E., Kemmler, M., and Denzler, J. (2013). Kernel null space methods for novelty detection. In *CVPR*.

Brun, L., Conte, D., Foggia, P., and Vento, M. (2011). People re-identification by graph kernels methods. In *International Workshop, GbRPR*.

Chang, X., Hospedales, T. M., and Xiang, T. (2018). Multi-level factorisation net for person re-identification. In *CVPR*.

Chen, J., Zhang, Z., and Wang, Y. (2015). Relevance metric learning for person re-identification by exploiting listwise similarities. *IEEE Transactions on Image Processing*, 24(12):4741–4755.

Chen, W., Chen, X., Zhang, J., and Huang, K. (2017). Beyond triplet loss: a deep quadruplet network for person re-identification. In *CVPR*.

Das, A., Chakraborty, A., and Roy-Chowdhury, A. K. (2014). Consistent re-identification in a camera network. In *ECCV*.

Farenzena, M., Bazzani, L., Perina, A., Murino, V., and Cristani, M. (2010). Person re-identification by symmetry-driven accumulation of local features. In *CVPR*.

Gou, M., Camps, O., and Sznaier, M. (2017). moM: Mean of Moments Feature for Person Re-identification. In *CVPR*.

Günther, M., Cruz, S., Rudd, E. M., and Boult, T. E. (2017). Toward Open-Set Face Recognition. *CoRR*, abs/1705.01567.

Hermans, A., Beyer, L., and Leibe, B. (2017). In Defense of the Triplet loss for Person Re-Identification. *CoRR*, abs/1703.07737.

Jain, A. K. and Li, S. Z. (2011). *Handbook of face recognition*. Springer.

Jose, C. and Fleuret, F. (2016). Scalable metric learning via weighted approximate rank component analysis. In *ECCV*.

Kliger, M. and Fleishman, S. (2018). Novelty Detection with GAN. cite arxiv:1802.10560.

Li, W., Zhu, X., and Gong, S. (2018a). Harmonious attention network for person re-identification. In *CVPR*.

Li, X., Wu, A., and Zheng, W.-S. (2018b). Adversarial open-world person re-identification. In *ECCV*.

Liao, S., Hu, Y., Zhu, X., and Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. In *CVPR*.

Lisanti, G., Masi, I., Bagdanov, A. D., and Del Bimbo, A. (2015). Person re-identification by iterative re-weighted sparse ranking. *PAMI*, 37(8):1629–1642.

Liu, J., Lian, Z., Wang, Y., and Xiao, J. (2017). Incremental Kernel Null Space Discriminant Analysis for Novelty Detection. In *CVPR*.

Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov):2579–2605.

Martinel, N., Micheloni, C., and Foresti, G. L. (2015). Kernelized saliency-based person re-identification through multiple metric learning. *IEEE Transactions on Image Processing*, 24(12):5645–5658.

Masana, M., Ruiz, I., Serrat, J., van de Weijer, J., and Lopez, A. M. (2018). Metric learning for novelty and anomaly detection. *arXiv preprint arXiv:1808.05492*.

Matsukawa, T., Okabe, T., Suzuki, E., and Sato, Y. (2016a). Hierarchical gaussian descriptor for person re-identification. In *CVPR*.

Matsukawa, T., Okabe, T., Suzuki, E., and Sato, Y. (2016b). Hierarchical Gaussian Descriptor for Person Re-identification. In *CVPR*, pages 1363–1372.

Matsukawa, T. and Suzuki, E. (2016). Person re-identification using CNN features learned from combination of attributes. In *ICPR*.

Perera, P. and Patel, V. M. (2018). Learning deep features for one-class classification. *arXiv preprint arXiv:1801.05365*.

Qian, X., Fu, Y., Wang, W., Xiang, T., Wu, Y., Jiang, Y.-G., and Xue, X. (2017). Pose-Normalized Image Generation for Person Re-identification. *arXiv preprint arXiv:1712.02225*.

Ruff, L. et al. (2018). Deep one-class classification. In *ICML*.

Sarfraz, M. S., Schumann, A., Eberle, A., and Stiefelhagen, R. (2018). A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In *CVPR*.

Schölkopf, B., Smola, A. J., Williamson, R. C., and Bartlett, P. L. (2000). New support vector algorithms. *Neural computation*, 12(5):1207–1245.

Shi, Z., Hospedales, T. M., and Xiang, T. (2015). Transferring a semantic representation for person re-identification and search. In *CVPR*.

Song, C., Huang, Y., Ouyang, W., and Wang, L. (2018). Mask-guided contrastive attention model for person re-identification. In *CVPR*.

Su, C., Li, J., Zhang, S., Xing, J., Gao, W., and Tian, Q. (2017). Pose-driven Deep Convolutional Model for Person Re-identification. In *ICCV*, pages 3980–3989. IEEE.

Su, C., Zhang, S., Xing, J., Gao, W., and Tian, Q. (2016). Deep attributes driven multi-camera person re-identification. In *ECCV*.

Wang, H., Zhu, X., Xiang, T., and Gong, S. (2016). Towards unsupervised open-set person re-identification. In *ICIP*.

Wang, J., Wang, Z., Gao, C., Sang, N., and Huang, R. (2017). Deeplist: Learning deep features with adaptive listwise constraint for person reidentification. *TCSVT*, 27(3):513–524.

Wang, J., Zhou, S., Wang, J., and Hou, Q. (2018). Deep ranking model by large adaptive margin learning for person re-identification. *Pattern Recognition*, 74:241–252.

Wang, T., Gong, S., Zhu, X., and Wang, S. (2014). Person re-identification by video ranking. In *ECCV*.

Wu, S., Chen, Y.-C., Li, X., Wu, A.-C., You, J.-J., and Zheng, W.-S. (2016). An enhanced deep feature representation for person re-identification. In *WACV*.

Xu, J., Zhao, R., Zhu, F., Wang, H., and Ouyang, W. (2018). Attention-aware compositional network for person re-identification. In *CVPR*.

Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., and Li, S. Z. (2014). Salient color names for person re-identification. In *ECCV*.

Ye, M., Ma, A. J., Zheng, L., Li, J., and Yuen, P. C. (2017). Dynamic label graph matching for unsupervised video re-identification. In *ICCV*.

Yu, H.-X., Wu, A., and Zheng, W.-S. (2017). Cross-view asymmetric metric learning for unsupervised person re-identification. In *ICCV*.

Zhang, L., Xiang, T., and Gong, S. (2016). Learning a discriminative null space for person re-identification. In *CVPR*.

Zhao, H., Tian, M., Sun, S., Shao, J., Yan, J., Yi, S., Wang, X., and Tang, X. (2017). Spindle Net: Person re-identification with Human Body Region Guided Feature Decomposition and Fusion. In *CVPR*.

Zheng, L., Huang, Y., Lu, H., and Yang, Y. (2017a). Pose invariant embedding for deep person re-identification. *arXiv preprint arXiv:1701.07732*.

Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., and Tian, Q. (2015a). Scalable Person Re-identification: A Benchmark. In *ICCV*.

Zheng, L., Wang, S., Tian, L., He, F., Liu, Z., and Tian, Q. (2015b). Query-adaptive late fusion for image search and person re-identification. In *CVPR*.

Zheng, L., Yang, Y., and Hauptmann, A. G. (2016a). Person Re-identification: Past, Present and Future. *CoRR*, abs/1610.02984.

Zheng, W.-S., Gong, S., and Xiang, T. (2016b). Towards open-world person re-identification by one-shot group-based verification. *PAMI*, 38(3):591–606.

Zheng, Z., Zheng, L., and Yang, Y. (2017b). Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro. In *ICCV*.

Zhong, Z., Zheng, L., Cao, D., and Li, S. (2017). Re-ranking person re-identification with k-reciprocal encoding. In *CVPR*.

Zhou, S., Wang, J., Wang, J., Gong, Y., and Zheng, N. (2017). Point to set similarity based deep feature learning for person re-identification. In *CVPR*.

Zhu, X., Wu, B., Huang, D., and Zheng, W.-S. (2017). Fast Open-World Person Re-Identification. *IEEE Transactions on Image Processing*.