Online Joint Multi-Metric Adaptation from Frequent Sharing-Subset Mining for Person Re-Identification

Jiahuan Zhou[1], Bing Su[2], Ying Wu[1]
[1]Department of Electrical and Computer Engineering, Northwestern University, US
[2]Institute of Software, Chinese Academy of Science, China
zhoujh09@gmail.com, subingats@gmail.com, yingwu@northwestern.edu

Abstract

Person Re-IDentification (P-RID), as an instance-level recognition problem, still remains challenging in computer vision community. Many P-RID works aim to learn faithful and discriminative features/metrics from offline training data and directly use them for the unseen online testing data. However, their performance is largely limited due to the severe data shifting issue between training and testing data. Therefore, we propose an online joint multi-metric adaptation model to adapt the offline learned P-RID models for the online data by learning a series of metrics for all the sharing-subsets. Each sharing-subset is obtained from the proposed novel frequent sharing-subset mining module and contains a group of testing samples which share strong visual similarity relationships to each other. Unlike existing online P-RID methods, our model simultaneously takes both the sample-specific discriminant and the set-based visual similarity among testing samples into consideration so that the adapted multiple metrics can refine the discriminant of all the given testing samples jointly via a multi-kernel late fusion framework. Our proposed model is generally suitable to any offline learned P-RID baselines for online boosting, the performance improvement by our model is not only verified by extensive experiments on several widely-used P-RID benchmarks (CUHK03, Market1501, DukeMTMC-reID and MSMT17) and state-of-the-art P-RID baselines but also guaranteed by the provided in-depth theoretical analyses.

1. Introduction

Person Re-Identification (P-RID), aiming to retrieve the same identity images of a query probe from a gallery set, is not only an attractive research task in computer vision community, but also a critical link to the practical applications such as public camera surveillance. A popular solution to P-RID is to perform supervised feature/metric learning [2, 29, 18, 30, 6, 40, 12] from the *offline* training data, then directly apply them to the *online* unsupervised testing data for evaluation. However, due to the severe training-testing data distribution shifting (testing data are drawn from totally different classes against the training data as shown in Fig. 1) caused by large variations in visual appearance, human pose, camera viewpoint, illumination change, and background clutter, the performance of offline learned models is limited indeed.

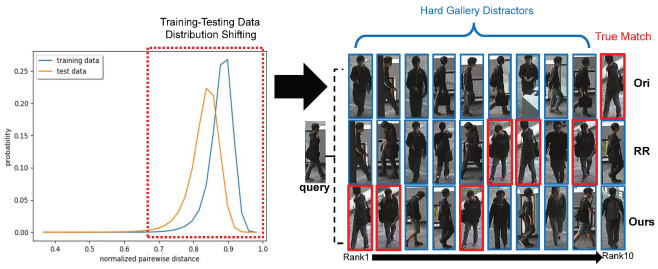

Figure 1. The normalized pair-wise distance distributions of both training and testing samples based on the well-trained HA-CNN model on Market1501 dataset demonstrate the severe training-testing data distribution shifting issue, where the extremely challenging hard negative distractors (in blue box) will significantly influence the retrieval accuracy (the *Original* top-10 retrieval results). Even using the state-of-the-art online re-ranking method [45] (RR), the ground-truth (in red box) still has a lower rank than the distractors. Our method succeeds in handling the distractors so that the true-match is successfully re-ranked to the top position in the list (Ours).

The root of such a limited performance is its treatment regardless of the information of online testing data themselves. So a straightforward solution is adapting the offline learned models for the online testing data to narrow the distribution gap. Recently, various online P-RID methods are proposed which can be roughly categorized into two branches. The *set-centric* re-ranking approaches [35, 9, 43, 3, 1] focus on optimizing the ranking list of queries based on the similarity relationships among

testing samples. Their performances totally rely on the offline learned models from training data while treat different testing samples equally ignoring the individual characteristics, hence the improvement is neither significant nor stable. The other category is *query-specific* metric adaptation [17, 38, 45] which aims to enhance the discriminant of each query individually. The generic offline learned metric is adapted to an instance-specific local metric for each query. Compared with the *set-centric* ones, the individual discriminant of queries is enhanced while the visual similarity relationships among given testing samples are ignored. Moreover, existing query-specific models [17, 38, 45] completely ignore the counterpart gallery data during adaptation. Even a discriminative probe-specific metric can be learned, the "hard" gallery samples with large intra-class and small inter-class variances will tremendously degrade its performance since they are still indistinguishable under the learned query-specific metric (Fig. 1).

In order to tackle the aforementioned issues, we propose a novel online joint multi-metric adaptation algorithm which not only takes individual characteristics of testing samples into consideration but also fully explore the visual similarity relationships among both query and gallery samples. As shown by Fig. 2, at the online P-RID testing stage, the redundant intrinsic visual similarity relationships among unlabeled query (gallery) set are utilized by our proposed *frequent sharing-subsets mining* model to automatically mine the concise and strong visual sharing associations of samples. Since a sharing-subset contains a group of queries (galleries) sharing strong visual similarity to each other, their local distributions will be jointly adjusted by efficiently learning a Mahalanobis metric for all of them. Once a series of such kind of *sharing-subset* based Mahalanobis metrics are learned, for each query (gallery), its instance-specific local metric is obtained via a multi-metric late fusion of all the *sharing-subset* based Mahalanobis metrics. Therefore, our proposed online joint **M**ulti-**M**etric adaptation model based on the frequent sharing-subsets **M**ining (denoted as $\mathbf{M}^3$) is able to refine the ranking performance online. The success of learning from sharing relies on discovering the latent sharing relationships among samples, which cannot be found by treating each instance independently [4]. Learning from sharing is good at handling such condition that only a limited number of learning data are available by taking the sharing relationships as data augmentation. Therefore the sharing strategy is particularly suitable for online P-RID learning in where each testing sample itself is the only positive sample available for learning.

The main contributions of this paper are as follows: (1) To handle the severe shifted training-testing data distribution issue in P-RID, we leap from offline global learning to online instance-specific metric adaptation. We propose a general and flexible learning objective to simultaneously enhance the local discriminant of testing query and gallery data. (2) By mining various frequent sharing-subsets, the intrinsic visual similarity sharing relationships are fully explored. Therefore the online time cost of learning metrics from sharing is much more smaller than learning local metrics independently. (3) To fulfill the time-efficient requirement of online testing, a theoretical sound optimization solution is proposed for efficient learning which is also proven to guarantee the improvement of performance. (4) Our proposed model can be readily applied to any existing offline P-RID baselines for online performance improvement. The efficiency and effectiveness of our method are further verified by the extensive experiments on four challenging P-RID benchmarks (CUHK03, Market1501, DukeMTMC-reID and MSMT17) based on various state-of-the-art P-RID models.

2. Related Work

Online Re-Ranking in P-RID: In recent years, increasing efforts have been paid to online P-RID re-ranking. Ye *et al.* [35] revised the ranking list by considering the nearest neighbors of both the global and local features. An unsupervised re-ranking model proposed by Garcia *et al.* [9] takes advantage of the content and context information in the ranking list. Zhong *et al.* [43] proposed a k-reciprocal encoding approach for re-ranking, which relies on a hypothesis that if a gallery image is similar to the probe in the k-reciprocal nearest neighbors, it is more likely to be a true-match. Zhou *et al.* [45] proposed to learn an instance-specific Mahalanobis metric for each query sample by using extra negative learning samples at online stage. Barman *et al.* [3] focused on how to make a consensus-based decision for retrieval by aggregating the ranking results from multiple algorithms, only the matching scores are needed. Bai *et al.* [1] concentrated on re-ranking with the capacity of metric fusion for P-RID by proposing an Unified Ensemble Diffusion (UED) framework. However, the aforementioned online re-ranking methods either simply treat different testing samples equally without considering the instance-specific characteristics or completely ignore the intrinsic visual similarity relationships among testing samples, so that the performance improvement is neither stable nor significant.

CNN-based Feature Extraction in P-RID: CNN-based feature extraction has achieved the state-of-the-art performance in P-RID. A novel Harmonious Attention CNN (HA-CNN) proposed by Li *et al.* [18] tries to jointly learn attention selection and feature representation in a CNN by maximizing the complementary information of different levels of visual attention (soft attention and hard attention). Wang *et al.* [30] proposed a novel deeply supervised fully attentional block that can be plugged into any CNNs to solve

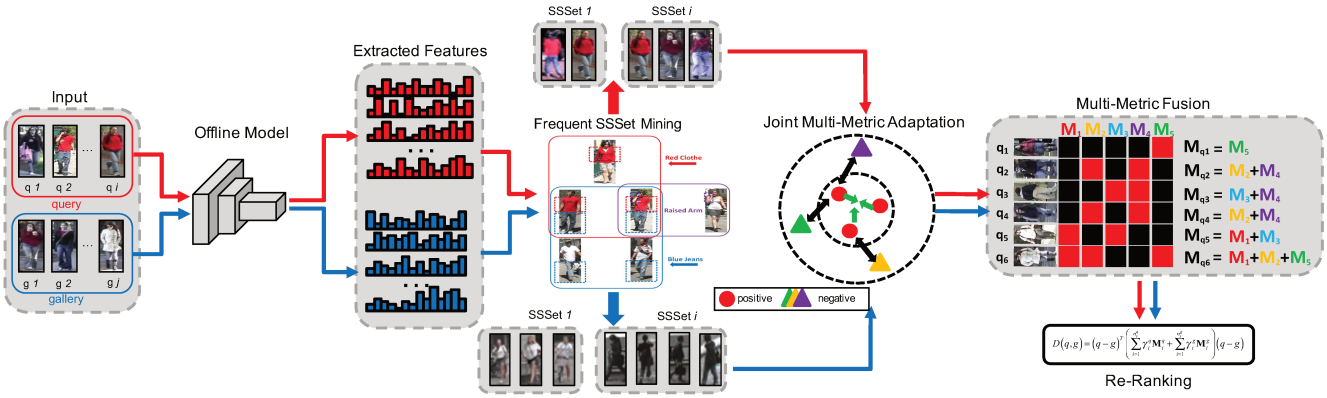

Figure 2. The online testing query and gallery samples are fed into the offline learned baseline model to obtain the feature descriptors firstly. The proposed frequent sharing-subset (SSSet) mining model is performed to the extracted features to generate multiple sharing-subsets which are further utilized by the proposed joint multi-metric adaptation model (The same sample may be contained by multiple SSSets since it shares different visual similarity relationships with different samples.). By fusing the learned matching metrics for each query and gallery sample, our final ranking list is obtained by a bi-directional retrieval matching (Sec. 3.5).

P-RID problem, and a novel deep network called Mancs is designed to learn stable features for P-RID. Hou *et al.* [12] proposed the Spatial Interaction-and-Aggregation (SIA) and Channel Interaction-and-Aggregation (CIA) modules to improve the representational capacity of deep convolutional networks. Chen *et al.* [6] proposed an Attentive but Diverse Network (ABD-Net) which integrates attention modules and diversity regularizations throughout the entire network to learn features that are representative, robust, and more discriminative for P-RID. Zheng *et al.* [40] aimed at improving the learned P-RID features by better leveraging the generated data by designing a joint learning framework that couples P-RID learning and data generation end-to-end. However, these well-trained networks are directly applied to the testing data for feature extraction and evaluation, the data distribution shifting between training and testing samples definitely limits the performance of these models. Therefore, our proposed method is suitable for any CNNs for sample-specific local metric adaptation at inference stage aiming to address the data shifting issue well and gain a further performance improvement.

3. M^3: Online Joint Multi-Metric Adaptation from Frequent Sharing-Subset Mining

3.1. Problem Settings and Notations

At the online testing stage of P-RID, two disjoint datasets, a **query set** $\mathcal{Q}$ and a **gallery set** $\mathcal{G}$ are given as:

$$\mathcal{Q} = \{(q_i, l_i^q)\}_{i=1}^{n_q} \quad \mathcal{G} = \{(g_i, l_i^g)\}_{i=1}^{n_g}$$

that $q_i, g_i \in \mathbb{R}^d$ are the extracted feature representations from an offline baseline model, either handcraft features or learned deep features. $l_i^q, l_i^g \in \{1, 2, ..., c\}$ are the labels from c classes which are totally different from the training

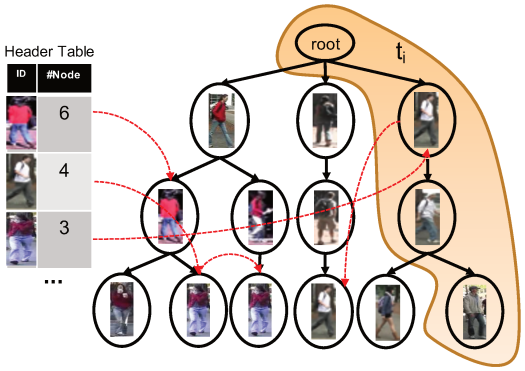

Figure 3. A CFI-Tree is constructed based on $\mathcal{T}$. The same identity may be contained by multiple t_i so that there may be multiple nodes for the same identity.

sample classes. P-RID aims to rank $\mathcal{G}$ for a query probe q based on the pair-wise similarity distance to a gallery g, $d(q, g) = \|q, g\|^2$. Our goal is to re-rank $\mathcal{G}$ for q by refining $d(q, g)$ to improve the rank of true-matches for q.

3.2. Unsupervised Frequent Sharing-Subset Mining

Although the identity label $\{l_i^q\}$ ($\{l_i^g\}$) is unknown during testing, the visual similarity relationships of $\mathcal{Q}$ ($\mathcal{G}$) are intrinsic and verified to be effective in investigating the underlying similarity structure of samples by previous online re-ranking methods [9, 43]. However, due to the large-scale sample size (especially for $\mathcal{G}$), the redundancy and repeatability of visual similarity relationships significantly limit the performance of previous online P-RID methods. Inspired by the well-established frequent itemset mining technique [8], we propose an unsupervised frequent sharing-subset (SSSet) mining algorithm to automatically mine frequent SSSets $\{\mathcal{S}_i\}_{i=1}^{n_s}$ from $\mathcal{Q}$, that all the samples in $\mathcal{S}_i$ share a *Strong Association Rule* on visual similarity [8].

Therefore, the mined SSSets not only keep the strong and reliable visual similarity sharing information but also significantly alleviate the redundancy. Compared with the original combinatorial problem suffering from exponential computation complexity $O(2^n)$, the time complexity of our proposed algorithm is $O(n^2)$ which is much more efficient when a large scale of testing samples are given.

Considering $\mathcal{Q}$ as the given *Item* set, we firstly prepare a *Transaction* set $\mathcal{T} = \{t_i\}_{i=1}^{n_t}$ from $\mathcal{Q}$ where each t_i is a subset of $\mathcal{Q}$. The affinity matrix $\mathbf{A} \in \mathbb{R}^{n_q \times n_q}$ of $\mathcal{Q}$ is defined as:

$$A_{i,j} = \begin{cases} \exp\left(\frac{-d(q_i,q_j)}{2\sigma}\right) / \sum_j \exp\left(\frac{-d(q_i,q_j)}{2\sigma}\right), & j \neq i \\ 0, & j = i \end{cases} \tag{1}$$

where σ is the variance parameter of distance matrix from $\mathcal{Q}$ so that $A_{i,j}$ represents the soft-max normalized visual similarity between q_i and q_j. The i-th row of $\mathbf{A}$ represents the similarity distribution between q_i and the other samples in $\mathcal{Q}$. To keep only the most reliable sharing relationships, a threshold Θ defined as the average affinity of $\mathcal{Q}$ is used for outlier filtering: $\Theta = \sum_{i=1}^{n_q}\sum_{j=1}^{n_q} A_{i,j}/n_q \cdot n_q$. Therefore, a binary index map $\mathbf{B}$ is obtained by:

$$B_{i,j} = \begin{cases} 1, & A_{i,j} \geq \Theta \\ 0, & A_{i,j} < \Theta \end{cases} \tag{2}$$

The non-zero $B_{i,j}$ implies the strong similarity sharing relationship between q_i and q_j. Therefor each non-zero row $\mathbf{B}_j$ of $\mathbf{B}$ can be considered as a *Transaction* t_i.

$$\mathcal{T} = \{t_i\} = \{\mathbf{B}_j\}, \ \forall \sum \mathbf{B}_j \geq 1 \tag{3}$$

Once the transaction set $\mathcal{T}$ is obtained, we propose to mine the frequent sharing-subsets from $\mathcal{T}$ that each sharing-subset is represented by a mined frequent pattern from a classical FP-Close mining algorithm [10]. To do so, a Closed Frequent Itemset Tree (CFI-Tree) is firstly constructed based on $\mathcal{T}$ under a *minimum support* 5 (Fig. 3), then the FP-Close mining algorithm in [10] is performed to the constructed CFI-Tree to obtain all the closed frequent patterns $\{\mathcal{S}_i\}_{i=1}^{n_s}$ that each $\mathcal{S}_i$ represents a sharing-subset.

3.3. Joint Multi-Metric Adaptation From SSSets

Once all the frequent SSSets $\{\mathcal{S}_i\}_{i=1}^{n_s}$ are obtained, our goal is to jointly learn n_s SSSets-based local Mahalanobis metrics for $\{\mathcal{S}_i\}_{i=1}^{n_s}$ by optimizing Eqn. 4:

$$\begin{aligned} &\arg\min_{\{\mathbf{M}_i\}} \frac{1}{2}\sum_{i=1}^{n_s} \|\mathbf{M}_i\|^2 \\ &w.r.t: \ \mathbf{M}_i \succeq 0 \\ &\left(s_u^i - s_v^j\right)^T (\mathbf{M}_i + \mathbf{M}_j)\left(s_u^i - s_v^j\right) \geq 2, \ \ \forall s_u^i \in \mathcal{S}_i, s_v^j \in \mathcal{S}_j \\ &\left(s_u^i - s_v^i\right)^T \mathbf{M}_i \left(s_u^i - s_v^i\right) = 0, \ \ \forall s_u^i \in \mathcal{S}_i, s_v^i \in \mathcal{S}_i \end{aligned} \tag{4}$$

The learned metric $\mathbf{M}_i$ from Eqn. 4 is shared by all the samples in $\mathcal{S}_i$. Suppose we have n_s SSSets and $O(n)$ samples in each $\mathcal{S}_i$, there are totally $O(n_s^2 n^2)$ inequality constraints and $O(n_s n^2)$ equality constraints in Eqn. 4 which are too difficult to deal with, so that we aim to reduce the constraint size in Eqn. 4. We find out that Eqn. 4 has an exactly equivalent form by only keeping the constraints related to one anchor sample s^i in $\mathcal{S}_i$, that s^i can be any sample in $\mathcal{S}_i$. Therefore the equivalent form is shown by Eqn. 5:

$$\begin{aligned} &\arg\min_{\{\mathbf{M}_i\}} \frac{1}{2}\sum_{i=1}^{n_s} \|\mathbf{M}_i\|^2 \\ &w.r.t: \ \mathbf{M}_i \succeq 0 \\ &\left(s^i - s_v^j\right)^T (\mathbf{M}_i + \mathbf{M}_j)\left(s^i - s_v^j\right) \geq 2, \ \ \forall s^i \in \mathcal{S}_i, s_v^j \in \mathcal{S}_j \\ &\left(s^i - s_v^i\right)^T \mathbf{M}_i \left(s^i - s_v^i\right) = 0, \ \ \forall s^i \in \mathcal{S}_i, s_v^i \in \mathcal{S}_i \end{aligned} \tag{5}$$

Revisit Eqn. 4, its equality constraints propose to collapse all $s_u^i \in \mathcal{S}_i$ together. Therefore keeping only the equality constraints related to the anchor sample s^i achieves the same collapsing performance. So as to the inequality constraints in Eqn. 4. Finally, we can reduce the constraint size by only keeping the constraints related to s^i as in Eqn. 5. The re-formed objective Eqn. 5 has only $O(n_s^2 n)$ and $O(n_s n)$ inequality and equality constraints respectively. An important merit of Eqn. 5 is that it can be efficiently optimized:

Theorem 1 *All the vectors $s^i - s_v^i$ in Eqn. 5 form a spanning space $\boldsymbol{H} = span(\sum_v \lambda_v (s^i - s_v^i))$. Eqn. 5 is equivalent to replace $s^i - s_v^j$ by $h_v^\perp$, the projection of $s^i - s_v^j$ in $\boldsymbol{H}^\perp$, that $\boldsymbol{H}^\perp$ is the orthogonal space of $\boldsymbol{H}$.*

Proof 1 *Since $\boldsymbol{M}_i$ is positive semi-definite, we have $\left(s^i - s_v^i\right)^T \mathbf{M}_i \left(s^i - s_v^i\right) = 0 \ \Leftrightarrow \mathbf{M}_i \left(s^i - s_v^i\right) = 0 \ \Leftrightarrow \boldsymbol{M}_i h = 0, \ \forall h \in \boldsymbol{H}$. Projecting $s^i - s_v^j$ to $\boldsymbol{H}$ and $\boldsymbol{H}^\perp$ generates two orthogonal bases h_v and $h_v^\perp$ respectively, so $s^i - s_v^j = h_v + h_v^\perp$. Replace the inequality constraints in Eqn. 5 by $h_v + h_v^\perp$:*

$$\begin{aligned} &\left(s^i - s_v^j\right)^T (\mathbf{M}_i + \mathbf{M}_j)\left(s^i - s_v^j\right) \\ &= \left(h_v + h_v^\perp\right)^T (\mathbf{M}_i + \mathbf{M}_j)\left(h_v + h_v^\perp\right) \\ &= {h_v^\perp}^T (\mathbf{M}_i + \mathbf{M}_j) h_v^\perp \end{aligned} \tag{6}$$

Now Eqn. 5 has an equivalent form as:

$$\begin{aligned} &\arg\min_{\{\mathbf{M}_i\}} \frac{1}{2}\sum_{i=1}^{n_s} \|\mathbf{M}_i\|^2 \\ &w.r.t: \ \boldsymbol{M}_i \succeq 0 \\ &{h_v^\perp}^T (\mathbf{M}_i + \mathbf{M}_j) h_v^\perp \geq 2, \ \ \forall s^i \in \mathcal{S}_i, s_v^j \in \mathcal{S}_j \\ &\boldsymbol{M}_i h = 0, \ \ \forall h \in \boldsymbol{H} \end{aligned} \tag{7}$$

Finally, we prove that Eqn. 7 has the same solution to Eqn. 4 by eliminating its PSD and equality constraints.

Theorem 2 *The solution to Eqn. 4 is exactly the same as solving the Eqn. 7 by relaxing its equality and PSD constraints, since they are indeed off-the-shelf.*

Proof 2 *If we get rid of the PSD and equality constraints in Eqn. 7, the new form is:*

$$\begin{aligned} &\arg\min_{\{\mathbf{M}_i\}} \frac{1}{2}\sum_{i=1}^{n_s}\|\mathbf{M}_i\|^2 \\ &w.r.t: \quad {h_v^\perp}^T\left(\mathbf{M}_i+\mathbf{M}_j\right)h_v^\perp \geq 2, \quad \forall s^i\in\mathcal{S}_i, s_v^j\in\mathcal{S}_j \end{aligned} \tag{8}$$

Eqn. 8 is exactly in the same form of a multi-kernel SVM problem so that it can be efficiently solved.

Thus the positive semi-definiteness of $\boldsymbol{M}_i$ is guaranteed since $\boldsymbol{M}_i = \sum \alpha_v \varphi(h_v^\perp) = \sum \alpha_v h_v^\perp \cdot {h_v^\perp}^T \succeq 0$. For the equality constraints in Eqn. 7, given a member s of $\boldsymbol{S}$, we have:

$$\boldsymbol{M}_i h = \left(\sum \alpha_v h_v^\perp \cdot {h_v^\perp}^T\right) h = \sum \alpha_v h_v^\perp \cdot \left({h_v^\perp}^T h\right) = 0 \tag{9}$$

which proves that the solution to Eqn. 8 satisfies the equality constraints as well.

3.4. Bi-Directional Discriminant Enhancement

At online testing stage, the gallery set $\mathcal{G}$, the counterpart of query set $\mathcal{Q}$, also plays an important role. As shown by Fig. 1, the re-ranking performance by using only the query-centric metric adaptation may suffer from ambiguous gallery distractors. The similar gallery images from different identities will significantly degrade the discriminant of $\mathbf{M}_p$ since these gallery distractors are still indistinguishable under $\mathbf{M}_p$. Therefore, we aim to handle these indistinguishable gallery samples by performing a gallery-centric local discriminant enhancement method as Eqn. 4. The SSSets of $\mathcal{G}$ and the corresponding joint metrics are obtained via Sec. 3.2 and Eqn. 4 respectively.

3.5. Multi-Metric Late Fusion For Re-Ranking

For one query probe q, it may be contained by multiple SSSets so that there will be multiple learned metrics $\mathbf{M}_i$ associated to q. The final metric $\mathbf{M}_q$ for q is obtained via a boosting-form multi-metric late fusion [24, 23]:

$$\mathbf{M}_q = \left(\sum_{i=1}^{n_s}\gamma_i\mathbf{M}_i\right) / \sum\gamma_i \tag{10}$$

where $\gamma_i = 1$ if $q \in \mathcal{S}_i$. For a gallery sample g, a similar fused metric $\mathbf{M}_g$ can be obtained likewise. Therefor the refined distance between q and g is defined as Eqn. 11 based on which the re-ranking list of q_i is obtained.

$$d\left(q,g\right) = \left(q-g\right)^T\left(\mathbf{M}_q+\lambda\mathbf{M}_g\right)\left(q-g\right) \tag{11}$$

4. Theoretical Analyses and Justifications

As demonstrated by Theorem. 2, the solution of our joint multi-metric adaptation objective can be readily transformed to the equivalent form as [45]. Therefore, the appealing theoretical properties in [45] can be inherited by our learned $\mathbf{M}_i$ as presented in Theorem. 3. Moreover, our late multi-kernel fusion metric Eqn. 10 will guarantee a further reduction of generalization error bound as in Theorem. 4.

Theorem 3 ***(The reduction of both asymptotic and practical error bound by the learned $\boldsymbol{M}_i$)**: As demonstrated by the Theorem.2 in [45], for an input x, its asymptotic error $\mathbb{P}^a(e|x)$ by using extra negative data $\mathcal{D}^a$ is:*

$$\mathbb{P}^a(e|x) = \frac{(2-q)\mathbb{P}(e|x)}{2-2q\mathbb{P}(e|x)} \leq \mathbb{P}(e|x) \tag{12}$$

where q is a probability scalar that $0 \leq q \leq 1$ and $\mathbb{P}(e|x)$ is the Bayesian error. Moreover, the asymptotic error $\mathbb{P}^a(e|x)$ can be best approximated by the practical error rate $\mathbb{P}_n(e|x)$ (n is finite) by finding a local metric $\boldsymbol{M}_x$ which turns out to be the one for our Eqn. 4.

Theorem 4 ***(The reduction of generalization error bound by using $\boldsymbol{M}_{q/g}$ in Eqn. 10)**: Our fused multi-kernel metric $\boldsymbol{M}_q = (\sum_{i=1}^{n_s}\gamma_i\boldsymbol{M}_i) / \sum\gamma_i$ is a linear combinations of several base kernels $\boldsymbol{M}_i$ from the family of finite Gaussian kernels: $\mathcal{K}_G^d := \{K_{\boldsymbol{M}} : (x_1,x_2) \mapsto e^{-(x_1-x_2)^T\boldsymbol{M}(x_1-x_2)} \mid \boldsymbol{M} \in \mathbb{R}^{d\times d}, \ \boldsymbol{M}\succeq 0\}$ which is bounded by B_k. Therefore, for a fixed $\delta \in (0,1)$, $n_s < n_k$ is the number of metrics (kernels) involved in our final joint multi-metric learning solution. With probability at least $1-\delta$ over the choice of a random training set $\mathcal{X} = \{x_i\}_{i=1}^n$ of size n we have:*

$$\mathcal{E}_{est}(\boldsymbol{M}_i) \simeq \mathcal{O}\left(\sqrt{\frac{n_k+B_k}{n}}\right) \tag{13}$$

$$\mathcal{E}_{est}(\boldsymbol{M}_q) \simeq \mathcal{O}\left(\sqrt{\frac{\log n_k+B_k+2n_s}{n}}\right) \tag{14}$$

In our work, we have $n_s \ll n_k$, that the selected number of kernels is much fewer than the total kernel number, so that $\mathcal{E}_{est}(\boldsymbol{M}_q) \simeq \mathcal{O}\left(\sqrt{\frac{\log n_k+B_k}{n}}\right) \ll \mathcal{E}_{est}(\boldsymbol{M}_i) \simeq \mathcal{O}\left(\sqrt{\frac{n_k+B_k}{n}}\right)$. The generalization error by using $\boldsymbol{M}_q$ is much smaller than using only any $\boldsymbol{M}_i$. The same conclusion can be obtained for $\boldsymbol{M}_g$ likewise.

Proof 3 *The classification rule of our learned $\boldsymbol{M}_i$ can be defined as $\zeta_j\left(\tilde{q}^T\boldsymbol{M}_i\tilde{x}_j - 1\right) \geq 1$ so that the margin is 1. Motivated by [25], the generalization error $\mathcal{E}_{est}(\boldsymbol{M}_i)$ of using kernel $\boldsymbol{M}_i$ is bounded by $\mathcal{O}\left(\sqrt{\frac{n_k+B_k}{n}}\right)$. While by*

Figure 4. The visualization of rank improvement on CUHK03 (1th, 2nd) and Market1501 (3rd, 4th) based on HA-CNN. For each case, its top-10 (left to right) matches are presented and the true-match is labeled by the red box. The 1st row is the baseline result, the 2nd row is the result only using $\mathbf{M}_q$ and the 3rd row is the result using both $\mathbf{M}_q$ and $\mathbf{M}_g$.

using $\boldsymbol{M}_q$, which is a linear combination of all $\boldsymbol{M}_i$ from the family of finite Gaussian kernel $\mathcal{K}_G^d$, its generalization error $\mathcal{E}_{est}(\boldsymbol{M}_q)$ is bounded by $\mathcal{O}\left(\sqrt{\frac{\log n_k + B_k + 2n_s}{n}}\right)$ which is guaranteed by the Theorem.2 in [14]. For the kernel family $\mathcal{K}_G^d$, $n_k \simeq \mathcal{O}(d^2)$ and in our work, $d \approx 10^3$ so that $n_k \approx 10^6$. The selected kernels for combination is about 20 in average so that $n_s \ll n_k$ which means $\mathcal{E}_{est}(\boldsymbol{M}_q) \ll \mathcal{E}_{est}(\boldsymbol{M}_i)$.

5. Experiments

5.1. Experimental Settings

Datasets. We evaluate our proposed M^3 model on CUHK03 [17], Market1501 [39], DukeMTMC-reID [41] and MSMT17 [33] benchmarks. The statistic details of the above datasets are summarized in Table. 1. For CUHK03 [1], the new splitting protocol proposed by [43] is adopted in our experiment so that 767 identities are used for training as well as the left 700 identities are used for testing. As for the other three benchmarks, Market1501, DukeMTMC-reID and MSMT17, the pre-determined probe and gallery sets are directly utilized with no modification.

Dataset	cuhk03	market	duke	msmt17
#T-IDs	767	751	702	1040
#Q-IDs	700	750	702	3060
#G-IDs	700	751	1110	3060
#cam	2	6	8	15
#images	28192	32668	36411	126441

Table 1. The statistics of P-RID benchmarks. #T/Q/G-IDs denote the number of training/query/gallery ids.

Baselines. Our proposed M^3 method is evaluated based on several state-of-the-art CNN-based P-RID models: ResNet50 [11], DenseNet121 [13], HA-CNN [18], MLFN [5] and ABDNet[6]. The general CNN models, ResNet50 and DenseNet121, are well trained on each benchmark for feature extraction. HA-CNN, MLFN and ABDNet are the P-RID specific CNNs so that the original works are directly utilized in our experiments. Besides, the other state-of-the-art P-RID methods [15, 21, 37, 27, 28, 5, 26, 40, 46, 6] are further compared. Moreover, related online P-RID methods including [45] (OL) and [43] (RR) are compared with our M^3 method.

Evaluation. We follow the same official evaluation protocols in [39, 41, 17, 33], the single-shot evaluation setting is adopted and all the results are shown in the form of Cumulated Matching Characteristic (CMC) at several selected ranks and mean Average Precision (mAP). Various ablation studies of our proposed model are explored in Sec. 5.5.

5.2. Comparison with the State-of-the-arts

Evaluation on CUHK03: The comparison results on CUHK03 (767/700 splitting protocol) are presented in Table. 2. Our M^3 model significantly boosts the baseline Rank@1(mAP) performance of ResNet50, DenseNet12, HA-CNN and MLFN to 66.9%(60.7%), 61.6%(54.4%), 69.8%(63.5%) and 73.4%(71.2%) with a 40.0%(29.7%), 50.2%(35.7%), 45.4%(33.4%) and 34.2%(44.7%) relative improvement respectively. Even compared with the state-of-the-art method MGN [31], our results outperform it by 5% at Rank@1. The reason for such a large improvement is that the "hard" gallery distractors which are still indistinguishable under $\mathbf{M}_q$ is well handled by our M^3 method (Fig. 4), so the ranking of true-match gallery targets is significantly improved.

Evaluation on Market1501: The superiority of our M^3 method is further verified by the experiments on Market1501. Table. 2 demonstrates that although the state-of-the-art approach ABDNet [6] has achieved a pretty high performance ($\geq$ 94%) on Market1501, the improvement of our M^3 is still over 3.7%(10%) on Rank@1(mAP) based on ABDNet (visualization results in Fig. 4).

Evaluation on DukeMTMC-reID: DukeMTMC-reID is a recent benchmark proposed for P-RID, but the latest methods have obtained promising performances. As shown in Table. 2, the recently published OSNet [46] has raised the state-of-the-art to 87.0%(70.2%). Our ABDNet+M^3 improves the Rank@1(mAP) result to 87.5%(73.3%), which beats OSNet by a large margin on mAP.

Evaluation on MSMT17: MSMT17 is the latest and largest benchmark so far which is pretty challenging due to the extreme large-scale identities and distractors. We eval-

[1] In our experiment, the CUHK03 detected dataset is utilized.

CUHK03(767/700)			Market1501			DukeMTMC-reID		
Method	**R@1**	**mAP**	**Method**	**R@1**	**mAP**	**Method**	**R@1**	**mAP**
ResNet50[11]	47.9	46.8	ResNet50[11]	88.5	71.3	ResNet50[11]	77.7	58.8
DenseNet121[13]	41.0	40.1	DenseNet121[13]	88.2	69.2	DenseNet121[13]	78.6	58.5
HA-CNN[18]	48.0	47.6	HA-CNN[18]	90.6	75.3	HA-CNN[18]	80.7	64.4
MLFN[5]	54.7	49.2	MLFN[5]	90.1	74.3	MLFN[5]	81.0	62.8
ABDNet[6]	N/A	N/A	ABDNet[6]	93.7	85.5	ABDNet[6]	84.1	67.7
OSNet[46]	N/A	N/A	OSNet[46]	94.2	82.6	OSNet[46]	87.0	70.2
PCB[28]	63.7	67.5	PCB[28]	83.3	69.2	PCB[28]	83.3	69.2
SVDNet[27]	41.5	37.3	SVDNet[27]	82.3	62.1	SVDNet[27]	76.7	56.8
DPFL[7]	40.7	37.0	DNSL[36]	61.0	35.6	DuATM[22]	81.8	64.6
PAN[42]	36.3	34.0	Part-aligned[26]	91.7	79.6	Part-aligned[26]	84.4	69.3
ResNeXt[34]	43.8	38.7	PN-GAN[19]	77.1	63.6	PAN[42]	71.6	51.5
DaRe[32]	55.1	51.3	DeepCC[20]	89.5	75.7	GAN[41]	67.7	47.1
MGN[31]	68.0	67.4	Mancs[30]	93.1	82.3	SPreID[16]	85.9	73.3
M^3+ResNet50	**66.9**	**60.7**	**M^3+ResNet50**	**95.4**	**82.6**	**M^3+ResNet50**	**84.7**	**68.5**
M^3+DenseNet121	**61.6**	**54.4**	**M^3+DenseNet121**	**95.3**	**81.2**	**M^3+DenseNet121**	**84.9**	**68.0**
M^3+HA-CNN	**69.8**	**63.5**	**M^3+HA-CNN**	**96.5**	**85.2**	**M^3+HA-CNN**	**87.1**	**72.2**
M^3+MLFN	**73.4**	**71.2**	**M^3+MLFN**	**96.4**	**85.0**	**M^3+MLFN**	**86.5**	**71.5**
M^3+ABDNet	**N/A**	**N/A**	**M^3+ABDNet**	**97.9**	**92.6**	**M^3+ABDNet**	**87.5**	**73.3**

Table 2. Compared with the state-of-the-arts on CUHK03, Market1501, and DukeMTMC-reID.

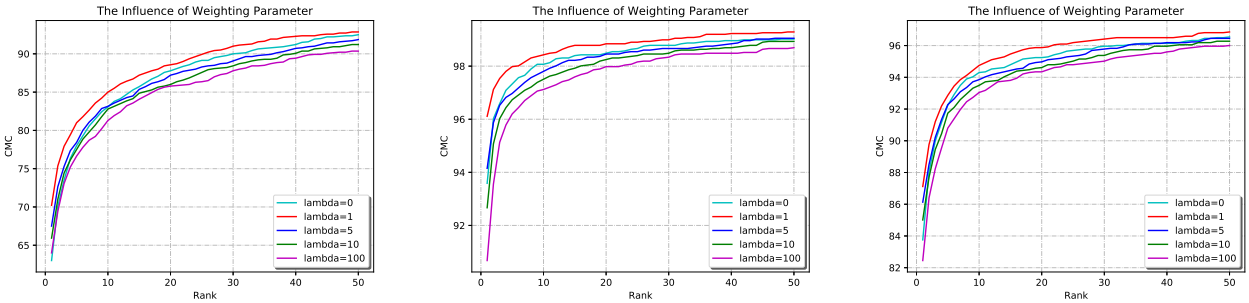

Figure 5. The influence of λ on (left) CUHK03, (mid) Market1501 and (right) DukeMTMC-reID based on HA-CNN.

MSMT17	Baseline		Baseline+M^3	
Method	**R@1**	**mAP**	**R@1**	**mAP**
ResNet50[11]	63.4	34.2	**72.8**	**55.0**
DenseNet121[13]	66.0	34.6	**75.5**	**43.1**
HA-CNN[18]	64.7	37.2	**74.3**	**43.8**
MLFN[5]	66.4	37.2	**72.8**	**43.4**
ABDNet[6]	82.3	60.8	**85.7**	**64.2**

Table 3. Compared with the state-of-the-arts on MSMT17.

uate the performance of selected baselines on the MSMT17 dataset with(w/) and without(w/o) our M^3 model in Table. 3. For all the baselines, our M^3 model significantly improves their Rank@1(mAP) performance. The performance of ABDNet is boosted from 82.3%(60.8%) to a state-of-the-art level of 85.7%(64.2%). Table. 3 verifies the scalability of our proposed M^3 model, even for the extremely large-scale query/gallery sets, our method is still able to consistently improve the baseline performance.

Method	**CUHK03**	**Market**	**Duke**
HA-CNN[18]	48.0(47.6)	90.6(75.3)	80.7(64.4)
HA-CNN+RR [43]	54.8(55.7)	91.4(79.0)	82.5(69.9)
HA-CNN+OL [45]	62.3(56.5)	92.7(78.9)	83.7(67.8)
HA-CNN+M^3	**69.8(63.5)**	**96.5(85.2)**	**87.1(72.2)**
Dense121[13]	41.0(40.1)	88.2(69.2)	78.6(58.5)
Dense121+RR [43]	48.1(51.5)	90.2(**85.0**)	83.7(**76.9**)
Dense121+OL [45]	53.1(49.3)	90.4(74.0)	80.2(64.1)
Dense121+M^3	**61.6(54.4)**	**95.3**(81.2)	**84.9**(68.0)

Table 4. Compared with online P-RID refinement methods.

5.3. Comparison with Online P-RID Re-ranking

Two state-of-the-art online P-RID re-ranking methods, OL [45] and RR [43], are compared with our M^3 since all the three methods can be readily utilized at online testing stage for further performance improvement. The comparison results in Table. 4 show that the query-specific method OL [45] works better on improving Rank@1 per-

Method	Market1501 → DukeMTMC					DukeMTMC → Market1501				
	R@1	R@5	R@10	R@20	mAP	R@1	R@5	R@10	R@20	mAP
MLFN[5]	45.8	63.9	71.6	78.1	20.3	30.4	47.5	53.9	59.5	17.1
MLFN+M^3	**67.6**	**78.8**	**83.0**	**86.6**	**32.7**	**43.7**	**57.0**	**62.6**	**68.2**	**24.7**
DenseNet121[13]	41.0	56.6	62.8	68.5	23.2	55.0	71.3	78.5	84.3	25.3
DenseNet121+M^3	**53.1**	**67.1**	**72.1**	**75.7**	**32.7**	**76.9**	**85.6**	**89.1**	**91.9**	**40.4**
HA-CNN[18]	43.3	59.7	66.7	74.6	18.9	24.0	39.0	45.1	51.6	13.5
HA-CNN+M^3	**61.6**	**73.6**	**78.7**	**82.9**	**25.8**	**37.6**	**51.9**	**56.8**	**62.8**	**20.5**

Table 5. Cross-dataset validation results with(**+M^3**) our M^3 model on Market1501 and DukeMTMC-reID. **Market1501 → DukeMTMC** mean using the model trained on Market1501 to evaluate DukeMTMC-reID.

Method	CUHK03			Market1501			DukeMTMC-reID			MSMT17		
	R@1	R@20	mAP	R@1	R@20	mAP	R@1	R@20	mAP	R@1	R@20	mAP
HA-CNN[18]	48.0	85.4	47.6	90.6	98.3	75.3	80.7	94.3	64.4	64.7	87.1	37.2
Our only w/ $\mathbf{M}_q$	63.4	87.6	63.5	93.8	98.8	81.2	83.9	95.3	69.0	68.7	88.7	40.6
Our only w/ $\mathbf{M}_g$	65.4	86.2	57.3	94.2	98.4	79.1	83.6	94.4	65.7	66.3	86.4	37.5
Our-Full	**69.8**	**88.8**	**63.5**	**96.5**	**98.9**	**85.2**	**87.1**	**95.8**	**72.2**	**74.3**	**90.0**	**43.8**

Table 6. The influence of each component in our M^3 algorithm.

formance but has little improvement on mAP due to the lack of gallery-specific local discriminant enhancement. In contrast, since RR [43] considers the k-reciprocal nearest neighbors of both query and gallery data, it achieves a large improvement on mAP but with limited improvement on Rank@1 owing to the lack of instance-specific local adaptation. Our M^3 outperforms the other two approaches significantly at both Rank@1 and mAP due to the fully utilization of both the group-level visual similarity sharing information and instance-specific local discriminant enhancement.

5.4. Cross-Set Generalization Ability Validation

We explore the generalization ability of our proposed M^3. We claim the improvement by M^3 is from the testing sample itself which is independent of how the baseline models are trained. Therefore we conduct a cross-set generalization ability validation experiment as shown in Table. 5. Following the setting in [44], the baseline model trained by Market1501 with our M^3 is evaluated on DukeMTMC-reID and vice versa. The results show our M^3 model is able to consistently and significantly improve the baseline performance regardless of whether the baseline is trained by the same-source data or not.

5.5. Ablation Study

The Influence of Model Components: The final retrieval performance of Eqn. 11 relies on a bi-directional retrieval matching, so the influence of each component is shown in Table. 6. As can be seen, by only keeping the query-specific metric adaptation $\mathbf{M}_q$ or the gallery-centric one $\mathbf{M}_g$, we still can achieve a significant improvement. While by performing a full-model bi-directional matching, the performance is further boosted by a large margin which demonstrates the necessity of bi-directional local discriminant enhancement. More visualizations are shown in Fig. 4.

The Influence of λ in Eqn. 11: The weighting parameter λ in Eqn. 11 aims to balance the importance of $\mathbf{M}_q$ and $\mathbf{M}_g$. The full CMC curves w.r.t λ of HA-CNN on CUHK03, Market1501 and DukeMTMC-reID are plotted in Fig. 5 respectively. As can be seen, setting $\lambda = 1$ gives the best performance since we perform a $\max$-normalization to both $\mathbf{M}_q$ and $\mathbf{M}_g$, over-weighting either side is prone to suppress the other side's impact.

6. Conclusion

Unlike previous online P-RID works, in this paper, we propose a novel online joint multi-metric adaptation algorithm which not only takes individual characteristics of testing samples into consideration but also fully utilizes the visual similarity relationships among both query and gallery samples. Our M^3 method can be readily applied to any existing P-RID baselines with the guarantee of performance improvement, and a theoretical sound optimization solution to M^3 keeps a low online computational burden. Compared with the other state-of-the-art online P-RID refinement approaches, our method achieves significant improvement on Rank@1(mAP) performance. Moreover, by implementing our method to the state-of-the-art baselines, their performance is further boosted by a large margin on four challenging large-scale P-RID benchmarks.

Acknowledgements

This work was supported in part by National Science Foundation grant IIS-1619078, IIS-1815561, and the Army Research Office ARO W911NF-16-1-0138, in part by National Natural Science Foundation of China under Grant No. 61976206, No. 61603373, Youth Innovation Promotion Association CAS No. 2019110.

References

[1] Song Bai, Peng Tang, Philip HS Torr, and Longin Jan Latecki. Re-ranking via metric fusion for object retrieval and person re-identification. In *CVPR*, 2019. 1, 2

[2] Sławomir Bak and KP Carr. One-shot metric learning for person re-identification. In *CVPR*, 2017. 1

[3] Arko Barman and Shishir K Shah. Shape: A novel graph theoretic algorithm for making consensus-based decisions in person re-identification systems. In *ICCV*, 2017. 1, 2

[4] R Caruna. Multitask learning: A knowledge-based source of inductive bias. In *ICML*, 1993. 2

[5] Xiaobin Chang, Timothy M Hospedales, and Tao Xiang. Multi-level factorisation net for person re-identification. In *CVPR*, 2018. 6, 7, 8

[6] Tianlong Chen, Shaojin Ding, Jingyi Xie, Ye Yuan, Wuyang Chen, Yang Yang, Zhou Ren, and Zhangyang Wang. Abd-net: Attentive but diverse person re-identification. 2019. 1, 3, 6, 7

[7] Yanbei Chen, Xiatian Zhu, and Shaogang Gong. Person re-identification by deep learning multi-scale representations. In *ICCV*, 2017. 7

[8] Basura Fernando, Elisa Fromont, and Tinne Tuytelaars. Effective use of frequent itemset mining for image classification. In *ECCV*, 2012. 3

[9] Jorge Garcia, Niki Martinel, Christian Micheloni, and Alfredo Gardel. Person re-identification ranking optimisation by discriminant context information analysis. In *ICCV*, 2015. 1, 2, 3

[10] Gösta Grahne and Jianfei Zhu. Efficiently using prefix-trees in mining frequent itemsets. In *FIMI*, volume 90, 2003. 4

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 6, 7

[12] Ruibing Hou, Bingpeng Ma, Hong Chang, Xinqian Gu, Shiguang Shan, and Xilin Chen. Interaction-and-aggregation network for person re-identification. In *CVPR*, 2019. 1, 3

[13] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *CVPR*, 2017. 6, 7, 8

[14] Zakria Hussain and John Shawe-Taylor. Improved loss bounds for multiple kernel learning. In *AISTATS*, 2011. 6

[15] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. *arXiv*, 2016. 6

[16] Mahdi M Kalayeh, Emrah Basaran, Muhittin Gökmen, Mustafa E Kamasak, and Mubarak Shah. Human semantic parsing for person re-identification. In *CVPR*, 2018. 7

[17] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *ACCV*, 2012. 2, 6

[18] Wei Li, Xiatian Zhu, and Shaogang Gong. Harmonious attention network for person re-identification. In *CVPR*, 2018. 1, 2, 6, 7, 8

[19] Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yu-Gang Jiang, and Xiangyang Xue. Pose-normalized image generation for person re-identification. In *ECCV*, 2018. 7

[20] Ergys Ristani and Carlo Tomasi. Features for multi-target multi-camera tracking and re-identification. In *CVPR*, 2018. 7

[21] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *CVPR*, 2018. 6

[22] Jianlou Si, Honggang Zhang, Chun-Guang Li, Jason Kuen, Xiangfei Kong, Alex C Kot, and Gang Wang. Dual attention matching network for context-aware feature sequence based person re-identification. 7

[23] Sören Sonnenburg, Gunnar Rätsch, and Christin Schäfer. A general and efficient multiple kernel learning algorithm. In *NIPS*, 2006. 5

[24] Sören Sonnenburg, Gunnar Rätsch, Christin Schäfer, and Bernhard Schölkopf. Large scale multiple kernel learning. *JMLR*, 2006. 5

[25] Nathan Srebro and Shai Ben-David. Learning bounds for support vector machines with learned kernels. In *COLT*, 2006. 5

[26] Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, and Kyoung Mu Lee. Part-aligned bilinear representations for person re-identification. In *ECCV*, 2018. 6, 7

[27] Yifan Sun, Liang Zheng, Weijian Deng, and Shengjin Wang. Svdnet for pedestrian retrieval. In *ICCV*, 2017. 6, 7

[28] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling. 6, 7

[29] M Feroz T Ali and Subhasis Chaudhuri. Maximum margin metric learning over discriminative nullspace for person re-identification. In *ECCV*, 2018. 1

[30] Cheng Wang, Qian Zhang, Chang Huang, Wenyu Liu, and Xinggang Wang. Mancs: A multi-task attentional network with curriculum sampling for person re-identification. In *ECCV*, 2018. 1, 2, 7

[31] Guanshuo Wang, Yufeng Yuan, Xiong Chen, Jiwei Li, and Xi Zhou. Learning discriminative features with multiple granularities for person re-identification. In *ACM Multimedia Conference*, 2018. 6, 7

[32] Yan Wang, Lequn Wang, Yurong You, Xu Zou, Vincent Chen, Serena Li, Gao Huang, Bharath Hariharan, and Kilian Q Weinberger. Resource aware person re-identification across multiple resolutions. In *CVPR*, 2018. 7

[33] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person trasfer gan to bridge domain gap for person re-identification. In *CVPR*, 2018. 6

[34] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *CVPR*, 2017. 7

[35] Mang Ye, Jun Chen, Qingming Leng, Chao Liang, Zheng Wang, and Kaimin Sun. Coupled-view based ranking optimization for person re-identification. In *ICMM*, 2015. 1, 2

[36] Li Zhang, Tao Xiang, and Shaogang Gong. Learning a discriminative null space for person re-identification. In *CVPR*, 2016. 7

[37] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *CVPR*, 2018. 6

[38] Ying Zhang, Baohua Li, Huchuan Lu, Atshushi Irie, and Xiang Ruan. Sample-specific svm learning for person re-identification. In *CVPR*, 2016. 2

[39] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *ICCV*, 2015. 6

[40] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In *ICCV*, 2019. 1, 3, 6

[41] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *ICCV*, 2017. 6, 7

[42] Zhedong Zheng, Liang Zheng, and Yi Yang. Pedestrian alignment network for large-scale person re-identification. *TCSVT*, 2018. 7

[43] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. 2017. 1, 2, 3, 6, 7, 8

[44] Zhun Zhong, Liang Zheng, Shaozi Li, and Yi Yang. Generalizing a person retrieval model hetero-and homogeneously. In *ECCV*, 2018. 8

[45] Jiahuan Zhou, Pei Yu, Wei Tang, and Ying Wu. Efficient online local metric adaptation via negative samples for person re-identification. In *ICCV*, 2017. 1, 2, 5, 6, 7

[46] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. In *ICCV*, 2019. 6, 7