Person Re-ID through Unsupervised Hypergraph Rank Selection and Fusion

Lucas Pascotti Valem, Daniel Carlos Guimarães Pedronette

Department of Statistics, Applied Mathematics and Computing (DEMAC),
São Paulo State University (UNESP), Rio Claro, Brazil

Abstract

Person Re-ID has been gaining a lot of attention and nowadays is of fundamental importance in many camera surveillance applications. The task consists of identifying individuals across multiple cameras that have no overlapping views. Most of the approaches require labeled data, which is not always available, given the huge amount of demanded data and the difficulty of manually assigning a class for each individual. Recently, studies have shown that re-ranking methods are capable of achieving significant gains, especially in the absence of labeled data. Besides that, the fusion of feature extractors and multiple-source training is another promising research direction not extensively exploited. We aim to fill this gap through a manifold rank aggregation approach capable of exploiting the complementarity of different person Re-ID rankers. In this work, we perform a completely unsupervised selection and fusion of diverse ranked lists obtained from multiple and diverse feature extractors. Among the contributions, this work proposes a query performance prediction measure that models the relationship among images considering a hypergraph structure and does not require the use of any labeled data. Expressive gains were obtained in four datasets commonly used for person Re-ID. We achieved results competitive to the state-of-the-art in most of the scenarios.

Keywords: Person Re-ID, Unsupervised, Hypergraph, Rank, Selection, Fusion

1. Introduction

Person Re-ID is of critical importance in the majority of modern security and surveillance applications [1–4]. The task consists in, given a query image of one person, to identify the same individual across different cameras which have no overlapping views. There are many difficulties for Re-ID retrieval [5], among them: *(i)* different viewpoints, *(ii)* possible low-image resolutions, *(iii)* illumination changes, *(iv)* occlusions, *(v)* difficulty of manually labeling data for training, *(vi)* large amount of data to be processed. The challenge of improving the effectiveness of these systems, especially in open-world scenarios, has attracted a lot of research efforts from the scientific community [4–6].

Initially, person Re-ID retrieval systems were mainly based on the use of hand-crafted feature representations [7–15]. Besides that, other strategies commonly used to generic image retrieval, like bag of visual words [13, 16], have also been employed. Aiming at further improving the quality of the results, rather than using traditional distance measures, researchers have proposed metric learning approaches for Re-ID [11, 17–20], most of them based on supervised models. In [6], an extensive evaluation of multiple combinations of feature extractors and metric learning approaches is discussed.

Due to the significant impact of deep learning on common image retrieval and machine learning [5], the Convolutional Neural Networks (CNN) also have gained a lot of attention and have been widely employed to solve Re-ID tasks in the recent years [5, 21–29]. Among the works, existing networks have been trained for Re-ID [30, 31], and some new architectures have been proposed solely focusing on Re-ID [21–25].

Despite the success of deep learning, one of the main difficulties resides in the lack of large publicly available Re-ID datasets for training [5], mainly be-

cause manually labeling images is a very difficulty task, especially in open-world scenarios [6], where the datasets may increase dynamically. In order to mitigate this problem, many authors have proposed multi-source training, which consists in joining multiple available datasets for training [26, 32, 33], usually increasing the network capacity for generalization.

However, despite several advances in multi-source approaches, it is inevitable the need for intensive manual annotation for obtaining training data. In practice, the demand for extensive training data restricts the generalization and scalability of supervised approaches, especially on person Re-ID tasks, which are not only resource-intensive to acquire identity annotation but also impractical for large-scale data [34, 35].

Usually, there are two main steps associated to Re-ID labeling process [5, 36, 37]: *(i)* intra-camera annotation that requires to compare a person with all the other unlabeled persons in a single camera with multiple views; and *(ii)* inter-camera that requires to match a person across different cameras with multiple views. Let P be the number of persons and S the number of camera views. The intra-camera annotation complexity is $O(S \times P^2)$ and inter-camera annotation complexity ranges from $O(S \times P^2)$ and $O(S^2 \times P^2)$. The worst case of inter-camera occurs because not all persons appear in every camera view in the majority of cases, which makes the association requiring to repeat for all V camera views [36, 37]. Commonly, inter-camera association significantly increases standard annotation costs. There are approaches that mitigate these issues; among them there is the Intra-Camera Supervised (ICS) [36, 37], Multi-Task Multi-Label (MATE) [36, 37], and Cross-camera Feature Prediction [38], which were recently proposed.

Due to the challenge of obtaining large amounts of strongly labeled data, semi-supervised methods have been employed, which is a typical strategy for supervision minimization. Based on information learned from a small set of labeled data,

the idea is to generate labels from unlabeled training data. Some research has been made in this direction [39–42]. However, these methods often suffer from performance degradation and often require a large proportion of expensive cross-view pairwise labeling [36].

There are also weakly supervised strategies that replace accurate labels with inaccurate annotations. In [43], the authors proposed the idea of obtaining multiple bounding boxes of the same person from untrimmed videos. This is done by training a deep learning model capable of extracting multiple bounding boxes of the same person in a video. Recently, [44] proposed to replace image-level annotations with bag-level annotations. Weakly supervised Re-ID is very challenging, since it is rather difficult to model the considerable variances across camera views (e.g. occlusion and illumination) without using strong label data [36, 43].

As an alternative solution, unsupervised approaches [34, 35, 45–51] have been attracting a lot of attention from the research community, especially because, once labeled data are not required, the methods become more suitable for real-world scenarios. In a promising research direction, to address the lack of labels issue, there are works proposed to post-process person Re-ID results by analyzing similarity relationships encoded in the datasets. Several authors have proposed unsupervised re-ranking approaches for Re-ID [52–60]. In [52], the original distances among images are improved by calculating Jaccard correlation scores of each ranked list. Various approaches exploit the reciprocal neighborhood information and other co-occurrence indexes aiming at improving the ranking results.

Another strategy commonly employed on generic unsupervised image retrieval and few exploited on Re-ID tasks consists in fusion approaches [61]. In general, the both broad categories of fusion have been successfully used in generic image retrieval: *(i)* early fusion, which combines the feature vectors; and *(ii)* late fusion, which usually combines ranked lists. Significant results have been achieved

4

based on fusion of different ranked lists and features [62, 63], with the purpose of obtaining more effective results by exploiting the complementarity of each input.

On person Re-ID tasks, some authors have proposed late fusion strategies based on rank aggregation [64, 65]. In [65], the aggregation is performed by attributing weights for each query of each ranker, but no pre-selection of features is performed, and all the features are used as input for the fusion step. There are also early fusion approaches for Re-ID, as [66], that propose a supervised multi-hypergraph fusion model for early fusion of feature extractors. It learns a hypergraph for each feature through a star expansion strategy and they are fused according to weights that the method has learned from training. In [67], a hypergraph structure is used with a deep learning model to improve the performance of the acquired features for Re-ID.

Although fusing different features can represent a significant advantage due to extra information available, how to choose what features to fuse can be a challenging task. Even for supervised approaches, selecting high-effective combinations of visual features remains a complex task, since it is necessary to consider various aspects, such as diversity and complementarity of results. Therefore, selecting features in an unsupervised way, without any labeled data is even more challenging, since no information about the effectiveness of individual visual features is available.

This paper addresses the challenging task of unsupervised selection and fusion of different features for more effective person re-identification. We propose a novel Hypergraph Rank Selection and Fusion (HRSF) framework, which combines an unsupervised rank-based formulation for feature selection [62] with a robust hypergraph model [68] for query performance prediction and rank aggregation based on manifold learning. Among our main contributions, we can highlight:

5

- The proposed HRSF framework uses a rank-based late fusion model, suitable for selection and fusion of a broad diversity of features. The selection is performed by exploiting an unsupervised measure for query performance prediction;

- A hypergraph rank-based formulation is used to encode the high-order relationship among images. This strategy is exploited for both selection and fusion tasks. Hypergraph models were little exploited in Re-ID literature [66, 67];

- The proposed technique is able to learn representations through the hypergraph structure that encodes multiple features from different rankers. The manifold learning based on a hypergraph model [68] allows effective fusion and final ranking. In addition, our approach innovates by fusing different feature extractors trained on different datasets;

- Different from most of fusion approaches for Re-ID which often consider ad hoc selections or combine all the features of the input set [64, 65, 67], our approach is capable of dealing with various features in a completely unsupervised scenario, selecting combinations in a very large search space. To the best of our knowledge, this is the first work which performs an explicit selection and subsequently fusion of features on person Re-ID tasks in a completely unsupervised way.

In fact, the proposed Hypergraph Rank Selection and Fusion (HRSF) approach is based on both [62] and [68]. However, while there are some aspects in common, also there are crucial differences. Among them, we can mention:

- **The problem of selecting and fusing rankers in unsupervised scenarios is very challenging. Hence, the use of an effective algorithm for**

selection is fundamental. While USRF [62] employs traditional query performance prediction approaches (Authority and Reciprocal scores), we propose a new measure named *Hypergraph Query Performance Prediction* (HQPP);

- Originally, USRF [62] uses the CPRR [69] as a method for fusion tasks. Differently, the proposed HRSF uses the LHRR [68] method, which in combination with the HQPP, makes **both the selection and fusion based on hypergraph structures.** The LHRR [68] method is also more robust than the CPRR [69], achieving superior retrieval results in most datasets;

- Both [62] and [68] were originally proposed and evaluated only on general purpose image retrieval scenarios. The proposed HRSF is employed and validated for person Re-ID tasks;

- In unsupervised scenarios, the optimal neighborhood size (k parameters) can be challenging to define. The experiments revealed that HRSF is **more robust than HQPP to different neighborhood sizes**, leading to the most effective results in the majority of scenarios.

A wide experimental evaluation was conducted on 4 different datasets with sizes ranging from 14,097 to 39,902 images. Up to 28 different rankers where considered in each case, resulting in millions of possible combinations. Experiments indicated that our approach was capable of selecting and fusing the rankers, achieving high-effective results superior to all rankers in isolation and competitive to state-of-the-art, when more than 20 recent Re-ID approaches are considered.

The remainder of this paper is organized as follows. Section 2 discusses a formal definition of problem setting. Section 3 presents the proposed rank selection and fusion for Re-ID. Section 4 discusses the conducted experimental evaluation and, finally, Section 5 presents our conclusions.

2. Problem Formulation

This section presents the formal definition of the rank model used in this work. Let $C=\{x_1,\, x_2,\, \ldots, x_N\}$ be an image collection, where N denotes the collection size. Let us consider a retrieval task where, given a query image, returns a list of images from the collection C.

Formally, given a query image x_q, a ranker denoted by R_j computes a ranked list $\tau_q=(x_1,\, x_2,\, \ldots,\, x_k)$ in response to the query. The ranked list τ_q can be defined as a permutation of the k-neighborhood set $\mathcal{N}(q,k)$, which contains the k most similar images to image x_q in the collection C. The permutation τ_q is a bijection from the set $\mathcal{N}(q,k)$ onto the set $[k] = \{1,2,\ldots,k\}$. The $\tau_q(i)$ notation denotes the position of image x_i in the ranked list τ_q.

The ranker R_j can be defined based on diverse approaches, including feature extraction or learning methods. In this paper, a feature-based approach is considered, defining R as a tuple (ε, ρ), where $\varepsilon : C \to \mathbb{R}^d$ is a function that extracts a feature vector v_x from an image $x \in C$; and $d\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a distance function that computes the distance between two images according to their corresponding feature vectors. Formally, the distance between two images x_i, x_j is defined by $\rho(\varepsilon(x_i),\, \varepsilon(x_j))$. The notation $\rho(i,j)$ is used for readability purposes.

A ranked list can be computed by sorting images in a crescent order of distance. In terms of ranking positions we can say that, if image x_i is ranked before image x_j in the ranked list of image x_q, that is, $\tau_q(i) < \tau_q(j)$, then $\rho(q,i) \le \rho(q,j)$. Taking every image in the collection as a query image x_q, a set of ranked lists $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_N\}$ can be obtained.

Different features and distance functions give rises to different rankers which, in turn, produce distinct ranked lists. Let $\mathcal{R} = \{R_1, R_2, \ldots, R_m\}$ be a set of rankers and $R_j \in \mathcal{R}$, we denote by $\mathcal{T}_j$ the set of ranked lists produced by R_j. A ranked list computed by the ranker R_j in response to a query x_q is denoted by $\tau_{j,q}$.

Our proposed method (HRSF) intends to select the most effective rankers from $\mathcal{R}$ based on their respective set of ranked lists, without the need of any labeled data. The selection function (f_s) is formally defined by Equation (1).

$$\mathfrak{X}_n^* = f_s(\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \ldots, \mathcal{T}_m), \tag{1}$$

where $\mathfrak{X}_n^*$ is the set of selected rankers, whose cardinality is n, such that $|\mathfrak{X}_n^*| = n$.

With the objective of facilitating the understanding of the reader, Table 1 presents the main symbols used in this paper.

3. Unsupervised Hypergraph Rank Selection and Fusion

This paper proposes a framework named Hypergraph Rank Selection and Fusion (HRSF) for unsupervised person Re-ID tasks. Our model is inspired by a recent approach [62] proposed for rank selection and fusion on general image retrieval tasks. The method is based on effectiveness estimations and correlation among features computed by a rank-based analysis. In [62], reciprocal references are exploited for effectiveness estimation and feature selection, while rank correlation measures are used for analyzing complementary and diversity aspects. The selected features are fused through a rank-based similarity learning method [69].

The proposed approach differs from previous work [62] on four main aspects: (*i*) the unsupervised measure used to estimate the quality of individual features, which is based on hypergraph structures; (*ii*) a more robust method to fuse the selected features; (*iii*) the proposed approach is evaluated and validated for person Re-ID and; (*iv*) it is more robust to different neighborhood sizes, leading to the most effective results in the majority of scenarios. We innovate by employing a robust hypergraph model for both tasks: query performance prediction and rank fusion. A recent manifold learning approach based on a rank-based hypergraph formulation [68] is exploited.

Table 1: Table of symbols.

Type	Symbol	Description
Retrieval Model	$\mathcal{C}$	Image collection.
	N	Image collection size.
	x_i	Image of index i.
	$\mathcal{N}(q,k)$	Neighborhood set for a query image x_q of size k.
	τ_q	Ranked list for the query image x_q.
	$\tau_q(j)$	Position of the image x_j in the ranked list of the image x_q.
	L	Size of the ranked lists.
	$\mathcal{T}$	Set of ranked lists for all the images in the dataset.
Selection Model	f_s	Function for ranker selection.
	R_i	Ranker of index i.
	$\tau_{i,q}$	Ranked list of the image x_q computed by the ranker x_i.
	$\mathcal{T}_i$	Set of ranked lists produced by the ranker R_i.
	$\mathcal{R}$	Set of rankers.
	m	Size of the set $\mathfrak{R}$.
	$\mathfrak{X}_n$	Candidate combination composed by n rankers.
	$\mathfrak{X}_n^*$	Selected combination composed by n rankers.
	$\mathfrak{X}^*$	Selected combination among all sizes.
	n	Size of a combination.
	w_p	Selection measure for pairs of rankers.
	β	Weight or relevance of the correlation.
	k	Neighborhood size.
	γ	Effectiveness estimation measure (HQPP).
	λ	Correlation measure (RBO).
	α	Constant used in RBO correlation measure.
Hypergraph Model	V	Set of vertexes.
	v_i	Vertex of index i.
	E	Set of hyperedges.
	e_i	Hyperedge of index i.
	$h(e_i,v_j)$	Reliance of vertex v_j to belong to a hyperedge e_i.
	$r(e_i,v_j)$	Density of ranking references to v_j in ranking of x_i and its neighbors.
	$\mathbf{H_G}$	Hypergraph model.
	$\mathbf{H}$	Incidence matrix.
	$\eta_r(i,x)$	Function that assigns a weight to image x according to its position in τ_i.
	η_f	Fused affinity measure used for rank aggregation.

The hyperedges weights, used for estimating the confidence of hyperedge associations, are exploited in our approach to predict the effectiveness of the different person Re-ID rankers in an unsupervised fashion. Additionally, we use a manifold learning algorithm for fusion tasks, the Log-based Hypergraph of Ranking References (LHRR) [68]. This keeps our approach completely unsupervised, and more robust, mostly based on hypergraph structures.

Figure 1 presents an overview of the proposed approach, where the main steps are illustrated and numerated. Given a set of different rankers provided by diverse features extractors and distance measures, (1) a hypergraph estimation measure is employed in order to predict the performance of each ranker without using data labels. In (2), a correlation measure is applied for each pair of rankers. The computed measures are used in the equation presented in the step (3), which computes the equation for each combination. The rankers selected in stage (3) are fused in stage (4), which uses LHRR [68] for rank aggregation.

3.1. Unsupervised Ranker Selection

Given a set of available rankers for person Re-ID and no labeled data, we aim to select a combination which produces the most effective results. The selection measure proposed on [62] relies on the idea that rankers can be analyzed in pairs using an effectiveness estimator and a correlation measure. It consists in attributing a weight to each pair (w_p), in such way that, the ones composed by the most effective rankers and the highest/lowest correlated ones should receive a higher score. This is presented in Equation (2).

$$w_p(\{R_1, R_2\}) = \frac{\gamma(R_1) \times \gamma(R_2)}{(1 + \lambda(R_1, R_2))^\beta},$$

(2)

where γ and λ are used to measure the effectiveness and correlation of rankers, respectively. The γ corresponds to our proposed Hypergraph Query Performance
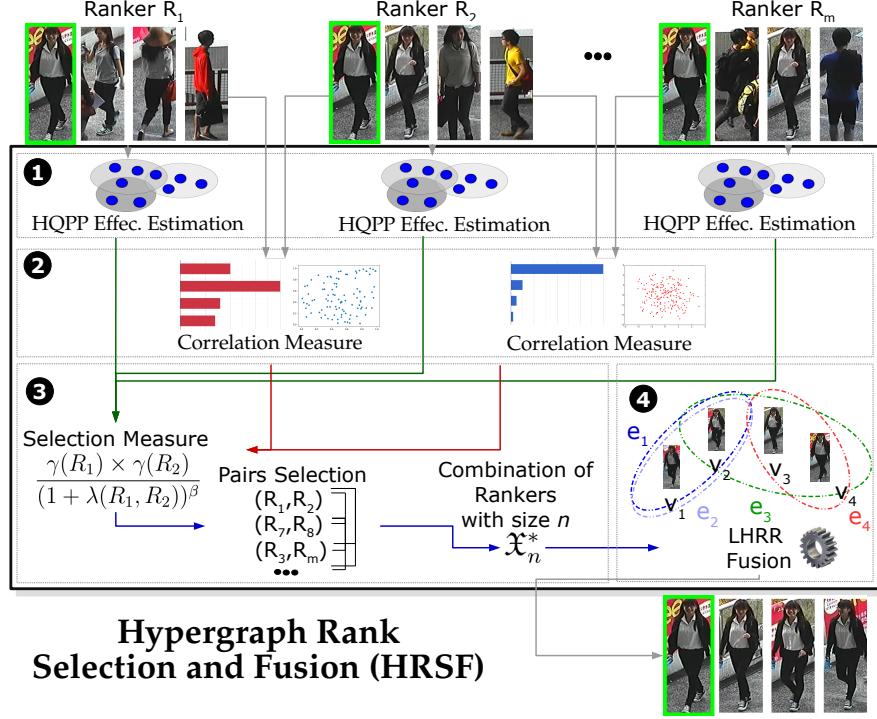
Figure 1: Overview of the HRSF proposed approach.

Prediction (HQPP), which is described in Section 3.2. While the effectiveness can be individually estimated for each ranker, the correlation measure is applied to pairs. The exponent β can be employed to decide if the selection favors the most correlated or diverse rankers. We adopt $\beta = -1$, since in [62] it was used for scenarios with a higher number of features.

Therefore, in our approach, a ranker pair is selected based on the score obtained by w_p. All ranker pairs are ranked according to w_p (in descending order). Only the pairs with the highest scores are selected. The user can choose the number of top combinations to be selected.

After the selection of a pair of rankers, which is denoted by $\mathfrak{X}_2^*$, combinations of other sizes are selected by performing intersection and union operations, in a

procedure detailed described in [62].

For computing the correlation between rankers, the Rank-Biased Overlap (RBO) [70] measure is used. This measure considers the overlap between top-k lists at increasing depths. The weight of the overlap is calculated based on probabilities defined at each depth. It can be formally defined as follows:

$$\lambda(\tau_i, \tau_j, k, \alpha) = (1 - \alpha) \sum_{d=1}^{k} \alpha^{d-1} \times \frac{|\mathcal{N}(i,k) \cap \mathcal{N}(j,k)|}{d},$$ (3)

where $\mathcal{N}(i,k)$ denotes the natural neighborhood of the top-k images for x_i and α is a constant ($\alpha = 0.9$ was used for all the experiments).

In this work, rather than the effectiveness estimations and the fusion method employed in [62], we used a query performance prediction measure (HQPP) and a recent manifold ranking aggregation method (LHRR) that model the ranked lists through hypergraph structures [68]. Both are discussed in the next sub-sections.

3.2. Hypergraph Query Performance Prediction

Query Performance Prediction (QPP) can be broadly defined as the task of estimating the effectiveness of a search/retrieval operation performed in response to a query, where no labeled data is available [71]. Initially proposed for textual retrieval systems [72, 73], the task assumed a diversified taxonomy in the literature and has been establishing as a promising approach in image retrieval systems [74].

In this work, we propose to use a Hypergraph Query Performance Prediction (HQPP) score for predicting the effectiveness of rankings produced by Person Re-ID features. The HQPP score uses a hypergraph formulation recently proposed [68] for manifold ranking on multimedia retrieval. Hypergraphs are a robust generalization of graphs, providing a powerful tool for capturing high-order relationships in several domains [75–77]. In opposition to traditional graph-based approaches, which represent only pairwise relationships, hypergraphs allow con-

13

necting any number of nodes in order to represent similarity among sets of objects [78].

This work and the HQPP score use a hypergraph model mainly based on the following main hypotheses and ideas:

- Similar objects present similar ranked lists and, therefore, similar hyperedges. Once the hyperedges are represented by an incidence matrix, the product of the hyperedges can be exploited to compute a more effective similarity measure between nodes;

- Similar objects are expected to reference each other in the same hyperedge. Therefore, hyperedges that concentrate a high number of ranking references on a few nodes are expected to be more effective. The Hypergraph Query Performance Prediction (HQPP) is formally defined based on this conjecture.

Following the definition of [68], a hypergraph can be defined as a tuple $H_G = (V, E, h_p)$, where V represents a set of vertices and E denotes the hyperedge set. The set of hyperedges E can be defined as a family of subsets of V such that $\bigcup_{e \in E} = V$. To each hyperedge e_i, a positive score $h_p(e_i)$ denotes the confidence of relationships among a set of vertices established by the hyperedge e_i.

While graphs are commonly represented by adjacency matrices, hypergraphs are often represented by incidence matrices. The incidence of a hyperedge e_i on a vertice v_j is represented by an incidence matrix $\mathbf{H}$, defined as follows:

$$h(e_i, v_j) = \begin{cases} r(e_i, v_j), & \text{if } v_j \in e_i, \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

where $h(e_i, v_j)$ denotes the reliance of the vertex v_j to belong to a hyperedge e_i and $r(e_i, v_j)$ is a function with a codomain in the $\mathbb{R}^+$ that indicates the degree to

14

which the vertex v_j belongs to a hyperedge e_i. A hyperedge e_i is defined for each image $x_i \in C$ based on the k-neighborhood set of x_i and its respective neighbors. In this context, the function $r(e_i, v_j)$ is defined based on the density of ranking references to v_j in the ranking of x_i and its neighbors. Formally, the function is defined as:

$$r(e_i, v_j) = \sum_{y \in \mathcal{N}(i,k) \wedge j \in \mathcal{N}(y,k)} \eta_r(i,y) \times \eta_r(y,j), \qquad (5)$$

where $\eta_r(i,x)$ is a function that assigns a weight of relevance to image x according to its position in the ranked list τ_i. The weight assigned to x according to its position in the ranked list τ_i is defined as follows:

$$\eta_r(i,x) = 1 - \log_k \tau_i(x). \qquad (6)$$

The size of the hyperedges varies according to the number of co-occurrence of images. A high diversity of elements may indicate a high degree of uncertainty and this information will be exploited for defining the weights of hyperedges. The weight of a hyperedge $h_p(e_i)$ denotes the confidence of relationships established among vertices by the hyperedge.

In order to compute the weight $h_p(e_i)$, we use the Hypergraph Neighborhood Set $\mathcal{N}_h$, which contains the k vertices with the greatest $h(e_i, \cdot)$ scores in the hyperedge e_i. As such, the hyperedge weight $h_p(e_i)$ is defined as:

$$h_p(e_i) = \sum_{j \in \mathcal{N}_h(i,k)} h(i,j). \qquad (7)$$

A high-effective hyperedge is expected to present an elevated value of h_p, indicating a consistent co-occurrence of the same elements with high confidence of membership. Therefore, the hyperedge weight $h_p(e_i)$ is defined as the Hyper-

graph Query Performance Prediction (HQPP) for ranked list of image x_i, which is denoted by γ:

$$\gamma(\tau_i) = h_p(e_i). \tag{8}$$

For a given ranker R_i, the γ can be computed for all the ranked lists to obtain the value of $\gamma(R_i)$ in Equation (2). We highlight that, HQPP was used to define γ in this work, but our approach is flexible and capable of supporting other measures.

3.3. Hypergraph Manifold Rank Aggregation

Once the person Re-ID features are selected, we fuse the respective produced rankings through a recently proposed manifold learning algorithm [68]. The Log-based Hypergraph of Ranking References (LHRR) [68], briefly described in this section, captures the dataset manifold structure through a hypergraph-based similarity measure, which can be used to rank aggregation tasks.

LHRR [68] exploits the hypergraph formulation discussed in the last section to represent high-order similarity relationships encoded in the dataset manifold. Subsequently, pairwise similarity scores are computed, allowing more effective ranking results. The pairwise similarity is computed based on the conjecture that similar elements present similar hyperedge representations. The similarity between hyperedges is computed based on the product of the incidence matrix $\mathbf{H}$ and its transpose to encode reciprocal relationship. A pairwise similarity matrix $\mathbf{S}$ is computed as:

$$\mathbf{S} = (\mathbf{HH}^T) \circ (\mathbf{H}^T\mathbf{H}) \tag{9}$$

In addition to the product of hyperedges, a Cartesian product operation is conducted to extract useful pairwise relationships direct from the set of elements defined by the hyperedges. Given two hyperedges $e_q, e_i \in E$, the Cartesian product between them can be defined as:

16

$$e_q \times e_i = \{(v_x, v_y): \ v_x \in e_q \wedge v_y \in e_i\}. \tag{10}$$

Let $e_q{}^2$ denote the Cartesian product between the elements of the same hyper-edge e_q, for each pair of vertices $(v_i, v_j) \in e_q{}^2$ a pairwise similarity relationship cp is computed to define the membership degrees of v_i and v_j. The function is formally defined as:

$$cp(e_q, v_i, v_j) = h_p(e_q) \times h(e_q, v_i) \times h(e_q, v_j). \tag{11}$$

A similarity measure based on Cartesian product is defined through a matrix $\mathbf{C}$, with each position computed as follows:

$$c(i,j) = \sum_{e_q \in E \wedge (v_i, v_j) \in e_q{}^2} cp(v_i, v_j). \tag{12}$$

The pairwise similarity defined based on hyperedges and Cartesian product operations provides complementary information. Hence, an affinity matrix $\mathbf{W}$ is computed by combining both matrices as:

$$\mathbf{W} = \mathbf{C} \circ \mathbf{S}. \tag{13}$$

Based on the affinity measure defined by $\mathbf{W}$, a ranking procedure can be performed for each feature giving rise to a new set of ranked lists. Next, a multiplicative rank-based formulation is used to combine the features, exploiting an adaptive weight, which is assigned to each query/feature according to the weight of the respective hyperedge. Let η_f denote the fused affinity measure; each element is computed as follows considering the top-L positions of τ_q:

$$\eta_f(q,i) = \prod_{f=1}^{m} \frac{(1 + h_p(f, e_q))}{(1 + \log_L \tau_{q,f}(i))}, \tag{14}$$

where $h_p(f, e_q)$ is the weight of hyperedge e_q according to the feature f and $\tau_{q,f}(i)$ denote the position of x_i in the ranked list of x_q according to the feature f. The combined affinity measure $\eta_f(\cdot, \cdot)$ gives rise to a unique set of ranked lists which is re-processed by the LHRR [68] algorithm as a single feature.

4. Experimental Evaluation

This section presents the experimental results conducted for evaluating our proposed approach. The experimental evaluation was conducted on 4 datasets often used for person Re-ID with sizes ranging from 14,097 to 39,902. For each dataset, up to 28 rankers were considered of different modalities (e.g. traditional descriptors, bag of visual words, deep learning). The large number of rankers is used with the objective of evaluating the capacity of our selection approach. It is desirable that, if the selection is accurate, only the most effective are selected to be fused. Also, there is a very large number of possible combinations, when all the possible sizes are considered. With 28 rankers, there are 268,435,456 possible combinations. Since it is impractical to execute all of them, selection is of fundamental importance in this context. The experimental evaluation also considers a comparison with fusion baselines and with state-of-the-art Re-ID approaches.

4.1. Experimental Protocol

Our experimental protocol considers 4 datasets, which are detailed in Table 2. For CUHK03 [15], we considered the detected version of the dataset, which uses the bounding boxes extracted by a DPM detector, and adopted the experimental protocol proposed by [52]. Both Market1501 [13] and DukeMTMC [79] use the experimental protocol proposed by the original authors that published the datasets. In the case of the Airport [6], a different protocol was adopted, where all the images (training, probe, and gallery) are considered as queries (probes) and part of the gallery at the same time. For all of them, the MAP (Mean Average Precision)

was reported. The R1 corresponds to the first value of the CMC (Cumulative Matching Characteristics) curve, which indicates the number of ranked lists that have an image that corresponds to the same individual in the first position after the query image. All the evaluations are single-shot (single-query), this means that each query corresponds to only one image.

Table 2: Datasets considered in the experimental evaluation.

Dataset	People	BBox	Cam	Label	Evaluation
CUHK03 [15, 52]	1,467	14,097	2	DPM	R1/ MAP [52]
Market1501 [13]	1,501	32,217	6	DPM	R1/MAP [13]
DukeMTMC [79]	1,812	36,411	8	Manual	R1/MAP [79]
Airport [6]	9,651	39,902	6	ACF	MAP (all queries)

Table 3 presents all the descriptors used in the experimental evaluation (i.e., as the input of our method). The number of rankers per dataset ranges from 21 to 28, and emcompasses different modalities (e.g. statistical, bag of visual words, deep learning). For most of the extractions, the Euclidean distance was considered, no metric learning was applied aiming at keeping the protocol fully unsupervised (not using the labels from the dataset being evaluated). Except for OSNET-AIN, where cosine distance was used, as done in [22]. For the non-deep methods (GBI-COV, LOMO, GOG, WHOS, ELF, HLBP, SDC, and BOVW), we applied PCA to reduce the feature vectors to 100 dimensions before the distances were extracted, as also done by other authors [6]. In total, we considered 7 different Convolutional Neural Networks (CNN): MobileNetV2, RESNET50, HACNN, MLFN, OSNET, OSNET-IBN, and OSNET-AIN. For each of these models, three different trainings were performed: on Market dataset, which is indicated by (M); on DukeMTMC dataset, which is indicated by (D); and on MSMT17 dataset, which is indicated by (MT). Different trainings and models were used with the objective of evaluating our selection and fusion. Most of them were pre-trained on the ImageNet [80]

19

dataset, except HACNN which was trained on Re-ID datasets from scratch. The MSMT17 [81] was used to train some of the networks, since it is a large dataset (126,441 images of 4,101 people in 15 cameras) that can facilitate the capacity of generalization. The majority of the networks were trained considering only the training subset, but RESNET50 and OSNET variants used all of the images (train, probe, and gallery) when trained on MSMT17. The CNN features were extracted with the Torchreid [82] trained weights [1].

To keep the comparisons fair, **since our evaluation is unsupervised, the models that were trained on the target dataset were not used.** For example, train on Market and test on Market is removed. These cases are reported with "—".

4.2. Experimental Analysis

The neighborhood size, denoted by k, is used in multiple steps of our method: for calculating the effectiveness measure, for the correlation measure, and in the fusion stage. Figure 2 presents an experiment that was conducted to evaluate the impact of k on Market1501 dataset, where both R1 and MAP are shown for different values of k. Notice that the method is robust to different parameter settings. We used $k = 20$ for CUHK03, Market1501, DukeMTMC and $k = 10$ for Airport in all of the remaining experiments. For the Airport dataset, a smaller k seems to be more adequate, since it has less images per individual (around 4) compared to the other collections.

In this work, HQPP is proposed as a measure to estimate the quality of each ranker in the selection stage. Aiming at assessing the use of this measure in Re-ID scenarios, Figure 3 shows an experiment where each dot corresponds to a different ranker and the MAP (measure that uses labeled data) is compared to the HQPP

---

[1]https://kaiyangzhou.github.io/deep-person-reid/MODEL_ZOO.html

Table 3: Values of MAP (%) and R-01 (%) for the descriptors and datasets considered on the experiments. The dataset used to train the CNN is referenced between parentheses (M = Market, D = DukeMTMC, MT = MSMT17).

| | Datasets | | | | | | |
| | CUHK03 | | Market1501 | | DukeMTMC | | Airport |
Features	**R1**	**MAP**	**R1**	**MAP**	**R1**	**MAP**	**MAP**
GBICOV [9]	0.63	0.82	10.21	3.27	—	—	—
LOMO [11]	0.79	0.89	19.15	6.46	6.60	2.82	35.35
GOG [10]	0.49	0.77	21.56	7.55	10.82	4.40	34.11
WHOS [12]	0.39	0.56	20.01	6.23	7.50	2.65	34.75
ELF [8]	0.34	0.52	12.02	3.85	2.42	0.83	31.17
HLBP [14]	0.32	0.43	7.07	2.18	0.76	0.54	32.68
SDC [16]	0.18	0.34	11.02	3.78	2.96	1.18	31.57
BOVW-350 [13]	1.69	1.80	33.11	13.34	14.41	6.71	32.73
BOVW-500 [13]	1.56	1.81	32.33	12.94	14.14	6.68	33.09
MobileNetV2 (M) [31]	4.39	4.34	—	—	24.01	12.34	35.62
MobileNetV2 (D) [31]	4.30	4.30	37.80	15.63	—	—	37.23
MobileNetV2 (MT) [31]	8.87	8.51	37.86	16.56	42.59	23.79	38.84
RESNET50 (M) [30]	3.84	3.90	—	—	25.67	13.62	38.01
RESNET50 (D) [30]	5.84	5.85	42.64	18.39	—	—	40.25
RESNET50 (MT) [30]	13.68	13.08	46.59	22.82	52.29	32.00	41.95
HACNN (M) [24]	5.51	5.69	—	—	23.79	13.13	36.40
HACNN (D) [24]	3.11	3.28	43.74	18.87	—	—	38.85
HACNN (MT) [24]	9.71	9.68	49.23	23.30	42.19	25.57	42.94
MLFN (M) [23]	4.91	5.19	—	—	30.39	16.96	38.67
MLFN (D) [23]	4.72	4.74	45.55	20.26	—	—	40.15
MLFN (MT) [23]	10.58	10.19	46.59	21.98	48.70	28.98	41.17
OSNET (MT) [21]	20.83	19.84	65.94	37.36	65.98	45.20	45.47
OSNET-IBN (M) [22]	10.48	10.22	—	—	48.52	26.59	40.96
OSNET-IBN (D) [22]	8.01	7.85	57.48	26.01	—	—	40.65
OSNET-IBN (MT) [22]	21.70	20.78	66.45	37.13	67.41	45.52	45.37
OSNET-AIN (M) [22]	12.14	11.67	—	—	52.42	30.35	42.05
OSNET-AIN (D) [22]	9.54	9.24	61.10	30.64	—	—	42.79
OSNET-AIN (MT) [22]	28.49	27.00	69.95	43.30	71.14	52.69	52.26

performance prediction score. As can be seen in the graph, there is a high correlation between the measures. The Pearson correlation among the dots is 0.9678, which indicates the high effectiveness of the selection strategy.
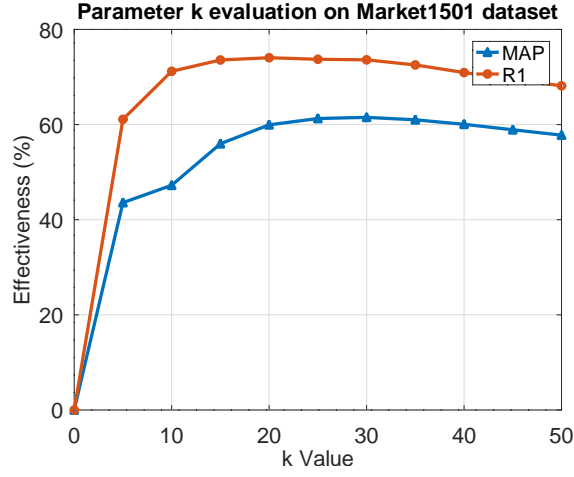
Parameter k evaluation on Market1501 dataset

Figure 2: Evaluation of the impact of parameter *k* on MAP and R1 for Market1501 dataset.

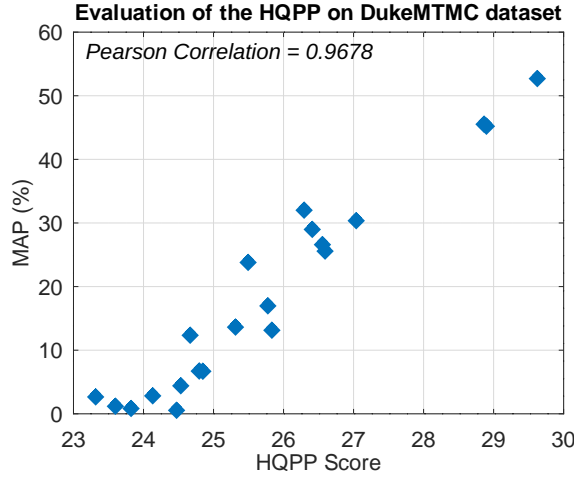**Evaluation of the HQPP on DukeMTMC dataset**

Figure 3: Evaluation of the HQPP measure compared to the MAP on DukeMTMC dataset.

The HRSF ranks the best combinations for each size. The user can choose the number of top combinations to be selected. We conducted an experiment on

CUHK03, Market, and DukeMTMC (Figures 4, 5, and 6) where the average MAP and R1 of the selected ranker pairs is presented as the number of selected pairs changes. Notice that the highest MAP and R1 values are in the first position (top-1), which evinces that the combination with the highest w_p (ranked in the first position) is also the one with highest effectiveness.
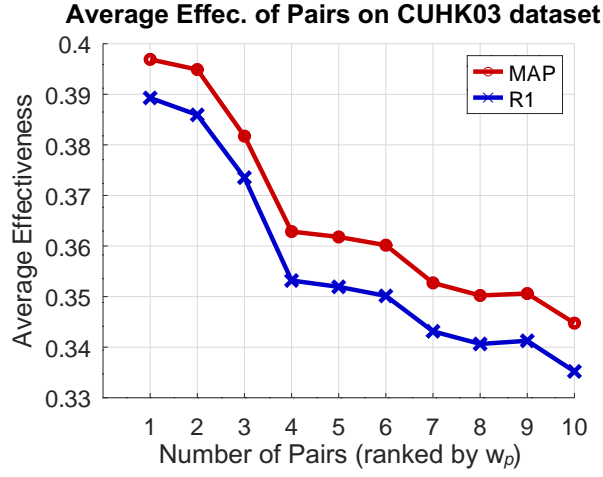

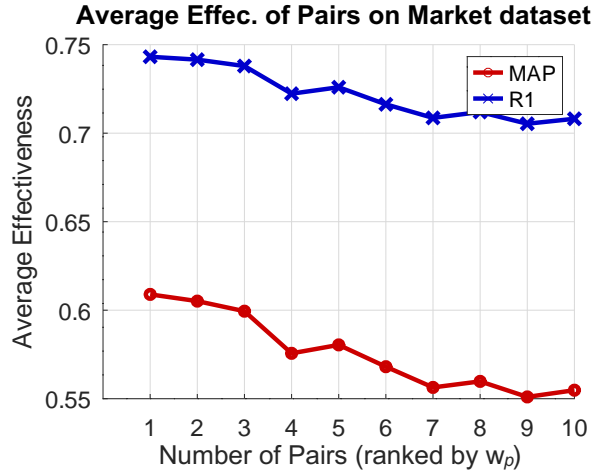Figure 4: Average MAP of top pairs on CUHK03 dataset.

Figure 5: Average MAP of top pairs on Market dataset.

23

The best combination available in the top-5 for each size is reported in Table 4. We report sizes from 1 to 6 ($\mathfrak{X}_2^*$, ..., $\mathfrak{X}_6^*$) and the best combination among them (which can be denoted just as $\mathfrak{X}^*$) is highlighted in bold. The best isolated ranker in each case is also listed for comparison purposes and to facilitate the visualization of the relative MAP gain. Notice that OSNET, OSNET-AIN, and OSNET-IBN are the most commonly selected rankers, which evinces the effectiveness of our selection, once these rankers are among the most effective ones. Additionally, in all cases the proposed selection and fusion **achieves better results the the best ranker in isolation**. The complementarity among the methods can be exploited by our approach achieving gains up to +47% (MAP) after the selection and fusion is performed. The results also indicate that gains can be obtained when networks trained on different datasets are combined, even when the same architecture is used.
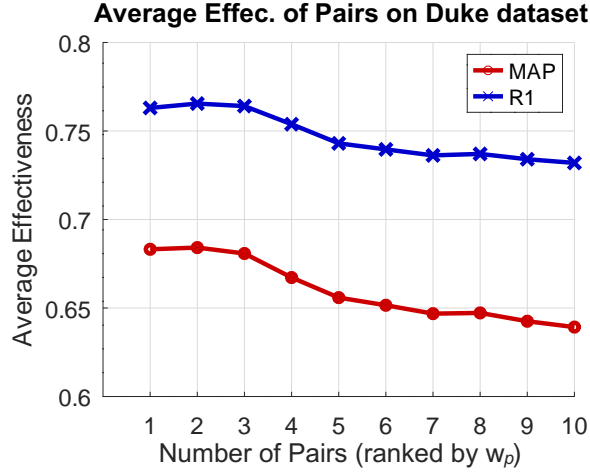

Figure 6: Average MAP of top pairs on Duke dataset.

Table 4: The best selected combination of each size (among top-5) is reported on each dataset.

Dataset	Comb. Size	Selected and Fused Rankers	R1 (%)	MAP (%)	MAP R. Gain
CUHK03	Best R	OSNET-AIN (MT)	28.49	27.00	—
	$\mathfrak{X}_2^*$	**OSNET-AIN (MT) + OSNET-IBN (MT)**	**39.04**	**39.69**	**+47.00%**
	$\mathfrak{X}_3^*$	OSNET-AIN (MT) + OSNET-IBN (MT) + OSNET (MT)	39.13	39.58	+46.59%
	$\mathfrak{X}_4^*$	OSNET-AIN (MT) + OSNET-IBN (MT) + OSNET (MT) + OSNET-AIN (M)	38.02	38.80	+43.70%
	$\mathfrak{X}_5^*$	OSNET-AIN (MT) + OSNET-IBN (MT) + OSNET (MT) + OSNET-AIN (M) + OSNET-IBN (M)	36.15	37.11	+37.44%
	$\mathfrak{X}_6^*$	HACNN (MT) + OSNET-AIN (D) + OSNET-AIN (M) + OSNET-AIN (MT) + OSNET-IBN (MT) + OSNET (MT)	35.46	36.17	+33.96%
Market1501	Best R	OSNET-AIN (MT)	69.95	43.30	—
	$\mathfrak{X}_2^*$	OSNET-AIN (MT) + OSNET (MT)	74.32	60.89	+40.62%
	$\mathfrak{X}_3^*$	OSNET-AIN (MT) + OSNET (MT) + OSNET-AIN (D)	75.56	62.64	+44.67%
	$\mathfrak{X}_4^*$	**OSNET-AIN (MT) + OSNET (MT) + OSNET-AIN (D) + OSNET-IBN (MT)**	**75.71**	**62.94**	**+45.36%**
	$\mathfrak{X}_5^*$	OSNET-AIN (MT) + OSNET (MT) + OSNET-AIN (D) + OSNET-IBN (MT) + HACNN (D)	74.00	60.69	+40.16%
	$\mathfrak{X}_6^*$	HACNN (MT) + OSNET-AIN (D) + OSNET-AIN (MT) + OSNET-IBN (D) + OSNET-IBN (MT) + OSNET (MT)	73.57	59.85	+38.22%
DukeMTMC	Best R	OSNET-AIN (MT)	71.14	52.69	—
	$\mathfrak{X}_2^*$	OSNET-AIN (MT) + OSNET-IBN (MT)	76.80	68.51	+30.02%
	$\mathfrak{X}_3^*$	**OSNET-AIN (MT) + OSNET-IBN (MT) + OSNET (MT)**	**77.24**	**68.88**	**+30.73%**
	$\mathfrak{X}_4^*$	OSNET-AIN (MT) + OSNET-IBN (MT) + OSNET (MT) + RESNET (MT)	76.89	68.56	+30.12%
	$\mathfrak{X}_5^*$	OSNET-AIN (MT) + OSNET-IBN (MT) + OSNET (MT) + RESNET (MT) + OSNET-AIN (M)	76.39	67.72	+28.53%
	$\mathfrak{X}_6^*$	OSNET-AIN (MT) + OSNET-IBN (MT) + OSNET (MT) + RESNET (MT) + OSNET-AIN (M) + MLFN (MT)	75.90	66.96	+27.08%
Airport	Best R	OSNET-AIN (MT)	—	52.26	—
	$\mathfrak{X}_2^*$	OSNET-AIN (MT) + OSNET (MT)	—	52.43	+0.33%
	$\mathfrak{X}_3^*$	OSNET-AIN (MT) + OSNET (MT) + OSNET-IBN (MT)	—	53.38	+2.14%
	$\mathfrak{X}_4^*$	OSNET-AIN (MT) + OSNET (MT) + OSNET-IBN (MT) + OSNET-AIN (M)	—	53.91	+3.16%
	$\mathfrak{X}_5^*$	**OSNET-AIN (MT) + OSNET (MT) + OSNET-IBN (MT) + OSNET-AIN (M) + HACNN (MT)**	—	**54.09**	**+3.50%**
	$\mathfrak{X}_6^*$	OSNET-AIN (MT) + OSNET (MT) + OSNET-IBN (MT) + OSNET-AIN (M) + HACNN (MT) + MLFN (MT)	—	54.02	+3.37%

4.3. Comparison with Fusion Baselines

In order to evaluate the proposed method compared to other approaches that both select and fuse the input features, Table 5 presents the proposed approach, HRSF, compared to both early and late fusion baselines on the four datasets. In all the cases, the same set of features, which were presented in the experimental protocol, were used. For the early fusion methods, the default parameters were used and all the features were processed with PCA to reduce the feature vectors to 100 components. From all the features, the top-1000 were selected to compose the new feature vector and the Euclidean distance was computed. As can be seen, the results of our approach are superior in most cases (CUHK03, Market1501, Airport) and comparable in others (DukeMTMC).

Table 5: Proposed approach compared to early and late fusion baselines.

Dataset	Category	Method	R1 (%)	MAP (%)
CUHK03	Early Fusion	Laplace [83]	9.56	10.19
		SPEC [84]	9.29	9.97
	Late Fusion	USRF [62]	38.24	39.03
		HRSF (ours)	**39.04**	**39.69**
Market1501	Early Fusion	Laplace [83]	**82.07**	61.26
		SPEC [84]	77.14	54.90
	Late Fusion	USRF [62]	75.97	62.69
		HRSF (ours)	75.71	**62.94**
DukeMTMC	Early Fusion	Laplace [83]	59.29	43.56
		SPEC [84]	59.29	43.56
	Late Fusion	USRF [62]	**77.82**	**68.98**
		HRSF (ours)	77.24	68.88
Airport	Early Fusion	Laplace [83]	—	45.33
		SPEC [84]	—	45.32
	Late Fusion	USRF [62]	—	39.75
		HRSF (ours)	—	**54.09**

An experiment was conducted with the objective of clarifying the robustness of HRSF to different values of k when compared to USRF. Figures 7 and 8 present the MAP of the best combination ($\mathfrak{X}^*$) among top-5 for different values of k on Market and Duke datasets, respectively. The MAP for $k = 20$ is the same of Table 5. Notice that, while both methods seem comparable for $k = 20$, **our method provided a significant higher MAP for other values of** k. This is fundamental for unsupervised scenarios, where the optimal k can be challenging to define.
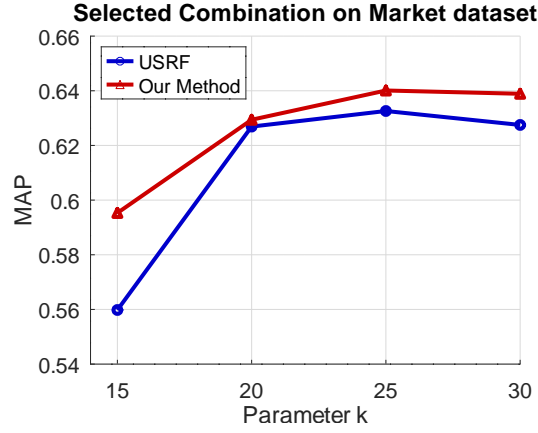

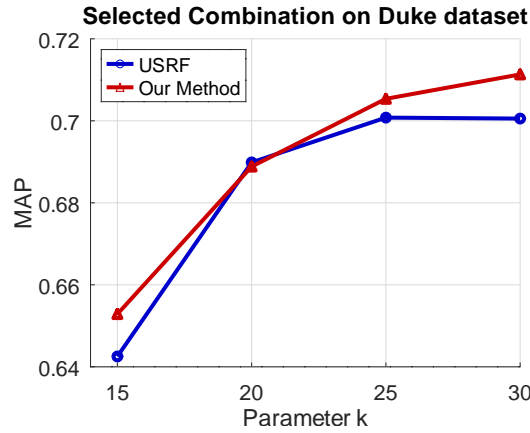Figure 7: Selected Combination (among top-5) on Market considering MAP.

Figure 8: Selected Combination (among top-5) on Duke considering MAP.

4.4. State-of-the-Art

This section presents comparisons with recent baselines. The taxonomy often varies in the literature; some of the subcategories of unsupervised Re-ID methods are: (*i*) **Cross-domain Methods:** they are trained on one (single-source) or more (multi-source) labeled source domains. The evaluation is conducted on the target dataset. It can be understood as a type of transfer learning; (*ii*) **Domain Adaptive Methods:** they use cross-domain training and perform a second step (adaptation) using unlabeled data from the target domain.

However, since they are all unsupervised and often there is overlap among the categories, we insert all of them into a single group named Unsupervised Domain Adaptation Methods [85]. Table 6 presents our results compared to around **20 state-of-the-art Unsupervised Domain Adaptation Methods [85]**.

In order to perform a fair comparison, it contains **only methods that did not use the labels from the target dataset for training** (train on CUHK03 and test on CUHK03 is not used, for example). Therefore, supervised and semi-supervised methods are not included. The abbreviations in parentheses indicate the datasets used for training [2]. For example, the use of (D, M) indicates that the reported result corresponds to a training done either on Duke or on Market dataset. The results reported on Market were trained on Duke and the results reported on Duke were trained on Market. **None of the presented methods were trained using labels from the target dataset.** The abbreviations were omitted for baselines that used more than 5 datasets as source for training (CAMEL [33] and baseline by [26]), but they can be consulted in [26], which used similar baselines and protocol.

We provided the best results for each method (considering the original papers) with the objective of keeping the evaluation as far as possible. Since the code and

---

[2]C02 = CUHK02, C03 = CUHK03, M = Market1501, D = DukeMTMC, MT = MSMT17

implementation are not available for the majority of methods, it is not possible to provide results considering the same sources in all cases.

Both MAP and R1 are reported in all the cases and our best results are presented in bold. The baselines do not perform any form of selection. We highlight that our method receives all the rankers as input and performs a wide selection among rankers with high and low effectiveness, which is a very challenging scenario. In contrast, **none of the baselines are required to perform any selection** and the features are chosen manually. **The selection stage is an important aspect and contribution of HRSF** that it is hard to replicate in the baselines. Notice that our approach achieved competitive or superior results for all the evaluated datasets.

With the objective to facilitate the visualization of the best results in the state-of-the-art, Table 7 presents the rank of the methods according to MAP and R1. The gray cells with bold values correspond to methods that have achieved a higher rank than HRSF in each case. Our method achieved the best MAP on DukeMTMC and had the second position in the other two datasets. For R1, HRSF is positioned among the top-4 in all cases. The mean of the rank on each dataset is presented in the most right columns. Notice that our method achieved one of the highest rank means among all of the methods, being only slightly behind ISSDA [86]. The comparisons show that, besides our results being among the best for all datasets, in some cases, other non-fusion based methods provided higher values than our approach. There are some possible explanations for this:

- Each dataset has different aspects (e.g., image resolution, picture angles, environment, number of images per person, dataset size). For this reason, **different methods may perform better or worse on distinct datasets**;

- The Baseline by [26] performs a multi-source training. The results reported by [26] are based on the transfer learning of a **training performed on 7 Re-**

ID datasets. It considers, by far, the **largest labeled source of all the baselines**, which leads to high results, especially on Market where it is ranked as the 2nd/3rd best R1/MAP (shown in Table 7);

- An idea that is exploited by some baselines is the generation of pseudo-labels. The most promising example is ISSDA [86], which has the best results on Market and is well ranked on DukeMTMC. Different from the others, ISSDA employs a self-supervised iterative pseudo-label generation and training. However, besides the effectiveness, **the authors [86] claim that the training stage is very time-consuming** since it requires, among other aspects, the execution of a clustering algorithm. Furthermore, ISSDA has an average ranking of 1.5 against 1.67 of our method (Table 7, MAP measure); **However, ISSDA does not report results on CUHK03. Our average ranking without considering CUHK03 is also 1.5**;

- The MAR [47] performs **soft label generation**. The applied strategy is capable of achieving promising results for improving the DukeMTMC dataset with R1 measure. However, apparently, the quality of the soft labels varies according to the dataset.

Since the proposed method is flexible, among future works, we intend to expand our approach by exploiting the use of pseudo-labels. This can be done by modifying HRSF or by providing methods like ISSDA [86] and MAR [47] as input, for example. In this work, only the rankers in Table 3 were used.

Table 6: State-of-the-art comparison considering MAP (%) and R-01 (%).

	Datasets					
	Market1501		DukeMTMC		CUHK03	
	R1	MAP	R1	MAP	R1	MAP
Unsupervised Domain Adaptation Methods						
ARN [45]	70.3	39.4	60.2	33.4	—	—
EANet [27]	66.4	40.6	45.0	26.4	51.4	31.7
ECN [46]	75.1	43.0	63.3	40.4	—	—
MAR [47]	67.7	40.0	87.1	48.0	—	—
TAUDL [48]	63.7	41.2	61.7	43.5	44.7	31.2
UTAL [49]	69.2	46.2	62.3	44.6	56.3	42.3
HHL (D,M) [87]	62.2	31.4	46.9	27.2	—	—
HHL (C03) [87]	56.8	29.8	42.7	23.4	—	—
ATNet (D,M) [28]	55.7	25.6	45.1	24.9	—	—
CSGLP (D,M) [88]	63.7	33.9	56.1	36.0	—	—
ISSDA (D,M) [86]	81.3	63.1	72.8	54.1	—	—
EANet (C03) [27]	59.4	33.3	39.3	22.0	—	—
EANet (D,M) [27]	61.7	32.9	51.4	31.7	—	—
SPGAN (D,M) [89]	43.1	17.0	33.1	16.7	—	—
DAAM (D,M) [90]	42.3	17.5	29.3	14.5	—	—
AF3 (D,M) [91]	67.2	36.3	56.8	37.4	—	—
AF3 (MT) [91]	68.0	37.7	66.3	46.2	—	—
PAUL (MT) [92]	68.5	40.1	72.0	53.2	—	—
EMTL (C02+D+M) [32]	52.8	25.1	39.7	22.3	—	—
CAMEL [33]	54.5	26.3	—	—	31.9	—
Baseline by [26]	80.5	56.8	67.4	46.9	29.4	27.4
Unsupervised Selection and Fusion (ours)						
HRSF ($\mathfrak{X}_2^*$)	74.32	60.89	76.80	68.51	**39.04**	**39.69**
HRSF ($\mathfrak{X}_3^*$)	75.56	62.64	**77.24**	**68.88**	39.13	39.58
HRSF ($\mathfrak{X}_4^*$)	**75.71**	**62.94**	76.89	68.56	38.02	38.80
HRSF ($\mathfrak{X}_5^*$)	74.00	60.69	76.39	67.72	36.15	37.11
HRSF ($\mathfrak{X}_6^*$)	73.57	59.85	75.90	66.96	35.46	36.17
HRSF ($\mathfrak{X}^*$, best result)	**75.71**	**62.94**	**77.24**	**68.88**	**39.04**	**39.69**

Table 7: State-of-the-art methods ranked by their results.

	Market1501		DukeMTMC		CUHK03		Mean	
	R1	MAP	R1	MAP	R1	MAP	R1	MAP
Unsupervised Domain Adaptation Methods								
ARN [45]	5	10	10	12	—	—	7.5	11
EANet [27]	11	7	16	15	**2**	3	9.67	8.34
ECN [46]	4	5	7	9	—	—	5.5	7
MAR [47]	9	9	**1**	4	—	—	5	6.5
TAUDL [48]	12	6	9	8	**3**	4	8	6
UTAL [49]	6	4	8	7	**1**	**1**	5	4
HHL (D,M) [87]	14	16	14	14	—	—	14	15
HHL (C03) [87]	17	17	17	17	—	—	17	17
ATNet (D,M) [28]	18	19	15	16	—	—	16.5	17.5
CSGLP (D,M) [88]	13	13	12	11	—	—	12.5	12
ISSDA (D,M) [86]	**1**	**1**	3	2	—	—	**2**	**1.5**
EANet (C03) [27]	16	14	19	19	—	—	11.67	16.5
EANet (D,M) [27]	15	15	13	13	—	—	14	14
SPGAN (D,M) [89]	21	22	20	20	—	—	20.5	21
DAAM (D,M) [90]	22	21	21	21	—	—	21.5	21
AF3 (D,M) [91]	10	12	11	10	—	—	10.5	11
AF3 (MT) [91]	8	11	6	6	—	—	7	8.5
PAUL (MT) [92]	7	8	4	3	—	—	5.5	5.5
EMTL (C02+D+M) [32]	20	20	18	18	—	—	19	19
CAMEL [33]	19	18	—	—	5	—	12	18
Baseline by [26]	**2**	3	5	5	6	5	4.34	4.34
Unsupervised Selection and Fusion (ours)								
HRSF ($\mathfrak{X}^*$, best result)	**3**	**2**	**2**	**1**	**4**	**2**	**3**	**1.67**

4.5. *Visual Results*

Some qualitative results were also elaborated to evince the quality of our obtained results. Two different queries (x_1, x_2) for the same person (ID) were chosen

from the DukeMTMC dataset. Figure 9 presents a graph for each ranker that composes the best combination ($\mathfrak{X}^*$) on DukeMTMC and the HRSF result. Each dot represents a gallery image, which is positioned in the graph according to its distance to the query images (x_1, x_2). The idea is that, since the query images are of the same person, the distance between them should be small and the images of the same ID should be closer to the bottom left corner. Images obtained from different camera views are presented in different symbols. It shows that the HRSF method was capable of reducing the distance of all the images belonging to the same class when compared to isolated rankers (OSNET, OSNET-IBN, OSNET-AIN).
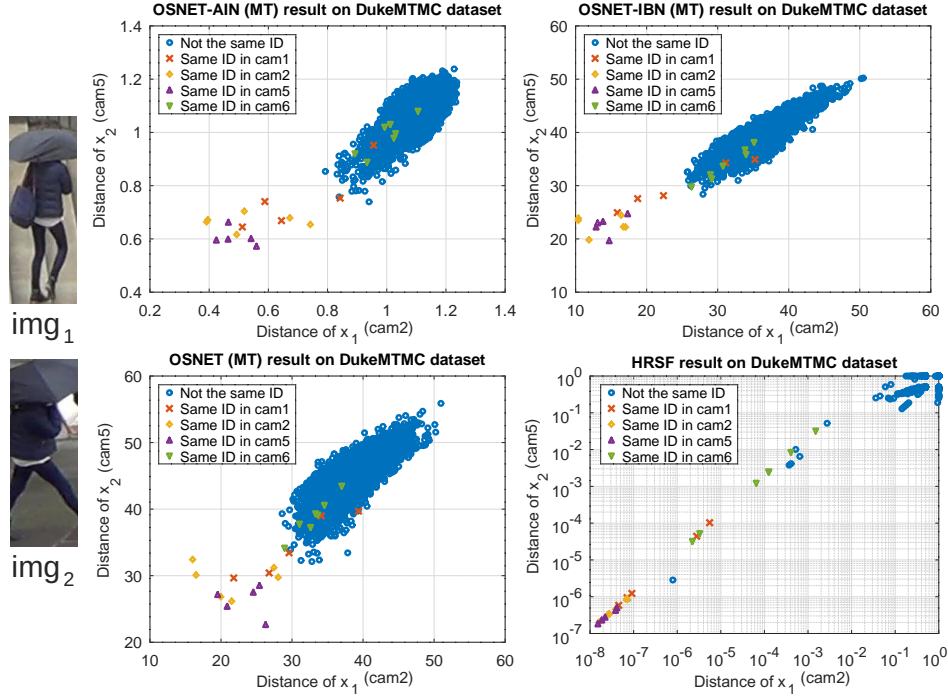

Figure 9: Distance distribution for two query images on DukeMTMC dataset.

Figures 10 and 11 present examples of visual queries on the DukeMTMC and CUHK03 datasets, respectively. The results are shown for the best combination obtained by HRSF ($\mathfrak{X}^*$) and the rankers that compose it. The query image is presented with green borders, and the wrong results with red borders. Notice that,

in these cases, beyond selecting the best results, our approach was also capable of removing most of the incorrectly retrieved images.

OSNET-AIN (MT)

OSNET-IBN (MT)

OSNET (MT)

HRSF Fusion ($\mathfrak{X}^*$)

Figure 10: Examples to illustrate the impact of HRSF selection and fusion on DukeMTMC dataset.

OSNET-AIN (MT)

OSNET-IBN (MT)

HRSF Fusion ($\mathfrak{X}^*$)
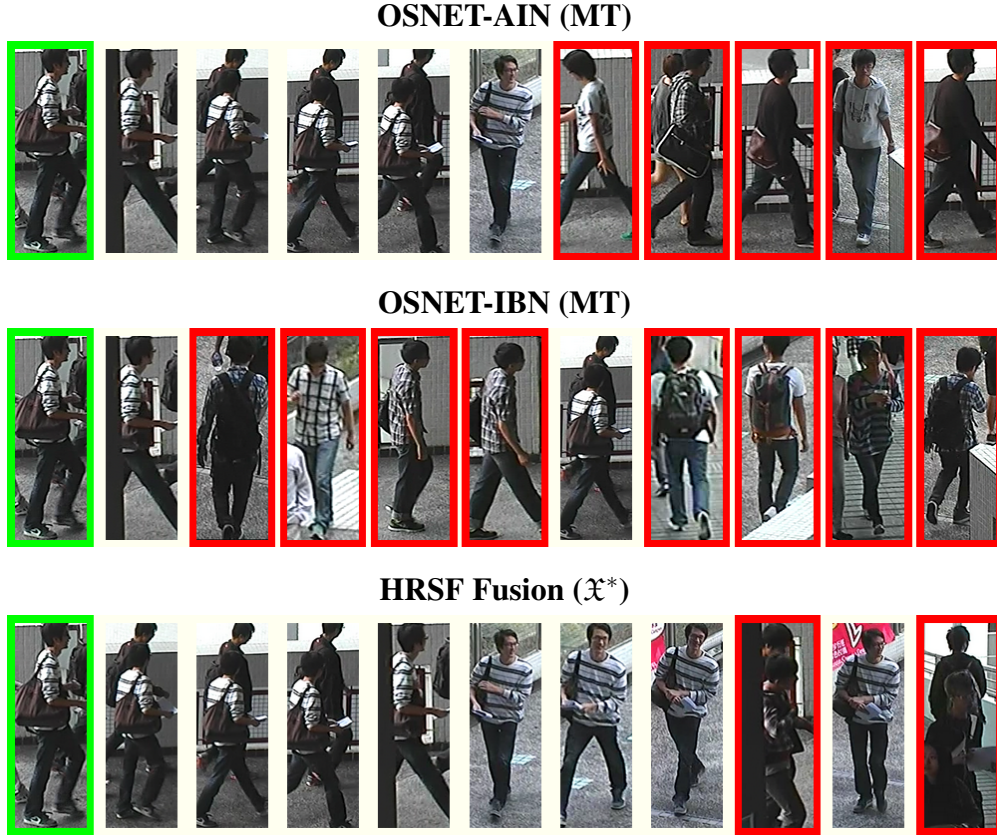
Figure 11: Examples to illustrate the impact of HRSF selection and fusion on CUHK03 dataset.

5. Conclusion

In this work, we have presented an approach to select and aggregate results from different Re-ID methods without using the target dataset training labels. The relationship among the dataset elements is modeled using hypergraph structures both for query performance prediction and fusion. In most of the cases, the best rankers were properly selected and fused. Our approach achieved significant gains in the majority of the scenarios and results competitive or superior to the state-of-the-art, including the most recent methods.

As future work, we intend to investigate the use of the proposed approach in multi-query scenarios and further expand the experimental protocol, including other datasets and feature extractors. Another idea is to expand our approach by exploiting the use of pseudo-labels. Our method, as currently presented, is ready to be executed in scenarios that do not require real time. In addition, we intend to address these real time scenarios in future work, where optimizations and parallelization techniques can be employed.

ACKNOWLEDGEMENTS

References

[1] O. Camps, M. Gou, T. Hebble, S. Karanam, O. Lehmann, Y. Li, R. J. Radke, Z. Wu, F. Xiong, From the lab to the real world: Re-identification in an airport camera network, IEEE Transactions on Circuits and Systems for Video Technology 27 (3) (2017) 540–553.

[2] M. Cristani, V. Murino, Chapter 10 - person re-identification, in: R. Chellappa, S. Theodoridis (Eds.), Academic Press Library in Signal Processing, Volume 6, Academic Press, 2018, pp. 365 – 394. `doi:https://doi.org/10.1016/B978-0-12-811889-4.00010-5`.

[3] Y. Li, Z. Wu, S. Karanam, R. Radke, Real-world re-identification in an airport camera network, 2014, pp. 1–6. `doi:10.1145/2659021.2659039`.

[4] L. Zheng, Y. Yang, A. G. Hauptmann, Person re-identification: Past, present and future, CoRR abs/1610.02984. `arXiv:1610.02984`.
URL `http://arxiv.org/abs/1610.02984`

[5] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, S. C. H. Hoi, Deep learning for person re-identification: A survey and outlook (2020). `arXiv:2001.04193`.

[6] S. karanam, M. Gou, Z. Wu, A. Rates-Borras, O. Camps, R. J. Radke, A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (3) (2019) 523–536.

[7] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, S. Z. Li, Salient color names for person re-identification, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, Cham, 2014, pp. 536–551.

[8] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: D. Forsyth, P. Torr, A. Zisserman (Eds.), Computer Vision – ECCV 2008, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 262–275.

[9] B. Ma, Y. Su, F. Jurie, Covariance descriptor based on bio-inspired features for person re-identification and face verification, Image and Vision Computing 32 (6) (2014) 379 – 390.

[10] T. Matsukawa, T. Okabe, E. Suzuki, Y. Sato, Hierarchical gaussian descriptor for person re-identification, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1363–1372.

[11] S. Liao, Y. Hu, Xiangyu Zhu, S. Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 2197–2206.

[12] G. Lisanti, I. Masi, A. D. Bagdanov, A. D. Bimbo, Person re-identification by iterative re-weighted sparse ranking, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (8) (2015) 1629–1642.

[13] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1116–1124.

[14] F. Xiong, M. Gou, O. Camps, M. Sznaier, Person re-identification using kernel-based metric learning methods, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, Cham, 2014, pp. 1–16.

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

[16] R. Zhao, W. Ouyang, X. Wang, Unsupervised salience learning for person re-identification, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3586–3593.

[17] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2288–2295.

[18] J. García, N. Martinel, A. Gardel, I. Bravo, G. L. Foresti, C. Micheloni, Modeling feature distances by orientation driven classifiers for person re-identification, Journal of Visual Communication and Image Representation 38 (2016) 115 – 129.

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

[20] L. Zhang, T. Xiang, S. Gong, Learning a discriminative null space for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[21] K. Zhou, Y. Yang, A. Cavallaro, T. Xiang, Omni-scale feature learning for person re-identification, in: The IEEE International Conference on Computer Vision (ICCV), 2019.

[22] K. Zhou, Y. Yang, A. Cavallaro, T. Xiang, Learning generalisable omni-scale representations for person re-identification, arXiv preprint arXiv:1910.06827.

[23] X. Chang, T. M. Hospedales, T. Xiang, Multi-level factorisation net for person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[24] W. Li, X. Zhu, S. Gong, Harmonious attention network for person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[25] Y. Sun, L. Zheng, Y. Yang, Q. Tian, S. Wang, Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline), in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Computer Vision – ECCV 2018, Springer International Publishing, Cham, 2018, pp. 501–518.

[26] D. Kumar, P. Siva, P. Marchwica, A. Wong, Fairest of them all: Establishing a strong baseline for cross-domain person reid, CoRR abs/1907.12016.
URL http://arxiv.org/abs/1907.12016

[27] H. Huang, W. Yang, X. Chen, X. Zhao, K. Huang, J. Lin, G. Huang, D. Du, Eanet: Enhancing alignment for cross-domain person re-identification, arXiv preprint arXiv:1812.11369.

[28] J. Liu, Z.-J. Zha, D. Chen, R. Hong, M. Wang, Adaptive transfer network for cross-domain person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[29] C. Tay, S. Roy, K. Yap, Aanet: Attribute attention network for person re-identifications, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 7127–7136.

[30] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[31] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.

[32] Y. Xian, H. Hu, Enhanced multi-dataset transfer learning method for unsupervised person re-identification using co-training strategy, IET Computer Vision 12 (8) (2018) 1219–1227.

[33] H.-X. Yu, A. Wu, W.-S. Zheng, Cross-view asymmetric metric learning for unsupervised person re-identification, in: The IEEE International Conference on Computer Vision (ICCV), 2017.

[34] Y. Lin, Y. Wu, C. Yan, M. Xu, Y. Yang, Unsupervised person re-identification via cross-camera similarity exploration, IEEE Transactions on Image Processing`doi:` `https://doi.org/10.1109/TIP.2020.2982826`.

[35] Y. Lin, L. Xie, Y. Wu, C. Yan, Q. Tian, Unsupervised person re-identification via softened similarity learning, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2020, 2020, pp. 3387–3396.

[36] X. Zhu, X. Zhu, M. Li, P. Morerio, V. Murino, S. Gong, Intra-camera supervised person re-identification, International Journal of Computer Vision 129 (5) (2021) 1580–1595. `doi:10.1007/s11263-021-01440-4`.

[37] X. Zhu, X. Zhu, M. Li, V. Murino, S. Gong, Intra-camera supervised person re-identification: A new benchmark, in: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 1079–1087. `doi:10.1109/ICCVW.2019.00138`.

[38] W. Ge, C. Pan, A. Wu, H. Zheng, W.-S. Zheng, Cross-Camera Feature Prediction for Intra-Camera Supervised Person Re-Identification across Distant Scenes, Association for Computing Machinery, New York, NY, USA, 2021, p. 3644–3653. URL `https://doi.org/10.1145/3474085.3475382`

[39] D. Figueira, L. Bazzani, H. Q. Minh, M. Cristani, A. Bernardino, V. Murino, Semi-supervised multi-feature learning for person re-identification, in: 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, 2013, pp. 111–116. `doi:10.1109/AVSS.2013.6636625`.

[40] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, J. Bu, Semi-supervised coupled dictionary learning for person re-identification, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3550–3557. `doi:10.1109/CVPR.2014.454`.

[41] H. Wang, X. Zhu, T. Xiang, S. Gong, Towards unsupervised open-set person re-identification, in: 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 769–773. `doi:10.1109/ICIP.2016.7532461`.

[42] X. Xin, J. Wang, R. Xie, S. Zhou, W. Huang, N. Zheng, Semi-supervised person re-identification using multi-view clustering, Pattern Recognition 88 (2019) 285–297. `doi:https://doi.org/10.1016/j.patcog.2018.11.025`. URL `https://www.sciencedirect.com/science/article/pii/S0031320318304126`

[43] J. Meng, S. Wu, W.-S. Zheng, Weakly supervised person re-identification, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 760–769. `doi:10.1109/CVPR.2019.00085`.

[44] G. Wang, G. Wang, X. Zhang, J. Lai, Z. Yu, L. Lin, Weakly supervised person re-id: Differentiable graphical learning and a new benchmark (2020). `arXiv:1904.03845`.

[45] Y.-J. Li, F.-E. Yang, Y.-C. Liu, Y.-Y. Yeh, X. Du, Y.-C. Frank Wang, Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018.

[46] Z. Zhong, L. Zheng, Z. Luo, S. Li, Y. Yang, Invariance matters: Exemplar memory for domain adaptive person re-identification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[47] H.-X. Yu, W.-S. Zheng, A. Wu, X. Guo, S. Gong, J.-H. Lai, Unsupervised person re-identification by soft multilabel learning, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[48] M. Li, X. Zhu, S. Gong, Unsupervised person re-identification by deep learning tracklet association, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 737–753.

[49] M. Li, X. Zhu, S. Gong, Unsupervised tracklet person re-identification, IEEE Transactions on Pattern Analysis and Machine Intelligence.

[50] D. Li, D. Li, Z. Zhang, L. Wang, T. Tan, Unsupervised cross-domain person re-identification: A new framework, 2019, pp. 1222–1226. `doi:10.1109/ICIP.2019.8804418`.

[51] H. Yu, A. Wu, W. Zheng, Unsupervised person re-identification by deep asymmetric metric embedding, IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (4) (2020) 956–973.

[52] Z. Zhong, L. Zheng, D. Cao, S. Li, Re-ranking person re-identification with k-reciprocal encoding, in: CVPR, 2017.

[53] W. Li, Y. Wu, M. Mukunoki, M. Minoh, Common-near-neighbor analysis for person re-identification, in: 2012 19th IEEE International Conference on Image Processing, 2012, pp. 1621–1624.

[54] V. Nguyen, T. D. Ngo, K. M. T. T. Nguyen, D. A. Duong, K. Nguyen, D. Le, Re-ranking for person re-identification, in: 2013 International Conference on Soft Computing and Pattern Recognition (SoCPaR), 2013, pp. 304–308.

[55] Q. Leng, R. Hu, C. Liang, Y. Wang, J. Chen, Person re-identification with content and context re-ranking, Multimedia Tools and Applications 74. `doi:10.1007/s11042-014-1949-7`.

[56] A. J. Ma, P. Li, Query based adaptive re-ranking for person re-identification, in: D. Cremers, I. Reid, H. Saito, M.-H. Yang (Eds.), Computer Vision – ACCV 2014, Springer International Publishing, Cham, 2015, pp. 397–412.

[57] J. García, N. Martinel, C. Micheloni, A. Gardel, Person re-identification ranking optimisation by discriminant context information analysis, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1305–1313.

[58] M. Ye, J. Chen, Q. Leng, C. Liang, Z. Wang, K. Sun, Coupled-view based ranking optimization for person re-identification, in: X. He, S. Luo, D. Tao, C. Xu, J. Yang, M. A. Hasan (Eds.), MultiMedia Modeling, Springer International Publishing, Cham, 2015, pp. 105–117.

[59] J. García, N. Martinel, A. Gardel, I. Bravo, G. L. Foresti, C. Micheloni, Discriminant context information analysis for post-ranking person re-identification, IEEE Transactions on Image Processing 26 (4) (2017) 1650–1665.

[60] R. Guo, C. Li, Y. Li, J. Lin, Density-adaptive kernel based re-ranking for person re-identification, in: 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 982–987.

[61] L. Piras, G. Giacinto, Information fusion in content based image retrieval: A comprehensive overview, Vol. 37, 2017, pp. 50 – 60.

[62] L. P. Valem, D. C. G. Pedronette, Unsupervised selective rank fusion for image retrieval tasks, Neurocomputing 377 (2020) 182 – 199.

[63] J. Wang, Z. Zhu, Image retrieval system based on multi-feature fusion and relevance feedback, in: 2010 International Conference on Machine Learning and Cybernetics, Vol. 4, 2010, pp. 2053–2058.

[64] M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen, R. Hu, Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing, IEEE Transactions on Multimedia 18 (12) (2016) 2553–2566.

[65] L. Zheng, S. Wang, L. Tian, Fei He, Z. Liu, Q. Tian, Query-adaptive late fusion for image search and person re-identification, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1741–1750.

[66] L. An, X. Chen, S. Yang, X. Li, Person re-identification by multi-hypergraph fusion, IEEE Transactions on Neural Networks and Learning Systems 28 (11) (2017) 2763–2774.

[67] Q. Zeng, H. Yu, Group affinity guided deep hypergraph model for person re-identification, Electronics Letters 55 (4) (2019) 186–188.

[68] D. C. G. Pedronette, L. P. Valem, J. Almeida, R. da S. Torres, Multimedia retrieval through unsupervised hypergraph-based manifold ranking, IEEE Transactions on Image Processing 28 (12) (2019) 5824–5838.

[69] L. P. Valem, D. C. G. Pedronette, J. Almeida, Unsupervised similarity learning through cartesian product of ranking references, Pattern Recognition Letters 114 (2018) 41 – 52.

[70] W. Webber, A. Moffat, J. Zobel, A similarity measure for indefinite rankings, ACM Transactions on Information Systems 28 (4) (2010) 20:1–20:38.

[71] A. Shtok, O. Kurland, D. Carmel, Query performance prediction using reference lists, ACM Trans. Inf. Syst. 34 (4).

[72] Y. Zhou, W. B. Croft, Ranking robustness: A novel framework to predict query performance, in: ACM Int. Conference on Information and Knowledge Management (CIKM'06), 2006, pp. 567–574.

[73] Y. Zhou, W. B. Croft, Query performance prediction in web search environments, in: ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'07), 2007, pp. 543–550.

[74] A. A. de Oliveira, E. Oakley, R. da Silva Torres, A. Rocha, Relevance prediction in similarity-search systems using extreme value theory, J. Vis. Commun. Image Represent. 60 (2019) 236–249.

[75] Y. Huang, Q. Liu, S. Zhang, D. N. Metaxas, Image retrieval via probabilistic hypergraph ranking, in: IEEE Conference on Conference on Computer Vision and Pattern Recognition (CVPR'10), 2010, pp. 3376–3383.

[76] B. Schölkopf, J. Platt, T. Hofmann, Learning with hypergraphs: Clustering, classification, and embedding, in: Advances in Neural Information Processing Systems (NIPS'07), 2007, pp. 1601–1608.

[77] L. Sun, S. Ji, J. Ye, Hypergraph spectral learning for multi-label classification, in: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08, 2008, pp. 668–676.

[78] A. Bretto, Hypergraph Theory: An Introduction, Springer Publish. Company, 2013.

[79] Z. Zheng, L. Zheng, Y. Yang, Unlabeled samples generated by gan improve the person re-identification baseline in vitro, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, p. 3754–3762.

[80] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25, Curran Associates, Inc., 2012, pp. 1097–1105.

[81] L. Wei, S. Zhang, W. Gao, Q. Tian, Person transfer gan to bridge domain gap for person re-identification, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 79–88.

[82] K. Zhou, T. Xiang, Torchreid: A library for deep learning person re-identification in pytorch, arXiv preprint arXiv:1910.10093.

[83] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, in: Proceedings of the 18th International Conference on Neural Information Processing Systems, NIPS'05, MIT Press, Cambridge, MA, USA, 2005, pp. 507–514.

[84] Z. Zhao, H. Liu, Spectral feature selection for supervised and unsupervised learning, in: Proceedings of the 24th International Conference on Machine Learning, ICML '07, ACM, New York, NY, USA, 2007, pp. 1151–1157.

[85] Z. Bai, Z. Wang, J. Wang, D. Hu, E. Ding, Unsupervised multi-source domain adaptation for person re-identification, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 12909–12918. doi: 10.1109/CVPR46437.2021.01272.

[86] H. Tang, Y. Zhao, H. Lu, Unsupervised person re-identification with iterative self-supervised domain adaptation, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019.

[87] Z. Zhong, L. Zheng, S. Li, Y. Yang, Generalizing a person retrieval model hetero- and homogeneously, in: The European Conference on Computer Vision (ECCV), 2018.

[88] C. Ren, B. Liang, Z. Lei, Domain adaptive person re-identification via camera style generation and label propagation, CoRR abs/1905.05382.
URL http://arxiv.org/abs/1905.05382

[89] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, J. Jiao, Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification, in: CVPR, 2018.

[90] Y. Huang, P. Peng, Y. Jin, J. Xing, C. Lang, S. Feng, Domain adaptive attention model for unsupervised cross-domain person re-identification, CoRR abs/1905.10529.
URL http://arxiv.org/abs/1905.10529

[91] H. Liu, J. Cheng, S. Wang, W. Wang, Attention: A big surprise for cross-domain person re-identification, CoRR abs/1905.12830.
URL http://arxiv.org/abs/1905.12830

[92] Q. Yang, H.-X. Yu, A. Wu, W.-S. Zheng, Patch-based discriminative feature learning for unsupervised person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.