Accepted version (pre-print) of paper published in IEEE International Conference on Image Processing (ICIP), 2022.
DOI: https://doi.org/10.1109/ICIP46576.2022.9898060

A NOVEL RANK CORRELATION MEASURE FOR MANIFOLD LEARNING ON IMAGE RETRIEVAL AND PERSON RE-ID

Lucas Pascotti Valem, Vinicius Atsushi Sato Kawai,
Vanessa Helena Pereira-Ferrero, Daniel Carlos Guimarães Pedronette
Department of Statistics, Applied Mathematics and Computing (DEMAC)
São Paulo State University (UNESP), Rio Claro, Brazil

ABSTRACT

Effectively measuring similarity among data samples represented as points in high-dimensional spaces remains a major challenge in retrieval, machine learning, and computer vision. In these scenarios, unsupervised manifold learning techniques grounded on rank information have been demonstrated to be a promising solution. However, various methods rely on rank correlation measures, which often depend on a proper definition of neighborhood size. On current approaches, this definition may lead to a reduction in the final desired effectiveness. In this work, a novel rank correlation measure robust to such variations is proposed for manifold learning approaches. The proposed measure is suitable for diverse scenarios and is validated on a Manifold Learning Algorithm based on Correlation Graph (CG). The experimental evaluation considered 6 datasets on general image retrieval and person Re-ID, achieving results superior to most state-of-the-art methods.

Index Terms— manifold learning, image retrieval, person Re-ID, correlation graph, rank correlation measures

1. INTRODUCTION

The production of visual content and wide dissemination of image-based applications is directly associated with technological advances of the last decades and has led to severe impacts in many different areas. As a consequence, applications grounded on image retrieval techniques face increasing challenges, given the diversity in the huge growth of collections [1]. Despite the remarkable advances in supervised techniques boosted by deep learning strategies, the difficulty in obtaining large labeled sets required by such approaches remains a major obstacle. In this direction, unsupervised approaches based on transfer learning for similarity learning tasks have been established as a promising solution.

Transfer learning strategies are commonly exploited to take advantage of training procedures conducted on large datasets. Convolutional Neural Networks (CNN) [2, 3] and Vision Transformers (ViT) [4, 5] models have been widely exploited through transfer learning for obtaining feature representations based on its last layers. For retrieval tasks, these representations are traditionally compared by pairwise distance functions, i.e., the Euclidean distance. However, such deep-based representations often lie on manifolds in a high dimensional space [6], such that pairwise similarity measures are insufficient to reveal the intrinsic relationship between images. In this scenario, diffusion and rank-based manifold learning approaches have been proposed in order to perform unsupervised similarity learning tasks, by considering contextual information and the structure of the dataset to compute more effective similarity/dissimilarity measures [7].

Recently, rank-based approaches [7, 8] have achieved high-effective retrieval results. In fact, rank structures provide a rich source of contextual similarity information, once the most relevant information are organized at the top of ranked lists. The Ranked-List Similarities (RL-Sim) algorithm [1] exploits rank correlation measures based on the conjecture that, if two images are similar, their respective ranked lists are expected to be similar as well. In this research direction, rank correlation measures and the overlap between the neighborhood sets have been successfully exploited [7, 8] to compute more effective similarity measures in retrieval tasks.

In this scenario, the relevance of effectively quantifying the similarities between ranked lists is latent, once many manifold learning methods are based on such correlation measures. The Rank-Biased Overlap (RBO) [9] measure, based on a probabilistic user model, uses a key parameter that determines the weight for the top positions in the ranking and has been widely used. However, most of the measures are dependent on the depth of ranked lists considered or the size of k-neighborhood set.

In this paper, a novel rank correlation measure is proposed and validated on an unsupervised manifold learning algorithm for image retrieval. We propose a measure based on the Jaccard index, which is capable of identifying maximum similarity indications at different depths of ranked lists. Therefore, the proposed measure is more robust to the definition of the size of the neighborhood set, which is essential in unsupervised scenarios and allows to achieve more effective results. The measure is used on a manifold learning algorithm based on a Correlation Graph (CG) and Strongly Connected Components (SCC) [10].

A wide experimental evaluation was conducted to assess the effectiveness of the proposed approach. General image retrieval and person Re-ID datasets were considered. CNN and ViT models were considered through transfer learning on unsupervised scenarios. The results demonstrated that the proposed JacMax measure achieves superior results than RBO measure in all evaluated scenarios. The proposed approach was also evaluated on the fusion of features, achieving results comparable or superior to the state-of-the-art in most datasets.

The authors are grateful to the São Paulo Research Foundation - FAPESP (grants #2021/07993-1, #2020/11366-0, #2020/02183-9, #2018/15597-6), Petrobras (grant #2017/00285-6), and the Brazilian National Council for Scientific and Technological Development - CNPq (grants #309439/2020-5 and #422667/2021-8) for the financial support.

2. BACKGROUND AND PROBLEM DEFINITION

This section formally defines the rank model considered (Section 2.1) and related rank correlation measures (Section 2.2).

2.1. Rank Model and Unsupervised Manifold Learning

Let $\mathcal{C} = \{o_1, o_2, \ldots, o_n\}$ be a collection of multimedia objects (images). Let $\mathbf{x}_i$ denote a vector representation for the object o_i in a space $\mathbb{R}^d$. Based on the comparison between representations, a ranked list can be obtained. Let d: $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a distance function that computes the distance between two images according to their corresponding representations (i.e., Euclidean distance). Formally, the distance between two images o_i, o_j is defined by $d(\mathbf{x}_i, \mathbf{x}_j)$.

Given a query image o_q, a ranked list τ_q=$(o_1, o_2, \ldots, o_L)$ can be computed in response to the query, where L denotes the length of the list. The ranked list τ_q can be defined as a permutation of a set $\mathcal{C}_L$ which contains the L most similar images to image o_q in the collection $\mathcal{C}$. The $\tau_q(o_i)$ notation denotes the position (or rank) of image o_i in the ranked list τ_q. Formally, we can define a ranked list τ_q such that, if image o_i is ranked before image o_j in the ranked list of image o_q, that is, $\tau_q(o_i) < \tau_q(o_j)$, then $d(\mathbf{x}_q, \mathbf{x}_i) \leq d(\mathbf{x}_q, \mathbf{x}_j)$.

Taking every image in the collection as a query image o_q, a set of ranked lists $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_n\}$ can be obtained. The objective of unsupervised manifold learning is to exploit the rich contextual similarity information encoded in $\mathcal{T}$ to define a more effective similarity measure and, therefore, a more effective set $\mathcal{T}_m$. Formally, it can be seen as a function $f_m(\cdot)$, such that $\mathcal{T}_r = f_m(\mathcal{T})$. The fusion problem is also considered, where different sets of ranked lists $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ are taken as input to compute a more effective set $\mathcal{T}_f$.

2.2. Rank Correlation Measures

A rank correlation measure defines a quantitative measure for assessing the similarity of two ranked lists. Given the broad use of top-k ranking analysis in retrieval applications, how to effectively compare such information assumes a fundamental relevance in many scenarios. Based on the model discussed in the previous section, a rank correlation measure can be defined as a function $r : \mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}$. Once most measures consider the top positions of ranked lists, a set $\mathcal{N}(o_q)$ is used to denote the k-neighborhood set which contains the top-k elements of the ranked list τ_q.

• **Jaccard:** the Jaccard index is a traditional statistic measure that computes the correlation between two ranked lists based on the size of the intersection and union of neighborhood sets. The index is formally defined as:

$$Jaccard(\tau_i, \tau_j, k) = \frac{|\mathcal{N}(o_i, k) \cap \mathcal{N}(o_j, k)|}{|\mathcal{N}(o_i, k) \cup \mathcal{N}(o_j, k)|}. \quad (1)$$

• **RBO:** the Rank-Biased Overlap (RBO) [9] also considers the overlap between top-k lists. However, different from the Jaccard index, it considers increasing depths through a weight computed based on probabilities defined at each depth. The RBO measure is defined as:

$$RBO(\tau_i, \tau_j, k) = (1-p) \sum_{d=1}^{k} p^{d-1} \times \frac{|\mathcal{N}(o_i, k) \cap \mathcal{N}(o_j, k)|}{d}, \quad (2)$$

where the parameter p is a value in the interval $[0, 1]$ which denotes the probability of examining the current depth.

3. PROPOSED APPROACH

This section presents the proposed approach. Section 3.1 discusses the main ideas and relations with competitive measures. Section 3.2 formally defines the proposed measure. Section 3.3 discusses its applications to unsupervised manifold ranking.

3.1. Overview and Main Ideas

There is a myriad of rank correlation measures proposed in the literature [7]. The most effective results on unsupervised manifold learning for retrieval tasks have been achieved by measures that consider the size of intersection/overlap at k-neighborhood sets (i.e., Jaccard). However, a challenging task consists in defining a proper value of k. Small values can lead to cuts that are unrepresentative in certain scenarios. Larger sizes, in turn, may bring in information that is not relevant.

An alternative is given by weighted measures which assign higher weights to overlaps at top positions (i.e., Intersection, RBO) [7, 1], or multi-level analysis [11]. In fact, assigning weights to top positions is a relevant strategy and improves the robustness of neighborhood size definition, but faces other difficulties on how to define the weights.

In this work, we propose to solve this challenge by identifying the depth which presents the maximum Jaccard index until a depth k. The main conjecture behind this approach is that a high overlap between ranked lists, at any depth, should be considered a strong indication of similarity. If it occurs at top positions, these are the most confident positions. If it occurs to depths closer to k, it requires a greater overlap.

3.2. Formal Definition

This section presents a formal definition for the proposed rank correlation measure. The JacMax measure can be defined as:

$$JacMax(\tau_i, \tau_j, k) = \max_{1 \leq k_d \leq k} \frac{|\mathcal{N}(o_i, k_d) \cap \mathcal{N}(o_j, k_d)|}{|\mathcal{N}(o_i, k_d) \cup \mathcal{N}(o_j, k_d)|}. \quad (3)$$

3.3. Application on Manifold Learning

The proposed rank correlation measure is validated on an unsupervised manifold learning algorithm for image retrieval. The manifold learning algorithm [10] is based on a Correlation Graph (CG) and Strongly Connected Components (SCCs). The correlation measures are exploited to encode contextual similarity information in the graph, by assigning weight to edges. The main idea of the algorithm consists in distinguishing high-effective edges and expanding relationships through these edges.

While the analysis of the graph edges aims to identify reliable similarity relationships, the SCCs are used to expand and identify novel relationships across the graph. Similar images are expected to be assigned to the same SCCs. In this way, the algorithm is able to take into account intrinsic inter-class geometry and can be more effective at measuring distances between images.

4. EXPERIMENTAL EVALUATION

This section presents the experimental evaluation. While Section 4.1 describes the datasets, features, and parameters; Section 4.2 shows the quantitative and qualitative results.

4.1. Experimental Protocol

A wide experimental evaluation was conducted, considering 6 different public image datasets, with distinct characteristics and size ranging from 1,491 to 36,411 images. Four of them are used for general image retrieval and two of them for person re-identification (Re-ID):

• **Corel5k** [12]: includes diverse scene content such as fireworks, bark, microscopy images, tiles, trees, and others. It is composed of 50 categories with 100 images for each class;

• **Dogs** [13]: contains images of 120 breeds of dogs from around the world. It includes 12,000 training images and 8,580 test images, totaling 20,580 images;

• **Holidays** [14]: a set of personal holidays photos divided into 500 groups, totalling 1,491 images;

• **UkBench** [15]: composed of 2,550 objects or scenes. Each object/scene is captured 4 times from different viewpoints, distances, and illumination conditions;

• **Market1501** [16]: a popular Re-ID dataset composed of 32,217 images of 1,501 individuals, where 3,368 images are considered as query images;

• **DukeMTMC** [17]: a Re-ID dataset composed of 36,411 images of 1,812 people, where 2,228 images are considered as query images.

In total, 6 different feature extractors (descriptors) were used, encompassing CNNs (ResNet152 [2], ODLFP [18], OSNet-AIN [3]) and vision transformers (VIT-B16 [4], SWIN-TF [5], TransReID [19]). The descriptors used on general image retrieval datasets (ResNet152, VIT-B16, SWIN-TF, ODLFP) were trained on ImageNet dataset and the Re-ID ones (OSNet-AIN, TransReID) were trained on MSMT17 dataset. All of our results use transfer learning and are completely unsupervised.

Regarding the protocol for effectiveness evaluation, all images are considered as query images, except for Re-ID, where only query images specified by the dataset protocol were considered, as done by most of the authors in the literature. All Re-ID evaluations are of single query type [20]. The MAP (Mean Average Precision) is used for all datasets. The Recall@1 (R1) is included since it is commonly used for person Re-ID. While R1 and MAP range from 0 to 1, the N-S Score ranges from 1 to 4 and corresponds to the average of correct images present in the top-4 positions for the UKbench dataset.

Concerning the parameters, we used $L = 1000$ for all general image retrieval datasets and $L = 2000$ for Re-ID datasets. The neighborhood size is $k = 50$ for Corel5k and Dogs, $k = 4$ for datasets with very few images per class (Ukbench and Holidays), and $k = 20$ for Re-ID (Market and Duke). For single feature executions, we used the default parameters of Correlation Graph: $th_{start} = 0.35$, $th_{inc} = 0.01$, and $th_{end} = 1$. For rank-aggregation, we considered: $th_{start} = 0.05$, $th_{inc} = 0.001$, and $th_{end} = 1$.

4.2. Experimental Results

We conducted an experiment with the objective of comparing the Correlation Graph with our proposed measure in contrast to the RBO. Table 1 presents the results for all the datasets and descriptors. Notice that our proposed measure achieved the highest MAP value in all cases.

Table 1: Re-ranking results considering MAP (%).

Datasets	Descriptors	Original MAP	Correlation Graph	
			RBO	JacMax
Corel5k	**ResNet [2]**	64.50	85.93	**86.15**
	VIT-B16 [4]	75.02	88.39	**89.92**
	SWIN-TF [5]	73.92	94.11	**95.15**
Dogs	**ResNet [2]**	63.73	80.93	**82.81**
	VIT-B16 [4]	79.83	86.67	**87.48**
	SWIN-TF [5]	45.54	68.24	**69.26**
Holidays	**ResNet [2]**	74.88	71.98	**75.66**
	VIT-B16 [4]	82.40	79.71	**83.44**
	SWIN-TF [5]	85.52	82.42	**85.21**
	CNN-ODLFP [18]	88.46	86.24	**90.25**
Ukbench	**ResNet [2]**	94.54	95.31	**97.17**
	VIT-B16 [4]	93.28	94.25	**96.29**
	SWIN-TF [5]	97.93	98.25	**99.01**
	CNN-ODLFP [18]	97.74	97.81	**98.92**
Market	**OSNet-AIN [3]**	43.27	42.89	**57.39**
	TransReID [19]	43.52	55.13	**55.64**
Duke	**OSNet-AIN [3]**	52.66	45.82	**68.39**
	TransReID [19]	55.42	29.39	**70.77**

An experiment was conducted performing the rank-aggregation of the best descriptors using our proposed JaccardMax. The results are shown for different effectiveness measures (NS Score, R1, and MAP) in Table 2. Notice that combining features provided even higher results than the previous single descriptor experiment (Table 1).

We used the best results obtained by our approach for comparing with the state-of-the-art. Table 3 presents a comparison for Holidays dataset (MAP), where the value of 91.12% is among the best results. In Table 4 is shown the comparison for UKBench dataset (N-S Score). The proposed method surpasses all the results presented with 3.97 which is very close to 4 (maximum value).

Table 5 shows a comparison with recent baselines for person Re-ID considering MAP (%) and R-01 (%). The methods are divided into 3 categories: Unsupervised, Domain Adaptive, and Cross-Domain. Each result corresponds to the highest reported by the authors of the methods. For baselines that used transfer learning, the reader can consult the corresponding papers to check the datasets used for training. We highlight that none of the methods were trained using labels of the target dataset with the objective of keeping the protocol unsupervised. The results show that our approach obtained very significant results, superior to most of the baselines for the Market dataset. For Duke, the method achieved the best MAP and the second-best R1.

Table 2: Rank-aggregation results for different measures.

Dataset	Features	NS Score	R1 (%)	MAP (%)
Corel5k	Best Isolated Feature	—	—	75.02
	RESNET + VIT	—	—	94.96
	RESNET + SWIN-TF	—	—	95.86
	VIT + SWIN-TF	—	—	**96.32**
Dogs	Best Isolated Feature	—	—	79.83
	RESNET + SWIN-TF	—	—	81.18
	VIT + SWIN-TF	—	—	85.44
	RESNET + VIT	—	—	**88.24**
Holidays	Best Isolated Feature	—	—	88.46
	VIT + SWIN-TF	—	—	86.02
	CNN-OLDFP + SWIN-TF	—	—	90.31
	CNN-OLDFP + SWIN-TF + VIT	—	—	**91.12**
UKbench	Best Isolated Feature	3.85	—	97.93
	RESNET + SWIN-TF	3.94	—	99.05
	CNN-OLDFP + VIT	3.95	—	99.13
	CNN-OLDFP + SWIN-TF	**3.97**	—	**99.55**
Market	Best Isolated Feature	—	69.57	43.52
	OSNET-AIN + OSNET-IBN	—	73.25	59.84
	OSNET-AIN + OSNET-IBN + TReID	—	73.40	60.82
	OSNET-AIN + TReID	—	**75.42**	**63.53**
Duke	Best Isolated Feature	—	71.81	55.42
	OSNET-AIN + OSNET-IBN	—	76.21	0.69.27
	OSNET-AIN + OSNET-IBN + TReID	—	**78.77**	73.39
	OSNET-AIN + TReID	—	78.59	**73.96**

Table 3: State-of-the-art on Holidays dataset (MAP).

MAP for the state-of-the-art methods				
Sun *et al.* [21]	Zheng *et al.* [22]	Pedronette *et al.* [23]	Li *et al.* [24]	Liu *et al.* [25]
85.50%	85.80%	86.19%	89.20%	90.89%
Yu *et al.* [26]	Gordo *et al.* [27]	Valem *et al.* [28]	Berman *et al.* [29]	**Our Result**
91.40%	90.30%	90.51%	**91.80%**	**91.12%**

Table 4: State-of-the-art on UKBench dataset (N-S Score).

***N-S scores* for the state-of-the-art methods**				
Lv *et al.* [30]	Liu *et al.* [25]	Pedronette *et al.* [23]	Bai *et al.* [31]	Liu *et al.* [32]
3.91	3.92	3.93	3.93	3.93
Bai *et al.* [33]	Valem *et al.* [28]	Valem *et al.* [8]	Chen *et al.* [34]	**Our Result**
3.94	3.94	3.95	3.96	**3.97**

With the objective of visualizing the effectiveness of our approach, some ranked lists are presented where the query image is shown in green borders and the wrong results in red borders. For comparing RBO with our proposed measure, Figure 1 presents an example of a query. Different from our approach, notice that RBO included many wrong results among the top positions.

Similarly, we present a visualization for person Re-ID. Figure 2 presents three ranked lists of a same query obtained for fusion on Market dataset considering OSNET-AIN + TransReID. Notice that our approach removed the wrong images present in the isolated descriptors.

5. CONCLUSION

In this work, we proposed a new correlation measure, the JaccardMax. This metric was evaluated as part of the Correlation Graph (CG) method. A wide experimental evaluation was conducted on 6 datasets encompassing both general image retrieval and person Re-ID. Several descriptors were used covering both recent CNNs and vision transformers. The experiments revealed that the proposed measure is not only capable of obtaining results superior to the ones considering the default measure (RBO), but also comparable or superior to most baselines. Among future works, we intend to investigate the measure as part of other re-ranking and feature selection methods.

Table 5: Comparison with recent person Re-ID baselines.

Method	Year	Datasets			
		Market1501		DukeMTMC	
		R1	MAP	R1	MAP
Unsupervised Methods					
EANet [35]	2018	66.4	40.6	45.0	26.4
ECN [36]	2019	75.1	43.0	63.3	40.4
UTAL [37]	2019	69.2	46.2	62.3	44.6
CAP [38]	2021	**91.4**	**79.2**	**81.1**	67.3
Domain Adaptive Methods					
HHL [39]	2018	62.2	31.4	46.9	27.2
CSGLP [40]	2019	63.7	33.9	56.1	36.0
ECN++ [41]	2020	**84.1**	**63.8**	74.0	54.4
MMCL [42]	2020	**84.4**	60.4	72.4	51.4
Cross-Domain Methods (single-source* and multi-source)**					
*EANet [35]	2018	61.7	32.9	51.4	31.7
*AF3 [43]	2019	67.2	36.3	56.8	37.4
*AF3 [43]	2019	68.0	37.7	66.3	46.2
*PAUL [44]	2019	68.5	40.1	72.0	53.2
**EMTL [45]	2018	52.8	25.1	39.7	22.3
**Baseline by [20]	2019	**80.5**	56.8	67.4	46.9
Our Proposed Approach					
Our Result		**75.42**	**63.53**	**78.59**	**73.96**

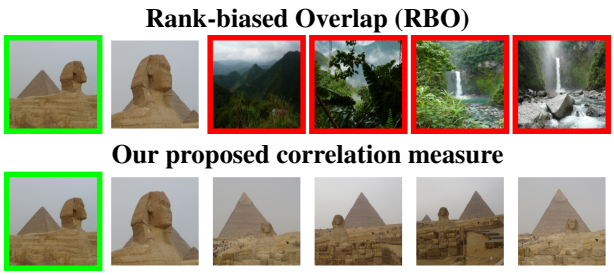

Fig. 1: Query on Holidays with results for RBO and JacMax.

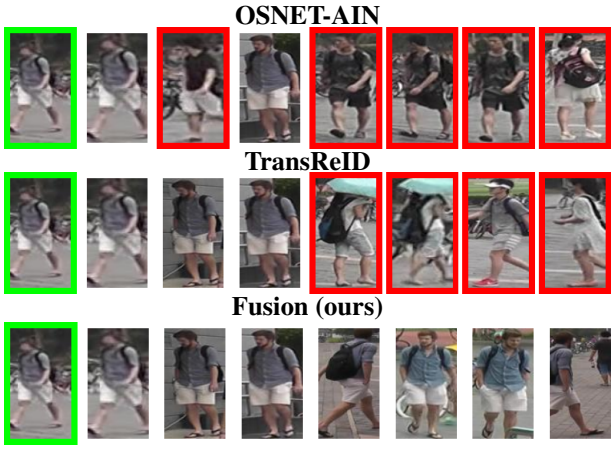

Fig. 2: Visual example of fusion result on Market dataset.

6. REFERENCES

[1] D. C. G. Pedronette and R. da S. Torres, "Image re-ranking and rank aggregation based on similarity of ranked lists," *Pattern Recognition*, vol. 46, no. 8, pp. 2350–2360, 2013.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, June 2016, pp. 770–778.

[3] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Learning generalisable omni-scale representations for person re-identification," *TPAMI*, pp. 1–1, 2021.

[4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *ICLR*, 2021.

[5] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," *ICCV*, 2021.

[6] A. Iscen, Y. Avrithis, G. Tolias, T. Furon, and O. Chum, "Fast spectral ranking for similarity search," in *CVPR*, 2018, pp. 7632–7641.

[7] L. P. Valem, C. R. D. Oliveira, D. C. G. Pedronette, and J. Almeida, "Unsupervised similarity learning through rank correlation and knn sets," *TOMM*, vol. 14, no. 4, pp. 1–23, 2018.

[8] L. P. Valem and D. C. G. a. Pedronette, "An unsupervised genetic algorithm framework for rank selection and fusion on image retrieval," in *ICMR*, New York, NY, USA, 2019, p. 58–62.

[9] W. Webber, A. Moffat, and J. Zobel, "A similarity measure for indefinite rankings," *TOIS*, vol. 28, no. 4, pp. 20:1–20:38, 2010.

[10] D. C. G. Pedronette and R. da S. Torres, "A correlation graph approach for unsupervised manifold learning in image retrieval tasks," *Neurocomputing*, vol. 208, no. Supplement C, pp. 66 – 79, 2016, sI: BridgingSemantic.

[11] N. G. De Sá, L. P. Valem, and D. C. G. Pedronette, "A multi-level rank correlation measure for image retrieval." in *VISIGRAPP (5: VISAPP)*, 2021, pp. 370–378.

[12] G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognition*, vol. 46, no. 1, pp. 188 – 198, 2013.

[13] A. Khosla, N. Jayadevaprakash, B. Yao, and L. Fei-Fei, "Novel dataset for fine-grained image categorization," in *Workshop on Fine-Grained Visual Categorization, CVPR*, June 2011.

[14] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *ECCV*, ser. ECCV '08, 2008, pp. 304–317.

[15] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *CVPR*, vol. 2, 2006, pp. 2161–2168.

[16] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *ICCV*, 2015, pp. 1116–1124.

[17] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *ICCV*, 2017, p. 3754–3762.

[18] K. Reddy Mopuri and R. Venkatesh Babu, "Object level deep feature pooling for compact image representation," in *CVPR*, June 2015.

[19] S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, "Transreid: Transformer-based object re-identification," in *ICCV*, October 2021, pp. 15 013–15 022.

[20] D. Kumar, P. Siva, P. Marchwica, and A. Wong, "Fairest of them all: Establishing a strong baseline for cross-domain person reid," *CoRR*, vol. abs/1907.12016, 2019.

[21] S. Sun, Y. Li, W. Zhou, Q. Tian, and H. Li, "Local residual similarity for image re-ranking," *Information Sciences*, vol. 417, no. Sup. C, pp. 143 – 153, 2017.

[22] L. Zheng, S. Wang, Z. Liu, and Q. Tian, "Packing and padding: Coupled multi-index for accurate image retrieval," in *CVPR*, June 2014, pp. 1947–1954.

[23] D. C. G. Pedronette, F. M. F. Gonçalves, and I. R. Guilherme, "Unsupervised manifold learning through reciprocal kNN graph and Connected Components for image retrieval tasks," *Pattern Recognition*, vol. 75, pp. 161 – 174, 2018.

[24] X. Li, M. Larson, and A. Hanjalic, "Pairwise geometric matching for large-scale object retrieval," in *CVPR*, June 2015, pp. 5153–5161.

[25] Z. Liu, S. Wang, L. Zheng, and Q. Tian, "Robust imagegraph: Rank-level feature fusion for image search," *TIP*, vol. 26, no. 7, pp. 3128–3141, 2017.

[26] W. Yu, K. Yang, H. Yao, X. Sun, and P. Xu, "Exploiting the complementary strengths of multi-layer cnn features for image retrieval," *Neurocomputing*, vol. 237, pp. 235–241, 2017.

[27] A. Gordo, J. Almazán, J. Revaud, and D. Larlus, "End-to-end learning of deep visual representations for image retrieval," *IJCV*, vol. 124, 2017.

[28] L. P. Valem and D. C. G. Pedronette, "Unsupervised selective rank fusion for image retrieval tasks," *Neurocomputing*, vol. 377, pp. 182 – 199, 2020.

[29] M. Berman, H. Jégou, V. Andrea, I. Kokkinos, and M. Douze, "MultiGrain: a unified image embedding for classes and instances," *arXiv e-prints*, Feb 2019.

[30] Y. Lv, W. Zhou, Q. Tian, S. Sun, and H. Li, "Retrieval oriented deep feature learning with complementary supervision mining," *TIP*, vol. 27, no. 10, pp. 4945–4957, 2018.

[31] S. Bai, Z. Zhou, J. Wang, X. Bai, L. J. Latecki, and Q. Tian, "Ensemble diffusion for retrieval," in *ICCV*, Oct 2017, pp. 774–783.

[32] G. Lao, S. Liu, C. Tan, Y. Wang, G. Li, L. Xu, L. Feng, and F. Wang, "Three degree binary graph and shortest edge clustering for re-ranking in multi-feature image retrieval," *JVCIR*, vol. 80, p. 103282, 2021.

[33] S. Bai, X. Bai, Q. Tian, and L. J. Latecki, "Regularized diffusion process for visual retrieval," in *AAAI*, 2017, pp. 3967–3973.

[34] X. Chen and Y. Li, "Deep feature learning with manifold embedding for robust image retrieval," *Algorithms*, vol. 13, no. 12, 2020.

[35] H. Huang, W. Yang, X. Chen, X. Zhao, K. Huang, J. Lin, G. Huang, and D. Du, "Eanet: Enhancing alignment for cross-domain person re-identification," *arXiv preprint arXiv:1812.11369*, 2018.

[36] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, "Invariance matters: Exemplar memory for domain adaptive person re-identification," in *CVPR*, 2019.

[37] M. Li, X. Zhu, and S. Gong, "Unsupervised tracklet person re-identification," *TPAMI*, 2019.

[38] M. Wang, B. Lai, J. Huang, X. Gong, and X.-S. Hua, "Camera-aware proxies for unsupervised person re-identification," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

[39] Z. Zhong, L. Zheng, S. Li, and Y. Yang, "Generalizing a person retrieval model hetero- and homogeneously," in *ECCV*, September 2018.

[40] C. Ren, B. Liang, and Z. Lei, "Domain adaptive person re-identification via camera style generation and label propagation," *CoRR*, vol. abs/1905.05382, 2019.

[41] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, "Learning to adapt invariance in memory for person re-identification," *TPAMI*, pp. 1–1, 2020.

[42] D. Wang and S. Zhang, "Unsupervised person re-identification via multi-label classification," in *CVPR*, 2020, pp. 10 978–10 987.

[43] H. Liu, J. Cheng, S. Wang, and W. Wang, "Attention: A big surprise for cross-domain person re-identification," *CoRR*, vol. abs/1905.12830, 2019.

[44] Q. Yang, H.-X. Yu, A. Wu, and W.-S. Zheng, "Patch-based discriminative feature learning for unsupervised person re-identification," in *CVPR*, June 2019.

[45] Y. Xian and H. Hu, "Enhanced multi-dataset transfer learning method for unsupervised person re-identification using co-training strategy," *IET Computer Vision*, vol. 12, no. 8, pp. 1219–1227, 2018.