Computer Science Review 53 (2024) 100657

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Review article

Unsupervised affinity learning based on manifold analysis for image retrieval: A survey

V.H. Pereira-Ferrero [a,*], T.G. Lewis [b], L.P. Valem [a], L.G.P. Ferrero [c], D.C.G. Pedronette [a], L.J. Latecki [d]

[a] Department of Statistics, Applied Mathematics and Computing, São Paulo State University (UNESP), Rio Claro, Brazil
[b] Center for Homeland Defense and Security, Naval Postgraduate School (NPS), Monterey, CA, United States
[c] School of Applied Sciences, University of Campinas (UNICAMP), Campinas, Brazil
[d] Computer & Information Sciences, Temple University (TU), Philadelphia, United States

ARTICLE INFO

Keywords:
Affinity learning
Manifold learning
Diffusion process
Ranking
Image retrieval
Multimedia retrieval
Unsupervised

ABSTRACT

Despite the advances in machine learning techniques, similarity assessment among multimedia data remains a challenging task of broad interest in computer science. Substantial progress has been achieved in acquiring meaningful data representations, but how to compare them, plays a pivotal role in machine learning and retrieval tasks. Traditional pairwise measures are widely used, yet unsupervised affinity learning approaches have emerged as a valuable solution for enhancing retrieval effectiveness. These methods leverage the dataset manifold to encode contextual information, refining initial similarity/dissimilarity measures through post-processing. In other words, measuring the similarity between data objects within the context of other data objects is often more effective. This survey provides a comprehensive discussion about unsupervised post-processing methods, addressing the historical development and proposing an organization of the area, with a specific emphasis on image retrieval. A systematic review was conducted contributing to a formal understanding of the field. Additionally, an experimental study is presented to evaluate the potential of such methods in improving retrieval results, focusing on recent features extracted from Convolutional Neural Networks (CNNs) and Transformer models, in 8 distinct datasets, and over 329.877 images analyzed. State-of-the-art comparison for Flowers, Corel5k, and ALOI datasets, the Rank Flow Embedding method outperformed all state-of-art approaches, achieving 99.65%, 96.79%, and 97.73%, respectively.

Contents

* Corresponding author.
E-mail address: nessahelena@gmail.com (V.H. Pereira-Ferrero).

https://doi.org/10.1016/j.cosrev.2024.100657
Received 26 February 2024; Received in revised form 10 July 2024; Accepted 18 July 2024

1. Introduction

The correlation between technological advancements and the diminished costs regarding capture and storage devices is inherently tied to the fast generation and pervasive dissemination of images. As a corollary, the field of image retrieval confronts mounting challenges, due to the huge expansion of image collections, whether in diversity or amount of images [1]. Strategies such as Content-Based Image Retrieval (CBIR) established itself as a crucial tool in this scenario, exploiting advances associated with fundamental issues of image understanding. The field aggregates researchers from different areas: computer vision, machine learning, information retrieval, human–computer interaction, database systems, data mining, information theory, and statistics [2,3].

CBIR organizes digital picture archives by visual content. It retrieves images from a collection based on a visual query, crucially relying on assessing similarity between images in high-dimensional spaces using metrics like Euclidean distance [3]. Nevertheless, approaches relying solely on Euclidean distance often struggle to capture non-linear and intricate similarity relationships. As a result, many machine-learning approaches have been employed to improve the effectiveness of image retrieval tasks. Considering image retrieval approaches, a novel Triplet-learning method with Opponent Class Adaptive Margin (OCAM) loss significantly enhances content-based image retrieval (CBIR) performance. OCAM outperforms existing methods, especially medical imaging applications [4]. To enhance video action recognition without extensive annotation, a vision-language multimodal embedding space can be used in a semisupervised learning framework [5]. Utilizing a siamese network, trained on both labeled and unlabeled data, another method employs a CNN-F architecture with hyperbolic tanh activation [6], using a robust hash code generation, tailored for extensive medical datasets. A Deep Collaborative Embedding model combines end-to-end learning with collaborative factor analysis, utilizing contextual information to refine tagging matrices and address out-of-sample problems [7]. It is also possible to mention a novel Deep Semantic Multimodal Hashing Network (DSMHN) designed for scalable image–text and video–text retrievals. It offers flexibility by supporting various loss functions with minimal modifications to the hash layer, demonstrating scalability [8].

The TRCaptionNet, a novel deep learning model, automatically generates accurate Turkish image captions using an image encoder, vision transformer-based feature projection module, and text decoder, leveraging both image and caption features [9]. Another study introduces MSViT, employing vision transformer-based image retrieval. It integrates global and local feature information using a two-branch transformer network and a multiscale feature fusion strategy, enhancing feature representation [10]. One weakly-supervised approach introduces a novel deep distance metric learning method for content-based image retrieval in social media, leveraging community-contributed images and user-provided tags to preserve semantic and visual structures [11]. The second approach introduces a novel hashing method for social image retrieval, utilizing user-provided tags to uncover semantic information and optimize the tagging matrix through binary matrix factorization [12].

Unsupervised learning techniques offer effective similarity measures without needing labeled data, leveraging contextual information and dataset structure [13].

Context-sensitive similarities enhance ranking and retrieval by capturing complex geometric characteristics within the dataset manifold [14]. The relevance of considering the dataset manifold was firstly evidenced for hand-crafted features, based on visual attributes like color, texture, and shape [15]. However, the positive potential of unsupervised manifold learning was also confirmed for different representations, such as mid-level representations (based on Bag of Visual Words) [16] and deep learning-based features trained by transfer learning. Such methods can achieve relevant effectiveness gains [17] for features based on Convolutional Neural Networks (CNNs) and models based on Transformers. Unsupervised manifold learning approaches in image retrieval offer global and local consistency analysis, contrasting with conventional methods relying on Euclidean distance for local structure [18].

During the last decades, a variety of approaches have been exploited for context-sensitive similarity learning [14–17,19–21]. From a broad perspective, most approaches can systematized into three main categories: diffusion process [14,15,22], rank-based [17,19], and deep learning approaches [21]. Although referenced in the literature under diversified taxonomies (distance/affinity/similarity learning [23–25], re-ranking [1,26,27], manifold learning [19]), distinct methods keep in common the objective of post-processing an initial similarity/dissimilarity measure to obtain a more global and effective measure. In general, the central role of unsupervised manifold structure acquisition and integration is underscored by the utilization of contextual similarity measures that go beyond pairwise comparisons. It is worth mentioning that these categories were defined to facilitate the organization of the survey, considering predominant strategies in the works, which, at the same time, are not limited to them.

In this work, we propose to survey the literature to present and organize the vast spectrum of approaches related to unsupervised similarity learning methods. To the best of our knowledge, this is the first survey to gather and organize research papers focused on unsupervised strategies based on manifold learning involving image retrieval tasks. Academic and scientific image retrieval is the task of finding relevant images from a database based on specific criteria or queries. It involves using computational methods to search, analyze, and retrieve images according to their visual features or associated textual metadata. While various surveys [3,28,29] have been dedicated to the more general theme related to CBIR, none of them focused on post-processing or context-aware similarity approaches. This work aims to bridge this gap, presenting a broad survey of the area, including both qualitative and quantitative perspectives. The state-of-the-art comparison also encompasses the Flowers, Corel5k, and ALOI datasets; Rank Flow Embedding method, for example, outperformed all state-of-art approaches, achieving the best results, 99.65%, 96.79%, and 97.73%, respectively [30]. This review of the field is original, and it is important to highlight its technical contribution. Therefore, the main contributions of this survey are three-fold:

- A qualitative analysis, which presents a general discussion about the evolution and categorizations of the area;
- A systematic review conducted on various relevant digital libraries, identifying a significant number of works;
- An experimental study focusing on recent features based on Convolutional Neural Networks (CNN) and Transformer-based models.

The remainder of this paper is organized as follows: Section 2 reviews the evolution of the area, history, timeline, concepts, and network analysis; Section 3 brings the systematic review and the adopted methodology, the main papers are organized in tables and discussed in specific subsections; Section 4 presents the experimental study; and Section 5 presents the final remarks and discussions.

2. Evolution and organization of the area

This section presents a comprehensive discussion of the field, with a specific focus on qualitative analysis and organizational perspectives. A brief retrospective is presented Section 2.1. Section 2.2 presents a general overview and organization of the area. A timeline and representative studies are discussed in 2.3. A visual perspective based on a keyword networked analysis is presented in Section 2.4.

2.1. Brief history and open challenges

Manifold learning remains a persistent theme in the realm of data representation and similarity measurement tasks. Nevertheless, arriving at a unanimous concept and definition for it proved to be a non-trivial endeavor. This section aims to discuss such concepts and definitions, the main advances achieved, and the open challenges in the area. In the context of image retrieval research, image collections are typically represented and manipulated using features extracted from the images. High-dimension representations often require changes in the dimensionality level due to both effectiveness and efficiency aspects, leading to the central purpose of dimensionality reduction. One of the formal definitions of manifold learning is also a recent popular approach to dimensionality reduction. Algorithms are built on the idea that many datasets' dimensionality can be described as a function of just a few underlying parameters. Manifold learning algorithms try to discover these parameters to find a low-dimensional representation of the data [31].

In this field, there are two most commonly used linear techniques for dimensionality reduction. PCA (Principal Components Analysis) and MDS (Multidimensional Scaling). Both introduced solutions based on matrices and their associating eigenvalues and eigenvectors. One of the most widespread types of dimensionality reduction was the PCA - Principal Component Analysis [32], however, limited to learning linear representations. Recent applications of Principal Component Analysis (PCA) have extended across diverse fields, including computer vision, contributing to interesting tasks. They were: facial and object recognition, image compression, and finding relevance in astronomy and bioinformatics. Conversely, Multidimensional Scaling (MDS), originating in the domain of psychology, has undergone a conceptual expansion. Contemporary implementations in bioinformatics have enabled endeavors such as the creation of a comprehensive portrayal of the protein structural landscape [33].

Nonetheless, over the years, the demand for non-linear and complex representations has grown. Two research published in Science Magazine used Locally Linear Embedding - LLE [34], in 2000, and ISOMAP (Isometric Feature Mapping) [35] in 2002, and have been considered major milestones in this theme. The first [34] highlights that most areas of science depend on exploratory data analysis and visualization. The fundamental problem of dimensionality reduction has been that it is challenging to analyze large amounts of multivariate data. Therefore, discovering compact representations of high-dimensional data is a growing and fundamental challenge. Locally Linear Embedding (LLE) is an unsupervised learning algorithm that computes low-dimensional, neighborhood-preserving embeddings of high-dimensional inputs. LLE maps input into a single global coordinate system of lower dimensionality and are therefore different from clustering methods. LLE can learn the global structure of nonlinear manifolds, such as those generated by images, for example. The second [35], published in 2002, is an enhancement of Tenenbaum's algorithm [36] which generalizes to arbitrary dimensionality if the connectivity and metric information of the manifold are correctly supplied. The main concern is improving dimensionality reduction algorithms' robustness to noise and developing innovative approaches due to constraints in reducing neighborhood size. This area is now crucial for future research exploration.

In 2003 [37], an interesting work highlighted a new universal ranking algorithm for data situated in Euclidean space, such as text or image data. The classification of the data concerning the structure of the intrinsic variety is collectively revealed by a large amount of data. In 2008, [38] the graph transduction method was one of the first to exclusively address the problem of image retrieval with a diffusion process in an unsupervised scenario. Since then, several approaches based on diffusion processes have been proposed [14].

Diffusion-based approaches were followed by methods based on ranking, more explored from 2011 onwards [39]. In 2013, the Ranked-List-Similarity (RL-Sim) algorithm was proposed, taking into account the rank correlation measures and the overlap between the neighborhood sets aiming at computing a more effective distance measure [40]. Also, the Rank-Biased Overlap (RBO) approach, based on a probabilistic user model, uses a key parameter that determines the strength weight for the top positions in the ranking [41].

More recently, various deep learning methods have been employed. These methods utilize an affinity graph to depict the overall structure of the data [22]. They also incorporate affinity diffusion across neighborhoods, which aids in identifying clusters of samples sharing similar semantics. This process contributes to the creation of a progressive model that incorporates an objective loss function aware of the group structure. Known as the multistage procedure [22], this approach guarantees that at each phase, the model focuses exclusively on dependable data groups that have been identified in the affinity graph up to that point.

Significant advances have been achieved by different approaches, including diffusion process, rank-based, and deep learning techniques. However, despite the significant effectiveness gains reached in image retrieval tasks, the area also includes some important open problems. A relevant challenge common to the different approaches consists in the capacity to deal with queries outside of the dataset, called unseen queries. Although some methods have also proposed approaches to handle the problem [17,42], it remains a challenge for most of the approaches. It is partially associated with efficiency and scalability aspects, which also constitute another important challenge in the area. While effectiveness is widely evaluated, efficiency is often neglected and considered by only a few works [42,43]. From another perspective, an open opportunity consists of the broad use of robust Graph Neural Network (GNN) models, few exploited [44,45], especially in some recent works [46].

2.2. Overview and organization

The task of image retrieval has proved to be a broad field of research with different levels of applications. Commonly, affinity values are analyzed and similarities between elements are evaluated. However, the structure of the underlying data manifold ends up not being considered, hence the need to obtain sensitive contextual similarities. Such similarities must explore the context, usually the geometry, of the underlying manifold. However, different approaches have been exploited to represent and analyze the similarity information encoded in the dataset manifold. Among the most representative works, it was possible to organize the strategies into 3 major groups: (a) Diffusion processes, (b) Ranking strategies; and (c) Deep approaches. It is imperative to underscore that the division of the three groups herein was meticulously undertaken with the intent of accentuating a prevailing strategic approach within each evaluated work. The objective is not to prescribe a conclusive or ultimate form of categorization but rather to provide readers with enhanced access to research pertinent to the highlighted strategy in each respective instance, without confining these works exclusively to the delineated categories.

Diffusion methods interpret the affinity matrix as a weighted graph, where nodes represent elements and connections reflect pairwise affinity values. Affinities are reassessed through graph similarity diffusion, often via random walks guided by a transition matrix. This process iteratively updates affinity matrices until convergence, improving retrieval performance continuously [14].

Considering the need for improvement strategies and computational costs, rank-based approaches have attracted attention. The use of ranking information may bring relevant advantages, as lowers computational efforts and independence of distance measures. In opposition to distance measures, which compare only pairs of images, ranked lists establish a deeper relationship, involving the comparison of the query image with all dataset images. The ranked lists constitute a rich source of similarity information, including the neighborhood set, which can be modeled in terms of top-k rank positions [1,47].

A recent strategic approach, termed the deep approach, uses contour features to represent views and infer 3D object characteristics. Zhu et al. [48] exemplify this approach, which leverages deep learning-based unified representations increasingly replacing traditional methods in retrieval tasks. Their contour-based representation has proven successful in 3D model-based and shape retrieval tasks, representing a recent advancement in the field.

It is possible to observe in Fig. 1 the cut of the literature that this survey proposes to delimit and comment on. Focusing on all the works that involved unsupervised learning and image retrieval, those that used well-known datasets were included in the survey; next, the papers reviewed chose some strategy for extracting features from the images, which were transformed into treatable data; then a distance function; and the works that included some type of context-aware similarity measures that brought final gains in the effectiveness of the retrieval task. We followed the organization in (a) Diffusion processes, (b) Ranking strategies, and (c) Deep approaches.

To exemplify each one, the works representing each strategy were cited; the graphs of (a) the Diffusion process proposed by Yang et al. (2009) [49], the data points can be embedded into Euclidean space by Diffusion Maps (DM), which can then reorganize the data points according to their geometric relation as revealed by the diffusion process. The first graph shows the colors of the points coded according to their second diffusion coordinate using Diffusion Maps, and the last graph shows the same plot using the Locally Constrained Diffusion Process (LCDP). The second strategy, (b) Ranking strategies, was exemplified using the work of Pedronette et al. (2018) [47], where the two graphs show the capacity of exploiting the geometry of the dataset manifold for computing new distances, where a query sample is selected in each represented moon, represented by a labeled point marked with a triangle. The color of other points is determined according to the closest labeled point. Firstly, the Two-Moons dataset considered the Euclidean distance, and once the geometry of the dataset is not considered, a large number of points are misclassified. Then, in the second graph, the classification was computed by the proposed manifold learning algorithm. The third strategy shows a qualitative analysis of the Oxford dataset done by Liu et al. (2019) [50]. Generalized-Mean (GeM) [51] and Guided Similarity Separation (GSS) descriptors are plotted using PCA followed by t-distributed stochastic neighbor embedding (t-SNE) [52] projection to two dimensions. In this case, it is presented three examples of queries with corresponding relevant database images colored red, green, and blue. Each query image was displayed with an Average Precision (AP) score; and a hard relevant database image.

2.3. Timeline and representative studies

The studies mentioned in the sections and which have stood out over the years can be summarized in an image that represents the timeline of the evolution of the area, in Fig. 2 The selected works are representative of the 3 main approaches they focus on: Diffusion process, Ranking, and Deep. Two research published in Science Magazine used Locally Linear Embedding - LLE [34], in 2000, and ISOMAP [35] in 2002, and have been considered the major milestones in this theme. In 2003, one of the seminal papers [37] was published, albeit with few experiments involving images. In 2008, the graph transduction method [38] was one of the first to exclusively address the problem of image recovery with a diffusion process in an unsupervised scenario. Then, several approaches based on diffusion processes were proposed: Locally Constrained Diffusion Process (LCDP) [49], Self-Smoothing Operator (SSO) [53], and Regularized Diffusion Process (RDP) [20]. Ranking-based methods were more explored from 2011 onwards [39]. Some representative examples are the Ranked List Similarities (RL-Sim) [1], the Sparse Contextual Activation (SCA) [54] and the Reciprocal kNN Graph and Connected Components [47] algorithms. During this decade, especially after 2011–2012, there was a development of strategies based on ranking on parallel fronts: at the same time that they gained greater notoriety in the literature, they were also investigated and developed more effective, robust, and efficient approaches. It is also worth mentioning interesting strategies based on Deep approaches. Mining on manifolds - Deep [21], a strategy that includes the attracts points that lie on the same manifold and repels different manifolds; a Deep Neural Network, which explores the capability of deep neural networks to generate explicitly better feature representation for image retrieval [55]; the Multi Domain 3D Shape, which proposed contour-based representation for successful 3D model-based shape retrieval [48]; and also, Guided Similarity Separation, an approach that uses the encoding neighbor information into image descriptors, generating cluster assignments and greatly optimization [50].

2.4. The big picture: the literature through keywords network analysis

This section aims to provide a broad perspective of the area using visual representations based on the main keywords of surveyed papers. The proposed visualization is grounded on formal network analysis and is described in the following.

After performing searches including specific queries for the theme detailed in this paper (details in Section 3), the *Web Of Science* and *Scopus* tools allow the export of research papers details as files. It was possible to identify and gather the keywords of all the research papers related to the occurrences. It was also possible to organize the papers by relevance, to obtain the first two thousand occurrences and their respective keywords. Through this identification, it was possible to elaborate on the construction of complex networks to analyze both the number of occurrences and the number of times that two words appear at the same time in a work.

For this purpose, we subdivided the years of publication according to the main keywords utilized in the papers. Also, the data was exported in the *VOSViewer* software to visualize each network for a period. This software allows visualization of the complex network working with different sizes of nodes (vertices) and connections (links). Before explaining each network model and each period analyzed, it is important to detail the concepts behind network model creation. Any network can be represented by a graph. Any graph can be represented by its adjacency matrix, from which other matrices such as Laplacian can be derived.

This kind of complex network analysis refers to the analysis of a mathematical graph. The measure of the degree of the nodes (parameters under analysis) of a complex network (graph) is related to the total number of edges (relations between the nodes) incident to this node. Nodes with a higher number of edges to its incidents are called hubs. Only the measure of nodes' degrees may not adequately reflect the complex importance of these node models.

A node is a measurable attribute, as shown in Fig. 7. The structure portion of a network is easily modeled by graph theory. Specifically, the network itself can be defined in terms of a set, $\mathbf{G} = \{\mathbf{N}, \mathbf{L}, \mathbf{f}\}$, where $\mathbf{N}$ is a set of nodes, $\mathbf{L}$ a set of links, and $\mathbf{f} : \mathbf{N} \times \mathbf{N}$ a mapping function that defines the structure of $\mathbf{G}$, how nodes are connected to each other through links. The mapping function contains enough information to draw the graph on a planar piece of paper using dots as nodes and lines as links. But the set $\mathbf{G}$ is inadequate to define the second part of a network, its dynamic behavior.

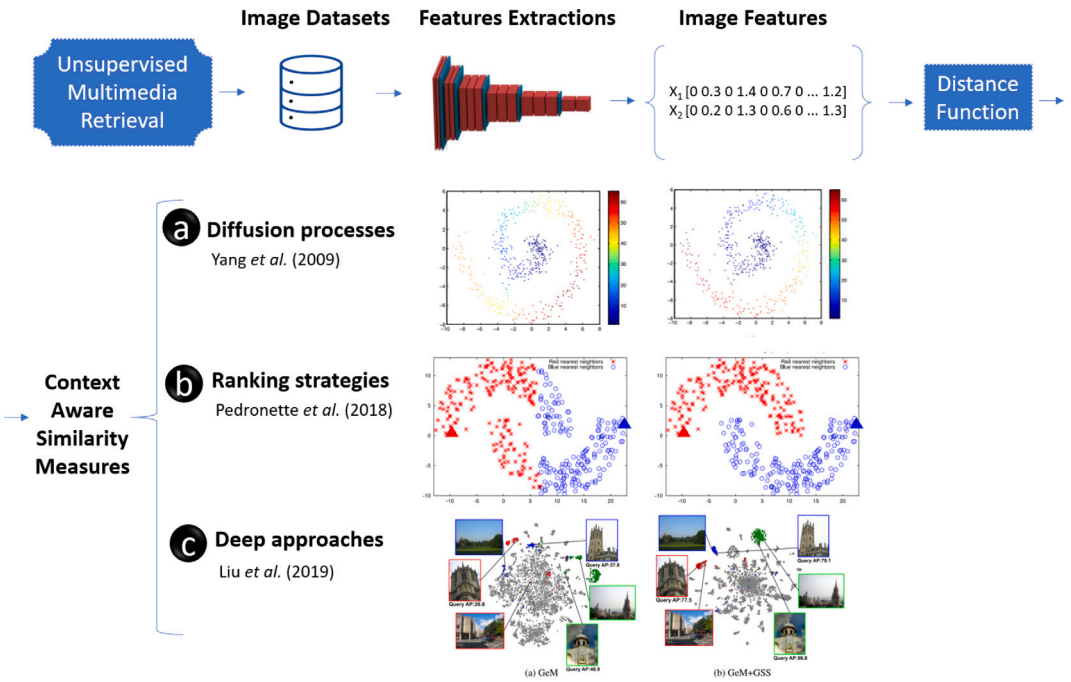

Fig. 1. Steps of selection of literature available research papers. Focusing on all the works that involved *Unsupervised learning* and image retrieval, research papers that used well-known *Image Datasets* were included in the survey; usually, each paper involved some strategy for *Feature Extractions* from the images, which were transformed into treatable data (*Image Features*), followed by an appropriated *Distance function*; interestingly, some type of *Context Aware Similarity Measures* were necessary to bring final accuracy gains in the retrieval task; it was possible to subdivide the strategies into 3 major groups: *(a) Diffusion processes*, *(b) Ranking strategies*; and *(c) Deep approaches*.

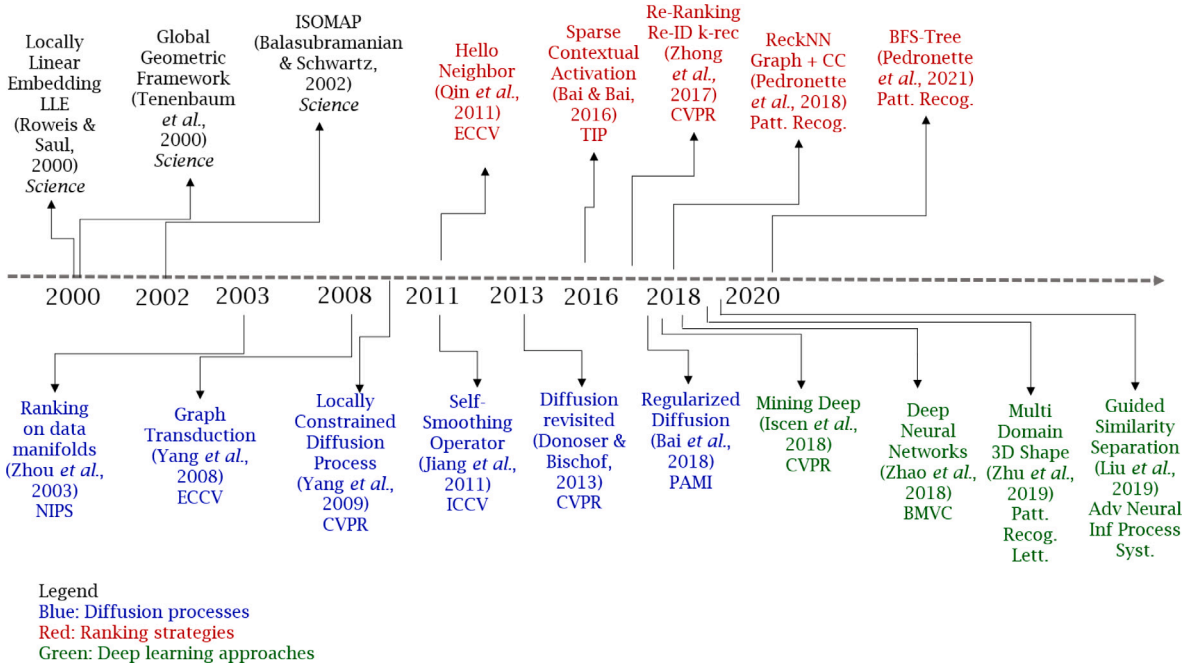

Fig. 2. The studies mentioned in the sections and which have stood out over the years can be summarized in an image that represents the timeline of the evolution of the area. The selected research papers are representative of the 3 main approaches they focus on The diffusion process, Ranking, and Deep.

Definition of a Complex Network

$$\mathbf{G(t)} = \{\mathbf{N(t)}, \mathbf{L(t)}, \mathbf{f(t)} : \mathbf{J(t)}\} \quad (1)$$

where,

t = time, simulated or real,

N = nodes, also known as vertices,

L = links, also known as edges,

f : **N** × **N** = mapping function that connects node pairs, yielding topology,

J = algorithm for describing behaviors of nodes and links versus time.

The elements of **N** are called nodes, the elements of **L** are called links, and the mapping function **f** is called the topology of **G**. The cardinality or size of **N**, denoted small **n**, is the number of nodes in **N**, and **m** is the number of links in **L**. Mathematically:

$\mathbf{G} = [\mathbf{N}, \mathbf{L}, \mathbf{f}]$ is a graph composed of three sets:

$\mathbf{N} = [\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}]$ are nodes; and $\mathbf{n} = |\mathbf{N}|$ is the number of nodes in $\mathbf{N}$,

$\mathbf{L} = [\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_m}]$ are links; and $\mathbf{m} = |\mathbf{L}|$ is the number of links in $\mathbf{L}$:

$\mathbf{f} : \mathbf{L} \rightarrow \mathbf{N} \times \mathbf{N}$ maps links onto node pairs:

Additionally, $\mathbf{v}$ and $\mathbf{e}$ are used to designate an element of $\mathbf{N}$ and $\mathbf{L}$, respectively, and enumerate them with subscripts: $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$ and $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_m}$.

A link is a mutual influence: node **A** is linked to node **B** if **A** influences **B**, and vice-versa, denoted $\mathbf{A} \leftrightarrow \mathbf{B}$. The correlation coefficient of link $\mathbf{A} \leftrightarrow \mathbf{B}$ is a measure of the influence of node **A** on node **B**. Correlations were normalized by dividing them by the maximum correlation value of overall links. Connection matrix **C**: matrix $\mathbf{N} \times \mathbf{N}$ of links connecting nodes: $\mathbf{C_{(i,j)}}$ = correlation result calculated between two measurements (nodes). **C** is symmetric when links are bidirectional, e.g. $(\mathbf{i}, \mathbf{j})$. Then $\mathbf{C_{(i,j)}} = \mathbf{C_{(j,i)}}$. If **C** is non-singular, its eigenvector $\mathbf{V} = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}\}$ where $\mathbf{v_i}$ are eigenvalues corresponding with nodes $\mathbf{n_i}$. Then the solution to $[\mathbf{C} - \mathbf{VI}] = \mathbf{0}$, where **I** is the identity matrix, yields the eigenvalues **V**. The degree of a node is the number of connecting links. The betweenness centrality of node **A** is the number of shortest paths passing through node **A** as determined by counting all the shortest paths from all nodes to all other nodes [56].

A complex network was constructed from keyword co-occurrences, with nodes sized by word frequency. Four separate networks were created to analyze different publication periods. Initially, all data from the search performed with the defined query are exported. In the VoSViewer software, the data type for creating the complex network is chosen. The delimited search between the years 2000 and 2005 was saved as a bibliographic file. Then, the creation of the network based on bibliographic data (Data Type) is chosen, since the idea is to analyze co-occurrences of keywords. In the Data Source step, the reading of data from reference files, RIS, EndNote, or RefWorks is selected, and the RIS file is then selected. The type of analysis is then predetermined, which is by "co-occurrences". Co-occurrences mean that relationships between items (nodes) are established based on the number of documents in which keywords (units of analysis) appear together (at the same time). The method of counting these words and establishing the link was defined as "Full counting", where the relationship between the words is a link with equal weight. Then the "threshold" is defined, where the minimum number of occurrences of a word is determined as equal to 2. In this network, of 191 words, 13 were above the threshold.

The spatial distance between nodes points to another important piece of information: the number of co-citations. The more distant the nodes (words) the fewer citations of works in common they have. The closer the nodes (words) are, the more citations of works in common they have. When larger gaps (spaces) are observed between groups of clusters, work opportunities that relate to them are observed. Clusters between 2 others can point to an interface (intermediation) between 2 large areas.

The first network is shown in Fig. 3 and represents the results of keywords found in articles published from 2000 to 2005. In total, there are 12 nodes with at least one connection and a maximum of 6 connections. It is observed that it is the least dense network (in terms of numbers of nodes and links) while showing that occurrences such as "CBIR" and "classification" (hub) gain relevance around 2003 and "diffusion" at the end of the interval (2005). For example, the classification node appears in 5 occurrences in this complex network. Furthermore, it was possible to calculate the link strength of this node, which is related to its degree (number of incident links) and is equal to 8. Therefore, in addition to being a hub node, the node classification is the most frequent keyword compared to the others in this range of searches.

The second network built is shown in Fig. 4 covers the years 2006 to 2011. In it, an increase in occurrences of keywords is observed, as well as an increase in the number of nodes and links, pointing to an increase in the number of scientific works and the use of related keywords. In total, there are 19 nodes with at least one connection and a maximum of 14 connections. "image retrieval", "relevance feedback" and "CBIR" are highlighted in the middle of the period and it is observed that from the middle to the end the use of "manifold learning" and "re-ranking" is more frequent since they appear matched above 20 co-occurrences in this network. Such a choice was necessary for better visualization of the nodes that stood out in the period.

The third network built is shown in Fig. 5 includes the years 2012 to 2017. In this network, an increase in occurrences of keywords is observed, as well as an increase in the number of nodes and links, pointing to an increase in the number of scientific works and the use of related keywords. In total, there are 28 nodes with at least 7 connections and import hubs (most connected nodes). They are "image retrieval", "classification" and "retrieval". Also, "manifold learning", "dimensionality reduction", "re-ranking" and "manifold ranking" are highlighted in the middle to the end of the period, pointing to a growing number of research papers involving such subjects.

The last network generated included the result of searches for articles in magazines and journals within the established query, with the filter from the year 2018 to 2022 (Fig. 6). The network analysis and model construction were generated with 15 co-occurrences and one new complex network. 39 nodes and 2 main clusters were obtained. The first year's cluster has keywords such as CBIR, re-ranking, and diffusion; the last few years are deep learning and deep metric learning. It is the densest network among all those built and compared, pointing to a growth in the amount of work in the area, in addition to the importance of themes related to images and new manipulation strategies.

Note that words like "information retrieval", "information storage" and "image retrieval" are hubs, that is, nodes (keywords) that appear more frequently considering the network as a whole, and specifically, each time interval. The network also enabled us to verify different clusters for each period: from 2005 to 2010 (predominantly blue), where diffusion process works predominated; from 2010 to 2015 (predominantly green), research with manifold ranking, re-ranking, and pattern recognition grown; and from 2015 onward (predominantly yellow), where deep strategies opened space and were associated with metric learning, deep neural networks, and retrieval methods. This interesting complex network analysis helps to understand the evolution of the research focuses in the area and its different related words, where the degrees of the nodes (degree) denote the frequency with which words and searches grow in a given period, also the strength of the link that clusters the most frequently associated words in searches. In this sense, network analysis is an abstraction of real facts. If abstraction can help in the explanation of the behavior of a real system, then network analysis is not only highly interesting but useful as well. It is interesting to note that there are specific micro rules in the complex network organization, one of them called *preferential attachment*, where links are attracted to nodes that already have a lot of links (in this keywords network analysis, this is related to pairs of keywords (nodes) commonly used that tended to be even more used in the papers). The fact that this network has the majority of nodes with the same average number of links and the specific number of nodes with many links (hubs) describes its topology structure as close to a Scale-free network [57], which follows a degree sequence distribution as a power-law. This characteristic is also called a macro level rule [56].

3. Systematic review

In this section, we take into consideration investigations from the literature involving systematic reviews and surveys. In general, the research involved keywords and strong search tools to find the papers associated with the topic. We also tried to prioritize literature selection from well-known established publishers.

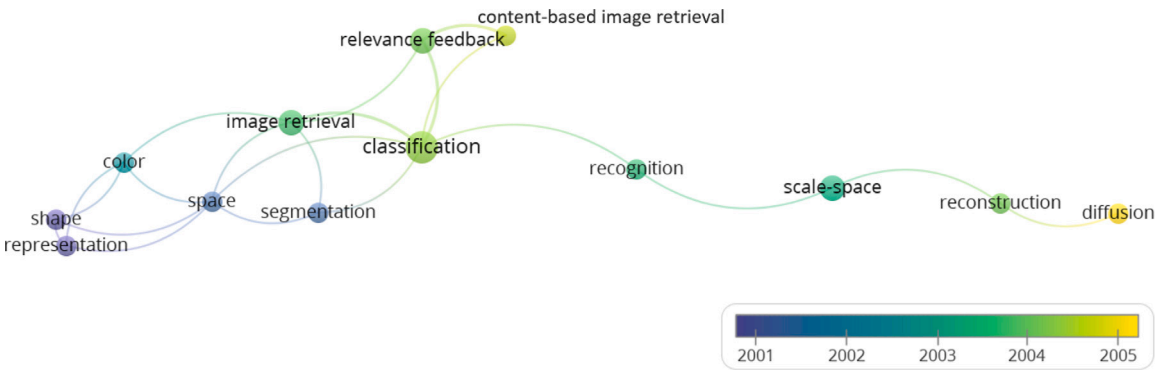

Fig. 3. A complex network built to analyze both the number of occurrences and the number of times that two words appear at the same time in a work, from the year 2000 until 2005.

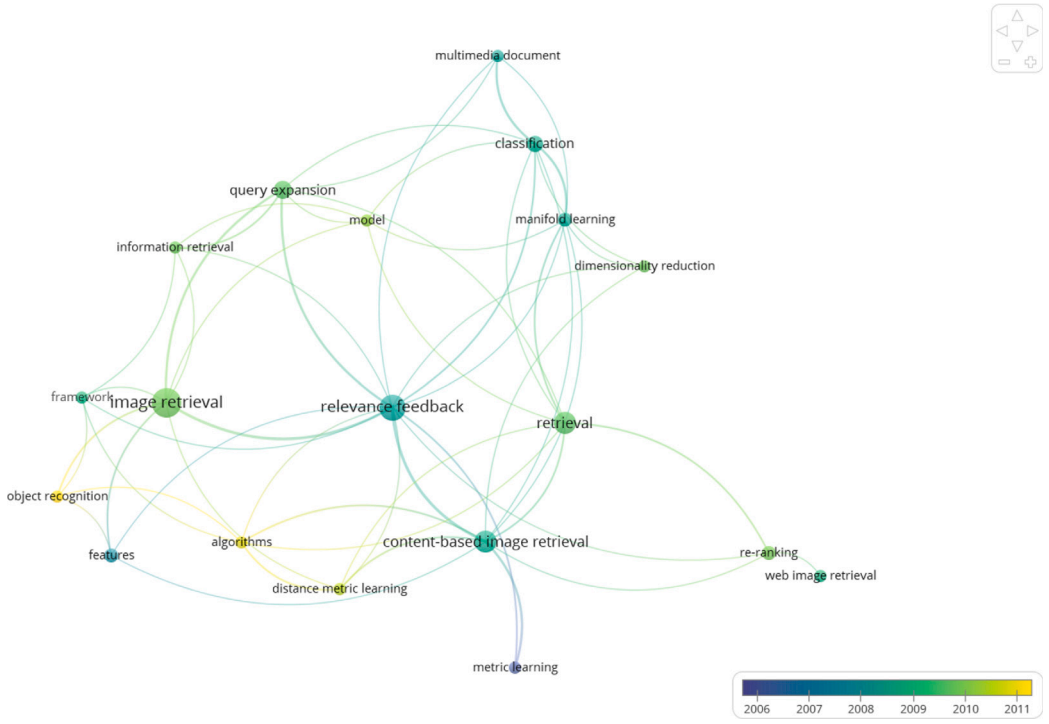

Fig. 4. A complex network built to analyze both the number of occurrences and the number of times that two words appear at the same time in a work, from the year 2006 until 2011.

3.1. Methodology

The central idea of this survey is to strategically gather the works available in the literature that involve research with unsupervised image recovery using manifold learning methods. It was necessary to study the main synonymous terms before defining the words that would compose the search. Considering that the target articles are predominantly from Computer and Exact Sciences, reputable search tools were chosen in available indexed databases: *Scopus* and the *Web of Science*. *Scopus* is a search engine from Elsevier's abstract and citation database launched in 2004. *Scopus* covers nearly 36,377 titles from approximately 11,678 publishers. *Web of Science* is a comprehensive platform that allows one to track ideas across disciplines and time from almost 1.9 billion cited references from over 171 million records, which provides access to multiple databases that provide comprehensive citation data for many academic disciplines, maintained by Clarivate Analytics. After defining the databases to be consulted, it was possible to define the words for the search (query) in both tools. The final query was defined as follows:

((“multimedia” or “image” or “shape” or “object”) and (“manifold learning” or “manifold” or “diffusion” or “re-ranking” or “reranking” or “re-rank” or “metric learning” or “affinity learning” or “distance learning” or “similarity learning” or “query expansion” or “contextual-sensitive similarity measures” or “contextual similarity” or “graph transduction” or “co-transduction”) and (“retrieval” or “ranking”)).

The first series of searches focused only on articles published in journals and magazines. In the *Web Of Science*, 1519 occurrences of works were obtained, and 1798 occurrences in the *Scopus* database. The search filtered “Articles” and “Review articles”, considering the occurrence of words in the “Topics” fields, which include: “Title”, “Abstract” and “Keywords”, both from the publisher and the tool. The occurrences included works from the year 1984 to 2022. The second series of searches focused only on articles published at international conferences. In the *Web Of Science*, 1122 occurrences of works were obtained, and 2391 occurrences in the *Scopus* database. The search filtered “Conference Papers” and “Conference Reviews”, considering the occurrence of words in the “Topics” fields, which include: “Title”, “Abstract” and “Keywords”, both from the publisher and the search

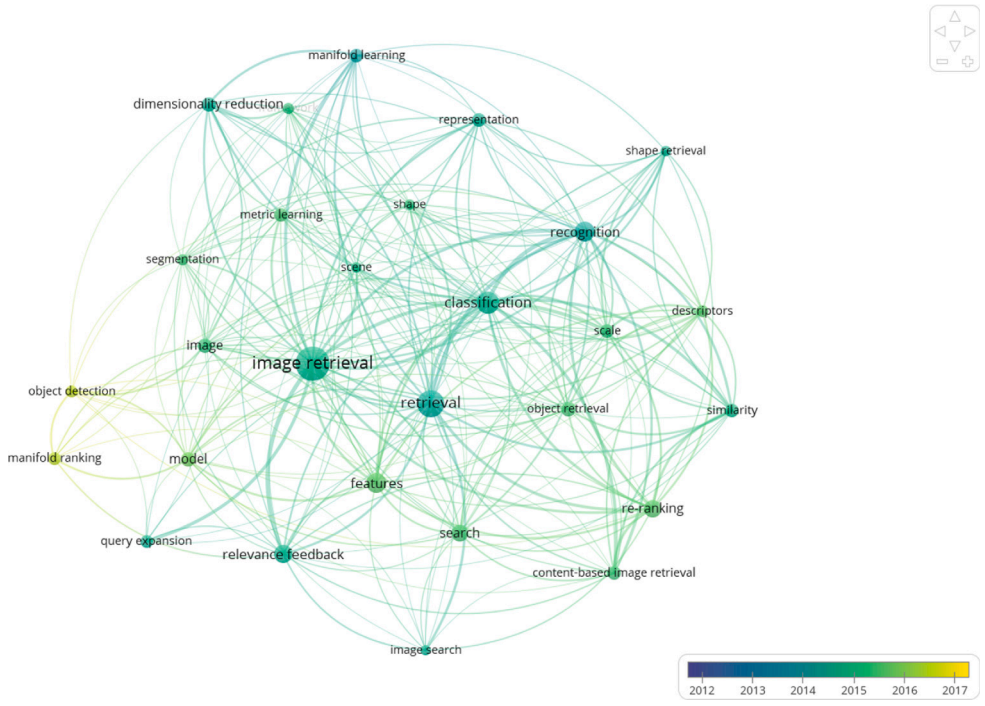

Fig. 5. A complex network built to analyze both the number of occurrences and the number of times that two words appear at the same time in a work, from the year 2012 until 2017.

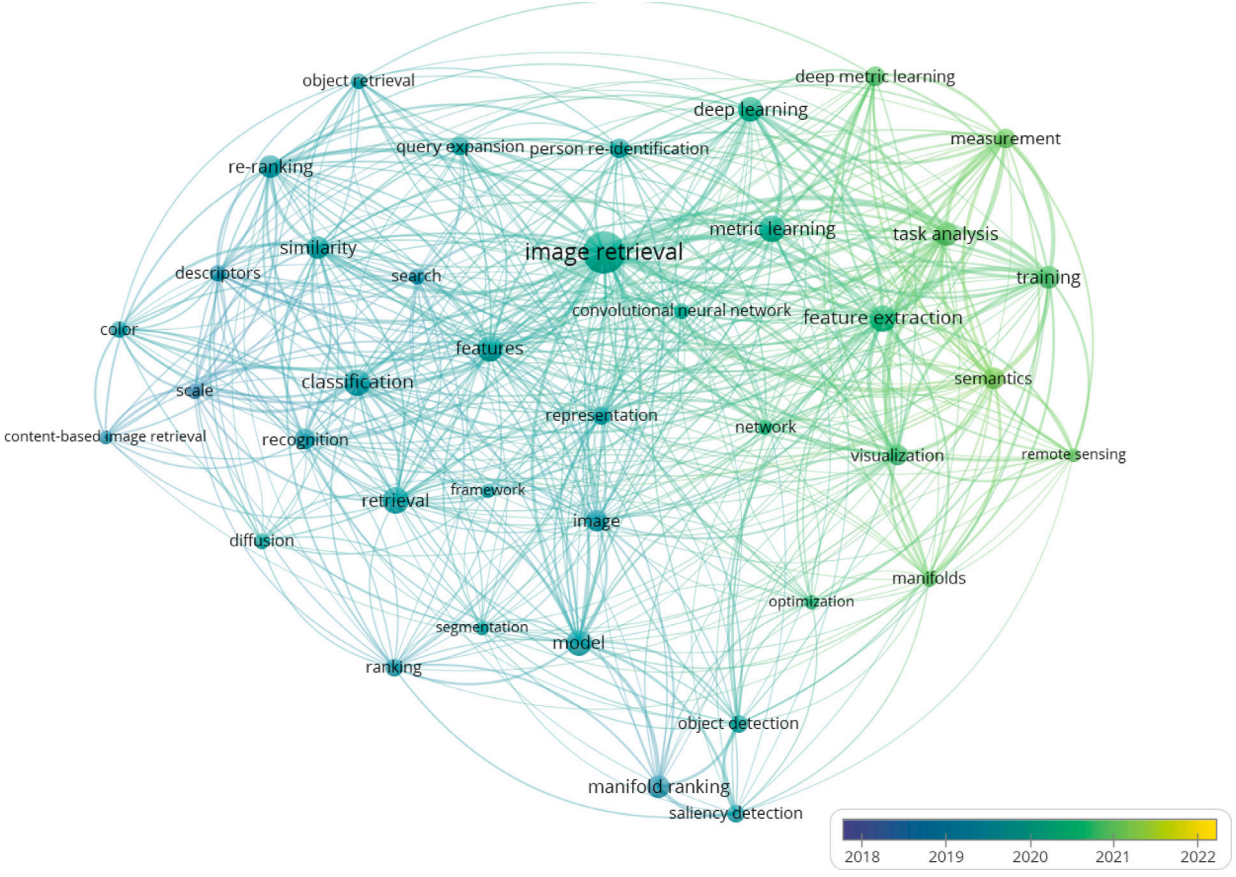

Fig. 6. The latest period complex network build to analyze both the number of co-occurrences, where two keywords appear at the same time in a work, from the year 2018 until 2022.

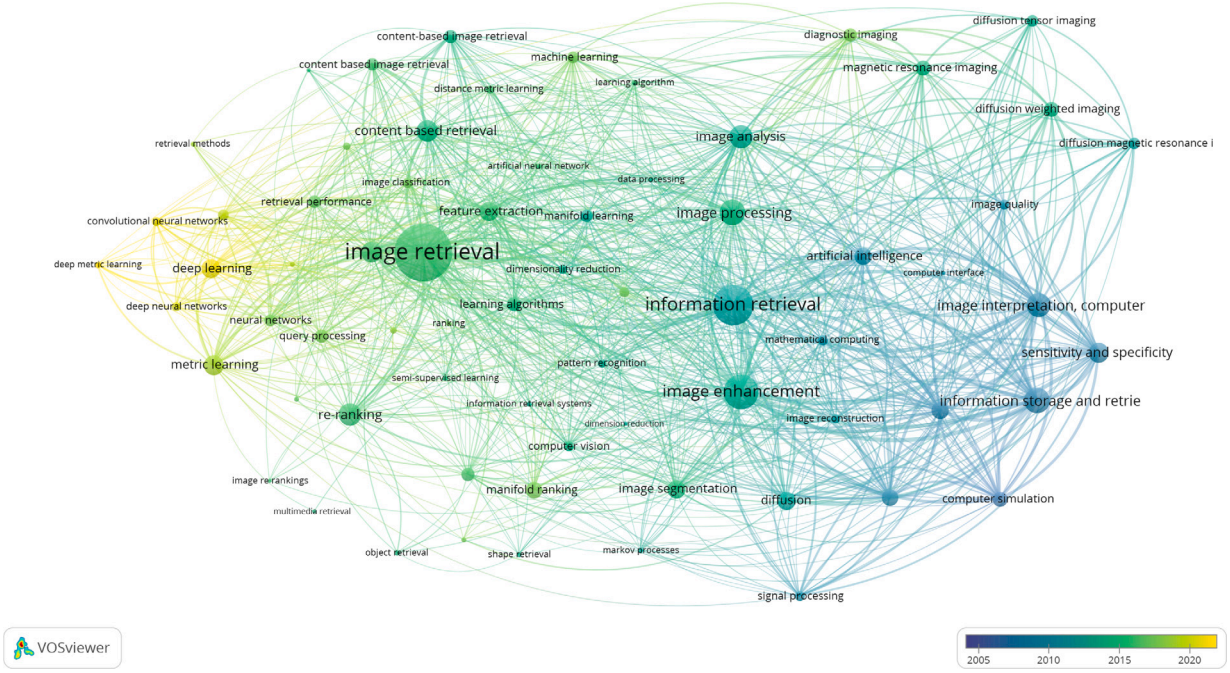

Fig. 7. A complex network built to analyze both the number of occurrences and the number of times that two words appear at the same time in a work.

tool. The occurrences included works from the year 1986 to 2022. Considering that there could be duplicate works when the lists were joined, it was necessary to use a methodological strategy that involves an algorithm in R language, called: "Bibliometrix: An R-tool for comprehensive science mapping analysis" [58]. The script reference is to an extremely well-cited article (over 1600 citations) often used in bibliometric analyses using the Bibliometrix package and an R script, the access to the exported occurrence files was adapted to remove duplicates in 4 results lists (2 from conferences, 2 from articles).

This step was extremely important to sequence the order of the lists of articles that would enter the Survey. Instead of ordering the lists according to their criteria, a renowned MCDA (Multiple Criteria Decision Analysis) method was considered for the classification of papers. The so-called "Methodi Ordinatio" is a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citations, and year of publication [59]. The method $InOrd$ involves the following equation:

$$InOrd = \alpha IF + \beta[10 - (Ry - Py)] + \gamma C_i, \quad (2)$$

where α is the weight to be given to the Impact Factor (IF), originally equal to 1. The IF is the Journal of Citation Reports (JCR) Impact Factor of the journal where the paper was published. β is the weight to be given to how much more recent the publication is, usually 1, varying from 1 to 10 (authors' suggestion). Ry is the Research Year and Py is the Publication Year. γ is the weight to be given to the number of citations in the paper, originally 1. Ci is the number of Citations in the paper.

After exporting the lists of articles and conferences in spreadsheets, another filter was applied to divide the articles. Since the Ordinatio Method gives more weight to the most cited articles, it would not be fair to keep recent articles (from 2019 onwards) on the same list as the others. In this way, 4 ordered lists were obtained. 2 lists of conference articles with different yearly intervals and 2 lists of articles from journals and journals with different yearly intervals. List 1: journal and journal articles published up to 2018. List 2: journal and journal articles published from 2019 onwards. List 3: conference papers published up to 2018. List 4: conference papers published from 2019 onward. In each list, the method ordinatio for sorting was applied. In each worksheet, titles and abstracts were read to define which ones would be downloaded and read. It was possible to select from List 1, 43 articles; List 2, 30 articles; List 3, 38 articles; List 4, 35 articles. Then, 146 full articles were downloaded.

It is possible to observe in Fig. 8 the search steps and strategies utilized to build this survey. It is a combination of well-established methodologies for surveys, meta-analysis, and review organization.

3.2. Prisma flow

This section includes Prisma diagrams that represent the visual summaries of the searches performed using the aforementioned methodological tools (Figs. 9 and 10). Two large databases *Web of Science* and *Scopus* (Identification) were accessed. Then, the works were gathered, and exported in files, to remove the duplicities via an algorithm in the software R (Screening). To apply the multi-criteria decision methodology, considering citations, we opted for the division of the occurrences of works in two-time intervals: until 2018 and from 2018 to 2022 (Eligibility). Then, exclusion and inclusion criteria were followed by access to the abstract, full texts, and quantitative and qualitative analyses.

3.3. Selected works

This work aimed to provide a concise and accessible synthesis of the primary research findings in the literature. Employing a rigorously established search methodology, we systematically organized these findings into tables. The tables are split by magazines and journal papers (Tables 1, 2, 3, 4), and conference papers (Tables 5, 6, 7). They highlight the predominant strategies used by authors. It is important to note that the methodological diversity across the included works transcends the categories used for organization. Nonetheless, this categorization serves as a useful tool to enhance accessibility, offering insights into the focal themes and prevalent strategies within each study.

3.4. Research categorizations and strategies

3.4.1. Diffusion

The works summarized in the tables according to the methodology of selection performed, can be gathered in this section according to their similarities in approaches.

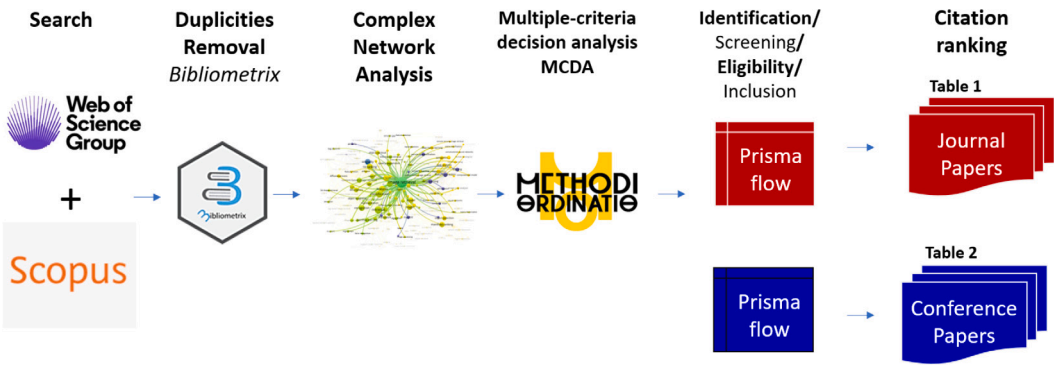

Fig. 8. The search sequence and strategies utilized to build this survey. It is a combination of well-established methodologies for surveys, meta-analysis, and review organization. Firstly, it is conducted a large search in computer science databases (*Web of Science* and *Scopus*); an algorithm for duplicity removal is then applied (*R-tool Bibliometrix*) [58]; a Complex Network analysis of the keywords according to time; an organization of the papers through a Multiple-criteria decision analysis, MCDA, Methodi Ordinatio [59]; 2 Prisma flows [60] for each type of publication, defining eligibility; and 2 final tables summarized and organized by a ranking of the most cited papers.

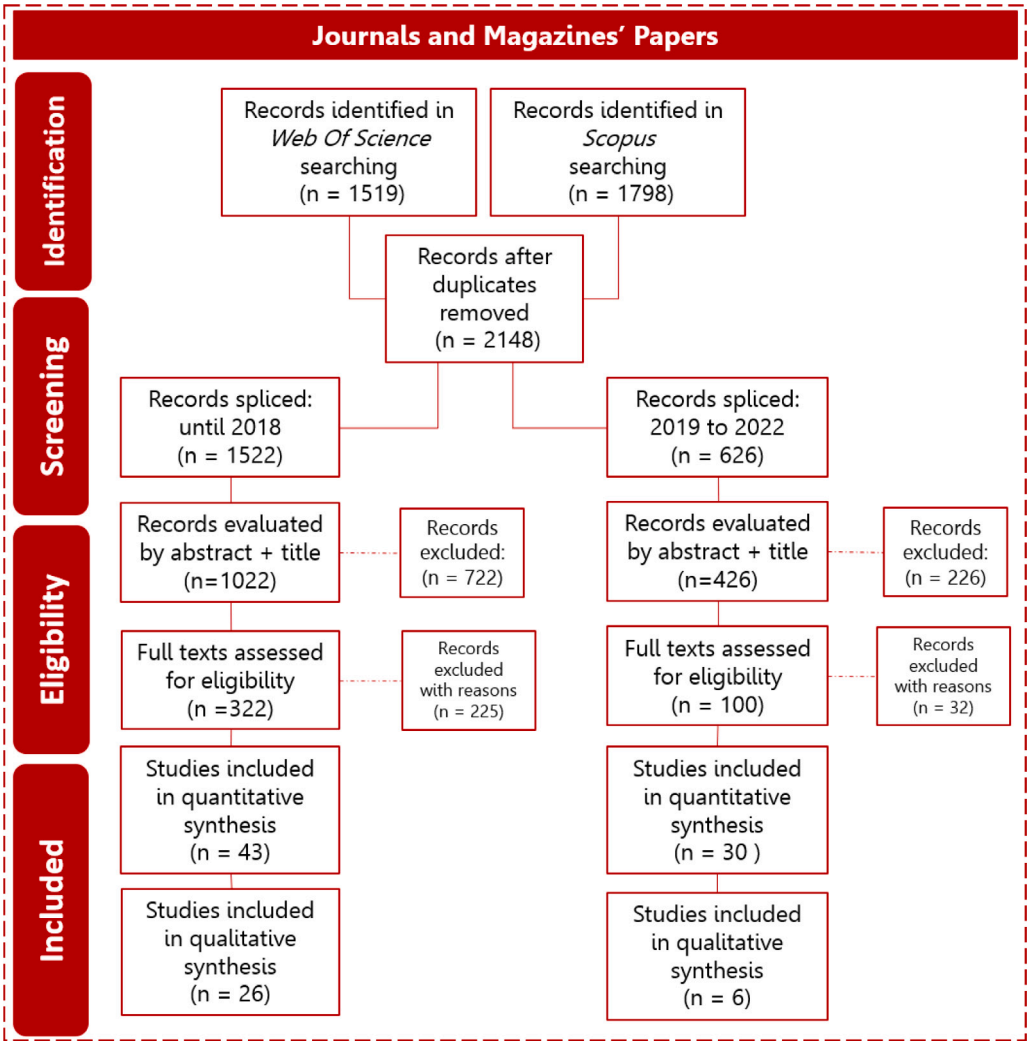

Fig. 9. Prisma flow that depicts the flow of information through the different phases of a systematic review. It was chosen the articles related to the topic, according to this well-known methodology for reviews [60].

Bai et al. [20] proposed a new affinity learning algorithm, the Regularized Diffusion Process (RDP) [20]. It performs affinity learning on tensor product hypergraphs, where hyperedges are utilized to capture the complex relationships. This way, the high-order information is brought by both the hypergraph and the tensor-order learning. The second is the use of contextual information to capture the geometry of the underlying manifold, with the time complexity of NSS being much lower than most diffusion process literature available. Bai et al. [100,114] also proposed Regularized Ensemble Diffusion (RED) related to the smoothness of graph-based manifolds, also reducing time complexity and outperforming other recent algorithms. In a similar approach, Iscen et al. [110] proposed a regional diffusion mechanism,

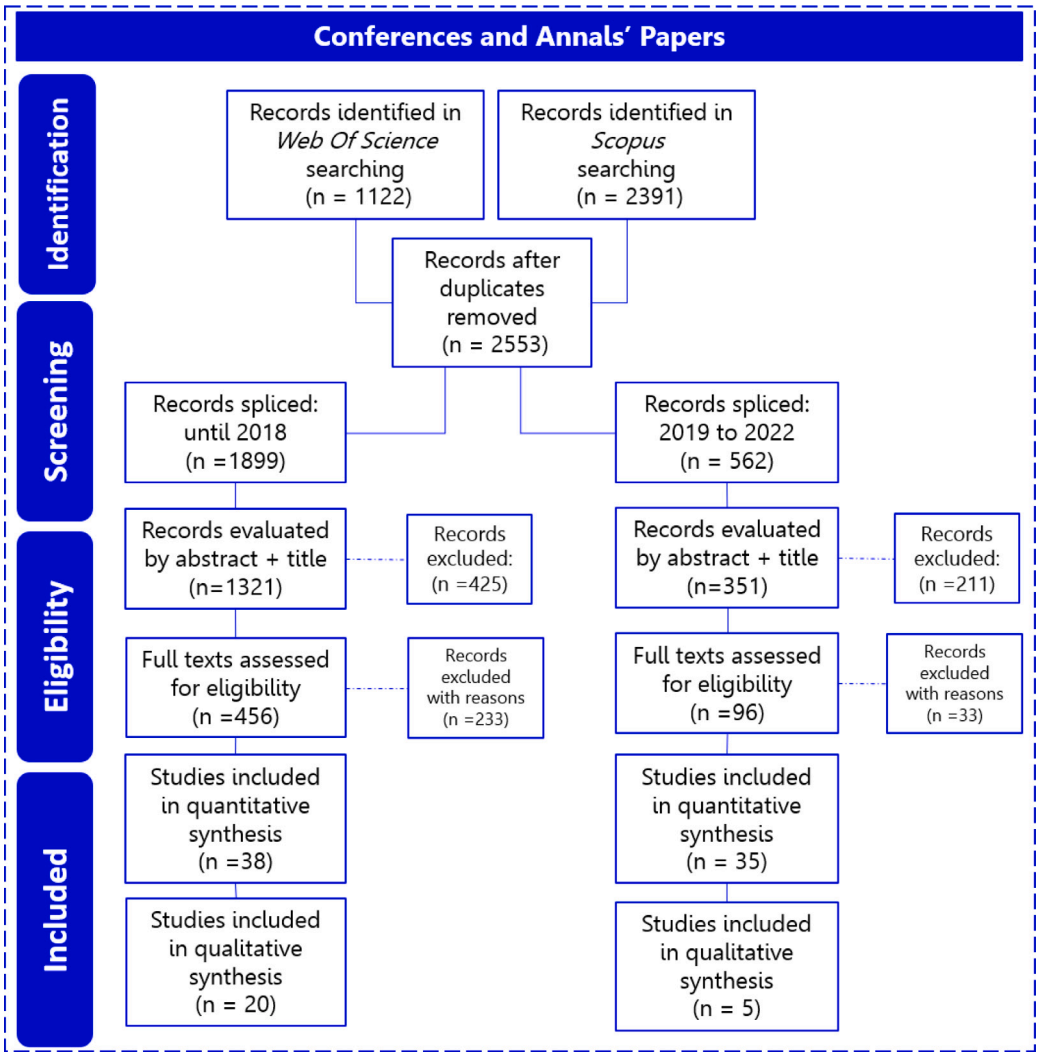

Fig. 10. Prisma flow that depicts the flow of information through the different phases of a systematic review. It was chosen the conference papers related to the topic, according to this well-known methodology for reviews [60].

which handles one or more query vectors at the same cost, performing diffusion through a sparse linear system solver, yielding practical query times well below one second. Jiang et al. [53] proposed a diffusion process that propagates similarity mass along the intrinsic manifold of data points, called Self-Smoothing Operator (SSO), directly improving a given similarity metric.

Wang et al. [105] proposed a fusion algorithm that combines multiple metrics through a diffusion process in an unsupervised way.

Bai et al. [15] studied similarity measures using graph transduction. The strategy includes new similarity measures learned, with significant improvements in retrieval results if compared to existing shape-matching methods, taking advantage of the manifold formed by the existing shapes. Diffusion/Ranking Using graph transduction, Bai et al. [75] propose a new shape retrieval algorithm, to fuse different similarity measures for robust shape retrieval through a semi-supervised learning framework.

Diffusion/Ranking Yang et al. [49] showed that other shapes influence the similarity measure of each pair of shapes, and the influence may be beneficial even in the unsupervised setting. A locally constrained diffusion process showed stability in noise presence and, it is possible to densify the shape space by adding synthetic points, called 'ghost points'.

Donoser & Bischof [14] revisited diffusion processes on affinity graphs for capturing the intrinsic manifold structure. Interestingly, automatically selecting a reasonable local neighborhood size is still an open issue.

Below, it is possible to mention other important works, not yet mentioned in the tables, but with relevant contributions to this field.

Yang et al. [117] investigated an innovative approach that adds synthetic points directly to distance spaces. To define the distances of ghost points to all other data points and insert ghost points to densify the data manifold using the context information to significantly improve the accuracy of retrievals. Through a diffusion process, and testing datasets like MPEG-7 [61], Nister and Stewenius (N-S) [65], and Caltech 101 [66], they find interesting results when compared with the state of art methods.

Pedronette & Torres [118] developed a novel hybrid method, named rank diffusion, which uses a diffusion process based on ranking information, which propagates contextual information through a diffusion process defined in terms of top-ranked objects, reducing the computational complexity of the proposed algorithm. A novel low-complexity method is proposed exploiting characteristics of both diffusion and rank-based approaches. High effectiveness gains can be obtained in several well-known datasets, like MPEG-7 [61], Soccer [90], Brodatz [91], ETH-80 [92], Holidays [67], UKBench [71], and at the same time, it is a low-complexity algorithm.

Probabilistic distribution and deep features were explored by Alemu & Pelillo (2020) [119]. To find a computationally efficient approach,

Table 1
Summary of Articles (part 1) published in journals and magazines. The chosen order is the most relevant number of citations received.

Journal	Year	Authors	Title	Main contribution	Strategy	Datasets
IEEE Trans. Pattern Anal. Mach. Intell	2009	Bai, Yang, Latecki, Liu & Tu [15]	Learning Context-Sensitive Shape Similarity by Graph Transduction	Similarity measures to the query shape in a graph structure are studied. A new similarity is learned through graph transduction, iteratively so that the neighbors of a given shape influence its final similarity to the query. Graph transduction and PageRank ranking yield significant improvements in both shape classification and shape clustering	Diffusion	MPEG-7 [61], Kimia's 99 [62], Face (all) [63], Swedish leaf [64]
IEEE Trans. Pattern Anal. Mach. Intell	2012	Yang, Prasad & Latecki [23]	Affinity Learning with Diffusion on Tensor Product Graph	Pairwise similarities and affinities are unreliable due to noise or intrinsic issues. Tensor product graph (TPG) by the tensor product of the original graph brings greater retrieval scores.	Diffusion	MPEG-7 [61], N-S [65], Caltech 101 [66], INRIA Holidays [67]
IEEE Trans. Image Process	2016	Bai & Bai [54]	Sparse Contextual Activation for Efficient Visual Re-Ranking	An extremely efficient algorithm for visual re-ranking with a feature vector called sparse contextual activation (SCA) that encodes the local distribution of an image, vector comparison under the generalized Jaccard metric. SCA improves retrieval performance in an unsupervised manner. Local Consistency Enhancement (LCE) is also being developed to improve the performance of SCA. The average time cost of re-ranking for a certain query can be controlled.	Ranking	PSB [68], WM-SHREC07 [69], YALE [70], MPEG-7 [61], UKBench [71]
Int. Journal of Computer Vision	2015	Tolias, Avrithis & Jégou [72]	Image search with selective match kernels: aggregation across single and multiple images	A match kernel that takes the best of existing techniques by combining an aggregation procedure with a selective match kernel. After performing a feature set augmentation, enjoy savings in memory requirements. A novel model to further incorporate matching kernels sharing the best properties of Hamming Embedding and Vector of Locally Aggregated Descriptors. The results include a significant increase in performance while enjoying a slight decrease in memory usage.	Ranking	Holidays [67], Ox. Buildings [73], Paris [74]

Table 2
Summary of Articles (part 2) published in journals and magazines. The chosen order is the most relevant number of citations received.

Journal	Year	Authors	Title	Main contribution	Strategy	Datasets
IEEE Trans. Image Process	2011	Bai, Wang, Yao, Liu & Tu [75]	Co-Transduction for Shape Retrieval	To improve the accuracy of adopted similarity measures, considering large intraclass variation. Fuse different similarity measures for robust shape retrieval through a semisupervised learning framework. Co-transduction and Tri-transduction algorithms to do a re-ranking for a novel shape retrieval framework.	Diffusion Ranking	MPEG-7 [61], N-S [65], Tari's shape [76], Wei's trademark [77]
IEEE Trans. Multim	2017	Bai, Bai, Zhou, Zhang, Tian & Latecki [78]	GIFT: Towards Scalable 3D Shape Retrieval	A real-time 3D shape search engine based on the projective images of 3D shapes brings: (1) efficient projection and extraction of preview features using GPU acceleration; (2) the first inverted file (FIF), is used to speed up the multiview matching procedure; and (3) the second inverted file, which captures a local distribution of 3D shapes in the resource collector, as an efficient context-based re-classification.	Ranking	ModelNet [79], SHREC14 LSGTB [80], ShapeNet Core55 [81], PSB [68], WM-SHREC07 [68], McGill [82]
IEEE Trans. Pattern Anal. Mach. Intell	2018	Bai, Bai, Tian & Latecki [20]	Regularized Diffusion Process on Bidirectional Context for Object Retrieval	Tensor product diffusion can be able to reveal the intrinsic relationship between objects. A new affinity learning algorithm is proposed, the Regularized Diffusion Process (RDP), which measures the smoothness of the manifold and simultaneously regularizes vertices in the affinity graph. The work is a generic tool for object retrieval, with the capacity of learning more faithful similarities.	Diffusion	MPEG-7 [61], YALE [70], ORL face [83], UKBench [71], Holidays [67], Oxford [73], TU Berlin Sketch [84], Wikipedia [85], PSB [86]
Pattern Recognition	2014	Chen, Li, Dick & Hill [87]	Ranking consistency for image matching and object retrieval	An image matching framework is proposed exploring ranking relationships. A list-wise min-hash scheme is developed, showing flexibility and efficacy.	Ranking	Oxford [73], Paris [74], Caltech [66], Flickr [88]

Table 3
Summary of Articles (part 3) published in journals and magazines. The chosen order is the most relevant number of citations received.

Journal	Year	Authors	Title	Main contribution	Strategy	Datasets
Inf. Sciences	2012	Pedronette & Torres [89]	Exploiting pairwise recommendation and clustering strategies for image re-ranking	A re-ranking method that considers relationships among images, the quality of ranked lists, and incorporates a clustering step for improving the final effectiveness.	Ranking	MPEG-7 [61], Soccer [90], Brodatz [91]
Pattern Recog.	2018	Pedronette, Gonçalves & Guilherme [47]	Unsupervised manifold learning through reciprocal kNN graph and Connected Components for image retrieval tasks	A novel manifold learning approach that exploits the intrinsic dataset geometry. The dataset manifold is modeled and analyzed in terms of a Reciprocal kNN Graph and its Connected Components. This method yields better effectiveness results than various methods recently proposed.	Ranking	MPEG-7 [61], Soccer [90], Brodatz [91], ETH-80 [92], Holidays [67], UKBench [71]
Image and Vision Comp.	2014	Pedronette, Penatti & Torres [27]	Unsupervised manifold learning using Reciprocal kNN Graphs in image re-ranking and rank aggregation tasks	A novel unsupervised manifold learning algorithm using Reciprocal kNN Graphs. With a subset of ranked lists as input, the computational and storage requirements are minimal. The re-ranking and rank aggregation algorithms yield better results in terms of effectiveness.	Ranking	MPEG-7 [61], Brodatz [91], UKBench [71]
Inf. Sciences	2015	Bai, Bai & Wang [93]	Beyond diffusion process: Neighbor set similarity for fast re-ranking	A simple yet effective method called Neighbor Set Similarity (NSS) is proposed, making use of contextual information to capture the geometry of the underlying manifold. A powerful fusion process to utilize the complementarity of different descriptors. The time complexity of NSS is much lower than the diffusion process; it is precise, faster, and proper for commercial purposes.	Ranking	MPEG-7 [61], N-S [65], ORL face [83]

Table 4
Summary of Articles (part 4) published in journals and magazines. The chosen order is the most relevant number of citations received.

Journal	Year	Authors	Title	Main contribution	Strategy	Datasets
Neuro-computing	2016	Pedronette & Torres [94]	A Correlation Graph Approach for Unsupervised Manifold Learning in Image Retrieval Tasks	An unsupervised manifold learning algorithm that takes into account the intrinsic dataset geometry can significantly improve the effectiveness of image retrieval systems with low computational efforts	Ranking	MPEG-7 [61], Soccer [90], Brodatz [91], ETH-80 [92], UW [95], UKBench [71]
Neuro-computing	2018	Zhu, Tang, Wang, Xu, Wang, Chen & Tian [96]	Saliency detection via affinity graph learning and weighted manifold ranking	A bottom-up saliency detection approach by affinity graph learning and weighted manifold ranking. An unsupervised learning approach is introduced to learn the affinity graph based on image data self-representation. An algorithm universally surpasses other unsupervised graph-based saliency detection methods.	Ranking	SOD [97], ECSSD [98], DUT-OMRON [99]
IEEE Trans. Image Process	2018	Bai, Zhou, Wang, Bai, Latecki & Tian [100]	Automatic Ensemble Diffusion for 3D Shape and Image Retrieval	Considering that many works are sensitive to noisy similarities, Regularized Ensemble Diffusion (RED) is proposed, with weights positively related to the smoothness of graph-based manifolds (tensor product). RED significantly reduces its proposed time complexity and outperforms algorithms focused on feature fusion or similarity diffusion, in addition to setting new performances.	Diffusion	ModelNet [79], Holidays [67], UKBench [71]

they introduced an incremental nearest neighbor (NN) selection method, considering the intrinsic manifold structure of a graph, the method shows its effectiveness in quantifying the discriminating power of given features in datasets like UKBench [71], INRIA Holidays [67], Oxford [73] and Paris [74].

The quadratic growth of the kNN graph size due to the high quantity of new connections between nodes in the graph is mentioned by Magliani et al. (2019) [120]. They propose Locality-Sensitive Hashing (LSH) projections, which obtain the same performance as a kNN graph after diffusion. The experiments involved Oxford datasets [73] with the advantage of less time, pointing to a better computational cost.

The geometry of data manifolds is an important aspect of diffusion processability. However, the selection of neighbors tends to be local [121]. Smooth Neighborhood (SN) is a proposal that mines the neighborhood structure to satisfy the manifold assumption, by imposing a weight learning paradigm. Through the MPEG7 dataset [61] and UKBench [71], the proposal achieved better performance when compared with state-of-art.

Another unsupervised deep learning approach is presented by Huang et al. (2020) [22] focused on deriving discriminative feature representations. Through a progressive affinity diffusion process, the experiments involved the datasets CIFAR [122], ImageNet [123], and MNIST [124] among others. The object image classification and clustering showed the performance superiority of the proposed approach.

Dou et al. (2020) [125] mention the advantages of diffusion processes and at the same time their limitations. They then proposed a novel method, Graph Diffusion Networks (GRAD-Net), which learns semantic representations by exploiting both local and global structures

Table 5
Summary of Conference Papers (part 1) published in most relevant conferences reviewed.

Journal	Year	Authors	Title	Main contribution	Strategy	Datasets
CVPR 2017	2017	Zhong, Zheng, Cao & Li [26]	Re-ranking person re-identification with k-reciprocal encoding	Considering Re-ID and the re-ranking under the Jaccard distance, a more robust k-reciprocal feature can capture similarity relationships from similar samples, producing effective improvements.	Ranking	Market1501 [101], CUHK03 [102], MARS [103], PRW [104]
CVPR 2011	2011	Qin, Gammeter, Bossard, Quack, & Van Gool [39]	Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors	An analysis of the k-reciprocal nearest neighbor is used and different parts of the ranked retrieval effectively re-ranking the retrieved images, demonstrating a significant improvement in comparative results.	Ranking	Oxford5k, Ox-ford105k [73], Paris [74], Un. of Kentucky [71], INRIA Holidays [67]
CVPR 2012	2012	Wang, Jiang, Wang, Zhou, & Tu [105]	Unsupervised metric fusion by cross diffusion	A fusion algorithm can output enhanced metrics by combining multiple similarity measures, through a diffusion process in an unsupervised way.	Diffusion	MPEG-7 [61], AT&T Face Image [83], Caltech 101 [66], N-S [65]
IEEE TPAMI	2008	Jegou, Schmid, Harzallah, & Verbeek [16]	Accurate image search using the contextual dissimilarity measure	A contextual dissimilarity measure can improve the accuracy in image search, considering the local distribution of the vectors and modifying the neighborhood structure. The approach showed better results than standard distance	Ranking	N-S & Lola [71]
IEEE CVPR	2012	Shen, Lin, Brandt, Avidan, & Wu [106]	Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking	A novel spatial constraint similarity measure is proposed, that considers object rotation, scaling, or viewpoint change. Object retrieval and localization significantly outperform other methods.	Ranking	Ox. Buildings [73], Paris [74], Un. of Kentucky [71], INRIA Holidays [67]

Table 6
Summary of Conference Papers (part 2) published in most relevant conferences reviewed.

Journal	Year	Authors	Title	Main contribution	Strategy	Datasets
IEEE Com. Soc. Conf. Com. Vis. Pat. Recog.	2009	Yang, Koknar-Tezel, & Latecki [49]	Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval	A locally constrained diffusion process that is more stable even if noise is present. The addition of ghost points to densify sparse data spaces demonstrates a significant increase in the retrieval rates.	Diffusion Ranking	MPEG-7 [61], Swedish Leaf [64]
IEEE CCVPR 2013	2013	Donoser & Bischof [14]	Diffusion processes for retrieval revisited	Considering retrieval applications, a generic framework for diffusion processes could be evaluated in different combinations of the transition matrix. Constraining the diffusion locally achieves the most promising boost in performance.	Diffusion	MPEG7 [61], Yale [70], ORL faces [83]
Comp. Vis. ECCV	2012	Zhang, Yang, Cour, Yu & Metaxas [107]	Query specific fusion for image retrieval	A graph-based query-specific fusion in ordered retrieval sets, by multiple retrieval methods, may enhance the retrieval precision. The retrieval quality is based on the consistency of the top candidates' nearest neighborhoods.	Ranking	UKbench [71], Corel-5K [108], Holidays [67] San Francisco Landmarks [109]
IEEE CVPR	2017	Iscen, Tolias, Avrithis, Furon, Chum & Rennes [110]	Efficient diffusion on region manifolds: Recovering small objects with compact CNN representations	Focused on diffusion, a mechanism that captures the image manifold in the feature space, brings a significant boost in performance of image retrieval with compact CNN descriptors on standard benchmarks.	Diffusion	Ox. Buildings [73], Paris [74], Instre [111]
Comp. Vis. ACCV	2009	Kontschieder, Donoser & Bischof [112]	Beyond pairwise shape similarity analysis	A modified mutual kNN graph as the underlying representation showed great results on shape retrieval tasks, and also an efficient unsupervised clustering method.	Ranking	MPEG-7 [61], KIMIA99 [62]

Table 7
Summary of Conference Papers (part 3) published in most relevant conferences reviewed.

Journal	Year	Authors	Title	Main contribution	Strategy	Datasets
IEEE CVPR	2018	Iscen, Tolias, Avithris & Chum [21]	Mining on manifolds: Metric learning without labels	A novel unsupervised learning method attracts points that lie on the same manifold and repels different manifolds, showing results on par or outperforming prior models.	Deep	Ox. Buildings [73], Paris [74], Instre [111], Holidays [67], CUB200 [113]
IEEE ICCV	2017	Bai, Zhou, Wang, Bai, Latecki, & Tian [114]	Ensemble diffusion for retrieval	Regularized Ensemble Diffusion (RED) proposed is bundled with an automatic weight learning paradigm, then negative impacts of noisy similarities are suppressed.	Diffusion	UKbench [71], Holidays [67], ModelNet40 [79]
IEEE Int. Conf. Comp. Vis.	2011	Jiang, Wang, & Tu [53]	Unsupervised metric learning by self-smoothing operator	The smoothing kernel is induced from an input similarity matrix, which will be directly improved through a smoothing/diffusion process along the data manifold. Its effectiveness has been demonstrated on tasks of image retrieval, clustering, segmentation, and classification	Diffusion	MPEG-7 [61]
IEEE CVPR	2018	Iscen, Avrithis, Tolias, Furon & Chum [115]	Fast spectral ranking for similarity search	The image retrieval as linear filtering over a graph is applied with fast spectral ranking and reproduces the excellent results of the online linear system solution.	Ranking	Ox. Buildings [73], Paris [74], Instre [111]
IEEE WCACV	2015	Yang, Matei aa & Davis [116]	Re-ranking by multi-feature fusion with diffusion for image retrieval	A re-ranking algorithm that fuses multi-feature information, pairwise similarity scores between images, and a diffusion process to the fused graph to reduce noise, which consistently improves the performance of baselines.	Ranking	UKbench [71], Oxford [73], Paris [74], Holidays [67]

of the image manifold in an unsupervised fashion. Using datasets like ORL face [83], Oxford [73], and Paris [74], they showed increased final performance.

3.4.2. Ranking

Following the structure established in the preceding section, the works delineated in the tables after the qualitative analysis can be systematically categorized within this section, elucidating commonalities in their methodological approaches.

Bai et al. [93] also used contextual information to capture the geometry of the underlying manifold, with the time complexity of NSS being much lower than most diffusion processes combined with re-ranking literature available.

Additionally, other work from Bai et al. [126] introduces a new algorithm called Smooth Neighborhood (SN) that mines the neighborhood structure to satisfy the manifold assumption. SN is adjusted to tackle multiple affinity graphs by imposing a weight learning paradigm; brings the theoretical guarantee of the underlying manifold structure and the capacity to deal with multiple affinity graphs. Testing different datasets, like MPEG-7 [61], Uk-Bench [71], PSB [86] and WM-SHREC07 [69], an integrated SN with Sparse Contextual Activation (SCA) showed to be a representative context-sensitive similarity, that can yield state-of-the-art performances on shape retrieval, image retrieval, and 3D model retrieval.

Also using graphs, Yang et al. [23] proposed to utilize the Tensor Product Graph (TPG) which takes into account higher-order information, for more reliable similarities. However, a graph diffusion process on TPG is equivalent to a novel iterative algorithm, and affinities are learned in an unsupervised setting.

Yang et al. [116] also joined graphs from multiple features with a mixture Markov model, with a probabilistic model of similarity scores, to determine the weight for each graph. Ranked lists of different queries receive different weights.

Still working with graphs, Zhang et al. [107] modeled retrieval ranks as graphs of candidate images and proposed a graph-based query-specific fusion approach, where multiple graphs are merged and re-ranked by conducting a link analysis on a fused graph.

Bai & Bai [54] developed the sparse contextual activation (SCA) as a features vector, which encodes the local distribution of an image. The vector comparison under the generalized Jaccard metric establishes the re-ranking step. Additionally, a local consistency enhancement, in an unsupervised manner, improved the retrieval. In a similar approach, Zhong et al. [26] a k-reciprocal encoding method to re-rank the re-ID. A k-reciprocal feature is calculated by encoding its k-reciprocal nearest neighbors into a single vector, used for re-ranking also under the Jaccard distance. Also, using the k-reciprocal nearest neighbor structure, Qin et al. [39] treat different parts of the ranked retrieval list with different distance measures, with benefits to both dimensionality problems and uneven distribution of images.

Jegou et al. [16] also considered the local distribution of the vectors and modified the neighborhood structure, with optimal parameter choice shown to be quite context-dependent. Shen et al. [106] proposed a new spatially-constrained similarity measure (SCSM) in addition to a novel and robust re-ranking method with the k-nearest neighbors of the query for automatically refining the initial search results.

Tolias et al. [72] proposed a vector aggregation method with selective kernel and vector binarization, optimizing classifier efficiency without increasing storage or query time.

Bai et al. [78] proposed a 3D shape search engine, which combines GPU acceleration and inverted file, as GIFT. In online processing, once a user submits a query shape, GIFT reacts and presents the retrieved shapes within one second (without the off-line preprocessing operations, such as CNN model training and inverted file establishment). While preserving high time efficiency, GIFT outperforms state-of-the-art methods.

Chen et al. [87] proposed a framework for exploring intrinsic ranking relationships for object retrieval tasks. Ranking consistency is an

image similarity criterion, used as a verification method to efficiently refine an existing ranking list. The results showed the flexibility and efficacy of the proposed image-matching framework.

Pedronette & Torres [89] also investigated a re-ranking method, which takes into account relationships among images and the quality of ranked lists. In this similar line, Pedronette et al. also have successfully investigated manifold learning exploring datasets intrinsic geometry [27,47,94], often with the manifold being modeled and analyzed in terms of a reciprocal kNN graph.

Besides, Kontschieder et al. [112] proposed a modified mutual kNN graph as the underlying representation and demonstrated its performance for the task of shape retrieval. Iscen et al. [115] proposed a fast spectral ranking for similarity search, using a linear graph filtering of a sparse signal in the frequency domain.

Below, it is possible to mention other important works, not yet mentioned in the tables, but with relevant contributions to this field.

Pedronette, Valem, Almeida & Torres [19] indicated that manifold learning methods can take into account the intrinsic global manifold structure. A novel algorithm could be proposed based on the hypergraphs for unsupervised multimedia retrieval tasks. Different datasets were tested, including MPEG-7 [61], Soccer [90], Brodatz [91], Holidays [67], UKBench [71], Corel5K [108], ALOI [127], MediaEval [128], and FCVID [129]. Interestingly, the LHRR algorithm exploits the capacity of hypergraphs for modeling high-order similarity relationships and achieves highly effective results in diverse multimedia retrieval scenarios.

Pedronette & Torres [130] introduced a novel unsupervised manifold learning algorithm based on the correlation graph and Strongly Connected Components (SCCs). The proposed algorithm computes a new distance that takes into account the intrinsic geometry of the dataset manifold. The tested datasets included MPEG-7 [61], Soccer [90], Brodatz [91], ETH-80 [92]. The results demonstrated the high effectiveness of the proposed method in several image retrieval tasks.

Pedronette, Torres & Calumby [131] used contextual spaces for image re-ranking and rank aggregation. Two novel re-ranking approaches that take into account contextual information were defined by the KNN. They include contextual spaces for encoding contextual information; two new re-ranking algorithms; and the evaluation of the proposed algorithms in several CBIR tasks; testing datasets like MPEG-7 [61], Brodatz [91], UW Dataset [95]. They used a combination of visual and textual descriptors and a post-processing (re-ranking) method with improved final results.

Iscen, Avrithis, Tolias, Furon, & Chum [132] proposed a new hybrid filtering method, based on temporal filtering and spectral–temporal graph, that allows for the first time to strike a reasonable balance between the two extremes of manifold ranking. It delivers great results in datasets like Oxford Buildings [73], and Paris [74], comparable with the state of the art, with the advantage of lower memory demands.

Local Residual Similarity (LRS) was proposed by Sun et al. [133] using the local neighborhood and the top-ranked images. The effectiveness of LRS was demonstrated on two benchmark datasets, UKBench [71], INRIA Holidays [67]. The authors showed that there could be a significant improvement in the final performance, particularly considering computational costs.

Still, considering image retrieval, Pang et al. [134] showed that a graph defined by a set of deep image features can constitute a heat transfer system. They proposed a practical solution to derive image vectors, in addition to a heat equation-based image re-ranking method. Such unsupervised-based methods showed to be compatible with different CNNs with interesting results for datasets like Oxford [73], Paris [74], Holidays [67] and Flickr [88].

The same research group [135] proposed another generalized strategy for image retrieval, using similarity propagation followed by a re-ranking of image vectors. Among the advantages, the strategy showed to be memory efficient, and did not require parameter tuning to achieve optimal performance; interesting experiments were conducted on the Oxford [73] and Paris [74] datasets.

Wang & Sun (2014) [136] focused on the problem of database retrieval and Graph Transduction for contextual information by the nearest neighbor graph. An optimal graph can be obtained and the model is unified by an objective function, and optimized by an iterative algorithm. By parameterizing the graph with combined weights, improved the contextual similarity learning method, using the interesting Open Access Series of Imaging Studies (OASIS) [137] in the experiments.

Pedronette, Valem & Torres (2021) [138] investigated the Breadth-First Search Tree (BFS) to exploit the similarity information encoded in the ranking references. In obtaining top-k ranking results, the BFS provided a hierarchical representation of the ranking results. The experiments involved interesting results in the datasets: MPEG-7 [61], Soccer [90], Brodatz [91], Holidays [67], UKBench [71], Corel5K [108], CIFAR 10 [122], and ALOI [127]. Significant effectiveness gains were obtained through the encoding of neighborhood relationships obtained by ranking references.

Arun et al. (2017) [139] unified rank aggregation and image re-ranking for more efficient retrieval results. Considering two-step clustering, an adaptive procedure updated the similarity scores among images, and the clusters played a key role in this process. Experiments with interesting results were conducted on the datasets: Holidays [67], Oxford [73], Corel [108], and Scene-15 [140].

Lao et al. (2021) [141] proposed a new approach in the selection of graph methods employed in the re-ranking process. The Three Degree Binary Graph (TDBG) was used to eliminate the outliers and a multi-feature fusion method was also proposed to enhance the retrieval performance for UKBench [71], and Corel [108] datasets, outperforming existing state-of-the-art manifold-based re-ranking methods.

Delviniotti et al. (2016) [142] present two re-ranking mechanisms for the improvement of image query results. The mechanisms seek to adjust the contents of the original query result, by measuring the degree to which the neighbor set of a result object agrees with that of the query object itself. The approach provides a simple and uniform framework for integrating structural information.

The problem of better feature representation is addressed by Shen et al. (2021) [143]. The meta-learn proposed works by re-ranking updates, and the similarity graph converges towards the target similarity graph induced by the image labels. They performed tests in datasets like Oxford [73], and Paris [74], and mentioned that the approach can work independently or in conjunction with classical re-ranking approaches for better image retrieval results.

3.4.3. Deep

Zhu, Rao, Bai, & Latecki [48] used contour features to represent views, once they could provide sufficient information to infer the characteristics of the whole 3D objects in a unified representation in retrieval tasks. Deep learning-based representations tend to gradually replace traditional learning or non-learning-based approaches. The proposed contour-based representation is successful in 3D model-based 3D shape retrieval tasking several datasets tested, such as SHREC'13 [144], SHREC'14 [145], SHREC'16 [146], ShapeNet Core55 [79], SHREC'16 and SHREC'17 Track [81].

Still considering deep learning approaches, Iscen et al. (2018) [21] presented a novel unsupervised framework for hard training. Initially, a set of images is defined with a significant initial representation (like a pre-trained CNN). The strategy included attraction points that lie on the same manifold and repel different manifolds and showed results on par with or outperforming prior models.

Zhao, Wang, Zhou, Shi, & Gao [55] proposed a modeling diffusion process by deep neural networks. Exploring a highly nonlinear diffusion process and the capability of deep neural networks, to generate explicit, better feature representation for image retrieval in complex datasets like Oxford5K and Oxford105K [73], Paris6k and Paris106K [74],

Instre [111], Sculpture [147]. Interestingly, the proposed approach achieved better retrieval results than the original diffusion process isolated.

Liu et al. (2019) [50] mention that the nearest neighbor graph in the exploration of information needed in image retrieval can have an alternative approach by directly encoding neighbor information into image descriptors. Their new approach is similar to the Deep Embedded Clustering (DEC) [148], which alternates between generating cluster assignments and optimizing the model. In addition, they used an unsupervised loss based on pairwise separation of image similarities and learned a new descriptor space that significantly improves retrieval accuracy. The datasets improved included Oxford [73], Paris [74], and Instre [111].

4. Experimental study

This section presents a brief experimental study on unsupervised affinity learning approaches in image retrieval tasks to provide an additional discussion on the topics presented in this survey. These approaches receive as input a set of ranked lists obtained for a given descriptor and post-process them to achieve more effective results. One of the objectives is to highlight the effectiveness gains obtainable by these methods with different datasets and descriptors.

A total of 8 datasets with diverse aspects were considered with sizes ranging from 5000 to 108,754 images and class sizes from 50 to 1812. The datasets were divided into two categories: *(i)* general-purpose, which includes a broad range of diverse categories, and *(ii)* person re-identification (Re-ID), which consists of images of people. For general-purpose datasets, it is common to have fewer classes with more elements, but in Re-ID (Re-identification), the opposite is true. To capture and analyze the diverse information contained in these images, both Convolutional Neural Networks (CNN) and Vision Transformer (ViT) models were used to extract features.

The methods used in this experimental analysis are all rank-based and enhance the ranked lists by redefining the similarity between elements, a process commonly known as similarity learning or re-ranking. We considered 5 methods, which are all unsupervised and briefly summarized as follows:

- *BFS-Tree of Ranking References (**BFSTREE**)* [138] uses a breadth-first tree structure to model and analyze the similarity and implicit relations between dataset elements based on rank correlations.
- *Cartesian Product of Ranking References (**CPRR**)* [24] performs Cartesian product of ranking references to learn similarities between data.
- *Log-based Hypergraph of Ranking References (**LHRR**)* [19] models the ranked lists as hypergraphs and exploits the relations between the elements in the dataset.
- *Rank Flow Embedding (**RFE**)* [17] employs a hypergraph to redefine the similarity between elements. From this hypergraph, it derives a graph and utilizes its connected components to identify groups of the most similar elements for re-ranking.
- *Rank-based Diffusion Process with Assured Convergence (**RDPAC**)* [149] performs a diffusion process to exploit the information contained in the ranked lists.

Most methods are recent or provide results comparable to the state-of-the-art according to their papers. All of them are available on open-source software, the Unsupervised Distance Leaning Framework (UDLF) [150]. In all methods, the default parameters of the UDLF [150][1] were considered. The only parameters that were changed are: *(i)* the neighborhood size (k) according to the particularities of each dataset[2]; and *(ii)* the ranked lists size (L). For all datasets, $L = 2000$ was used. Given a dataset, we applied the same parameters across all methods and descriptors to ensure a fair and consistent comparison.

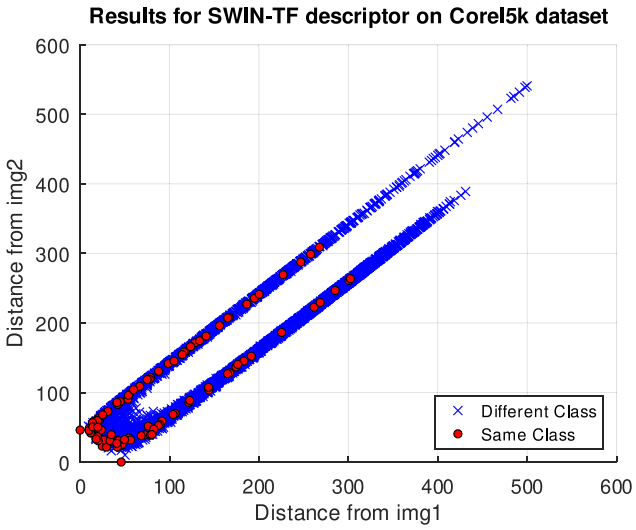

Fig. 11. Distances distribution for two query images on Corel5k dataset before (left) the execution of LHRR method.

Both quantitative and visual results were reported for the similarity learning approaches considering different datasets and features, which were mentioned in this survey. For general-purpose datasets, all the features were extracted considering deep learning models trained on the ImageNet dataset. While, for Re-ID, they were trained on MSMT17, a multi-scene multi-time person re-identification dataset. Since the protocol is unsupervised, descriptors always perform transfer learning.

Table 8 presents the results for 5 general-purpose datasets. Along with the MAP obtained, the relative gain is also reported for each pair of method and descriptor considering the improvement over the *"original"* (i.e., without any approach) value. The relative gains are highlighted in bold, and the best result in each row is shaded in gray. Notice that, in most cases, the methods provided results significantly higher than the standalone feature, with notorious gains. The LHRR consistently achieved the best results across most cases, frequently followed by the RFE. This is probably because they utilize hypergraph models to exploit first and second-order neighborhoods of elements, leveraging the relevant underlying structure of the dataset. Notably, the Dogs [151] dataset is an exception where CPRR achieved superior results. Unlike others, CPRR utilizes symmetrical similarities between elements, which can be particularly advantageous for this dataset.

Similarly, Table 9 presents the results for the same set of methods on 3 Re-ID datasets. Once again, all methods presented a significant gain over the original feature result. Notice that, for CUHK03 and Market, RDPAC and BFSTREE achieved the best results, while LHRR was the best for Duke. A hypothesis is that this is related to the Re-ID detector used to crop the images in each dataset.

We also conducted a visual analysis. In Figs. 11 12, there are two plots: one before the execution of LHRR (Fig. 11) and after (Fig. 12). Each image in the dataset is represented by a different point. The values on the axes correspond to the distances of each image when compared to images img_1 and img_2. Elements belonging to the same class are highlighted in red. It is worth noting that distance learning has led to a substantial enhancement, bringing elements within the same class closer together.

Taking into account the two images, previously denoted as img_1 and img_2, Fig. 13 displays their ranked lists before and after the execution of distance learning. The query images are marked with green borders, while the incorrect results are highlighted with red borders. The application of LHRR successfully eliminated all incorrect results in both scenarios.

[1] UDLF version: https://github.com/UDLF/UDLF/releases/tag/v1.60.

[2] Neighborhood sizes: $k = 90$ for Corel5k [108]; $k = 50$ for CUB200 [113]; $k = 100$ for Dogs [151], Food101 [152], and SUN397 [153]; $k = 20$ for CUHK03 [154]; $k = 30$ for Market [155] and Duke [156].

Table 8
Mean Average Precision (MAP %) results for 5 general-purpose datasets considering different similarity learning approaches and descriptors [157–161].

General-Purpose Datasets						
COREL5K [108] Dataset (5,000 images, 50 classes)						
Descriptors↓ Methods→	**ORIGINAL**	**BFSTREE** [138]	**CPRR** [24]	**RFE** [17]	**LHRR** [19]	**RDPAC** [149]
RESNET [157]	64.75%	84.76%	86.00%	87.51%	90.15%	86.21%
	R. Gain	**+30.93%**	**+32.81%**	**+35.13%**	**+39.19%**	**+33.17%**
DPNET [158]	65.11%	84.22%	85.19%	85.69%	87.40%	84.96%
	R. Gain	**+29.33%**	**+30.84%**	**+31.63%**	**+34.20%**	**+30.48%**
SENET [159]	56.71%	76.13%	77.19%	84.75%	88.90%	80.81%
	R. Gain	**+34.22%**	**+36.14%**	**+49.46%**	**+56.77%**	**+42.53%**
SWIN-TF [160]	74.21%	91.39%	92.20%	95.41%	96.98%	94.17%
	R. Gain	**+23.12%**	**+24.25%**	**+28.57%**	**+30.66%**	**+26.89%**
VIT [161]	75.19%	89.56%	90.21%	91.85%	92.75%	90.27%
	R. Gain	**+19.08%**	**+20.01%**	**+22.18%**	**+23.34%**	**+20.07%**
CUB200 [113] Dataset (11,788 images, 200 classes)						
Descriptors↓ Methods→	**ORIGINAL**	**BFSTREE** [138]	**CPRR** [24]	**RFE** [17]	**LHRR** [19]	**RDPAC** [149]
RESNET [157]	23.11%	35.32%	38.05%	36.20%	39.51%	36.89%
	R. Gain	**+52.81%**	**+64.71%**	**+56.53%**	**+70.75%**	**+59.67%**
DPNET [158]	26.47%	38.05%	39.06%	37.28%	40.36%	39.16%
	R. Gain	**+43.78%**	**+47.61%**	**+40.84%**	**+52.54%**	**+47.89%**
SENET [159]	18.75%	23.78%	23.89%	22.67%	24.72%	24.27%
	R. Gain	**+26.83%**	**+27.41%**	**+20.91%**	**+31.84%**	**+29.44%**
SWIN-TF [160]	58.37%	73.09%	75.54%	67.64%	74.09%	74.43%
	R. Gain	**+25.25%**	**+29.42%**	**+15.86%**	**+26.92%**	**+27.52%**
VIT [161]	60.79%	71.08%	71.92%	67.71%	71.65%	71.38%
	R. Gain	**+16.89%**	**+18.28%**	**+11.35%**	**+17.88%**	**+17.40%**
DOGS [151] Dataset (20,580 images, 120 classes)						
Descriptors↓ Methods→	**ORIGINAL**	**BFSTREE** [138]	**CPRR** [24]	**RFE** [17]	**LHRR** [19]	**RDPAC** [149]
RESNET [157]	64.16%	76.84%	81.28%	71.62%	79.18%	78.07%
	R. Gain	**+19.76%**	**+26.70%**	**+11.63%**	**+23.42%**	**+21.70%**
DPNET [158]	78.29%	85.87%	89.66%	77.57%	85.63%	85.68%
	R. Gain	**+9.68%**	**+14.53%**	**-0.92%**	**+9.36%**	**+9.43%**
SENET [159]	86.39%	88.05%	90.71%	80.67%	85.89%	87.05%
	R. Gain	**+1.92%**	**+4.99%**	**-6.62%**	**-0.58%**	**+0.76%**
SWIN-TF [160]	46.14%	57.78%	64.56%	48.79%	61.59%	59.07%
	R. Gain	**+25.24%**	**+39.95%**	**+5.75%**	**+33.47%**	**+28.04%**
VIT [161]	80.01%	83.21%	86.32%	76.20%	84.18%	82.67%
	R. Gain	**+4.00%**	**+7.89%**	**-4.76%**	**+5.21%**	**+3.32%**
FOOD101 [152] Dataset (101,000 images, 101 classes)						
Descriptors↓ Methods→	**ORIGINAL**	**BFSTREE** [138]	**CPRR** [24]	**RFE** [17]	**LHRR** [19]	**RDPAC** [149]
RESNET [157]	9.04%	11.64%	12.25%	15.84%	17.20%	11.75%
	R. Gain	**+28.76%**	**+35.51%**	**+75.22%**	**+89.82%**	**+29.98%**
SWIN-TF [160]	34.23%	40.40%	40.18%	42.10%	44.11%	41.17%
	R. Gain	**+18.01%**	**+17.38%**	**+23.03%**	**+28.90%**	**+20.26%**
VIT [161]	36.05%	40.10%	39.51%	41.20%	41.55%	40.63%
	R. Gain	**+11.25%**	**+9.62%**	**+14.29%**	**+15.25%**	**+12.70%**
SUN397 [153] Dataset (108,754 images, 397 classes)						
Descriptors↓ Methods→	**ORIGINAL**	**BFSTREE** [138]	**CPRR** [24]	**RFE** [17]	**LHRR** [19]	**RDPAC** [149]
RESNET [157]	18.16%	23.19%	24.24%	25.73%	28.18%	23.42%
	R. Gain	**+27.63%**	**+33.43%**	**+41.69%**	**+55.07%**	**+28.90%**
SWIN-TF [160]	37.05%	43.81%	44.89%	43.22%	47.46%	44.21%
	R. Gain	**+18.23%**	**+21.15%**	**+16.66%**	**+28.10%**	**+19.32%**
VIT [161]	39.42%	43.90%	44.68%	43.65%	45.87%	44.16%
	R. Gain	**+11.36%**	**+13.31%**	**+10.74%**	**+16.36%**	**+12.04%**

In future work, we intend to investigate other methods, descriptors, and parameter sensitivity; and analysis on even larger datasets. On the other hand, larger-scale datasets may provide challenges: user-tagged annotations, potentially introducing noise into data [165]; significant pre-processing, which increases computational costs [166]; and the amplification of class imbalance [167], with impact on the effectiveness [168]. In summary, considering both sets of analyses, there were 216 distinct MAP results and over 329,877 images analyzed. This exploratory and combinatorial approach can assist in developing more adaptable and effective proposals capable of addressing the variability and complexity of real-world data.

5. Conclusions

Despite the disruptive advances in visual data representation strategies supported by CNN and Transformer-based models, the similarity assessment between images remains a challenging task. Represented as points in high-dimensional spaces, images are commonly compared based on pairwise measures, which neglect more global similarity relationships. Unsupervised approaches capable of exploiting contextual information encoded in the dataset manifold going beyond pairwise analysis represent an effective way to obtain more effective similarity measures and, therefore, more effective retrieval results.

In this work, we performed an organization of published research of unsupervised methods focused on post-processing similarity measurement on image retrieval tasks. It involved the diverse taxonomy of the area, with terms like manifold learning, diffusion process, distance/similarity/affinity learning, and re-ranking methods. A systematic review of the literature was conducted using well-established searching tools (*Web Of Science*; *Scopus*), depicting the evolution of the field over the years. A network analysis was also performed, involving

Table 9
Mean Average Precision (MAP %) results for person Re-ID datasets considering different similarity learning approaches and descriptors [162–164].

Person Re-ID Datasets						
CUHK03 [154] Dataset (14,097 images, 1,467 classes)						
Descriptors↓ Methods→	**ORIGINAL**	**BFSTREE [138]**	**CPRR [24]**	**RFE [17]**	**LHRR [19]**	**RDPAC [149]**
RESNET [157]	13.06%	21.48%	20.00%	19.89%	20.76%	21.44%
	R. Gain	**+64.32%**	**+53.14%**	**+52.22%**	**+58.88%**	**+64.01%**
HACNN [162]	9.65%	15.20%	13.92%	14.47%	14.85%	15.39%
	R. Gain	**+57.51%**	**+44.15%**	**+49.84%**	**+53.89%**	**+59.48%**
MLFN [163]	10.16%	16.04%	14.74%	15.43%	15.24%	16.02%
	R. Gain	**+57.87%**	**+45.08%**	**+51.97%**	**+50.20%**	**+57.68%**
OSNET-AIN [164]	26.99%	41.67%	38.47%	39.23%	39.67%	41.30%
	R. Gain	**+54.38%**	**+42.53%**	**+45.35%**	**+47.04%**	**+53.01%**
OSNET-IBN [164]	20.77%	34.01%	31.36%	32.27%	32.84%	34.54%
	R. Gain	**+63.68%**	**+51.01%**	**+55.35%**	**+58.21%**	**+66.42%**
Market [155] Dataset (32,217 images, 1,501 classes)						
Descriptors↓ Methods→	**ORIGINAL**	**BFSTREE [138]**	**CPRR [24]**	**RFE [17]**	**LHRR [19]**	**RDPAC [149]**
RESNET [157]	22.78%	35.67%	34.25%	35.41%	36.15%	36.02%
	R. Gain	**+56.57%**	**+50.35%**	**+55.49%**	**+58.61%**	**+57.95%**
HACNN [162]	23.26%	33.80%	31.05%	32.64%	32.76%	34.20%
	R. Gain	**+45.34%**	**+33.43%**	**+40.29%**	**+40.84%**	**+46.99%**
MLFN [163]	21.95%	32.06%	29.68%	31.06%	31.62%	32.45%
	R. Gain	**+46.11%**	**+35.22%**	**+41.55%**	**+44.06%**	**+47.83%**
OSNET-AIN [164]	43.27%	59.97%	57.74%	59.01%	60.12%	60.64%
	R. Gain	**+38.61%**	**+33.44%**	**+36.42%**	**+39.01%**	**+40.15%**
OSNET-IBN [164]	37.10%	54.14%	51.93%	53.41%	54.94%	55.59%
	R. Gain	**+45.91%**	**+39.95%**	**+43.92%**	**+48.12%**	**+49.89%**
Duke [156] Dataset (36,411 images, 1,812 classes)						
Descriptors↓ Methods→	**ORIGINAL**	**BFSTREE [138]**	**CPRR [24]**	**RFE [17]**	**LHRR [19]**	**RDPAC [149]**
RESNET [157]	31.97%	49.40%	48.91%	51.21%	51.69%	50.33%
	R. Gain	**+54.56%**	**+52.99%**	**+60.12%**	**+61.60%**	**+57.45%**
HACNN [162]	25.52%	40.38%	38.56%	40.90%	40.95%	40.67%
	R. Gain	**+58.33%**	**+51.14%**	**+60.31%**	**+60.47%**	**+59.35%**
MLFN [163]	28.94%	44.79%	43.59%	45.71%	46.27%	45.60%
	R. Gain	**+54.76%**	**+50.59%**	**+57.91%**	**+59.88%**	**+57.63%**
OSNET-AIN [164]	52.66%	68.32%	67.39%	68.77%	69.12%	68.48%
	R. Gain	**+29.74%**	**+27.95%**	**+30.57%**	**+31.23%**	**+30.02%**
OSNET-IBN [164]	45.49%	64.03%	62.84%	64.92%	65.96%	65.14%
	R. Gain	**+40.75%**	**+38.15%**	**+42.67%**	**+45.02%**	**+43.25%**

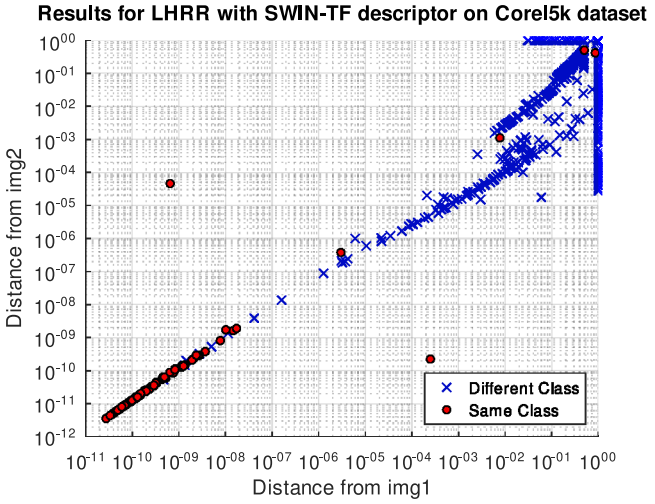

Fig. 12. Distances distribution for two query images on Corel5k dataset after (right) the execution of LHRR method.

interesting keywords and the temporal weight of each collaboration. To select the surveyed works, an ordination according to a Multiple Criteria Decision Analysis *Method Ordinatio* was considered, splitting conferences and journals types into different groups and years (once recent works received fewer citations), defining eligibility, quantitative, and qualitative criteria (through *Prisma Flow*). Such gathered strategies made possible a final ranking with the most cited papers organized by the most cited. The selected works were discussed and summarized according to their predominant category and methodology employed (Diffusion, Ranking, and Deep). It is worth noting that the diversity of methodologies employed in the included works surpasses the categorizations used for section organization. Nevertheless, these categories serve as a practical means to facilitate accessibility, shedding light on thematic focus and prevailing strategies within each study. In this way, the survey provides a comprehensive organization and reference of the research area, especially considering the absence of other surveys focused on this theme. Finally, a brief experimental study demonstrated the potential of effectiveness gains of recent approaches applied to recent CNN and Transformer-based features.

As a broad perspective trend, it was possible to identify that over the years' diffusion methods were profoundly investigated, giving rise to many different approaches followed by rank-based strategies. More recently, deep learning-related strategies also have been applied. However, a promising research direction in the area points to new ways of combining such mentioned and discussed approaches. In another relevant direction, it is interesting to note that, despite the focus of this survey being unsupervised affinity learning, there is a broad range of applications associated with image retrieval. It is worth mentioning, for example, the work of Iglesias et al. (2011) [169] that introduces a classification system for Alzheimer's disease through similarity measures enhanced by the Self-Smoothing Operator (SSO). Using the enhanced metric in nearest neighborhood classification, they showed significantly improved accuracy for Alzheimer's Disease over Diffusion Maps. Person Re-ID represents another relevant application. A representative work is the re-ranking method proposed by Zhong et al. [26]. In addition, manifold learning strategies are not restricted to image retrieval applications and have been successfully exploited in other machine

(a) Ranked lists before and after LHRR for img1.

(b) Ranked lists before and after LHRR for img2.

Fig. 13. Ranked lists before and after the LHRR execution for SWIN-TF descriptor on Corel5k dataset.

learning scenarios, such as clustering [170] and weakly-supervised classification [171].

Considering retrieval tasks, the recent strategies include: Structural Embedding Network (SENet), which captures the internal structure of the images and compresses them into dense self-similarity descriptors [172]; Rank Flow Embedding (RFE) for unsupervised and semi-supervised scenarios [30]; Universal and Compact Representation Learning for Image Retrieval (Unicom), effective for universal and compact feature embedding [173]; and Graph Convolution based Re-ranking (GCR) for visual retrieval tasks via feature propagation [174]. It is worth mentioning recent trends in the area. Embeddings and representation learning are common themes. Considering correlated tasks, there is research involving semisupervised subspace learning with adaptive pairwise graph embedding (APGE) [175]; and structure-aware deep spectral embedding [176], which demonstrates the excellent clustering performance. It is worth highlighting that traditional approaches, such as clustering and diffusion, still have recent advances, based on a fusion-and-diffusion strategy, in which multiple affinity graphs are fused via a weight learning [177].

CRediT authorship contribution statement

V.H. Pereira-Ferrero: Investigation, Software, Visualization, Data curation, Writing – original draft, Writing – review & editing. **T.G. Lewis:** Investigation, Network Analysis, Visualization, Data curation, Writing – review & editing. **L.P. Valem:** Investigation, Visualization, Data curation, Writing – review & editing. **L.G.P. Ferrero:** Investigation, Software, Visualization, Data curation, Writing – review & editing. **D.C.G. Pedronette:** Conceptualization, Methodology, Investigation, Writing – review & editing. **L.J. Latecki:** Investigation, Software, Visualization, Data curation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

The authors are grateful to São Paulo Research Foundation - FAPESP (grants #2020/02183-9 and #2018/15597-6), Brazilian National Council for Scientific and Technological Development - CNPq (grants 313193/2023-1, and 422667/2021-8). This work was also partially supported by National Science Foundation - NSF (grant III-2107213).

References

[1] D.C.G. Pedronette, R. da S. Torres, Image re-ranking and rank aggregation based on similarity of ranked lists, Pattern Recognit. 46 (8) (2013) 2350–2360.
[2] L.C.C. Bergamasco, F.L. Nunes, Intelligent retrieval and classification in three-dimensional biomedical images—a systematic mapping, Comp. Sci. Rev. 31 (2019) 19–38.
[3] R. Datta, D. Joshi, J. Li, J.Z. Wang, Image retrieval: Ideas, influences, and trends of the new age, ACM Computing Surveys (Csur) 40 (2) (2008) 1–60.
[4] Ş. Öztürk, E. Çelik, T. Çukur, Content-based medical image retrieval with opponent class adaptive margin loss, Inform. Sci. 637 (2023) 118938.
[5] A. Çelik, A. Küçükmanisa, O. Urhan, Feature distillation from vision-language model for semisupervised action classification, Turk. J. Electr. Eng. Comput. Sci. 31 (6) (2023) 1129–1145.
[6] Ş. Öztürk, Hash code generation using deep feature selection guided siamese network for content-based medical image retrieval, Gazi Univ. J. Sci. 34 (3) (2021) 733–746.
[7] Z. Li, J. Tang, T. Mei, Deep collaborative embedding for social image understanding, IEEE Trans. Pattern Anal. Mach. Intell. 41 (9) (2018) 2070–2083.
[8] L. Jin, Z. Li, J. Tang, Deep semantic multimodal hashing network for scalable image-text and video-text retrievals, IEEE Trans. Neural Netw. Learn. Syst. 34 (4) (2020) 1838–1851.
[9] S. Yildiz, A. Memiş, S. Varli, TRCaptionNet: A novel and accurate deep Turkish image captioning model with vision transformer based image encoders and deep linguistic text decoders, Turk. J. Electr. Eng. Comput. Sci. 31 (6) (2023) 1079–1098.
[10] X. Li, J. Yu, S. Jiang, H. Lu, Z. Li, Msvit: training multiscale vision transformers for image retrieval, IEEE Trans. Multimed. (2023).
[11] Z. Li, J. Tang, Weakly supervised deep metric learning for community-contributed image retrieval, IEEE Trans. Multimed. 17 (11) (2015) 1989–1999.
[12] Z. Li, J. Tang, L. Zhang, J. Yang, Weakly-supervised semantic guided hashing for social image retrieval, Int. J. Comput. Vis. 128 (2020) 2265–2278.
[13] L.P. Valem, C.R.D. Oliveira, D.C.G. Pedronette, J. Almeida, Unsupervised similarity learning through rank correlation and knn sets, TOMM, ACM Trans. Multimed. Comput. Commun. Appl 14 (4) (2018) 1–23.
[14] M. Donoser, H. Bischof, Diffusion processes for retrieval revisited, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1320–1327.

[15] X. Bai, X. Yang, L.J. Latecki, W. Liu, Z. Tu, Learning context-sensitive shape similarity by graph transduction, IEEE Trans. Pattern Anal. Mach. Intell. 32 (5) (2009) 861–874.

[16] H. Jegou, C. Schmid, H. Harzallah, J. Verbeek, Accurate image search using the contextual dissimilarity measure, IEEE Trans. Pattern Anal. Mach. Intell. 32 (1) (2008) 2–11.

[17] L. Pascotti Valem, D.C.G. Pedronette, L.J. Latecki, Rank flow embedding for unsupervised and semi-supervised manifold learning, IEEE Trans. Image Process. 32 (2023) 2811–2826.

[18] Z. Huang, S. Liu, P. Du, X. Cheng, Ranking tweets with local and global consistency using rich features, in: Advances in Knowledge Discovery and Data Mining: 18th Pacific-Asia Conference, PAKDD 2014, Tainan, Taiwan, May 13-16, 2014. Proceedings, Part I 18, Springer, 2014, pp. 298–309.

[19] D.C.G. Pedronette, L.P. Valem, J. Almeida, R.d.S. Torres, Multimedia retrieval through unsupervised hypergraph-based manifold ranking, IEEE Trans. Image Process. 28 (12) (2019) 5824–5838.

[20] S. Bai, X. Bai, Q. Tian, L.J. Latecki, Regularized diffusion process on bidirectional context for object retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 41 (5) (2018) 1213–1226.

[21] A. Iscen, G. Tolias, Y. Avrithis, O. Chum, Mining on manifolds: Metric learning without labels, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7642–7651.

[22] J. Huang, Q. Dong, S. Gong, X. Zhu, Unsupervised deep learning via affinity diffusion, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (07) 2020, pp. 11029–11036.

[23] X. Yang, L. Prasad, L.J. Latecki, Affinity learning with diffusion on tensor product graph, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2012) 28–38.

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

[25] D.C.G. Pedronette, O.A.B. Penatti, R.T. Calumby, R. da Silva Torres, Unsupervised distance learning by reciprocal kNN distance for image retrieval, in: International Conference on Multimedia Retrieval, ICMR '14, 2014, p. 345.

[26] Z. Zhong, L. Zheng, D. Cao, S. Li, Re-ranking person re-identification with k-reciprocal encoding, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1318–1327.

[27] D.C.G. Pedronette, O.A. Penatti, R.d.S. Torres, Unsupervised manifold learning using reciprocal knn graphs in image re-ranking and rank aggregation tasks, Image Vis. Comput. 32 (2) (2014) 120–130.

[28] S.R. Dubey, A decade survey of content based image retrieval using deep learning, IEEE Trans. Circuits Syst. Video Technol. 32 (5) (2021) 2687–2704.

[29] J. Wan, D. Wang, S.C.H. Hoi, P. Wu, J. Zhu, Y. Zhang, J. Li, Deep learning for content-based image retrieval: A comprehensive study, in: Proceedings of the 22nd ACM International Conference on Multimedia, 2014, pp. 157–166.

[30] L.P. Valem, D.C.G. Pedronette, L.J. Latecki, Rank flow embedding for unsupervised and semi-supervised manifold learning, IEEE Trans. Image Process. (2023).

[31] L. Cayton, Algorithms for manifold learning, Technical Report CS2008-0923, University of California, San Diego, 2005.

[32] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometr. Intell. Lab. Syst. 2 (1–3) (1987) 37–52.

[33] A.J. Izenman, Introduction to manifold learning, Wiley Interdiscip. Rev. Comput. Stat. 4 (5) (2012) 439–446.

[34] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[35] M. Balasubramanian, E.L. Schwartz, The isomap algorithm and topological stability, Science 295 (5552) (2002) 7.

[36] J.B. Tenenbaum, V.d. Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.

[37] D. Zhou, J. Weston, A. Gretton, O. Bousquet, B. Schölkopf, Ranking on data manifolds, Adv. Neural Inf. Process. Syst. 16 (2003).

[38] X. Yang, X. Bai, L.J. Latecki, Z. Tu, Improving shape retrieval by learning graph transduction, in: European Conference on Computer Vision, Springer, 2008, pp. 788–801.

[39] D. Qin, S. Gammeter, L. Bossard, T. Quack, L. Van Gool, Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors, in: CVPR 2011, IEEE, 2011, pp. 777–784.

[40] C.Y. Okada, D.C.G. Pedronette, R. da S. Torres, Unsupervised distance learning by rank correlation measures for image retrieval, in: ACM International Conference on Multimedia Retrieval, ICMR'2015, 2015.

[41] W. Webber, A. Moffat, J. Zobel, A similarity measure for indefinite rankings, ACM Trans. Inf. Syst. (ISSN: 1046-8188) 28 (4) (2010) 20:1–20:38.

[42] F. Yang, R. Hinami, Y. Matsui, S. Ly, S. Satoh, Efficient image retrieval via decoupling diffusion into online and offline processing, in: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, AAAI '19, 2019.

[43] L.P. Valem, D.C.G. Pedronette, R.d.S. Torres, E. Borin, J. Almeida, Effective, efficient, and scalable unsupervised distance learning in image retrieval tasks, in: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, ICMR '15, 2015, pp. 51–58.

[44] X. Zhang, M. Jiang, Z. Zheng, X. Tan, E. Ding, Y. Yang, Understanding image retrieval re-ranking: A graph neural network perspective, 2020, arXiv:2012.07620 URL https://arxiv.org/abs/2012.07620.

[45] Z. Dou, H. Cui, L. Zhang, B. Wang, Learning global and local consistent representations for unsupervised image retrieval via deep graph diffusion networks, 2020, arXiv:2001.01284 URL https://arxiv.org/abs/2001.01284.

[46] Y. Zhang, Q. Qian, H. Wang, C. Liu, W. Chen, F. Wang, Graph convolution based efficient re-ranking for visual retrieval, 2023, arXiv:2306.08792 URL https://arxiv.org/abs/2306.08792.

[47] D.C.G. Pedronette, F.M.F. Gonçalves, I.R. Guilherme, Unsupervised manifold learning through reciprocal kNN graph and connected components for image retrieval tasks, Pattern Recognit. 75 (2018) 161–174.

[48] Z. Zhu, C. Rao, S. Bai, L.J. Latecki, Training convolutional neural network from multi-domain contour images for 3D shape retrieval, Pattern Recognit. Lett. 119 (2019) 41–48.

[49] X. Yang, S. Koknar-Tezel, L.J. Latecki, Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 357–364.

[50] C. Liu, G. Yu, M. Volkovs, C. Chang, H. Rai, J. Ma, S.K. Gorti, Guided similarity separation for image retrieval, Adv. Neural Inf. Process. Syst. 32 (2019).

[51] F. Radenović, G. Tolias, O. Chum, Fine-tuning CNN image retrieval with no human annotation, IEEE Trans. Pattern Anal. Mach. Intell. 41 (7) (2018) 1655–1668.

[52] L. Van der Maaten, G. Hinton, Visualizing data using t-sne., J. Mach. Learn. Res. 9 (11) (2008).

[53] J. Jiang, B. Wang, Z. Tu, Unsupervised metric learning by self-smoothing operator, in: 2011 International Conference on Computer Vision, IEEE, 2011, pp. 794–801.

[54] S. Bai, X. Bai, Sparse contextual activation for efficient visual re-ranking, IEEE Trans. Image Process. 25 (3) (2016) 1056–1069.

[55] Y. Zhao, L. Wang, L. Zhou, Y. Shi, Y. Gao, Modelling diffusion process by deep neural networks for image retrieval, in: BMVC, 2018, p. 161.

[56] T.G. Lewis, Network science: Theory and applications, John Wiley & Sons, 2011.

[57] R. Albert, H. Jeong, A.-L. Barabási, Error and attack tolerance of complex networks, Nature 406 (6794) (2000) 378–382.

[58] M. Aria, C. Cuccurullo, Bibliometrix: An R-tool for comprehensive science mapping analysis, J. Informetr. 11 (4) (2017) 959–975.

[59] R.N. Pagani, J.L. Kovaleski, L.M. Resende, Methodi ordinatio: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication, Scientometrics 105 (3) (2015) 2109–2135.

[60] M.J. Page, D. Moher, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, et al., PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews, Bmj 372 (2021).

[61] L.J. Latecki, R. Lakamper, T. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, in: Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662), vol. 1, IEEE, 2000, pp. 424–429.

[62] T.B. Sebastian, P.N. Klein, B.B. Kimia, Recognition of shapes by editing their shock graphs, IEEE Trans. Pattern Anal. Mach. Intell. 26 (5) (2004) 550–571.

[63] E. Keogh, The ucr time series classification/clustering home-page, 2006, http://www.cs.ucr.edu/~eamonn/time_series_data/.

[64] O. Söderkvist, Computer vision classification of leaves from swedish trees, 2001.

[65] H. Stewénius, D. Nistér, Object recognition benchmark, 2012.

[66] L. Fei-Fei, R. Fergus, P. Perona, One-shot learning of object categories, IEEE Trans. Pattern Anal. Mach. Intell. 28 (4) (2006) 594–611.

[67] H.J.D. Schmid, Hamming embedding and weak geometry consistency for large scale image search–extended version–, 2008.

[68] H. Jégou, M. Douze, C. Schmid, On the burstiness of visual elements, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 1169–1176.

[69] D. Giorgi, S. Biasotti, L. Paraboschi, Shape retrieval contest 2007: Watertight models track, SHREC Compet. 8 (7) (2007) 7.

[70] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: Illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.

[71] D. Nister, H. Stewenius, Scalable recognition with a vocabulary tree, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'06, 2, Ieee, 2006, pp. 2161–2168.

[72] G. Tolias, Y. Avrithis, H. Jégou, Image search with selective match kernels: aggregation across single and multiple images, Int. J. Comput. Vis. 116 (3) (2016) 247–261.

[73] J. Philbin, O. Chum, M. Isard, J. Sivic, A. Zisserman, Object retrieval with large vocabularies and fast spatial matching, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007, pp. 1–8.

[74] J. Philbin, O. Chum, M. Isard, J. Sivic, A. Zisserman, Lost in quantization: Improving particular object retrieval in large scale image databases, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8.

[75] X. Bai, B. Wang, C. Yao, W. Liu, Z. Tu, Co-transduction for shape retrieval, IEEE Trans. Image Process. 21 (5) (2011) 2747–2757.

[76] C. Aslan, A. Erdem, E. Erdem, S. Tari, Disconnected skeleton: Shape at its absolute scale, IEEE Trans. Pattern Anal. Mach. Intell. 30 (12) (2008) 2188–2203.

[77] C.-H. Wei, Y. Li, W.-Y. Chau, C.-T. Li, Trademark image retrieval using synthetic features for describing global shape and interior structure, Pattern Recognit. 42 (3) (2009) 386–394.

[78] S. Bai, X. Bai, Z. Zhou, Z. Zhang, Q. Tian, L.J. Latecki, GIFT: Towards scalable 3D shape retrieval, IEEE Trans. Multimed. 19 (6) (2017) 1257–1271.

[79] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao, 3D shapenets: A deep representation for volumetric shapes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1912–1920.

[80] B. Li, Y. Lu, C. Li, A. Godil, T. Schreck, M. Aono, M. Burtscher, Q. Chen, N.K. Chowdhury, B. Fang, et al., A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries, Comput. Vis. Image Underst. 131 (2015) 1–27.

[81] M. Savva, F. Yu, H. Su, M. Aono, B. Chen, D. Cohen-Or, W. Deng, H. Su, S. Bai, X. Bai, et al., Shrec16 track: largescale 3d shape retrieval from shapenet core55, in: Proceedings of the Eurographics Workshop on 3D Object Retrieval, vol. 10, 2016.

[82] K. Siddiqi, J. Zhang, D. Macrini, A. Shokoufandeh, S. Bouix, S. Dickinson, Retrieving articulated 3-D models using medial surfaces, Mach. Vis. Appl 19 (4) (2008) 261–275.

[83] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: Proceedings of 1994 IEEE Workshop on Applications of Computer Vision, IEEE, 1994, pp. 138–142.

[84] M. Eitz, J. Hays, M. Alexa, How do humans sketch objects? TOG, ACM Trans. Graph. 31 (4) (2012) 1–10.

[85] J.C. Pereira, E. Coviello, G. Doyle, N. Rasiwasia, G.R. Lanckriet, R. Levy, N. Vasconcelos, On the role of correlation and abstraction in cross-modal multimedia retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 36 (3) (2013) 521–535.

[86] P. Shilane, P. Min, M. Kazhdan, T. Funkhouser, The princeton shape benchmark, in: Proceedings Shape Modeling Applications, 2004., IEEE, 2004, pp. 167–178.

[87] Y. Chen, X. Li, A. Dick, R. Hill, Ranking consistency for image matching and object retrieval, Pattern Recognit. 47 (3) (2014) 1349–1360.

[88] M.J. Huiskes, M.S. Lew, The mir flickr retrieval evaluation, in: Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval, 2008, pp. 39–43.

[89] D.C.G. Pedronette, R.d.S. Torres, Exploiting pairwise recommendation and clustering strategies for image re-ranking, Inform. Sci. 207 (2012) 19–34.

[90] J.v. de Weijer, C. Schmid, Coloring local feature extraction, in: European Conference on Computer Vision, Springer, 2006, pp. 334–348.

[91] P. Brodatz, A. Textures, A photographic album for artists and designers. 1966, 2009, Images downloaded in July.

[92] B. Leibe, B. Schiele, Analyzing appearance and contour based methods for object categorization, in: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings., vol. 2, IEEE, 2003, pp. II–409.

[93] X. Bai, S. Bai, X. Wang, Beyond diffusion process: Neighbor set similarity for fast re-ranking, Inform. Sci. 325 (2015) 342–354.

[94] D.C.G. Pedronette, R.d.S. Torres, A correlation graph approach for unsupervised manifold learning in image retrieval tasks, Neurocomputing 208 (2016) 66–79.

[95] T. Deselaers, D. Keysers, H. Ney, Features for image retrieval: an experimental comparison, Inf. Retr. 11 (2) (2008) 77–107.

[96] X. Zhu, C. Tang, P. Wang, H. Xu, M. Wang, J. Chen, J. Tian, Saliency detection via affinity graph learning and weighted manifold ranking, Neurocomputing 312 (2018) 239–250.

[97] V. Movahedi, J.H. Elder, Design and perceptual validation of performance measures for salient object segmentation, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, IEEE, 2010, pp. 49–56.

[98] J. Shi, Q. Yan, L. Xu, J. Jia, Hierarchical image saliency detection on extended CSSD, IEEE Trans. Pattern Anal. Mach. Intell. 38 (4) (2015) 717–729.

[99] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3166–3173.

[100] S. Bai, Z. Zhou, J. Wang, X. Bai, L.J. Latecki, Q. Tian, Automatic ensemble diffusion for 3d shape and image retrieval, IEEE Trans. Image Process. 28 (1) (2018) 88–101.

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

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

[103] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, Q. Tian, Mars: A video benchmark for large-scale person re-identification, in: European Conference on Computer Vision, Springer, 2016, pp. 868–884.

[104] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, Q. Tian, Person re-identification in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1367–1376.

[105] B. Wang, J. Jiang, W. Wang, Z.-H. Zhou, Z. Tu, Unsupervised metric fusion by cross diffusion, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2997–3004.

[106] X. Shen, Z. Lin, J. Brandt, S. Avidan, Y. Wu, Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 3013–3020.

[107] S. Zhang, M. Yang, T. Cour, K. Yu, D.N. Metaxas, Query specific fusion for image retrieval, in: European Conference on Computer Vision, Springer, 2012, pp. 660–673.

[108] G.-H. Liu, J.-Y. Yang, Content-based image retrieval using color difference histogram, Pattern Recognit. 46 (1) (2013) 188–198.

[109] D.M. Chen, G. Baatz, K. Köser, S.S. Tsai, R. Vedantham, T. Pylvänäinen, K. Roimela, X. Chen, J. Bach, M. Pollefeys, et al., City-scale landmark identification on mobile devices, in: CVPR 2011, IEEE, 2011, pp. 737–744.

[110] A. Iscen, G. Tolias, Y. Avrithis, T. Furon, O. Chum, Efficient diffusion on region manifolds: Recovering small objects with compact cnn representations, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2077–2086.

[111] S. Wang, S. Jiang, Instre: a new benchmark for instance-level object retrieval and recognition, TOMM, ACM Trans. Multimed. Comput. Commun. Appl 11 (3) (2015) 1–21.

[112] P. Kontschieder, M. Donoser, H. Bischof, Beyond pairwise shape similarity analysis, in: Asian Conference on Computer Vision, Springer, 2009, pp. 655–666.

[113] C. Wah, S. Branson, P. Welinder, P. Perona, S. Belongie, The Caltech-UCSD Birds-200–2011 Dataset, Tech. Rep. CNS-TR-2011-001, California Institute of Technology, 2011.

[114] S. Bai, Z. Zhou, J. Wang, X. Bai, L. Jan Latecki, Q. Tian, Ensemble diffusion for retrieval, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 774–783.

[115] A. Iscen, Y. Avrithis, G. Tolias, T. Furon, O. Chum, Fast spectral ranking for similarity search, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7632–7641.

[116] F. Yang, B. Matei, L.S. Davis, Re-ranking by multi-feature fusion with diffusion for image retrieval, in: 2015 IEEE Winter Conference on Applications of Computer Vision, IEEE, 2015, pp. 572–579.

[117] X. Yang, X. Bai, S. Köknar-Tezel, L.J. Latecki, Densifying distance spaces for shape and image retrieval, J. Math. Imaging Vis. 46 (1) (2013) 12–28.

[118] D.C.G. Pedronette, R.d.S. Torres, Unsupervised rank diffusion for content-based image retrieval, Neurocomputing 260 (2017) 478–489.

[119] L.T. Alemu, M. Pelillo, Multi-feature fusion for image retrieval using constrained dominant sets, Image Vis. Comput. 94 (2020) 103862.

[120] F. Magliani, K. McGuinness, E. Mohedano, A. Prati, An efficient approximate kNN graph method for diffusion on image retrieval, in: International Conference on Image Analysis and Processing, Springer, 2019, pp. 537–548.

[121] S. Bai, S. Sun, X. Bai, Z. Zhang, Q. Tian, Smooth neighborhood structure mining on multiple affinity graphs with applications to context-sensitive similarity, in: European Conference on Computer Vision, Springer, 2016, pp. 592–608.

[122] A. Krizhevsky, G. Hinton, et al., Learning multiple layers of features from tiny images, 2009.

[123] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.

[124] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.

[125] Z. Dou, H. Cui, L. Zhang, B. Wang, Learning global and local consistent representations for unsupervised image retrieval via deep graph diffusion networks, 2020, arXiv preprint arXiv:2001.01284.

[126] S. Bai, S. Sun, X. Bai, Z. Zhang, Q. Tian, Improving context-sensitive similarity via smooth neighborhood for object retrieval, Pattern Recognit. 83 (2018) 353–364.

[127] J.-M. Geusebroek, G.J. Burghouts, A.W. Smeulders, The amsterdam library of object images, Int. J. Comput. Vis. 61 (1) (2005) 103–112.

[128] S. Schmiedeke, C. Kofler, I. Ferrané, Overview of mediaeval 2012 genre tagging task, in: MediaEval 2012 Workshop, Pisa, Italy, 2012.

[129] J. Huang, S.R. Kumar, M. Mitra, W.-J. Zhu, R. Zabih, Image indexing using color correlograms, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 1997, pp. 762–768.

[130] D.C.G. Pedronette, R.d.S. Torres, Unsupervised manifold learning by correlation graph and strongly connected components for image retrieval, in: 2014 IEEE International Conference on Image Processing, ICIP, Ieee, 2014, pp. 1892–1896.

[131] D.C.G. Pedronette, R. da Silva Torres, R.T. Calumby, Using contextual spaces for image re-ranking and rank aggregation, Multimedia Tools Appl. 69 (3) (2014) 689–716.

[132] A. Iscen, Y. Avrithis, G. Tolias, T. Furon, O. Chum, Hybrid diffusion: Spectral-temporal graph filtering for manifold ranking, in: Asian Conference on Computer Vision, Springer, 2018, pp. 301–316.

[133] S. Sun, Y. Li, W. Zhou, Q. Tian, H. Li, Local residual similarity for image re-ranking, Inf. Sci. 417 (2017) 143–153.

[134] S. Pang, J. Ma, J. Xue, J. Zhu, V. Ordonez, Deep feature aggregation and image re-ranking with heat diffusion for image retrieval, IEEE Trans. Multimed. 21 (6) (2018) 1513–1523.

[135] S. Pang, J. Ma, J. Zhu, J. Xue, Q. Tian, Improving object retrieval quality by integration of similarity propagation and query expansion, IEEE Trans. Multimed. 21 (3) (2018) 760–770.

[136] J.J.-Y. Wang, Y. Sun, From one graph to many: Ensemble transduction for content-based database retrieval, Knowl.-Based Syst. 65 (2014) 31–37.

[137] D.S. Marcus, A.F. Fotenos, J.G. Csernansky, J.C. Morris, R.L. Buckner, Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults, J. Cogn. Neurosci. 22 (12) (2010) 2677–2684.

[138] D.C.G. Pedronette, L.P. Valem, R.d.S. Torres, A BFS-tree of ranking references for unsupervised manifold learning, Pattern Recognit. 111 (2021) 107666.

[139] K. Arun, V. Govindan, S. Kumar, On integrating re-ranking and rank list fusion techniques for image retrieval, Int. J. Data Sci. Anal. 4 (1) (2017) 53–81.

[140] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'06, 2, IEEE, 2006, pp. 2169–2178.

[141] G. Lao, S. Liu, C. Tan, Y. Wang, G. Li, L. Xu, L. Feng, F. Wang, Three degree binary graph and shortest edge clustering for re-ranking in multi-feature image retrieval, J. Vis. Commun. Image Represent. 80 (2021) 103282.

[142] A. Delvinioti, H. Jégou, L. Amsaleg, M.E. Houle, Image retrieval with reciprocal and shared nearest neighbors, in: 2014 International Conference on Computer Vision Theory and Applications, VISAPP, 2, IEEE, 2014, pp. 321–328.

[143] X. Shen, Y. Xiao, S.X. Hu, O. Sbai, M. Aubry, Re-ranking for image retrieval and transductive few-shot classification, Adv. Neural Inf. Process. Syst. 34 (2021) 25932–25943.

[144] B. Li, Y. Lu, A. Godil, T. Schreck, M. Aono, H. Johan, J.M. Saavedra, S. Tashiro, SHREC'13 track: large scale sketch-based 3D shape retrieval, 2013.

[145] B. Li, Y. Lu, C. Li, A. Godil, T. Schreck, M. Aono, M. Burtscher, H. Fu, T. Furuya, H. Johan, et al., SHREC'14 track: Extended large scale sketch-based 3D shape retrieval, in: Eurographics Workshop on 3D Object Retrieval, vol. 2014, 2014, pp. 121–130.

[146] B. Li, Y. Lu, F. Duan, S. Dong, Y. Fan, L. Qian, H. Laga, H. Li, Y. Li, P. Lui, et al., SHREC'16 track: 3D sketch-based 3D shape retrieval, 2016.

[147] R. Arandjelović, A. Zisserman, Smooth object retrieval using a bag of boundaries, in: 2011 International Conference on Computer Vision, IEEE, 2011, pp. 375–382.

[148] J. Xie, R. Girshick, A. Farhadi, Unsupervised deep embedding for clustering analysis, in: International Conference on Machine Learning, PMLR, 2016, pp. 478–487.

[149] D.C. Guimarães Pedronette, L. Pascotti Valem, L.J. Latecki, Efficient rank-based diffusion process with assured convergence, J. Imaging 7 (3) (2021).

[150] L.P. Valem, D.C.G. Pedronette, An unsupervised distance learning framework for multimedia retrieval, in: ACM on International Conference on Multimedia Retrieval, ICMR 2017, 2017, pp. 107–111.

[151] A. Khosla, N. Jayadevaprakash, B. Yao, L. Fei-Fei, Novel dataset for fine-grained image categorization, in: First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, 2011.

[152] L. Bossard, M. Guillaumin, L. Van Gool, Food-101 – mining discriminative components with random forests, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, Cham, 2014, pp. 446–461.

[153] J. Xiao, J. Hays, K.A. Ehinger, A. Oliva, A. Torralba, SUN database: Large-scale scene recognition from abbey to zoo, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 3485–3492.

[154] 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.

[155] 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.

[156] 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, pp. 3754–3762.

[157] 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.

[158] H. Shi, Q. Zhou, Y. Ni, X. Wu, L.J. Latecki, DPNET: Dual-path network for efficient object detection with lightweight self-attention, in: 2022 IEEE International Conference on Image Processing, ICIP, 2022, pp. 771–775.

[159] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[160] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: 2021 IEEE/CVF International Conference on Computer Vision, ICCV, 2021, pp. 9992–10002.

[161] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, CoRR (2020) abs/2010.11929 arXiv:2010.11929.

[162] 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.

[163] 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.

[164] K. Zhou, Y. Yang, A. Cavallaro, T. Xiang, Learning generalisable omni-scale representations for person re-identification, 2019, arXiv preprint arXiv:1910.06827.

[165] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, Y. Zheng, Nus-wide: a real-world web image database from national university of Singapore, in: Proceedings of the ACM International Conference on Image and Video Retrieval, 2009, pp. 1–9.

[166] Z.-Q. Zhao, P. Zheng, S.-t. Xu, X. Wu, Object detection with deep learning: A review, IEEE Trans. Neural Netw. Learn. Syst. 30 (11) (2019) 3212–3232.

[167] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, 2020, in: Proceedings of the IEEE International Conference on Computer Vision, vol. 42, (2) pp. 386–397.

[168] X. Wang, J. Xu, J. Hua, Z. Hao, Multi-label image classification optimization model based on deep learning, in: Wireless Sensor Networks: 14th China Conference, CWSN 2020, Dunhuang, China, September 18– 21, 2020, Revised Selected Papers 14, Springer, 2020, pp. 269–285.

[169] J.E. Iglesias, J. Jiang, C.-Y. Liu, Z. Tu, A.D.N. Initiative, et al., Classification of alzheimer's disease using a self-smoothing operator, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2011, pp. 58–65.

[170] B. Rozin, V.H. Pereira-Ferrero, L.T. Lopes, D.C.G. Pedronette, A rank-based framework through manifold learning for improved clustering tasks, Inform. Sci. 580 (2021) 202–220.

[171] J.G.C. Presotto, S.F. dos Santos, L.P. Valem, F.A. Faria, J.P. Papa, J. Almeida, D.C.G. Pedronette, Weakly supervised learning based on hypergraph manifold ranking, J. Vis. Commun. Image Represent. 89 (2022) 103666.

[172] S. Lee, S. Lee, H. Seong, E. Kim, Revisiting self-similarity: Structural embedding for image retrieval, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 23412–23421.

[173] X. An, J. Deng, K. Yang, J. Li, Z. Feng, J. Guo, J. Yang, T. Liu, Unicom: Universal and compact representation learning for image retrieval, 2023, arXiv preprint arXiv:2304.05884.

[174] Y. Zhang, Q. Qian, H. Wang, C. Liu, W. Chen, F. Wan, Graph convolution based efficient re-ranking for visual retrieval, IEEE Trans. Multimed. (2023).

[175] H. Nie, Q. Li, Z. Wang, H. Zhao, F. Nie, Semisupervised subspace learning with adaptive pairwise graph embedding, IEEE Trans. Neural Netw. Learn. Syst. (2023).

[176] H. Yaseen, A. Mahmood, Learning structure aware deep spectral embedding, 2023, arXiv preprint arXiv:2305.08215.

[177] Q. Li, S. An, L. Li, W. Liu, Y. Shao, Multi-view diffusion process for spectral clustering and image retrieval, IEEE Trans. Image Process. (2023).