Large-Scale Image Indexing and Retrieval Methods: A PRISMA-Based Review

Abdelkrim Saouabe[1], Said Tkatek[2], Hicham Oualla[3], Carlos SOSA Henriquez[4]
Computer Science Research Laboratory, Faculty of Sciences, IbnTofail University Kenitra, Morocco[1, 2]
AKKODIS Research, Paris, France[1, 3, 4]

***Abstract*—Large-scale image indexing and retrieval are pivotal in artificial intelligence, especially within computer vision, for efficiently organizing and accessing extensive image databases. This systematic literature review employs the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology to thoroughly analyze and synthesise the current research landscape in this domain. Through meticulous research and a stringent selection process, this study uncovers significant trends, pioneering methodologies, and ongoing challenges in large-scale image indexing and retrieval. Key findings reveal a growing adoption of deep learning techniques, the integration of multimodal data to improve retrieval accuracy, and persistent challenges related to scalability and real-time processing. These insights offer a valuable resource for researchers and practitioners striving to enhance the efficiency and effectiveness of image indexing and retrieval systems.**

Keywords—Image indexing; image retrieval; similarity; PRISMA, computer vision

I. Introduction

In the aim of big data, the proliferation of digital images has created a need for efficient methods of indexing and retrieving large-scale visual information. With the rapid proliferation of digital content, it is estimated that by 2025, over 160 zettabytes of data will be generated annually, with a significant portion being image and video data. This exponential growth underscores the necessity for advanced indexing and retrieval systems. Large-scale image indexing and retrieval systems play an essential role in different contexts like image search algorithms, content-based image retrieval and multimedia data management. The rapid growth of image data from diverse sources such as social media, surveillance systems and scientific imagery has underlined the importance of robust and scalable techniques for organizing and accessing this vast visual content.

This systematic literature review uses the PRISMA method - a widely recognized approach for systematic reviews in healthcare and the social sciences - to carry out a comprehensive survey of existing literature in the field of large-scale image indexing and retrieval. The PRISMA method guarantees transparency, reproducibility and rigor in the synthesis of evidence from a wide range of studies. By adhering to this methodology, our review aims to provide an unbiased assessment of state-of-the-art techniques, identify gaps in current research and propose future directions for advancing this vital area of computer vision and information retrieval.

The remainder of this review document is organized as follows: Section II outlines the methodology utilized in the systematic literature review, detailing the search approach, criteria for study selection, and the process of extracting data. Section III discusses the outcomes of our analysis, emphasizing key discoveries, patterns, and obstacles uncovered in the chosen studies. Section IV examines the implications of these findings and suggests avenues for future research. Finally, Section V concludes the analysis with a summary of contributions and key lessons.

By synthesizing and analyzing the collective knowledge of large-scale image indexing and retrieval, this journal aims to inform researchers, developers and practitioners engaged in image-based information systems, offering insights to accelerate progress in this dynamic and rapidly evolving field.

II. Methodology of PRISMA

The research methodology used in this study, particularly the use of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) model, is of crucial importance in ensuring the rigor and transparency of the analysis of existing studies on large-scale image indexing and retrieval. The PRISMA process is based on specific guidelines that guide each stage of the systematic review, from initial planning to synthesis of results.

First, the PRISMA methodology requires a clear definition of the review's objectives, including the formulation of precise research questions. These questions guide the selection of relevant studies to be included in the analysis. Next, a detailed search protocol is drawn up, describing the inclusion and exclusion criteria for studies and the literature search strategy used to identify relevant articles.

The systematic search is carried out through academic databases and specialized search engines, using keywords and search terms appropriate to the field of large-scale indexing and image retrieval. The selected articles are then subjected to independent evaluation by two or more reviewers to ensure the quality and consistency of the selection, the methodology is presented in Fig. 1.

Once the included studies have been identified, relevant data are systematically extracted from each selected article. This includes information on the methodologies employed, the results obtained and the authors' conclusions. The extracted data is then synthesized and analyzed to identify trends, gaps and emerging recommendations in the field of large-scale image indexing and retrieval.

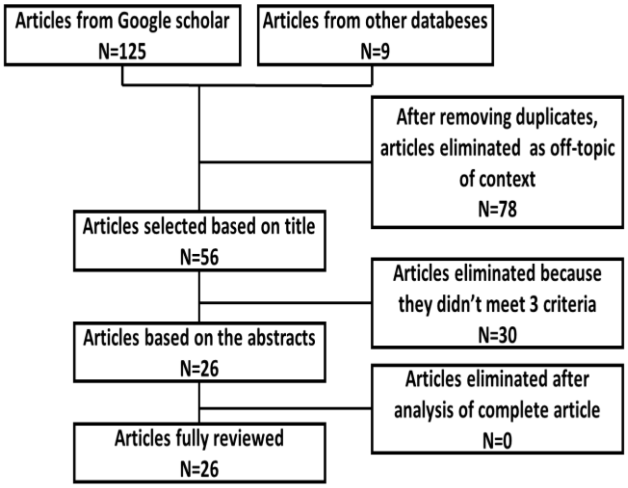
Articles from Google scholar
N=125
Articles from other databeses
N=9
After removing duplicates, articles eliminated as off-topic of context
N=78
Articles selected based on title
N=56
Articles eliminated because they didn't meet 3 criteria
N=30
Articles based on the abstracts
N=26
Articles eliminated after analysis of complete article
N=0
Articles fully reviewed
N=26

Fig. 1. Selection process based on PRISMA.

Therefore, the selection has been carried out by applying the search results for the combinations selected. The inclusion and exclusion criteria are described as follows:

- Inclusion criteria: The articles dealing with "Large-Scale Image Processing" and "Large-Scale Image indexing and retrieval", the contribution of the article, the approach, and the metric.
- Exclusion criteria: The document types other than scientific, the language other than French and English, the duplicates, and the articles which are not directly dealing with the research object, excluding the research subject.

III. Exploring Contemporary Image Indexing and Retrieval Families

This section is dedicated to the state of the art of image indexing and retrieval methods. We can distinguish four families of approaches, which are described in the following sections. Section A presents for Deep Learning-Based Image Indexing and Retrieval family. Section B presents the family of Hybrid Image Retrieval Systems. Section C describes the set of methods dedicated to Image Indexing with Textual Information. Section D the family of Large-Scale Image Retrieval and Indexing. The approaches listed below are presented in chronological order.

A. Deep Learning-Based Image Indexing and Retrieval

The first approach in this category is the one presented in [1] in 2012, which introduces a novel framework for attribute-based image retrieval, allowing users to describe search objects using intuitive attributes. It explores various research aspects related to this method, highlighting recent advances and challenges encountered. This framework extends existing search models to handle relationships between query attributes and weak attributes, improving expressiveness and scalability. To efficiently learn this dependency model without overfitting, the paper proposes a semi-supervised graphical model. This model uses latent trees to represent the joint distributions of query and weak attributes at each level and uses an alternating inference algorithm to estimate the conditional probability.

The approach in study [1] presents a comprehensive dataset for multi-attribute image retrieval, called a-TRECVID, comprising 126 fully labeled query attributes and 6,000 weak attributes of 0.26 million images. The evaluation is based on mean AUC (Area Under Curve), which is commonly used to evaluate the performance of binary classification task, using a different dataset, like a-TRECVID, a-Yahoo and a-PASCAL dataset. The experimental results demonstrate significant performance improvements over state-of-the-art techniques, with a higher AUC equal to 85% compared with other approaches. The semi-supervised model significantly improves the generalizability of the proposed method for cross-dataset searching and searching with a very small training dataset.

The paper in [2] focuses on large-scale partially duplicated image retrieval. Provided a reference image, the goal is to identify pictures featuring the identical object or scene within an extensive database instantly. The approach concerns the development of a coupled binary embedding method for large-scale image retrieval. Multi-index binary indexing is used to combine SIFT visual words with binary features at the indexing level. The correlations are modeled between different features, proposing the concept multi-IDF, which represents a weighted sum of the individual IDFs of each merged feature. The study also explores the integration of the local color descriptor into the retrieval process. The framework is extended to include a binary color feature, using binary features to check visual word match pairs and enhance discrimination capability. This method uses heterogeneous binary features, such as color features, and extends to other binary features. It also incorporates the Multiple Assignment (MA) technique to improve recall of candidate images. Databases used for evaluation include Ukbench (10,200 images), Holidays (1,491 images), DupImage (1,104 images) and MIR Flickr (1 million images).

The performance evaluation in study [2] is based on different metrics depending on the database. For UKBench, performance is measured by recall for the top 4 candidates, while for Holidays and DupImage, mAP (mean Average Precision) is used to assess the quality of image retrieval. In the Holidays dataset, the inverted SIFT files achieve a mAP of 73.9%, while the CN files yield a mAP of 50.5%. The article also analyzes the impact of different parameters on retrieval precision, such as the weighting parameter σ and the Hamming distance threshold κ. In addition, it compares the proposed method with other image retrieval approaches using metrics such as the N-S score for UKBench and the mAP for the Holidays and DupImage datasets.

Deep convolutional neural networks (CNNs) have been effectively utilized for image classification tasks. However, when utilized for image retrieval, the conventional assumption that the last layers of CNNs yield optimal performance, as observed in classification tasks, is often challenged. The Research presented in study [3] demonstrates that, for instance-level image retrieval, lower layers of CNNs frequently outperform the final layers. The presented methodology involves obtaining convolutional features from various layers of CNNs, specifically employing the OxfordNet and GoogLeNet architectures. These features are then encoded into a compact representation using Vector of Locally Aggregated

Descriptors (VLAD) encoding. The diverse layers and dimensions of input images are assessed for their impact on the effectiveness of convolutional features in image retrieval tasks. The experiments were conducted on three datasets for retrieving images at the instance level: Holidays, Oxford, and Paris. The Holidays dataset comprises 1491 personal holiday photos across 500 categories, with the first image of each category used as a query. The Oxford and Paris datasets contain images of famous landmarks, each with specified regions of interest for retrieval. Oxford includes 5062 images, while Paris includes 6412 images, both with multiple queries per landmark. The approach of [3] emphasizes on:

- Utilization of Deep Neural Networks: By employing the OxfordNet and GoogLeNet for feature extraction, exploring various layers to identify optimal performance for image retrieval.
- Feature Extraction: Convolutional features are extracted from selected layers of the networks, considering different dimensions of input images.
- Encoding with VLAD: Features are compressed into compact representations using VLAD encoding to facilitate efficient retrieval.

The experimental results show that middle or deeper layers with more refined resolutions often yield better outcomes in image retrieval when contrasted with using the final layer. Specifically, when employing compressed 128-dimensional VLAD descriptors, the method achieves state-of-the-art performance and outperforms existing VLAD and CNN-based approaches on two of the three test datasets (Holidays, Oxford and Paris). Computing times between 0.4 and 3.5 seconds depending on the sparsity of the user sketch have bene reported, for a database of 1.5 million images. Following the standard evaluation protocol, mAP is used to evaluate the performance of the proposed approach. The mAP initially increases as you move deeper into the network. The variation in mAP varies from one base to another, between 10% and 75%, and increases with the layers.

The adoption of Deep Learning for content-based image retrieval has surged in recent years. In the research presented in [4], a method for indexing Deep Convolutional Neural Network Features to facilitate efficient retrieval from extensive image databases has been introduced. The approach involves encoding these features into text representations, allowing the utilization of a text retrieval engine for image similarity searches. This led to the development of LuQ (name of the approach), a robust retrieval system that integrates full-text search with content-based image retrieval capabilities. The main idea behind LuQ is to index DCNN features using a text encoding that allows us to use a text search engine to perform an image similarity search. To enhance index efficiency and query response time, they conducted evaluations on various tuning parameters for text encoding. As a result, a web-based prototype capable of efficiently searching through a dataset containing 100 million images was developed.

To evaluate the efficiency of LuQ, the Yahoo Flickr Creative Commons 100 million (YFCC100M) dataset was employed [4]. This dataset was established in 2014 under the Yahoo Webscope program. YFCC100M comprises 99.2 million photos and 0.8 million videos that were uploaded to Flickr between 2004 and 2014. The authors have reported an average query time of less than four seconds, without parallelization, for the configuration LuQ and Cur = 10. This approach is 10 magnitudes quicker than the sequential scan using L2. The mAP is assessed based on the quantization factor Q. The findings indicate that a high mAP is achieved when Q is set to 30, yielding a value of 62%.

The paper in [5] discusses the development of a new method based on CNNs to improve the performance of content-based image retrieval (CBIR). This innovative approach uses deep CNN models, exploiting class information available in the data, as well as information provided by distractors, to improve search accuracy. It uses a deep CNN model to extract meaningful features from images, adjusting its weights to bring each image closer to its relevant representations and further away from those that are irrelevant. The experimental results are presented on public datasets for image retrieval, demonstrating its effectiveness. Its main contributions include the integration of multiple relevant and irrelevant samples in the training phase for each sample, as well as the definition of representation objectives for training samples and regression on hidden layers.

For the future, the authors of [5] plan to extend their approach to the use of cropped queries in the Oxford and Paris datasets for research. Evaluations were carried out on the Oxford 5k, Paris 6k, UKBench and UKBench-2 databases, containing 5062, 6412, 10200 and 7650 images respectively. The presented approach also uses the BVLC Reference CaffeNet model and re-train the pre-trained CNN on the dataset. It refines the distance between distractor representations and each specific image using Euclidean loss during training. Adjustments are made to the CaffeNet model, such as removing certain layers and replacing them with PReLU layers, making this method apply to all layers except FC6 and FC7. The performance is evaluated using the mAP for 55 queries, the mAP results vary from 22% to 98% depending on the database and Feature Representation. The approach is compared to other CNN-based methods as well as to manually designed methods on the Oxford 5k, Paris 6k, UKBench and UKBench-2 datasets in terms of precision and recall. It also compares this method to other approaches.

The approach presented in the paper [6], addresses the optimization of large-scale similarity search using matrix factorization. The method presented reformulates the image search problem into a matrix factorization problem, which can be solved by eigenvalue decomposition or dictionary learning. The database used includes various datasets such as Oxford5k, Paris6k, 100,000 Flickr distracting images, Oxford105k, Paris106k, Yahoo Flickr Creative Commons 100M, Holidays and UKB.

The performance is assessed by measuring retrieval performance, mainly mAP for most datasets, except for UKB where performance is measured by 4×recall@4. The mAP increases as a function of Complexity Ratio and reaches values between 60% and 80% depending on dataset used. A comparison is made between the eigenvalue decomposition method, Dictionary Learning (DL) and locality-sensitive hashing (LSH). Dictionary learning shows better performance, especially with large datasets such as Paris6k, Oxford5k, Oxford105k and Paris106k, using complexity ratio and mAP as performance measures [6]. The approach offers an efficient alternative for large-scale similarity search by reducing the number of vector operations required, while maintaining search performance comparable to exhaustive search.

The research presented in study [7], presents the DELF (DEep Local Feature), an attentive local feature descriptor designed specifically for large-scale image retrieval tasks. DELF is based on convolutional neural networks trained solely with image-level annotations from a landmark image dataset. To identify semantically meaningful local features for image retrieval, an attention mechanism for keypoint selection is incorporated, which shares most network layers with the descriptor. DELF can seamlessly replace existing keypoint detectors and descriptors in image retrieval systems, leading to improved feature matching and geometric verification. The approach provides reliable confidence scores to mitigate false positives, particularly excelling in scenarios where query images lack correct matches in the database. To assess the effectiveness of DELF, the Google-Landmarks dataset was introduced, a comprehensive large-scale dataset designed to challenge image retrieval systems with diverse scenarios including background clutter, partial occlusion, multiple landmarks, and objects at varying scales. The framework presented in this work represents a significant advancement in local feature descriptors for image retrieval, demonstrating robust performance across challenging real-world scenarios through the integration of CNN-based learning and attention mechanisms. In the first stage, the performance is evaluated using a modified precision (PRE) and Recall. In the second stage, the performance was evaluated using mAP on the Oxf5k, Oxf105k, Par6k and Par106K databases. The results show satisfactory values for the proposed approach, with mAP values equal to 90%, 88,5%, 95,7% and 92,8% for the four databases used.

Instance Search (INS) presents a significant challenge compared to traditional image search, as relevancy is defined at the instance level rather than image-wide. Prior research has relied on complex ensemble systems involving object proposal generation and subsequent feature extraction for matching, often resulting in a disjointed approach with decreased effectiveness [8]. Moreover, the sheer volume of proposals has hindered the matching speed, limiting these methods scalability to large datasets. To address these shortcomings, the paper presented by [8] introduces Deep Region Hashing (DRH), a novel approach for large-scale INS using image patches as queries. DRH is an end-to-end, deep neural network integrating object proposal, feature extraction, and hash code generation. Notably, it shares the full-image convolutional feature map with the region proposal network, making region proposals practically cost-free. Furthermore, it maps high-dimensional, real-valued region features into compact binary codes for efficient object-level matching in large-scale datasets.

The experimental results in study [8] across four datasets, Oxford5K, Oxford106k, Paris6k and Paris 106k, demonstrate that DRH outperforms state-of-the-art methods in terms of mAP while achieving an approximate 85% increase in efficiency for all the datasets. The algorithm performance is evaluated on the instance search task by studying the impact of different components within the presented framework. A comparative analysis is then conducted between DRH and existing algorithms, focusing on both efficiency and effectiveness, using two standard datasets as benchmarks. For the lDRH, the INS is improved by approximately 3% for the Oxford datasets, while the gQE improves the performance of DRH by 85% on the Oxford 5k dataset.

The paper in [9] discusses the development of a large-scale image hashing method called Semi-supervised Deep Hashing (SSDH), aimed at improving the efficiency of image retrieval. Firstly, it proposes a semi-supervised loss that simultaneously minimizes the empirical error on labeled data and the embedding error on labeled and unlabeled data, to preserve semantic similarity and capture meaningful relationships between data for efficient hashing. Secondly, a semi-supervised deep hashing network is designed to fully exploit both labeled and unlabeled data, enabling the simultaneous learning of hash functions and image representations in a semi-supervised manner. In addition, an online graph construction method is proposed to take advantage of deep features evolving during training, to better capture semantic relationships between images.

The experiments in study [9] are carried out on several datasets, including CIFAR-10, MNIST, NUS-WIDE and MIRFLICKR. Evaluation of the SSDH method is based on various metrics such as mAP, precision-recall curves, accuracy @ topk and accuracy @ top500. This evaluation compares SSDH to eight state-of-the-art methods, including unsupervised, semi-supervised and supervised approaches. The proposed method had a high recall value of 78% and a minimal retrieval time of 980 ms, which is in the order of 10% more satisfactory than the comparative methods. The mAP varies between 70% and 98% depending on different hash code lengths and dataset.

The paper in [10] discusses large-scale image retrieval using transductive support vector machines (TSVMs) combined with hierarchical binary trees (BHTs). This innovative approach comprises several key components. Firstly, it exploits TSVMs in combination with BHTs to facilitate efficient large-scale image retrieval. Secondly, it involves the creation of multiple binary hierarchical trees based on the separability of visual object classes, which contributes to a better organization of data for search. In addition, the TSVM classifier is trained using a stochastic gradient-based solver, enabling efficient scaling with large datasets. Finally, the approach includes a method for learning class hierarchies, using graph cutting and hierarchical binary trees, to ensure a meaningful margin between samples of different classes.

The experiments of [10] are conducted on various datasets including Cifar100, CLEF 2013 Dataset, NUS-WIDE dataset, Cifar10 dataset and MNIST 3-digit dataset. The evaluation of the proposed method, TSVMH-BHT, is done compared with other supervised and unsupervised hashing methods. Measures used for this evaluation include the mAP and Euclidean distance metrics. The mAP results range from 24.46% to

37.27% for Fisher vectors (FVs) and 41.48% to 46.78% for CNN. In addition, the approach's effectiveness is also evaluated against other large-scale image search methods, particularly on specific datasets such as Cifar100.

In the paper [11], other novel image indexing systems are introduced based on the composition of an inverted file index (IVF) and a structured binary encoding mechanism termed SUBIC (Structured Unifying Binary Encoding). Unlike traditional approaches that rely on unsupervised clustering for indexing, the proposed system leverages a unified neural framework to learn both indexing components. The IVF system partitions the dataset using Vector Quantization (VQ), facilitating efficient retrieval by restricting searches to relevant subsets. Concurrently, the SUBIC encoder embeds image features into a structured binary space, enabling rapid approximate distance computations during retrieval.

The approximate distance computation methods in the framework encompass hashing techniques and structured variants of VQ. The methodology in study [11] extends supervised learning approaches to enhance the efficiency and accuracy of distance computations. To evaluate the efficacy of the feature encoder, the performance of SUBIC encoding is compared against unsupervised Vector Quantization (VQ) on various test datasets. Furthermore, the baseline indexing systems (IVF-PQ, IMI-PQ, DSH-SUBIC) are assessed by comparing metrics such as the average number of retrieved images and mean Average Precision across the first T responses. mAP results vary between 46% and 93% depending on the dataset and method used. The evaluations are conducted against established benchmarks including Oxford5K, Oxford5K, Paris6K, Holidays, Oxford105K, and Paris106K. The presented study aims to demonstrate the superiority of the proposed indexing system over traditional approaches, highlighting improvements in retrieval performance across diverse datasets through the integration of supervised learning techniques into the indexing pipeline.

The study referenced in study [12] explores the creation of a system that can efficiently retrieve images instantaneously from extensive repositories, focusing on scalability and computational power. It specifically targets applications in remote detection and botany. The method involves processing images independently without considering relationships between subsets of images. A deep Convolutional Neural Network (CNN) is used to extract features and generate deep representations from the image data. Additionally, an optimized data structure is introduced to improve query speed by employing a structure organized in hierarchical levels and recursive similarity assessments. The study includes a comprehensive series of trials to assess the precision and computational effectiveness of the suggested image retrieval approach, which is tailored for botanical identification and high definition remotely detecting data. Comparative analysis is conducted against traditional content-based image retrieval (CBIR) methods like the bag of visual words (BOVW) and integrating multiple features (MFF) methods.

The experiments in study [12] aimed to assess fundamental aspects of Content-Based Image Retrieval (CBIR), focusing on accuracy and computational efficiency. Feature extraction was conducted using the Keras API in Python within a deep learning framework, while MATLAB was employed for feature indexing. The accuracy of the proposed method was evaluated against BOVW and MFF, traditional feature-based methods. Additionally, computational efficiency and retrieval time were compared with inverted index organization and flat structure search strategies. The experiments utilized the University of California Merced (UCM) Dataset, which includes 21 land cover classes with large-scale aerial images sourced from the USGS national map urban area imagery. Each class comprised of 100 images sized at 256 × 256 pixels, with a spatial resolution of 30 cm per pixel in RGB spectral space for assessing the effectiveness of high-resolution remotely detecting image scene classification.

The evaluation assesses performance using mean Average Precision (mAP) and average retrieval time metrics. The proposed approach achieves maximum precision exceeding 90% in mAP scores. On the MalayaKew (MK) and UCM image datasets, RL-CNN utilizing a hierarchical indexing scheme achieved average retrieval times of 0.039 and 0.025 seconds, respectively. In comparison, RL-CNN employing sequential searching ranked second with retrieval times of 0.164 and 0.142 seconds on the same datasets. It's important to highlight that sequential searching operates with an O (N) linear complexity, resulting in significantly longer execution times as the image count rises to hundreds of thousands or even millions, contrasting techniques such as BOVW using inverted index and BOVW without indexing showed slower retrieval times-0.29 and 1.8 seconds for the MK dataset, and 0.66 and 3.9 seconds for the UCM dataset, respectively. Moreover, RL-CNN employing sequential searching exhibited better efficiency compared to BOW with a comparable framework [11].

The development of the DSLL (Distribution Structure Learning Loss) algorithm for image retrieval based on deep metric learning is discussed in study [13]. DSLL preserves the structural information of positive samples by learning a hyperplane for each query sample in the model, while using the structural distribution-based entropy weight to assess the spatial relationship between negative samples and their environment. This method combines the eigenvectors with the weights to train the network, improving retrieval accuracy by preserving the structure of the image feature vectors and the consistency of the structural similarity ranking.

The DSLL performance is evaluated from different aspects by study [13], including the impact of choosing different selections of the boundary τ, comparing performance based on non-ranking, ED (Euclidean distance) and structural consistency methods, as well as the impact of choosing different selections of the threshold β. These evaluations are carried out with the AlexNet and VGG architectures on the Oxford 5k and Paris 6k datasets. In addition, the performance mAP of DSLL is compared with that of state-of-the-art image retrieval methods under VGG and ResNet deep networks on various datasets. The result of the evolution of mAP depending on training epoch shows that the mAP increases and achieves 59% for Oxford5K and 70% for Paris5k.

The proposed method by study [14], named OS2OS (Score Objects in Scene for Objects in Scene), aims to model object-level regions using image key points from an image index, enabling small significant objects to be accurately weighted in the results, without requiring costly object detections. Several datasets are used, including Oxford 5k, Paris 6k, Google-Landmarks, the NIST Media Forensics Challenge 2018 (MFC2018) and the Reddit dataset.

The evaluation of the method uses features such as SURF and DELF, as well as indexing techniques such as Optimized Product Quantization (OPQ) for nearest neighbor search. Performance is assessed by comparing the OS2OS score with other spatial verification methods. The OS2OS scoring provides competitive or superior performance, without the need for bounding boxes to pre-select regions of interest. The mAP varies between 74% and 86% depending on the techniques used and the Paris or Oxford dataset. The recall scores confirm the effectiveness of the approach comparing to other methods with higher values 47,9%, 54 ,8% and 59,3% respectively for top-50, 100, and 200 most related retrieved images.

The paper in [15] presents a new image retrieval method called CBIR-Similarity Measure through Artificial Neural Network Interpolation (CBIR-SMANN), aimed at improving image retrieval using artificial neural network (ANN) interpolation. The process involves resizing images, applying Gaussian filtering as pre-processing, and identifying key points using a Hessian detector. Features such as mean, kurtosis and standard deviation are extracted and fed to an artificial neural network for interpolation. The interpolated data are then stored in a database for later retrieval. The main contributions of this method are as follows:

- Acquisition and verification of image data including information on objects, colors, spatial information, textures, and shapes, leading to maximum retrieval rates and accuracy.
- Introduction of a weightless feature description and detection model that efficiently recovers appropriate results from complex and cluttered datasets.
- Introduction of a method to implement semantic variation with a similarity measure and color matching to highlight objects.
- Ability of the technique to extract only important image information from the anchor translation instead of iterating over complete images.
- Deploying a retrieval system optimized for storage, processing speed, and computational efficiency, ensuring search results are obtained within seconds.

The future perspective suggested in the article is to integrate the convolution network to obtain improved results. The experiments were carried out on a public dataset containing 1000 images divided into 15 classes. Evaluation metrics used included recovery time, accuracy, false positive rate, false negative rate, specificity, F1 score, error, precision, recall and negative predictive value [15]. The results indicate that CBIR-SMANN achieved a high recall rate of 78% with minimum retrieval time of 980ms has given a high precision with 82% compared to other approaches that were cited in the paper.

The paper in [16] presents the Super Global method, which transforms the conventional two-stage image retrieval model. This approach exclusively relies on global features for both initial retrieval and reranking, thereby improving efficiency without sacrificing accuracy. The method utilizes only global image features for both retrieval phases, eliminating the necessity for local features and implementing advancements in global feature extraction and reranking processes.

The scalability issues are addressed in image retrieval systems, specifically the substantial storage and computational costs associated with local feature matching during reranking. The effectiveness of the proposed method is evaluated using standard image retrieval benchmarks, demonstrating significant improvements over existing approaches. Notably, on the Revisited Oxford+1M Hard dataset, the single-stage performance improves by 7.1%, while the two-stage approach achieves a gain of 3.7% with a remarkable speedup of 64,865 times. The two-stage system outperforms the current single-stage state-of-the-art by 16.3% [16]. The mAP Results were performed on the ROxford and RParis datasets (and their large-scale versions ROxf+1M and RPar+1M), with Medium and Hard evaluation protocols. The best mAP results are obtained for RN50 and RN101 and vary between 68% and 90% depending on the database used.

The paper in [17] discusses image-based patent searching, highlighting its growing importance in the fields of intellectual property and information retrieval. The proposed method focuses on a simple yet robust approach, using a lightweight structure for feature extraction and a neck structure to obtain a low-dimensional representation to facilitate patent search. The network is trained with classification loss in a geometric angular space, accompanied by data augmentation specifically tailored to patent drawings, without scaling. The database used, DeepPatent, comprises a total of 45,000 different design patents and 350,000 drawing images, with a training set of 254,787 images from 33,364 patents and a validation set of 44,815 images from 5,888 patents.

The performance is evaluated using mAP and Rank-N metrics [17]. The results show that the proposed method outperforms traditional and deep learning methods, highlighting its robustness and scalability, for mAP the result obtained is the higher, with value equal to 71%, comparing with some other approaches, the same for Rank-1, Rank-5, Rank-20, the experiments show the performances of the proposed approach which equal respectively to 88,9, 95,8 and 98,1 and higher than the values obtained for other existing approaches in the literature. These results also suggest promising directions for exploring more robust loss functions in the context of image-based patent search, paving the way for future advances in this field.

B. Hybrid Image Retrieval Systems

In this family the first approach is the one presented in study [18] in 1996 which is the oldest approach in this state of art. A novel approach to object recognition in image databases is presented in study [18], focusing on a process of progressive clustering of coherent image regions that satisfy increasingly stringent constraints. This method is based on the aggregation of coherent image regions satisfying increasing color and texture criteria. It incorporates hierarchical clustering and learning techniques for classification, enabling general objects to be processed in uncontrolled environments. The results are evaluated using different metrics such as precision and recall measuring the method's effectiveness in retrieving objects from large image collections, with databases including QBIC and Photobook without mentioning their specific size.

The system's effectiveness was assessed using 4289 training images and 565 images for testing purposes. The evaluation was implemented in various configuration system and using a different metric (Precision, Recall, Response ratio, Test response….). The result of precision varies between 48% and 61%, recall varies also between 7% and 79% depending on the system configurations used.

In the presented work in study [19], a novel incremental and parallel Correspondence Factor Analysis (CFA) algorithm optimized for large-scale image retrieval using GPU processing was introduced. The CFA algorithm is adapted specifically for content-based image retrieval employing local image descriptors such as SIFT (Scale-Invariant Feature Transform). The primary objectives of this approach include dimensionality reduction and theme discovery to streamline image search processes and reduce query response times. To accommodate very large image databases, an incremental and parallelized version of the AFC algorithm has been presented. This adapted version is leveraged to construct inverse files based on extracted indicators, facilitating the identification of images sharing similar themes with the query image. Notably, this indexing step is also optimized for parallel execution on GPU hardware to achieve rapid query responses.

Experimental results in study [19] conducted on the Nister-Stewenius image database integrated with 1 million Flickr images demonstrate the substantial performance gains of the incremental and parallel algorithm compared to the standard version of AFC. To assess the scalability of the presented methods, the Nistér Stewénius database was extended by merging it with additional Flickr images (100,000, 200,000, 500,000, and one million). The vocabulary size was set of 5000 words. The algorithms were implemented in C++ utilizing LAPACK and ATLAS libraries for efficient computation and optimization. These experiments underscore the effectiveness and scalability of the proposed incremental and parallel CFA algorithm for large-scale image retrieval, offering significant improvements in retrieval speed and efficiency across expansive image datasets. The response time and precision are used to study the performance of the proposed method. The response time is the lowest (7.53ms) compared with other methods. In terms of precision, the results obtained show the quality of the proposed method, with a value equal to 62.5%.

As the volume of multimedia content grows rapidly, there is an increasing demand for efficient image retrieval systems. Content-based image retrieval (CBIR) systems are pivotal in addressing this challenge [20]. However, retrieving specific images from vast databases can be time-consuming. To address this challenge, methods for image organization are utilized to speed up how quickly images can be found. The study outlined by [3] highlights the advancement of a proficient image retrieval system that integrates various organizing methods to decrease retrieval time. The emphasis lies in developing a hybrid image retrieval system that utilizes texture, color, and shape characteristics of images. Particularly, the gray level co-occurrence matrix (GLCM) is employed to capture texture details, color moments are used for extracting color attributes, and the region props procedure is applied to shape feature extraction. By combining these diverse image attributes, the proposed system aims to enhance retrieval efficiency and enable faster access to relevant images within large databases. Following the feature fusion process using texture, color, and shape attributes, principal component analysis (PCA) is applied to optimize the selection of fused features. Next, two indexing methods—similarity-based indexing and cluster-based indexing—are evaluated in the hybrid image retrieval system to gauge their efficacy.

The study in [20] reveals that the hybrid color descriptor combined with cluster-based indexing yields significant improvements. The findings show mean precision percentages of 93.8%, 79.6%, 70%, 98.7%, 93.5%, and 79.5% across different datasets, such as Corel-1K, Corel-5K, Corel-10K, COIL-100, GHIM-10, and ZUBUD. These findings demonstrate the efficacy of the proposed hybrid image retrieval system, particularly when utilizing cluster-based indexing, in achieving enhanced retrieval performance across diverse datasets. The utilization of PCA for feature optimization further contributes to the system’s effectiveness in retrieving relevant images efficiently from large repositories.

C. Image Indexing with Textual Information

The paper in [21] presents an automatic image text alignment algorithm aimed at improving the indexing and retrieval of large-scale web images by aligning them with relevant auxiliary text terms or phrases. The algorithm operates in several stages:

- Web Crawling and Segmentation: A large collection of cross-media web pages containing both web images and associated text is crawled and segmented to create image‐text pairs. These pairs consist of informative web images and their corresponding text terms or phrases.

- Near-Duplicate Image Clustering: The web images are grouped into clusters of near duplicates based on visual similarities. Images within the same cluster share similar semantics and are associated with similar auxiliary text terms or phrases that frequently co-occur in relevant text blocks. This clustering process helps reduce uncertainty in determining the relationship between image semantics and auxiliary text.

- Random Walk on Phrase Correlation Network: A random walk is performed on a phrase correlation network to refine relevance scores between web images and their associated text terms or phrases. This step enhances the precision of image‐text alignment. The algorithm effectiveness is validated through experiments on large-scale cross-media web pages, demonstrating positive results in terms of image retrieval and indexing.

As a result of the [21] methodology, a database containing 5,000,000 image – text pairs are curated. This approach enhances the capability of image retrieval systems by leveraging text information associated with web images, thereby improving the accuracy and effectiveness of indexing and retrieval tasks on a large scale.

D. Large-Scale Image Retrieval and Indexing

The challenges posed by the rapid growth of online image repositories like Flickr, which house vast quantities of images requiring efficient indexing, searching, and browsing capabilities has been assessed by [22]. The presented approach leverages image content as valuable information for image retrieval. The Latent Dirichlet Allocation (LDA) models are adopted to represent images for content-based retrieval, involving learning image representations in an unsupervised manner, where each image is characterized as a mixture of topics or object parts depicted within the image. This modeling enables the placement of images into subspaces for higher-level reasoning, facilitating the discovery of similar images. The various similarity measures are explored based on this image representation to enhance retrieval accuracy. To validate the presented approach, it is evaluated on a real-world image database comprising over 246,000 images and compare it against image models based on probabilistic Latent Semantic Analysis (pLSA). The results demonstrate the effectiveness and scalability of the proposed LDA-based approach for large-scale image databases.

The active learning is integrated by [22] with user relevance feedback into our framework to further enhance retrieval performance. This incorporation of user feedback allows for iterative refinement of the retrieval process, adapting to user preferences and improving the relevance of retrieved images. Overall, the work presents the potential of LDA-based image representation for content-based retrieval, particularly in managing large-scale image datasets, and underscores the benefits of incorporating active learning mechanisms to optimize retrieval outcomes based on user interaction and feedback.

Large-scale image retrieval has demonstrated significant potential for real-life applications. The conventional method relies on Inverted Indexing, where images are represented using a Bag-of- Words model. However, a key drawback of this approach is the neglect of spatial information associated with visual words during image representation and comparison. This oversight leads to reduced retrieval accuracy. Earlier researchers in [23] investigated a technique for integrating spatial data into the Inverted Index, aiming to boost precision without compromising retrieval speed. Their methodology was tested on established datasets (Oxford Building 5K, Oxford Building 5K+100K, and Paris 6K), demonstrating the efficacy of their proposed method.

To evaluate the accuracy of the retrieval system, the study in [23] compared the mAP and processing time (in seconds) of four methods (Baseline 1, Baseline 2, II+SPM, II+SPM*) across three datasets: Oxford 5K, Oxford 105K, and Paris 6K. The method II+SPM* improves the mAP of the Baseline 2 by about 2.31%, 4.21% and 3.31% on Oxford 5K, Oxford 105K and Paris 6K, respectively. Furthermore, the study investigates the impact of background visual word weighting on the final mAP of the retrieval system across the three datasets. Additionally, the II+SPM* presents lower processing times than Baseline 2 and comparable magnitudes with Baseline 1 and II+SPM.

The exponential growth of online images necessitates efficient indexing for large-scale digital image retrieval. Designing a compact yet highly efficient image indexing system remains challenging, primarily due to the semantic gap between user queries and the complex semantics of vast datasets. In the paper [24], a novel approach that constructs a joint semantic-visual space by integrating visual descriptors and semantic attributes was presented. This integration aims to bridge the semantic gap by combining attributes and indexing within a unified framework. The proposed joint space enables coherent semantic-visual indexing, leveraging binary codes to enhance retrieval speed while preserving accuracy. To address this, the following contributions are made:

- Interactive Optimization Method: Proposed by an interactive optimization approach to discover the joint semantic and visual descriptor space efficiently.
- Convergence Analysis: Proved by the convergence of the optimization algorithm, ensuring convergence to a good solution after a finite number of iterations.
- Integration with Spectral Hashing: By integrating the semantic-visual joint space with spectral hashing, providing an efficient solution for searching billion-scale datasets.
- Online Cloud Service Design: By developing an online cloud service to deliver more efficient multimedia services based on the proposed indexing system.

Experimental evaluations on standard retrieval datasets (Holidays1M, Oxford5K) demonstrate the effectiveness of the presented method compared to state-of-the-art approaches. Moreover, the cloud system significantly enhances performance, presenting the practical utility and scalability of the presented semantic-visual joint space indexing framework. Overall, the presented work [24] contributes to advancing efficient image retrieval systems by addressing the semantic gap through a unified semantic-visual space, optimizing retrieval speed and accuracy, and facilitating scalable multimedia services in cloud environments.

The paper in [25] discusses the development of a large-scale image retrieval system for everyday scenery with typical items. This system uses advances in deep learning and natural language processing (NLP) to gain deeper insights into images by capturing the interconnections between objects within an image. The goal is to empower users to access highly pertinent images and receive recommendations for analogous image searches to delve deeper into the repository. The proposed method, named QIK (Querying Images Using Contextual Knowledge), utilizes forecasts generated by deep networks for tasks related to interpreting images, such as generating descriptions for images and identifying objects, instead of creating local/global image descriptors using CNN-based features. QIK uses contemporary natural language processing (NLP) frameworks for effective and precise image retrieval in daily situations. The QIK's structure comprises two primary elements: the Indexing module and the Query Handler. The Indexer creates a probabilistic image understanding (PIU) for each image in the database, employing cutting-edge captioning and object detection algorithms. There is an exhaustive survey of all the image contextualization methods mentioned in [26]. A PIU comprises the most probable descriptions and identifies items in an image, enabling the contextualization of ordinary scenarios and comprehension of the connections between items. The Query Handler can employ either image descriptions or identified items for image retrieval.

The method evaluation is based on metrics such as retrieval time, accuracy, false positive rate, false negative rate, specificity, F1 score, etc. The tests in study [25] were carried out on datasets including images of daily scenes from MSCOCO and Unsplash.

The Progressive Distributed and Parallel Similarity Retrieval (DPRS) method addresses the search for similarity between computed tomography (CTI) image sequences in resource-constrained cell phone networks (MTNs), while preserving the confidentiality of medical data. DPRS relies on four key techniques: a PCTI-based similarity measure, a lightweight privacy-preserving strategy, an SSL-based data distribution scheme, and a UDI framework. Various experiments [27] following these techniques have been conducted on a database comprising 50,000 CTIS used to diagnose various types of lesions. The experimental results show a significant improvement in response time compared to the state-of-the-art, using metrics such as response time, precision and recall evaluating the effectiveness of the DPRS method. In conclusion, DPRS represents a promising approach for similarity search in medical information systems, combining innovative techniques to optimize similarity between CTISs, preserve data confidentiality, distribute data efficiently and provide effective indexing for fast search.

E. Other Works

There are numerous other works addressing the issue of image indexing that we have not included in this article. These works were excluded because they did not meet the specific inclusion criteria defined by the PRISMA framework we employed. Our criteria were stringent to ensure the relevance and quality of the studies reviewed, focusing on specific methodologies, contexts, and outcomes pertinent to our research objectives. Some of the excluded studies, such as those cited in references [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59] and [60], provide valuable insights and contributions to the field but fell outside our defined scope. These studies might focus on different aspects of image indexing, employ alternative methodologies, or address broader contexts that, while important, were not directly aligned with our research parameters. By acknowledging these works, we recognize the breadth and diversity of research in image indexing, even though they were not part of our systematic review.

IV. Discussion

The systematic review conducted using the PRISMA method has provided a comprehensive synthesis and analysis of the literature on large-scale image indexing and retrieval. The results of this review highlight several important trends, methods, and challenges in this rapidly evolving field of image processing research.

One of the most significant trends identified in the review is the strong advance in deep learning-based approaches, particularly the use of convolutional neural networks (CNNs). These techniques have substantially improved the performance of image retrieval systems, offering superior accuracy and efficiency compared to traditional methods. The ability of CNNs to automatically learn hierarchical feature representations has been a key factor in their success.

Despite these advancements, several challenges remain. One of the primary issues is the generalization of these models to more diverse datasets. Current models often perform well on specific, well-curated datasets but struggle with variability in real-world data. This includes variations in scale, viewpoint, lighting conditions, and occlusions. Additionally, the efficient management of large volumes of data remains a significant hurdle. The storage, retrieval, and processing of massive datasets require robust and scalable solutions.

To address these challenges, promising avenues of research have emerged. Integrating semantic and contextual knowledge into image retrieval systems is one such direction. By understanding the context and semantics of images, retrieval systems can provide more accurate and relevant search results. For instance, combining visual information with textual data or metadata can enhance the retrieval process by adding another layer of information.

Another area of interest is the development of models that are robust to variations in scale and perspective. Techniques such as data augmentation, multi-scale feature extraction, and the use of generative adversarial networks (GANs) for synthetic data generation are being explored to improve model robustness.

The review also highlights the importance of interdisciplinary collaborations to advance research in large-scale image indexing and retrieval. Collaborations between computer scientists, data engineers, domain experts, and industry practitioners can drive innovation and address practical challenges. Such collaborations can lead to the development of more effective and efficient retrieval systems that are applicable to a wide range of real-world scenarios.

Looking ahead, the practical applications of advanced image retrieval systems are vast. From medical imaging and remote sensing to e-commerce and social media, the ability to efficiently and accurately retrieve images has significant implications. Future research should focus on creating more adaptable, robust, and scalable systems that can handle the complexities of real-world data.

We have structured our state-of-the-art in the form of tables to summarize the different approaches, techniques, databases and metrics used in image indexing. This organization makes it possible to present the essential information clearly and concisely, facilitating comparison and analysis of the various methods employed. By detailing the characteristics and performance of the techniques studied, we can better understand their respective advantages and disadvantages, helping us to identify the solutions best suited to our specific image indexing needs.

The summarized Table I provides an overview of various techniques, datasets, and evaluation metrics employed in image retrieval and recognition research, particularly within the family of deep learning-based image indexing and retrieval. The techniques covered include CNNs, DELF, binary encoding, VLAD encoding, DRH, SIFT, FV, among others. Commonly used datasets are Oxford 5k, Oxford 105k, Paris6k, Paris106k, Holiday, and Flickr. The primary evaluation metric is mean average precision (mAP), typically ranging from 70% to over 90%. Additionally, metrics such as retrieval time, precision (PRE), recall (REC), F1 score, Rank-N, and others are also utilized.

Table II offers a thorough summary of the methods, techniques, datasets, and metrics employed in various studies on hybrid image retrieval systems family, image indexing with textual information family, and large-scale image retrieval and indexing family. The table highlights several techniques such as SIFT, PCA, GLCM, DOM, SVM, NLP, and CTI, along with additional techniques mentioned in certain studies, showcasing a broad spectrum of methodologies applied in this field. The studies utilize diverse datasets including Nister-Stewenius, Corel, COIL, ZUBUD, Paris6K, Oxford5K, Oxford105K, Holidays, MSCOCO, and Unsplash, as well as other datasets unique to individual research projects. The evaluation metrics featured in these studies encompass retrieval time, mean Average Precision (mAP), recall (REC), and Rank-N, with mAP being the most frequently cited metric.

Most of these approaches concentrate on the family of Deep Learning-Based Image Indexing and Retrieval, highlighting the significance of deep learning architecture in this field. The decision to choose between these different families depends on the specific problem at hand, as well as various factors such as datasets, image quality, hardware resources, and more.

TABLE I. A SUMMARY OF METHODS, DATA SETS, AND EVALUATION METRICS IN THE DEEP LEARNING-BASED IMAGE INDEXING AND RETRIEVAL FAMILY

Approaches cited in		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]
Techniques	CNNs				×			×			×							
	DELF							×										
	Binary encoding											×						
	VLAD encoding			×														
	DRH								×									
	SIFT	×																
	FV										×							
	Other’s		×			×	×			×	×			×	×	×	×	×
Dataset	Oxford 5k			×		×	×	×	×			×		×			×	
	Oxford 105k						×	×	×								×	
	Paris6k			×		×	×	×	×			×		×			×	
	Paris106k						×	×	×								×	
	Holiday	×	×	×			×					×		×				
	Flickr	×	×		×		×											
	MK												×					
	UCM dataset												×					
	Cifar									×	×							
	a-PASCAL														×			
	Another dataset	×	×			×	×	×		×	×				×	×		×
Metrics	Retrievel time		×		×								×			×	×	
	mAP	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	PRE							×										

	REC	×				×	×	×		×					×	×		
	F1 score															×		
	Rank-N																	×
	Another metric	×	×						×	×	×			×	×	×		

TABLE II. EXAMINING APPROACHES, DATA SETS, AND EVALUATION METRICS FOR THE REMAINING FAMILIES

Approaches cited in		Hybrid Image Retrieval Systems			Image Indexing with Textual Information	Large-Scale Image Retrieval and Indexing				
		[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[27]
Techniques	SIFT		×							
	PCA			×						
	GLCM			×						
	DOM				×					
	SVM					×				
	NLP								×	
	CTI									×
	Other's						×	×		
Dataset	Nister-Stewenius		×							
	Corel			×						
	COIL			×						
	ZUBUD			×						
	Paris6K				×		×			
	Oxford5K						×	×		
	Oxford105K						×			
	Holidays							×		
	MSCOCO								×	
	Unsplash								×	
	Another dataset	×		×	×	×		×		×
Metrics	Retrievel time		×							
	mAP	×	×		×		×	×		
	PRE									
	REC	×								
	F1 score									
	Rank-N				×					
	Another metric	×			×	×	×	×		×

V. CONCLUSION

The systematic review using the PRISMA method has comprehensively synthesized and analyzed the literature on large-scale image indexing and retrieval. The results highlight current trends, methods, and challenges in this dynamic area of image processing research. We have seen a strong advance in deep learning-based approaches, notably convolutional neural networks, which have significantly improved the performance of image retrieval systems. However, challenges remain, such as generalization to more diverse datasets, robustness to variations in scale and view, and efficient management of large volumes of data. Promising avenues of research include the integration of semantic and contextual knowledge to improve the accuracy and relevance of search results.

Finally, while significant progress has been made in the field of large-scale image indexing and retrieval, this review underscores the importance of ongoing research and interdisciplinary collaborations to overcome existing challenges and drive further advancements. Highlighting the necessity of such collaborative efforts, this review provides a foundation for future work aimed at improving the accuracy, efficiency, and applicability of image retrieval systems, with a focus on practical applications and substantial progress in this constantly evolving field.

References

[1] Felix, X. Y., Ji, R., Tsai, M. H., Ye, G., & Chang, S. F. (2012, June). Weak attributes for large-scale image retrieval. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 2949-2956). IEEE..

[2] Zheng, Liang, Shengjin Wang, and Qi Tian. "Coupled binary embedding for large-scale image retrieval." IEEE transactions on image processing 23.8 (2014): 3368-3380.

[3] Yue-Hei Ng, Joe, Fan Yang, and Larry S. Davis. "Exploiting local features from deep networks for image retrieval." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2015.

[4] Amato, G., Debole, F., Falchi, F., Gennaro, C., Rabitti, F. (2016). “Large Scale Indexing and Searching Deep Convolutional Neural Network Features, ” In: Madria, S., Hara, T. (eds) Big Data Analytics and Knowledge Discovery. DaWaK 2016. Lecture Notes in Computer Science(), vol 9829. Springer, Cham 2016. https://doi.org/10.1007/978-3-319-43946-4_14

[5] Tzelepi, Maria, and Anastasios Tefas. "Exploiting supervised learning for finetuning deep CNNs in content based image retrieval." 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016.

[6] Iscen, Ahmet, Michael Rabbat, and Teddy Furon. "Efficient large-scale similarity search using matrix factorization." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[7] Noh, H., Araujo, A., Sim, J., Weyand, T., & Han, B. Large-scale image retrieval with attentive deep local features. In Proceedings of the IEEE international conference on computer vision, 2017.

[8] Song, J., He, T., Gao, L., Xu, X., & Shen, H. T. (2017). Deep region hashing for efficient large-scale instance search from images. arXiv preprint arXiv:1701.07901.

[9] Zhang, Jian, and Yuxin Peng. "SSDH: Semi-supervised deep hashing for large scale image retrieval." IEEE Transactions on Circuits and Systems for Video Technology 29.1 (2017): 212-225.

[10] Cevikalp, Hakan, Merve Elmas, and Savas Ozkan. "Large-scale image retrieval using transductive support vector machines." Computer Vision and Image Understanding 173 (2018): 2-12.

[11] Jain, H., Zepeda, J., Pérez, P., & Gribonval, R. (2018). Learning a complete image indexing pipeline. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4933-4941).

[12] P. Sadeghi-Tehran, P. Angelov, N. Virlet, and M. J. Hawkesford, “Scalable database indexing and fast image retrieval based on deep learning and hierarchically nested structure applied to remote sensing and plant biology,” J. Imag., vol. 5, no. 3, Art. no. 33, 2019.

[13] Fan, L., Zhao, H., Zhao, H., Liu, P., & Hu, H. (2019). Distribution structure learning loss (DSLL) based on deep metric learning for image retrieval. Entropy, 21(11), 1121.

[14] Brogan, J., Bharati, A., Moreira, D., Rocha, A., Bowyer, K. W., Flynn, P. J., & Scheirer, W. J. (2021). Fast local spatial verification for feature-agnostic large-scale image retrieval. IEEE Transactions on image processing, 30, 6892-6905.

[15] Ahmad, Faiyaz. "Deep image retrieval using artificial neural network interpolation and indexing based on similarity measurement." CAAI Transactions on Intelligence Technology 7.2 (2022): 200-218.

[16] Shao, S., Chen, K., Karpur, A., Cui, Q., Araujo, A., & Cao, B. (2023). Global features are all you need for image retrieval and reranking. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 11036-11046).

[17] Wang, Hongsong, and Yuqi Zhang. "Learning Efficient Representations for Image-Based Patent Retrieval." Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Singapore: Springer Nature Singapore, 2023.

[18] Forsyth, D. A., Malik, J., Fleck, M. M., Greenspan, H., Leung, T., Belongie, S., ... & Bregler, C. (1996). Finding pictures of objects in large collections of images. In Object Representation in Computer Vision II: ECCV'96 International Workshop Cambridge, UK, April 13–14, 1996 Proceedings 2 (pp. 335-360). Springer Berlin Heidelberg.

[19] Pham, N. K., Poulet, F., Morin, A., & Gros, P. (2010, January). Indexation et recherche d'images à très grande échelle avec une AFC incrémentale et parallèle sur GPU. In EGC (pp. 145-150).

[20] Bhardwaj S, Pandove G, Dahiya PK (2020) A futuristic hybrid image retrieval system based on an effective indexing approach for swift image retrieval. International Journal of Computer Information Systems and Industrial Management Applications 12:001–013.

[21] Zhou, Ning, and Jianping Fan. "Automatic image–text alignment for large-scale web image indexing and retrieval." Pattern Recognition 48.1 (2015): 205-219.

[22] Hörster, E., Lienhart, R., & Slaney, M. (2007, July). Image retrieval on large-scale image databases. In Proceedings of the 6th ACM international conference on Image and video retrieval (pp. 17-24).

[23] Nguyen, V. T., Ngo, T. D., Tran, M. T., Le, D. D., & Duong, D. A. (2015). A combination of spatial pyramid and inverted index for large-scale image retrieval. International Journal of Multimedia Data Engineering and Management (IJMDEM), 6(2), 37-51.

[24] Hong, R., Li, L., Cai, J., Tao, D., Wang, M., & Tian, Q. (2017). Coherent semantic-visual indexing for large-scale image retrieval in the cloud. IEEE Transactions on Image Processing, 26(9), 4128-4138.

[25] Zachariah, Arun, Mohamed Gharibi, and Praveen Rao. "A large-scale image retrieval system for everyday scenes." Proceedings of the 2nd ACM International Conference on Multimedia in Asia. 2021.

[26] Saouabe, A., Tkatek, S., Mazar, M., & Mourtaji, I. (2023, October). Evolution of Image Captioning Models: An Overview. In 2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM) (pp. 1-5). IEEE. https://doi.org/10.1109/WINCOM59760.2023.10322923

[27] Zhuang, Yi, Nan Jiang, and Yongming Xu. "Prdistributed and parallel similarity retrieval of large CT image sequences in mobile telemedicine networks." Wireless communications and mobile computing 2022 (2022): 1-13.

[28] Deng, Q., Wu, S., Wen, J., & Xu, Y. (2018). Multi-level image representation for large-scale image-based instance retrieval. CAAI Transactions on Intelligence Technology, 3(1), 33-39.

[29] JÉGOU, Hervé, DOUZE, Matthijs, et SCHMID, Cordelia. Représentation compacte des sacs de mots pour l'indexation d'images. In : RFIA 2010-Reconnaissance des Formes et Intelligence Artificielle. 2010.

[30] MOHR, Roger, GROS, Patrick, LAMIROY, Bart, et al. Indexation et recherche d'images. Actes du 16ecolloque gretsi sur le traitement du signal et des images, Grenoble, France, 1997.

[31] Pham, N. K., Poulet, F., Morin, A., & Gros, P. (2010, January). Indexation et recherche d'images à très grande échelle avec une AFC incrémentale et parallèle sur GPU. In EGC (pp. 145-150).

[32] Radenović, F., Iscen, A., Tolias, G., Avrithis, Y., & Chum, O. (2018). Revisiting oxford and paris: Large-scale image retrieval benchmarking. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5706-5715).

[33] Aly, Mohamed, Mario Munich, and Pietro Perona. "Indexing in large scale image collections: Scaling properties and benchmark." 2011 IEEE Workshop on Applications of Computer Vision (WACV). IEEE, 2011.

[34] Deng, Jia, Alexander C. Berg, and Li Fei-Fei. "Hierarchical semantic indexing for large scale image retrieval." CVPR 2011. IEEE, 2011.

[35] Eitz, M., Hildebrand, K., Boubekeur, T., & Alexa, M. (2009). A descriptor for large scale image retrieval based on sketched feature lines. SBIM, 9, 29-36.

[36] Ladhake, S. (2015). Promising large scale image retrieval by using intelligent semantic binary code generation technique. Procedia Computer Science, 48, 282-287.

[37] Perronnin, F., Liu, Y., Sánchez, J., & Poirier, H. (2010, June). Large-scale image retrieval with compressed fisher vectors. In 2010 IEEE computer society conference on computer vision and pattern recognition (pp. 3384-3391). IEEE.

[38] MOHEDANO, Eva, MCGUINNESS, Kevin, O'CONNOR, Noel E., et al. Bags of local convolutional features for scalable instance search. In : Proceedings of the 2016 ACM on international conference on multimedia retrieval. 2016. p. 327-331.

[39] Zhang, C., Lin, Y., Zhu, L., Liu, A., Zhang, Z., & Huang, F. (2019). CNN-VWII: An efficient approach for large-scale video retrieval by image queries. Pattern Recognition Letters, 123, 82-88.

[40] Karaman, S., Lin, X., Hu, X., & Chang, S. F. (2019, June). Unsupervised rank-preserving hashing for large-scale image retrieval. In Proceedings of the 2019 on International Conference on Multimedia Retrieval (pp. 192-196).

[41] Husain, Syed Sameed, and Miroslaw Bober. "Improving large-scale image retrieval through robust aggregation of local descriptors." IEEE transactions on pattern analysis and machine intelligence 39.9 (2016): 1783-1796.

[42] Wu, P., Wang, S., Dela Rosa, K., & Hu, D. (2024). FORB: a flat object retrieval benchmark for universal image embedding. Advances in Neural Information Processing Systems, 36.

[43] Talwalkar, A., Kumar, S., Mohri, M., & Rowley, H. (2013). Large-scale SVD and Manifold Learning. Journal of Machine Learning Research, 14.

[44] Quack, T., Mönich, U., Thiele, L., & Manjunath, B. S. (2004, October). Cortina: a system for large-scale, content-based web image retrieval. In Proceedings of the 12th annual ACM international conference on Multimedia (pp. 508-511).

[45] Li, W., Feng, C., Lian, D., Xie, Y., Liu, H., Ge, Y., & Chen, E. (2023, August). Learning balanced tree indexes for large-scale vector retrieval. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 1353-1362).

[46] Husain, Syed Sameed, and Miroslaw Bober. "Improving large-scale image retrieval through robust aggregation of local descriptors." IEEE

[47] Transactions on pattern analysis and machine intelligence 39.9 (2016): 1783-1796.

[48] Mohammed Alkhawlani, Mohammed Elmogy and Hazem Elbakry, “Content-Based Image Retrieval using Local Features Descriptors and Bag-of-Visual Words” International Journal of Advanced Computer Science and Applications(IJACSA), 6(9), 2015. http://dx.doi.org/10.14569/IJACSA.2015.060929

[49] Rashad, Metwally, Ibrahem Afifi, and Mohammed Abdelfatah. "RbQE: An efficient method for content-based medical image retrieval based on query expansion." Journal of Digital Imaging 36.3 (2023): 1248-1261.

[50] Wang, J., Liu, W., Kumar, S., & Chang, S. F. (2015). Learning to hash for indexing big data—A survey. Proceedings of the IEEE, 104(1), 34-57.

[51] Lin, K., Yang, H. F., Hsiao, J. H., & Chen, C. S. (2015). Deep learning of binary hash codes for fast image retrieval. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 27-35).

[52] Ali, Fathala. "Content based image retrieval (CBIR) by statistical methods." Baghdad Science Journal 17.2 (SI) (2020): 0694-0694.

[53] Lin, K., Yang, H. F., Hsiao, J. H., & Chen, C. S. (2015). Deep learning of binary hash codes for fast image retrieval. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 27-35).

[54] Gkelios, Socratis, Yiannis Boutalis, and Savvas A. Chatzichristofis. "Investigating the vision transformer model for image retrieval tasks." 2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS). IEEE, 2021.

[55] Majeed, S., Usman, M., Sattar, K., Iqbal, S., & Shabir, J. (2022). Optimization of Content Based Image Retrieval Using Hybrid Approach. Quaid-E-Awam University Research Journal of Engineering, Science & Technology, Nawabshah., 20(01), 110-120.

[56] Agrawal, S., Chowdhary, A., Agarwala, S., Mayya, V., & Kamath S, S. (2022). Content-based medical image retrieval system for lung diseases using deep CNNs. International Journal of Information Technology, 14(7), 3619-3627.

[57] Saouabe, A., Tkatek, S., Oualla, H., & Mourtaji, I. (2024, Juin). Image Indexing Approches for Enhacing Content-Based Image Retreival: An Overview. In 2024 10th International Conference on Ubiquitous Networks (UNet) . IEEE. (in press)

[58] El-Nouby, A., Neverova, N., Laptev, I., & Jégou, H. (2021). Training vision transformers for image retrieval. arXiv preprint arXiv:2102.05644.

[59] Saouabe, A., Tkatek, S., Oualla, H., & Mourtaji, I. (2024). To Improving Visual Search Capabilities via Content-Based Image Retrieval. In 2024 3rd International Conference on Embedded Systems and Artificial Intelligence (Esai) . IEEE. (unpublished)

[60] Charles Adjetey and Kofi Sarpong Adu-Manu, “Content-based Image Retrieval using Tesseract OCR Engine and Levenshtein Algorithm” International Journal of Advanced Computer Science and Applications(IJACSA),12(7),2021. http://dx.doi.org/10.14569/IJACSA.2021.0120776