INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

REVIEW ARTICLE

Check for updates

OPEN ACCESS

Received: 02-06-2023

Accepted: 02-08-2023

Published: 08-09-2023

Citation: Khan OR, Bhat JI (2023) Delving into the Depths of Image Retrieval Systems in the Light of Deep Learning: A Review. Indian Journal of Science and Technology 16(34): 2693-2702. https://doi.org/10.17485/IJST/v16i34.1341

[*] **Corresponding author**.

ovais.rashid@iust.ac.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Delving into the Depths of Image Retrieval Systems in the Light of Deep Learning: A Review

Ovais Rashid Khan[1]*, Javaid Iqbal Bhat[1]

1 Department of Computer Science, Islamic University of Science and Technology, Kashmir, India

Abstract

Objective: The objective of this study is to conduct a comprehensive review of existing research and literature in the field of Content-Based Image Retrieval (CBIR). This review highlights the key challenges associated with the extraction and representation of visual semantics of images. This paper discusses the measure used computing similarity and ranking of retrieved images by CBIR system. The review discusses limitation of traditional approaches and also highlights the challenges with the current deep learning methods in semantic feature representation, defining the similarity metrics and indexing. This paper also highlights scalability and generalization challenges in implementing real environment. **Methods:** A thorough literature review was conducted on well-established databases, including Scopus, Web of Science, IEEE Xplore, ACM, and Science Direct, employing appropriate keywords. Mention the period of coverage. Pertinent search terms encompassed local feature representation, global feature representation, low-level features, high level features, semantic gap, image embeddings, handcrafted features, deep learning, image descriptors, similarity, and image indexing, with the aim of exploring content-based image retrieval systems. Comparative analysis was performed on the chosen articles, taking into account factors such as algorithms, methodologies, datasets, and evaluation metrics. The results discussed using comparative analysis, ensuring a comprehensive overview of recent literature on content-based image retrieval, offering valuable insights and highlighting emerging trends in the field. **Findings :** The research uncovers the novelty in the realm of content-based image retrieval (CBIR) by highlighting the challenge of high-level visual semantics when comparing images, as perceived by humans. It emphasizes that feature extraction methods and choices significantly influence CBIR system performance, stressing the importance of selecting suitable features and similarity measures based on image dataset characteristics and application requirements. The study underscores the persistent obstacle of the semantic gap between low-level visual features and high-level semantic concepts, encouraging exploration of diverse approaches like deep learning, relevance feedback, and ontology-based methods to bridge this gap. Particularly, deep learning techniques, notably Convolutional Neural Networks (CNNs), have shown promising results in CBIR by automatically learning hierarchical representations capturing high-level semantic information. However, the review also highlights the challenges of scaling deep learning methods and the limited accessibility of precisely labelled datasets, which can hinder performance and generalization across diverse image datasets and real-world scenarios. Deep learning models pose interpretability challenges due to their complex, opaque nature and hierarchical semantic representations.

Keywords: CBIR; ContentBased Image Retrieval; Deep Learning; Convolutional Neural Network; Local Features; Global Features; Similarity Metric; Semantic Gap

1 Introduction

The influence of rapid technological development has provided people with new digital tools for exploration and experimentation of new applications. Digital image acquisition and capturing tools have been quite popular in many domains. Utilizations of images in different fields are in trend, be it using in clinical science for diseases detection[(1)], for diagnosis in radiology[(2)], using remote sensing[(3)], use case scenario in communication[(4)], document scrutiny by retrieving watermarks[(5)] are just few examples. The huge image repositories generated through this exploration and the constraints posed by conventional text-based database systems, underscoring the need to explore robust solutions for image databases and adopt efficient strategies to facilitate image retrieval. The limitations in text-based system are the representation of images by human annotated keywords and may not capture the semantics of images called semantic gap, a major challenge for image retrieval systems. Manual tagging of images is time-consuming and prone to errors, and language-specific annotations may not generalize the image effectively. Although neural networks have made significant advancements in pattern recognition and analysis, semantically interpretable feature representations in vision applications remains a challenge. As represented features in neural networks are often not well interpretable, posing challenges in generalizing these representations for downstream tasks[(6)].

Searching images by visual content is a well-known problem and has been studied for decades in the multimedia community[(7) (8)]. Content-based image retrieval (CBIR) is an image retrieval system that aims to find similar images within a repository based on their visual content. The fundamental concept underlying an image database system is to extract and represent images using computable characteristics known as features, which are represented as vectors. These features serve as a compact and meaningful representation of the image content, enabling various applications such as content-based image retrieval, image processing, image classification, and object detection. By extracting and utilizing these features, image database systems enable efficient and effective management, organization, and retrieval of images based on their visual properties. CBIR employs image descriptors, which are code representations that capture distinctive characteristics of the images. These descriptors encompass various features, which are information or properties essential for conducting computations relevant to the application. The usual implementation of a content-based image retrieval system is depicted in Figure 1.

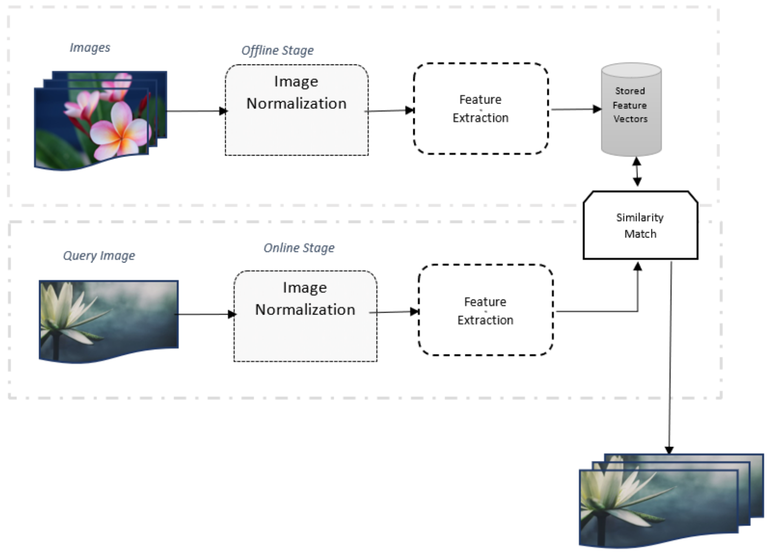

Fig 1. General Frame work of CBIR System

The development of image retrieval systems encompasses three fundamental processes that revolve around image analysis. The initial and pivotal step is image processing, where the system discerns and defines computable characteristics of images, termed features. These features encapsulate significant aspects of the images, including color, texture, shape, and other relevant visual information. Subsequently, the system captures and represents each image's features using feature descriptors, which are subsequently stored in a dedicated database. Once the database of features is created next step called online phase were user inputs image query for retrieval of similar images. The system utilizes pre-established feature extraction techniques that were developed during the offline phase, were system undergoes training to recognize and extract meaningful features from images, allowing for effective extraction of similar features from the query image during online retrieval. The system conducts a comparison between the extracted features of the query image and the stored features in the database. This comparison is commonly performed using distance metrics, such as Euclidean distance or Cosine similarity. Consequently, images possessing the highest similarity scores with the query image are retrieved from the database and presented to the user as the most relevant matches. The ranking process is crucial in presenting relevant images matching the query's content. Top-ranked images should exhibit visual appearance, context, and semantics similarities, ensuring accurate and satisfactory search results. Proper ranking enhances the user experience by quickly delivering relevant images, eliminating the need to sift through irrelevant ones.

Various datasets used in evaluating the CBIR systems include: Oxford 5k and 105k Datasets, CIFAR-10 and CIFAR-100, Caltech-101 and Caltech-256, MIRFLICKR-1M Dataset, ImageNet Dataset and INRIA Holidays Dataset.

2 Methodology

2.1. Global Features

Image analysis features refer to computable properties that are utilized to identify certain aspects of image. They are typically represented as vectors, allowing for efficient processing and analysis. Images, being stored as digital data consisting of pixel values, contain patterns and regularities that can be quantified for various identification and classification tasks. Features are categorized into two groups: global features and local features. Global features capture visual characteristics that span the entire image, such as color characteristics, shape information, and texture characteristics. Numerous methodologies have been devised to extract these global features to create an efficient representation of images. Color is an important visual characteristics recognition task and various techniques have been developed to extract color information for measuring the similarity content. Color histograms quantifies the distribution of color values and their intensities across the image and distributions measure similarity between image[9]. Color moment method relies on the moments of probabilistic of colors where unique moments of distribution are compared for similarity measure with advantage of being sale and rotation invariant[9]. Color Moments takes its assumption form central limit theory where each distribution is represented its moments (Means and Standard Deviations). Texture is the surface related property of objects; from computational point of view, it measures the intensity arranges in local regions but from structured perspective it measures the spatial relation of pixel intensity of local region. Texture proper define surface smoothness or roughness and patterns associated with it. Popular methods for texture analysis are: Gabor filter which analysis the frequencies and orientation in local regions to extract texture pattern in images and is argued to be similar to human vision system in its function[10], Markov random field uses probabilistic models to study the spatial correlation among neighbouring pixels[11], edge histogram method focuses on the edges evaluating edge distribution, orientation, width etc.[12]. Edge histogram descriptors capture the spatial relation in edge distribution and is useful in non-homogeneous texture. Shape is also very important feature of object and global visual characteristics for recognition of object. Shape is the outline of objects and is crucial for identification. Shape extraction can be grouped in to categories i.e., region-based scheme or contour representation approach[13]. Region-based scheme focuses interior points while contour makes used of shape boundary points. Various shape descriptors listed in Figure 2.

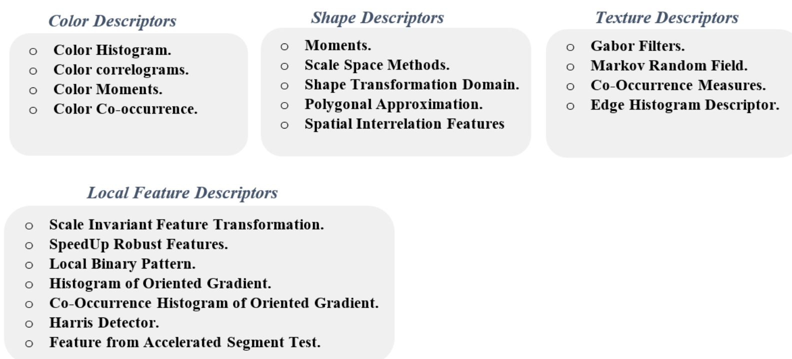

Fig 2. Feature Extraction Methods

2.2. Local Features

Local features are characterized by their distinctiveness within the image's neighbourhood and are computed over small regions. These features are crucial for achieving invariance to geometric and luminosity variations. A highly robust method to derive local features developed is Scale invariant feature transform (SIFT) which uses nearest neighbour algorithm and Hough transform to group features of object[14]. The Scale invariant key-points method was developed for feature invariance, while scale invariance depends on minimizing the mismatch of key points. The use of exposure feature and the hamming distance method improved similarity match efficiency. The text also discusses the use of Hough Transform methods to address clutter and occlusion problems, and the proposed use of Orthogonal Polynomials Transform for shorter length descriptors with simple computation for color images[15]. The use of multifeatured descriptors is more efficient in image retrieval systems, as demonstrated by a study on standard image datasets, which analysed local and global feature performances in terms of computational complexity, memory requirements, and accuracy[16]. Another local feature extraction technique is the Speed Up Robust Features (SURF), which is a distinctive and robust key point detector and feature vector feature descriptor that is fast and efficient due to its use of the Hessian matrix approximation for interest key point identification and the Image integral representation for fast computations of feature vectors. SURF is improved through the fusion of color histogram and Haar wavelet to construct feature vectors, resulting in an increase in the distinctiveness of descriptors and correct matches[17]. Various techniques used to minimize effect of noise on local features, features are combines across both special and frequency domain to improve the accuracy of similarity match in image retrieval systems. Normalized color histograms and Gabor filters are used for combining features in spatial and frequency domains[18]. Genetic optimization is used for weight optimization of features, which showed promising results[19]. Spatial resolution is improved by using Diagonal Texture Structure Descriptor, which derives feature vectors from hue, saturation, and texture properties. The text also introduces the concept of textons, which are distinguishable visual components of images studied to extract semantics. Smaller receptive fields increase the probability of getting patterns[20]. A new feature descriptor is proposed based on the relation between color, intensity, texture, and local special sensitivity, using histogram methods. The standard Euclidean distance is used for similarity match[21]. A model was proposed based on visual words dictionary-based scheme uses descriptors from SURF in combination with fast retina key points (FREAK) descriptors to form content-based image retrieval system to take advantage of computational efficiency of SURF and discriminative power of FREAK for classification [22]. Simplistic computational scheme for feature extraction of LBP has been experimented for image retrieval systems to improve the semantic analysis [23]. The above discussed methods are handcrafted and are specific to the domain which may not perform well for other image domains.

2.3. Machine Learning Methods

Machine learning methods have emerged as potent tools for uncovering patterns and relationships within data. They have garnered significant attention from researchers across various domains, particularly within the content-based image retrieval community. In this field, machine learning techniques are employed to extract features from images, surpassing conventional technologies in their ability to comprehend the semantic content of images.

Research efforts are focused on the development of image querying systems using diverse approaches in machine learning, including supervised learning, unsupervised learning, and reinforcement learning. These methodologies aim to enhance the performance and capabilities of image retrieval systems by leveraging advanced techniques in the field of machine learning. In the domain of Content-Based Image Retrieval (CBIR), extensive research has been conducted to explore and assess a wide range of machine learning algorithms [24,25]. Classifier-based methods combined with relevance feedback have been extensively studied, with promising results demonstrated by the generalizability of Support Vector Machine (SVM) classifiers. However, the performance of SVM classifiers can be compromised by small training datasets, imbalanced positive and negative feedback, and high-dimensional feature sets. To address these challenges, an approach involving asymmetric bagging methods has been proposed [26]. Additionally, decision tree-based strategies have been introduced to learn semantic information from features in SVM-based methods that require feedback for semantic learning. Decision tree methods have garnered attention for their performance in classification and semantic learning [27]. Supervised requires labelled data which is not always the case, with such data unsupervised machine learning algorithms are used for identifying of the similar images based on clusters having same type of features.

In unsupervised scenarios, the well-known Fuzzy C-Mean Clustering method is employed [28]. Inspired by natural language processing, a novel approach called Bag-of-words is used for image classification. This approach generates a dictionary list and employs machine learning techniques to compute relationships among words to predict semantics in sentences. The Bag-of-words model can also be applied to computer vision problems. Feature descriptors are extracted and converted into a dictionary of codewords using the Bag-of-words model, and k-means clustering is utilized to group similar visual words. For classification, SVM classifiers are trained on feature maps derived using the Hellinger kernel function [29].

The curse of dimensionality poses challenges for machine learning algorithms with high-dimensional data, resulting in reduced performance and increased computational cost. To mitigate these challenges, hash methods have been explored for dimensional reduction, producing compact codes for efficient retrieval. Hash-based methods have gained attention in image retrieval systems. Recent research focuses on reinforcement learning approaches to develop hash-based strategies, using policy gradient to generate binary hash codes and rewards based on similarity preservation [30]. The network architecture includes a convolutional neural network for feature extraction and a fully connected layer for binary representation. Another approach, Deep Reinforcement Hashing with Redundancy aims to minimize redundancy in feature vectors through hash codes for inference [31]. Convolutional neural models are also explored for creating a hash to improve the inference time [32].

2.4 Deep Learning Methods

Deep neural networks have revolutionized machine learning, making a significant impact across various domains. In particular, deep learning networks have been widely applied in financial analysis, natural language modelling and computer vision. In the field of computer vision, substantial progress has been made in tasks such as classification, object detection and segmentation [33]. Convolutional Neural Networks (CNNs) have emerged as the state-of-the-art approach for object classification, identification, and segmentation in computer vision. CNN networks consist of multiple layers, including convolutional, pooling, and fully connected layers, which work together to extract features from input data. The convolutional layers perform localized operations by convolving filters across the input, capturing spatial information and extracting relevant features. CNN networks are typically structured in a sequential manner, with the initial layers dedicated to extracting basic features such as edges and corners. As the network progresses deeper, subsequent layers are trained to recognize more advanced features by combining the previously extracted lower-level features. This hierarchical approach enables the network to identify complex structures and objects in the input data as is shown in the Figure 3. CNN networks commonly employ techniques such as pooling and down sampling to reduce spatial dimensions, extract significant features, and improve the network's robustness, efficiency and classification capabilities.

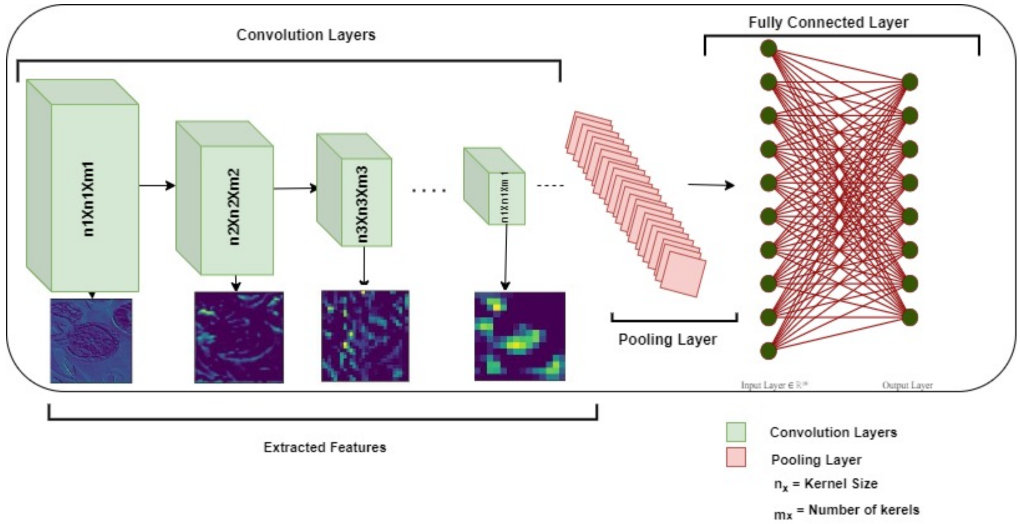

Fig 3. Architecture of Convolutional Neural Network

The effectiveness of deep convolutional networks lies in their ability to learn discriminative features, which can be leveraged to enhance the resilience and effectiveness of content-based image retrieval models. One method involves using a customized CNN architecture for content-based image retrieval, where two CNNs are employed consecutively to extract features. The extraction process operates at multiple levels to capture a more meaningful semantic representation of the image. To create concise descriptions from the high-dimensional features, the bilinear root pooling method is applied. This automatic feature extraction system has shown promising results on benchmark datasets[34]. Another approach to image retrieval involves utilizing the convolutions to extract intermediate features in CNNs and incorporates relevance information using feedback approaches, to enhance the image retrieval task by fine tuning the network on conceptual information provided by feedback. This model is specifically designed to cater to devices with limited computational capabilities[35]. Attention method enhances the ability learn more discriminative features for image representation[36]. Attention focuses on specific parts of an image which are most important for the identification. Researcher have also taken the strategy of fusion of low-level feature with high level to create a better feature representation and improve the efficiency of retrieval methods[37]. Social networks have revolutionized the way we connect and share information, and their potential extends beyond personal interactions. With the advent of Content-Based Image Retrieval (CBIR) technology, social networks can now be utilized as powerful platforms for visual search and discovery. By leveraging the vast amount of user-generated visual content, social networks can enhance CBIR algorithms by providing an extensive and diverse dataset for training and testing[38]. Users can upload and tag images, allowing the CBIR system to learn from the collective intelligence of the online community. Furthermore, social networks facilitate the formation of interest-based communities, enabling users to discover visually similar images and explore related content that aligns with their preferences. This integration of social networks and CBIR opens up exciting possibilities for visual search, recommendation systems, and personalized content curation, fostering a more engaging and interactive online experience[39].

3 Similarity Metrics

The concept of similarity metrics is the task of determining whether images have the same visual content is important and is achieved through similarity measures. The process involves feature extraction and generating descriptors. The concept of similarity needs to be defined, whether it is based on near-identical characteristics or specific comparability. The similarity function compares the feature sets of images and returns a result, which reflects the difference between the two sets. A smaller value indicates comparable images, while a larger value implies dissimilar images.

Geometric distance metrics are measures that are used to calculate the distance between two geometric objects. Some of the commonly used geometric distance metrics include:

3.1. Euclidean distance

This is the most commonly used distance metric. It is the straight-line distance between two points in a Euclidean space. Euclidean distance between two vectors is given by equation I.

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad \text{p, q are vectors} \tag{I}$$

3.2. Manhattan distance

Also known as taxicab distance, this metric measures the distance between two points as the sum of the absolute differences of their coordinates. Manhattan distance between vectors is given by equation II.

$$d(p,q) = ||p-q|| = \sum_{i=1}^{n} |p_i - q_i| \quad \text{p, q are vectors} \tag{II}$$

3.2 Minkowski distance

This is a generalization of both the Euclidean and Manhattan distance metrics. It is defined as the path root of the sum of the absolute differences of the coordinates of the two points raised to the power of m. Minkowski distance is calculated by equation III

$$d(p,q) = \left(\sum_{i=1}^{n} |p_i - q_i|^p\right)^{\frac{1}{m}}, \quad \text{p, q are vectors} \tag{III}$$

Statistical distance measures are a type of approach used for comparing the similarity of objects. They are defined on statistical objects such as variables or probability distribution samples. Unlike geometric distance metrics, statistical distance measures do not directly compute distances between objects. Instead, they measure the divergence between distributions. These methods are useful in cases where geometric distance metrics may not be appropriate or where statistical properties of objects are more important than their geometric properties.

3.3 Mahalanobis Distance

This metric takes into account the covariance structure of the data and is useful when dealing with high-dimensional data. Mahalanobis distance is calculated equation IV.

$$d(p,q) = ||\mathrm{p}-\mathrm{q}|| = \sum_{i=1}^{n} |p_i - q_i| \quad \text{p, q are vectors} \tag{IV}$$

3.4 Kullback–Leibler divergence

Kullback-Leibler distance is defined between two probability distribution. It estimates the difference between distributions. Equation V describes Kullback–Leibler divergence.

$$\mathrm{DKL(P||Q)} = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right) \quad P,\ Q\ are\ distributions,\ x = probability\ space. \tag{V}$$

3.5 Jaccard Index

Jaccard similarity metrics is defined on sets and computes the differences between sample sets. It is defined as intersection over union as given below. Jaccard index is given by equation VI.

$$\mathrm{J(A,B)} = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad A,\ B\ are\ sample\ sets. \tag{VI}$$

4 Performance Metrics

Performance metrics are used to evaluate the effectiveness of a content-based image retrieval system. There are several metrics that are commonly used to assess the performance of such systems and are defined below:

4.1. Precision (P)

Precision defined as the ratio of retrieved relevant images to the total number of retrieved images. It measures the accuracy of the system in returning relevant images. Precision is given by equation VII.

$$\mathrm{P} = \frac{\text{Number of relevent images retrieved}}{\text{Total number of images retrieved}} \tag{VII}$$

4.2 Average Precision (AP)

AP is the average value of precision over all recall values. It measures how well the system ranks the relevant images. The average precision is computed by equation VIII.

$$\mathrm{AP} = \sum_{m=0}^{n-1} [\mathrm{Recall}(m) - \mathrm{Recall}(m+1)] * \mathrm{Precession}(m) \quad n = number\ of\ thresh\ hold \tag{VIII}$$

4.3 Mean Average Precision (MAP)

MAP is the mean of the AP values over all queries. It provides a summary of the overall performance of the system. Mean average is computer by equation IX.

$$\mathrm{MAP} = \frac{1}{n} \sum_{m=1}^{n} \text{Average Precision of class } K \quad n = number\ of\ classes. \tag{IX}$$

4.4 Recall (R)

Recall is the ratio of retrieved relevant images to the total number of relevant images. It measures the completeness of the system in retrieving all relevant images. Recall is calculated by equation X.

$$\mathbf{R} = \frac{\text{Number of revelent images retrieved}}{\text{Total number of relevent images in the data set}} \tag{X}$$

4.5 F Score

F score is the harmonic mean of precision and recall. It provides a single metric that combines both precision and recall. F score is computed by equation XI.

$$F_\beta = \left(1 + \beta^2\right) \cdot \frac{Precision\ .\ Recall}{\beta^2\ .\ Precesion + Recall} \tag{XI}$$

It is important to note that no single metric can capture the overall performance of a content-based image retrieval system. Therefore, it is recommended to evaluate the system against multiple metrics to assess its overall robustness.

5 Conclusion

In conclusion, this review paper sheds light on the significant challenges encountered by deep learning in content-based image retrieval (CBIR). The heavy reliance on large labeled datasets poses efficiency concerns, particularly in domains where acquiring such datasets is arduous. Additionally, scalability issues in CBIR systems limit their ability to perform accurate retrieval across diverse image datasets in various domains, impacting their practical applicability. The importance of interpreting deep learning models cannot be overstated, especially in critical fields such as healthcare and finance, where decisions based on model outputs carry significant consequences. Understanding the internal workings of these models is vital for ensuring their trustworthiness and facilitating the analysis of their decision-making processes. As we look ahead, the future of CBIR hinges on the development of robust algorithms that tackle the challenges of huge datasets and model interpretability. Self-supervised learning, transfer learning, and interpretability techniques emerge as promising avenues to overcome these hurdles. By embracing these innovative approaches, we can enhance the efficiency and overall performance of CBIR systems, making them more adaptable, dependable, and well-suited to handle diverse image collections and real-world scenarios. These advancements hold the potential to revolutionize CBIR and expand its applications across various industries, enabling more effective and accurate image retrieval solutions.

References

1) Kumar RB, Marikkannu P. An Efficient Content Based Image Retrieval using an Optimized Neural Network for Medical Application. *Multimedia Tools and Applications*. 2020;79(31-32):22277–22292. Available from: https://doi.org/10.1007/s11042-020-08953-z.
2) Silva W, Gonçalves T, Härmä K, Schröder E, Obmann VC, Barroso MC, et al. Computer-aided diagnosis through medical image retrieval in radiology. *Scientific Reports*. 2022;12(1):20732. Available from: https://doi.org/10.1038/s41598-022-25027-2.
3) Babbar J, Rathee N. Satellite Image Analysis: A Review. *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. 2019;p. 1–6. Available from: https://doi.org/10.1109/ICECCT.2019.8869481.
4) Wu Z, Jiang S, Zhou X, Wang Y, Zuo Y, Wu Z, et al. Application of image retrieval based on convolutional neural networks and Hu invariant moment algorithm in computer telecommunications. *Computer Communications*. 2020;150:729–738. Available from: https://doi.org/10.1016/j.comcom.2019.11.053.
5) Pondenkandath V, Alberti M, Eichenberger N, Ingold R, Liwicki M. Cross-Depicted Historical Motif Categorization and Retrieval with Deep Learning. *Journal of Imaging*. 2020;6(7):71–71. Available from: https://doi.org/10.3390/jimaging6070071.
6) Guoping Q. Challenges and opportunities of image and video retrieval. *Frontiers in Imaging* . 2022. Available from: https://doi.org/10.3389/fimag.2022.951934.
7) Devareddi RB, Srikrishna A. Review on Content-based Image Retrieval Models for Efficient Feature Extraction for Data Analysis. In: 2022 International Conference on Electronics and Renewable Systems (ICEARS). IEEE. 2022;p. 969–980. Available from: https://doi.org/10.1109/ICEARS53579.2022.9752281.
8) Ibtihaal M, Hameed, Sadiq H, Abdulhussain, Basheera M, Mahmmod DT, et al. Content-based image retrieval: A review of recent trends. . Available from: https://doi.org/10.1080/23311916.2021.1927469.
9) Liu GH, Yang JY. Content-based image retrieval using color difference histogram. *Pattern Recognition*. 2013;46(1):188–198. Available from: https://doi.org/10.1016/j.patcog.2012.06.001.
10) Barbu T. Content-Based Image Retrieval Using Gabor Filtering. *2009 20th International Workshop on Database and Expert Systems Application*. 2009;p. 236–240. Available from: https://doi.org/10.1109/DEXA.2009.61.
11) Chan YK, Tsai MH, Wang JS, Guo SW, Wu JL. Color-texture-based image retrieval system using Gaussian Markov random field model. *Mathematical Problems in Engineering*. 2009;p. 1–17. Available from: https://doi.org/10.1155/2009/410243.
12) Won CSW, Park DKP, Park SJPK. Efficient Use of MPEG-7 Edge Histogram Descriptor. *ETRI Journal*. 2002;24(1):23–30. Available from: https://doi.org/10.4218/etrij.02.0102.0103.
13) Zhang D, Lu G. Review of shape representation and description techniques. *Pattern Recognition*. 2004;37(1):1–19. Available from: https://doi.org/10.1016/j.patcog.2003.07.008.
14) Lowe DG. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*. 2004;60(2):91–110. Available from: https://doi.org/10.1023/B:VISI.0000029664.99615.94.
15) Kalpana J, Krishnamoorthi R. Color image retrieval technique with local features based on orthogonal polynomials model and SIFT. *Multimedia Tools and Applications*. 2016;75(1):49–69. Available from: https://doi.org/10.1007/s11042-014-2262-1.
16) Alzu'bi A, Amira A, Ramzan N, Jaber T. Improving content-based image retrieval with compact global and local multi-features. *International Journal of Multimedia Information Retrieval*. 2016;5(4):237–253. Available from: https://doi.org/10.1007/s13735-016-0109-4.
17) Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*. 2008;110(3):346–359. Available from: https://doi.org/10.1016/j.cviu.2007.09.014.
18) Bani NT, Fekri-Ershad S. Content-based image retrieval based on combination of texture and colour information extracted in spatial and frequency domains. *The Electronic Library*. 2019;37(4):650–666. Available from: https://doi.org/10.1108/EL-03-2019-0067.
19) Phadikar BS, Phadikar AS, Maity GK. Content-based image retrieval in DCT compressed domain with MPEG-7 edge descriptor and genetic algorithm. *Pattern Analysis and Applications*. 2018;21(2):469–489. Available from: https://doi.org/10.1007/s10044-016-0589-0.
20) Song W, Zhang Y, Liu F, Chai Z, Ding F, Qian X, et al. Taking advantage of multi-regions-based diagonal texture structure descriptor for image retrieval. *Expert Systems with Applications*. 2018;96:347–357. Available from: https://doi.org/10.1016/j.eswa.2017.12.006.
21) Raza A, Dawood H, Dawood H, Shabbir S, Mehboob R, Banjar A. Correlated Primary Visual Texton Histogram Features for Content Base Image Retrieval. *IEEE Access*. 2018;6:46595–46616. Available from: https://doi.org/10.1109/ACCESS.2018.2866091.
22) Jabeen S, Mehmood Z, Mahmood T, Saba T, Rehman A, Mahmood MT. An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model. *PLOS ONE*. 2018;13(4):e0194526. Available from: https://doi.org/10.1371/journal.pone.0194526.
23) Srivastava P, Khare A. Utilizing multiscale local binary pattern for content-based image retrieval. *Multimedia Tools and Applications*. 2018;77(10):12377–12403. Available from: https://doi.org/10.1007/s11042-017-4894-4.
24) Chen L, Li S, Bai Q, Yang J, Jiang S, Miao Y. Review of Image Classification Algorithms Based on Convolutional Neural Networks. *Remote Sensing*. 2021;13(22):4712. Available from: https://doi.org/10.3390/rs13224712.
25) Oerlemans A, Guo Y, Lew MS, Chua TS. Special issue on deep learning in image and video retrieval. *International Journal of Multimedia Information Retrieval*. 2020;9(2):61–62. Available from: https://doi.org/10.1007/s13735-020-00194-y.
26) Tao D, Tang X, Wu X. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006;28(7):1088–1099. Available from: https://doi.org/10.1109/TPAMI.2006.134.
27) Liu Y, Zhang D, Lu G. Region-based image retrieval with high-level semantics using decision tree learning. *Pattern Recognition*. 2008;41(8):2554–2570. Available from: https://doi.org/10.1016/j.patcog.2007.12.003.
28) Rajam IF, Valli S. Content-based image retrieval using a quick SVM-binary decision tree - QSVMBDT. *Communications in Computer and Information Science*. 2011;p. 11–22. Available from: https://doi.org/10.1007/978-3-642-24055-3_2.
29) Yousuf M, Mehmood Z, Habib HA, Mahmood T, Saba T, Rehman A, et al. A Novel Technique Based on Visual Words Fusion Analysis of Sparse Features for Effective Content-Based Image Retrieval. *Mathematical Problems in Engineering*. 2018;2018:1–13. Available from: https://doi.org/10.1155/2018/2134395.
30) Yuan X, Ren L, Lu J, Zhou J. Relaxation-Free Deep Hashing via Policy Gradient. *Computer Vision – ECCV 2018*. 2018;p. 141–157. Available from: https://doi.org/10.1007/978-3-030-01225-0_9.
31) Yang J, Zhang Y, Feng R, Zhang T, Fan W. Deep reinforcement hashing with redundancy elimination for effective image retrieval. *Pattern Recognition*. 2020;100:107116. Available from: https://doi.org/10.1016/j.patcog.2019.107116.

32) Yang W, Wang L, Cheng S. Deep parameter-free attention hashing for image retrieval. *Scientific Reports*. 2022;12(1):7082. Available from: https://doi.org/10.1038/s41598-022-11217-5.
33) Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*. 2017;60(6):84–90. Available from: https://doi.org/10.1145/3065386.
34) Alzu'bi A, Amira A, Ramzan N. Content-based image retrieval with compact deep convolutional features. *Neurocomputing*. 2017;249:95–105. Available from: https://www.researchgate.net/publication/315968890_Content-Based_Image_Retrieval_with_Compact_Deep_Convolutional_Features.
35) Tzelepi M, Tefas A. Deep convolutional learning for Content Based Image Retrieval. *Neurocomputing*. 2018;275:2467–2478. Available from: https://doi.org/10.1016/j.neucom.2017.11.022.
36) Gao L, Wu Y, Yang T, Zhang X, Zeng Z, Chan CKD, et al. Research on Image Classification and Retrieval Using Deep Learning with Attention Mechanism on Diaspora Chinese Architectural Heritage in Jiangmen, China. *Buildings*. 2023;13(2):275. Available from: https://doi.org/10.3390/buildings13020275.
37) Lu X, Zhang L, Niu L, Chen Q, Wang JQ. A Novel Adaptive Feature Fusion Strategy for Image Retrieval. *Entropy*. 2021;23(12):1670. Available from: https://doi.org/10.3390/e23121670.
38) Rashid Y, Bhat JI. Unlocking the Power of Social Networks with Community Detection Techniques for Isolated and Overlapped Communities: A Review. *Indian Journal Of Science And Technology*. 2023;16(25):1857–1871. Available from: https://doi.org/10.17485/IJST/v16i25.841.
39) Rashid Y, Bhat JI. Topological to deep learning era for identifying influencers in online social networks :a systematic review. *Multimedia Tools and Applications*. 2023. Available from: https://doi.org/10.1007/s11042-023-16002-8.