

**CONTENT BASED IMAGE RETRIEVAL SYSTEM USING DEEP LEARNING CNN
WITH RESNET50 BACKBONE**

Kishor Rajendrakumar Shinde¹, Nilam Nimraj Ghuge², Alok Agarwal³

¹Research Scholar, Shri Jagdishprasad Jhabarmal Tibrewala University, Churu Vidyanagari,
Rajasthan

Email: kshinde3@gmail.com

²Professor, JSPM's Bhivarabai Sawant Institute of Technology & Research

Email: ghuge1974@gmail.com

³Professor, Shri Jagdishprasad Jhabarmal Tibrewala University, Churu Vidyanagari,
Rajasthan

Email: alokagarwal26aaa@gmail.com

Abstract:

Introduction: Content-Based Image Retrieval (CBIR) systems aim to retrieve relevant images from large databases based on visual content rather than metadata. Traditional CBIR techniques often struggle with accuracy due to limited feature representation. Recent advances in deep learning offer new opportunities for extracting high-level features. This work explores the integration of deep neural networks into CBIR. Specifically, it utilizes the ResNet50 architecture for enhanced image understanding and retrieval.

Objectives: The primary objective is to develop an efficient and accurate CBIR system using deep learning. It aims to leverage the ResNet50 model to extract meaningful image features. The system is designed to improve retrieval accuracy over traditional methods. It seeks to handle diverse image content with flexibility. The goal is to display the top 10 visually similar images for any given query image.

Methods: A pre-trained ResNet50 model is used to extract high-level features from a curated dataset of 1,000 images. These features are stored for comparison during retrieval. When a user submits a query image, features are extracted and compared using cosine similarity. This approach combines deep feature extraction with similarity-based ranking for effective image retrieval.

Results: The proposed system integrates Deep Learning using CNN and ResNet50 with cosine similarity to perform efficient image retrieval. The model accurately extracts and compares features, ensuring meaningful retrieval results. Figures 3–6 demonstrate sample trials, showcasing the system's ability to retrieve and rank images based on cosine similarity values.

Conclusions: This research presents a CBIR system enhanced with deep learning using the ResNet50 model for accurate image retrieval. When a user submits a query image, the system retrieves and ranks the top 10 most similar images using cosine similarity. By leveraging deep learning, the system captures complex visual features, offering improved accuracy over traditional CBIR methods.

Keywords: Content based Image retrieval, CNN, Cosine Similarity, Deep Learning, ResNet50.

1. Introduction

“Content Based Image Retrieval” (CBIR) is a method that allows users to search for images based on their visual content rather than textual descriptions [1]. This approach is particularly useful when dealing with large image databases where manual annotation is impractical or impossible. Key Components of a CBIR System include:

Feature Extraction:

a. **Low level features:** Colour, texture and figure. b. **High level features:** Semantic and contextual features.

- **Feature Representation:** Converting extracted features into a suitable format for comparison and retrieval [2].
- **Similarity Measurement:** Defining a metric to compare the similarity between query and database images [3].
- **Image Retrieval:** Ranking and retrieving images based on their resemblance to the query image.

The primary problem in ‘Content Based Image Retrieval’ is the semantic gap. This refers to the discrepancy between low level visual features (e.g., colour, texture, shape) extracted from images and high level semantic concepts (e.g., "cat," "sunset," "landscape") that humans associate with them [4]. Despite significant advancements in CBIR, the semantic gap remains a major challenge [5]. A CBIR system that can accurately bridge this gap would revolutionize various fields, including:

- **Medical Imaging:** Precise image retrieval could aid in diagnosis and treatment planning [6].
- **E-commerce:** Improved visual search capabilities can enhance online shopping experiences.
- **Law Enforcement:** Efficient image search can accelerate investigations and identify suspects.
- **Digital Libraries:** Enhanced image search can facilitate research and knowledge discovery.

To address the semantic gap and improve CBIR performance, researchers are exploring the following avenues:

- A. **Advanced Feature Extraction Techniques:** Deep Learning and Hybrid Approaches.
- B. **Semantic Similarity Measures:** Semantic Embedding’s and Cross-Modal Learning.
- C. **Interactive CBIR:** User Feedback and Active Learning.
- D. **Large-Scale CBIR:** Efficient Indexing Techniques and Distributed Computing.

Objectives

By tackling these research challenges, we can design more robust and effective CBIR systems that better meet the needs of users in various domains. The paper is organised as literature survey in section II, proposed system along with its architecture is presented in section III while section IV highlights the results which are further discussed on the basis of similarity and ranking of retrieved image.

2. Methods

Proposed Methodology

In today's digital age, the rapid growth of online images has made it increasingly challenging to efficiently search for visually similar images. Traditional image retrieval systems, such as those relying on SIFT [26] and BoVW [27] features, often struggle to handle large-scale image databases and complex visual patterns. Increased image retrieval performance has been caused by advancements in deep learning, especially in convolutional neural networks (CNNs). CNNs are trained on large datasets using supervised learning techniques. This training process encompasses adjusting constraints like learning rate, loss function, and the number of epochs to adjust the model's ability to extract meaningful features from images. By learning from vast amounts of labelled data, CNNs can effectively capture complex visual patterns and relationships, leading to accurate image classification and retrieval [28].

Neural networks typically consist of input, hidden, and output layers as shown in figure 1 [29]. Hidden layers, which can be multiple in number, play a vital role in haul out meaningful features from input data. In the case of image data, these hidden layers perform operations like filtering, edge detection, colour extraction, and texture recognition. By stacking multiple layers, Deep Convolutional Neural Networks (DCNNs) can learn increasingly complex representations of images. These DCNNs are powerful feature extractors, capable of capturing subtle visual patterns that traditional methods might miss. These learned features can be used to compare query and repository images for efficient image retrieval.

Deep neural networks, while powerful, face a significant challenge known as the "vanishing gradient problem." As the network depth increases, gradients flowing back through the layers can become extremely small, making it hard for the network to learn effectually. This can lead to degraded performance, even with deeper architectures. Residual Neural Networks (ResNets) were a ground-breaking innovation to address this issue. They introduce "skip connections" that allow information to bypass intermediate layers and directly reach deeper layers. This enables gradients to flow more easily, facilitating the training of much deeper networks. ResNets have been highly successful, demonstrating state-of-the-art performance in various computer vision tasks, including image classification as evidenced by their success in the ImageNet challenge in 2015 [30]. The architecture of the proposed system is shown in figure 2.

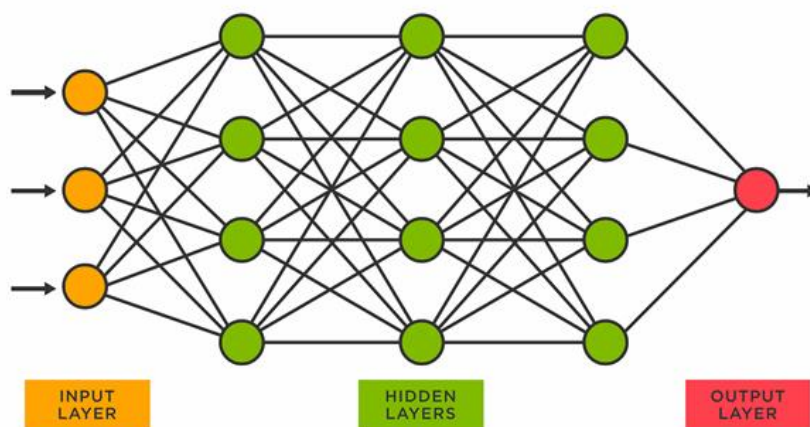


Figure 1. Structure of Convolutional neural network [29]

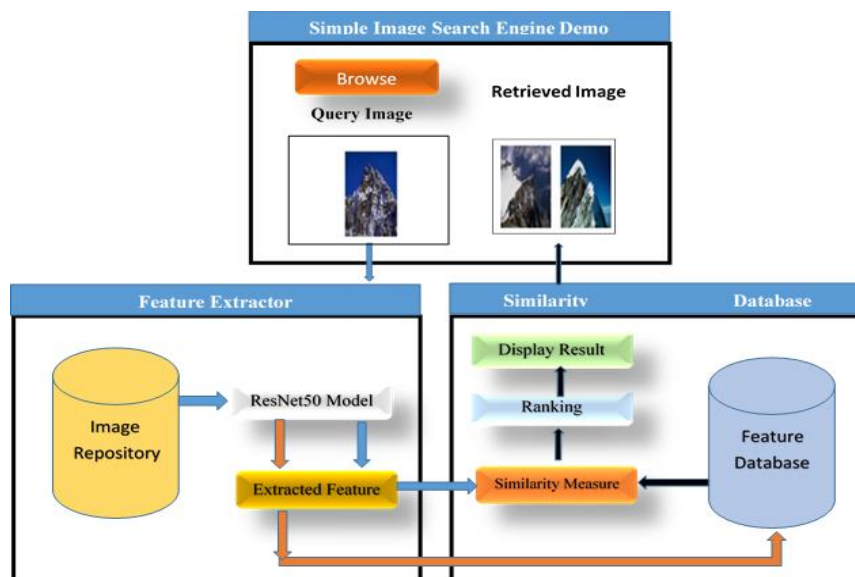


Figure 2. Architecture of proposed CBIR system

The architecture as shown in the figure is composed of the following:

- **Server** – This is the client-side module which is used for user interaction with the CBIR system. It provides the interface to the user to browse and select input query image from the repository to submit. The output of the resultant images retrieved by the system is displayed in the "simple image search engine Demo Page."
- **Training and feature extraction** – The feature extractor CNN model is built using the ResNet50 backbone and it trains images in order to extract features. The extracted features are stored in a features database and can be used in future.
- **Database** – All the features extracted from the image retrieval module are stored here.
- **Similarity** – Thus, this module compares its features with those already existing in the features database to find out the similarity of features in any query. Now based upon similarity, an image rank can be derived and finally shown on the demo image search engine page.

Feature-based similarity retrieval involves representing data objects as feature vectors and calculating similarity scores based on these vectors. Various methods that can be applied for measuring similarity are listed below;

- The Euclidean distance:** Calculates the distance between two straight line points using Euclidean space and is mathematically given as

$$\text{Euclidean distance} = \sqrt{\sum (A_i - B_i)^2} \tag{1}$$

Whereas similarity based on Euclidean distance is given as

$$\text{Similarity} = \frac{1}{1 + \sqrt{\sum (A_i - B_i)^2}} \tag{2}$$

- Manhattan Distance:** Capture the difference in the sum of the absolute differences of the elements of two vectors.

$$\text{Similarity} = \frac{1}{1 + \sum(|A_i - B_i|)}$$

(3)

C. **Jaccard Similarity:** Calculate the resemblance between two sets by dividing the size of their intersection by the size of their union.

$$\text{Similarity} = \frac{A \cap B}{A \cup B}$$

(4)

D. **Cosine Similarity:** Cosine comparison is a mathematical measure used to determine the resemblance between two non-zero vectors in a multi-dimensional space. It specifically measures the cosine of the angle between the two vectors. Cosine resemblance primarily focuses on the route of the vectors rather than their magnitude. This makes it particularly useful when dealing with sparse data, where the magnitude of the vectors can be misleading [31]. The cosine similarity score ranges from -1 to 1:

- 1: Specifies that the two vectors are perfectly aligned (same direction).
- -1: Specifies that the two vectors point in exactly opposite directions.
- 0: Specifies that the two vectors are orthogonal (perpendicular) to each other.

Cosine similarity is calculated as follows:

$$\text{Cosine similarity (A, B)} = \frac{(A \cdot B)}{(\|A\| * \|B\|)}$$

(5)

where:

A and B are the two vectors

A · B is the dot product of A and B.

||A|| and ||B|| are the magnitudes (lengths) of vectors A and B, respectively.

Let's consider two vectors:

$$A = (1, 2, 3), B = (4, 5, 6)$$

$$A \cdot B = (1 * 4) + (2 * 5) + (3 * 6) = 4 + 10 + 18 = 32$$

$$\|A\| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14} \quad ; \quad \|B\| = \sqrt{4^2 + 5^2 + 6^2} = \sqrt{77}$$

$$\text{Cosine similarity (A, B)} = \frac{32}{\sqrt{14 * 77}} \sim 0.972$$

This indicates that the two vectors are highly similar, pointing in nearly the same direction. Cosine similarity is a versatile metric for measuring the similarity between vectors, particularly in high-dimensional spaces. Its focus on direction makes it robust to variations in vector magnitudes, making it a valuable tool in various applications, including information retrieval, recommendation systems, and image analysis.

Algorithm for CBIR system

1. Train the CNN ResNet50 model with 1000 categories of images out of which 80% is used to train the model and the rest 20% to validate the results.

2. Load previously trained ResNet50 as feature extractor.
3. Input a query image A.
4. Extract features of A using feature extractor.
5. For every feature of B in feature database, obtain cosine similarity between A and B.
6. Display retrieved images on demo page in the order of similarity and ranking between query and retrieved images.

3. Results

The proposed system combines Deep learning CNN and ResNet50 with cosine similarity to retrieve images from query. The system has been modulated to produce 10 similar images based on query along with their cosine similarity and ranking. Figure 3-6 are a few sets of trials on the model which retrieves the images based on query along with the values of cosine similarity, and images are displayed as per their ranking.

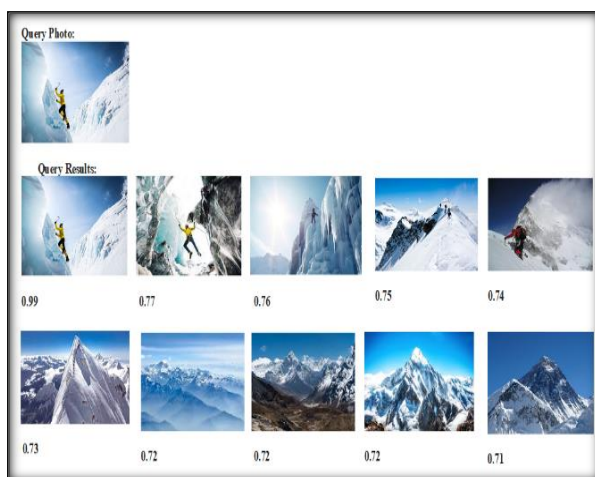


Figure 3. Query input 1

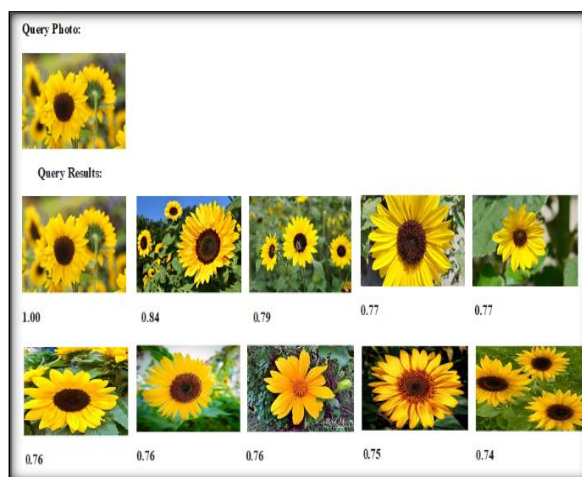


Figure 4. Query input 2

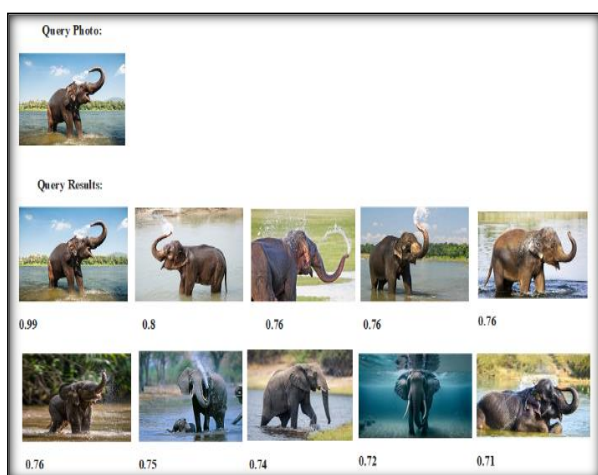


Figure 5. Query input 3



Figure 6. Query input 4

From the results above, it is clear that the proposed CBIR system is able to generate the query results per the query set by the user. It searches for identical images in the dataset and ranks them in the decreasing order of their similarity.

4. Discussion

This research introduces a Content-Based Image Retrieval (CBIR) system that has also included deep learning capability such as ResNet50 as a model for obtaining similar images. The working of this system will include training a model using 1,000 images into a database. The trained models can then feature extract a query image and match its features with those of images stored in the database. The user submits a query image to the system, extracts image features, and compares them with those of images in the database; finally, it retrieves and displays the top ten most similar images to the user. The proposed system outperforms traditional CBIR methods due to its ability to capture complex visual patterns using deep learning. This leads to more precise and relevant image retrieval system. This approach offers significant advantages over traditional CBIR methods:

- **Improved Accuracy:** The deep learning model can capture complex visual patterns, leading to more accurate image retrieval.
- **Efficient Retrieval:** The system can quickly process queries and return results in real-time.
- **Flexibility:** The model can be adapted to different datasets and image retrieval tasks.

Future Directions:

- **Hybrid Approaches:** Combining multiple feature extraction techniques, such as colour histograms and texture features, with deep learning can further enhance retrieval accuracy, especially for complex image queries.
- **Large-Scale Retrieval:** For large-scale datasets, efficient indexing and retrieval techniques, such as approximate nearest neighbour search, can be employed to accelerate the search process.
- **User-Centered Design:** Incorporating user feedback and iterative refinement of the system can improve user satisfaction and the overall effectiveness of the CBIR system.

By addressing these aspects, the proposed CBIR system can be further optimized to provide a robust and efficient solution for various image retrieval applications.

References

- [01] Hamed Qazanfari, Mohammad M. AlyanNezhadi, Zohreh Nozari Khoshdaregi, "Advancements in Content-Based Image Retrieval: A Comprehensive Survey of Relevance Feedback Techniques", *Computer vision and Pattern Recognition*, arXiv, 132-141, December 2023.
- [02] Sandra Jardim, Joao Antonio, Carlos Mora, Artur Almeida, "A Novel Trademark Image Retrieval System Based on Multi-Feature Extraction and Deep Networks", *Open Access, J. Imaging*, 8(9), 115-122, Aug 2022.
- [03] Yingnan Zheng, Zhizhong Kang, Cao Zhen, "An Image Retrieval Method for Lunar Complex Craters Integrating Visual and Depth Features", *Research gate, Electronics*, 13(7), 121-129, March 2024.

- [04] Lu. Y, Lei Zhang, Qi Tian, Wei-Yang Ma, “What are the high-level concepts with small semantic gaps?”, Research gate, Conference: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-32, June 2008.
- [05] Rong Zhao, William Grosky, “Bridging the semantic gap in Image Retrieval”, Research gate, chapter-II, Part-II, Content based Retrieval and image database techniques, 46-54, Oct 2001.
- [06] Camilo G Sotomayor, Marcelo Mendoza, et.al, “Content-Based Medical Image Retrieval and Intelligent Interactive Visual Browser for Medical Education, Research and Care”, Diagnostics (Basel), National Library of Medicine,11(8), 212-220, Aug 2021.
- [07] Shahbaz Sikandar, Rabbia Mahum, AbdulMalik Alsalman, “A Novel Hybrid Approach for a Content-Based Image Retrieval Using Feature Fusion”, Open Access, *Appl. Sci.* 13(7), 69-78, Feb 2023.
- [08] Liz Makena, “Cross-Modal Retrieval: Image-to-Text and Text-to-Image Search”, comet, Feb 2024.
- [09] Sitalakshmi Venkatraman, Siddhivinayak Kulkarni, “MapReduce neural network framework for efficient content-based image retrieval from large datasets in the cloud”, Research gate, Conference: 2012 12th International Conference on Hybrid Intelligent Systems (HIS), 149-158, December 2012.
- [10] Ji Wan, Dayong Wang, Steven Chu, Hong Hoi, Pengcheng Wu, Jintao Li, et.al, “Deep Learning for Content-Based Image Retrieval: A Comprehensive Study”, MM '14: Proceedings of the 22nd ACM international conference on Multimedia, 157-166, May 2014.
- [11] Chi Zhang, Jie Liu, “Content Based Deep Learning Image Retrieval: A Survey”, ACM, 9th International Conference on Communication and Information Processing (ICCIP), 158-163, December 2023.
- [12] Giuseppe Amato, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Lucia Vadicamo, “Large-scale instance-level image retrieval”. *Information Processing & Management* 57 (6), 107-115, November 2020.
- [13] Cong Bai, Hongkai Li, Jinglin Zhang, Ling Huang, Lu Zhang, “Unsupervised adversarial instance-level image retrieval”. *IEEE Transactions on Multimedia*, 23, 2199–2207, March 2021.
- [14] Ayan Kumar Bhunia, Yongxin Yang, Timothy M Hospedales, Tao Xiang, Yi-Zhe Song, “Sketch less for more: On-the-fly fine-grained sketch-based image retrieval”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 9779–9788, Feb 2020.
- [15] Jamil Ahmad, Khan Muhammad, Sambit Bakshi, and Sung Wook Baik. “Object-oriented convolutional features for fine-grained image retrieval in large surveillance datasets”. *Future Generation Computer Systems* 81, 314–330, April 2018.
- [16] Gang Chen, Xiang Cheng, Sen Su, Chongmo Tang, “Multiple-instance ranking based deep hashing for multi-label image retrieval”. *Neuro-computing*, 402, 89–99, Aug 2020.

- [17] Yun Ge, Zihong Yang, Zihan Huang, Famao Ye, “A multi-level feature fusion method based on pooling and similarity for HRRS image retrieval”, *Remote Sensing Letters*, 12 (11), 1090–1099, Feb 2021.
- [18] Xiaoxiao Guo, Hui Wu, Yu Cheng, Steven Rennie, Gerald Tesauro, Rogerio Feris, “Dialog-based interactive image retrieval” *Advances in neural information processing systems*, 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), 1-11, 2018.
- [19] Albert Gordo, Jon Almazan, Jerome Revaud, Diane Larlus, “End-to-end learning of deep visual representations for image retrieval”, *International Journal of Computer Vision*, 124(2), 237–254, June 2017.
- [20] Tian, D. “Support Vector Machine for Content-based Image Retrieval: A Comprehensive Overview”, *J. Inf. Hiding Multim. Signal Process.* 9 (6), 1464–1478, November 2018.
- [21] Xia, Z., Wang, X., Zhang, L., Qin, Z., Sun, X, Ren, K., “A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing”, *IEEE Transactions on Information Forensics and Security*, 11(11), 2594–2608, Nov 2016.
- [22] Wan J, Wang D, Hoi S.C.H, Wu P, Zhu J, Zhang Y, Li J, “Deep learning for content-based image retrieval: A comprehensive study”, In *Proceedings of the 22nd ACM International Conference on Multimedia*, Orlando, FL, USA, 3(7), 157-166, November 2014.
- [23] Krizhevsky A, Hinton G.E, “Using very deep auto encoders for content-based image retrieval”. In *Proceedings of the ESANN 2011, 19th European Symposium on Artificial Neural Network*, Bruges, Belgium, April 2011.
- [24] Mehmood Z, Mahmood T, Javid M.A., “Content-based image retrieval and semantic automatic image annotation based on the weighted average of triangular histograms using support vector machine”, *Springer*, 48, 166–181, June 2017.
- [25] Alsmadi M.K., “Query-sensitive similarity measure for content-based image retrieval using meta-heuristic algorithm”, *Journal of King Saud University, Computer and Information Science*, 30 (3), 373–381, July 2018.