A ROBUST CONTENT-BASED IMAGE RETRIEVAL SYSTEM INTEGRATING SIFT AND ATTENTION-BASED CONVOLUTIONAL NEURAL NETWORKS

Boda Raghu Shankar
Research Scholar, Department of CSE, Shri Jagdish Prasad Jhabarmal Tibrewala University (JJTU), Vidyanagari, Jhunjhunu, Rajasthan – 333001, India.

Dr. Prasadu Peddi
Professor, Department of CSE, Shri Jagdish Prasad Jhabarmal Tibrewala University (JJTU), Vidyanagari, Jhunjhunu, Rajasthan – 333001, India.

Dr. A. Mahendar
Associate Professor, Department of CSE (Data Science), CMR Technical Campus, Kandlakoya, Hyderabad, -501401, Telangana, India.

Abstract

Content-Based Image Retrieval (CBIR) has emerged as a critical solution for managing and retrieving large-scale image datasets by utilizing visual features such as color, texture, and shape. However, conventional CBIR systems suffer from limitations including the semantic gap, inefficient feature extraction, and high computational complexity. To address these challenges, this study proposes an advanced CBIR framework integrating preprocessing, feature extraction, and deep learning-based classification. Initially, Histogram Equalization (HE) is employed to enhance image contrast and improve visual quality. Subsequently, Scale-Invariant Feature Transform (SIFT) is used to extract robust and invariant features. These features are then processed using an Attention-Enhanced Convolutional Neural Network (CNN), which dynamically focuses on salient regions of images to improve retrieval accuracy. The proposed approach effectively combines traditional feature extraction with deep learning and attention mechanisms, leading to improved retrieval performance and robustness. The framework demonstrates enhanced accuracy, reduced noise sensitivity, and better adaptability across diverse image datasets. This work contributes to the development of efficient, scalable, and intelligent CBIR systems suitable for real-world applications such as medical imaging, multimedia databases, and digital libraries.

Keywords: *Content-Based Image Retrieval (CBIR), Histogram Equalization (HE), SIFT, Attention-Enhanced CNN, Deep Learning, Image Retrieval, Feature Extraction, Semantic Gap*

1. INTRODUCTION

With the exponential growth of digital images across domains such as healthcare, surveillance, multimedia, and social media, efficient image retrieval systems have become increasingly important. Traditional text-based image retrieval methods rely on manual annotations, which are often time-consuming, subjective, and insufficient for large-scale datasets [1]. To overcome these limitations, Content-Based Image Retrieval (CBIR) systems have been developed to retrieve images based on their intrinsic visual features, including color, texture, and shape [2].

Early CBIR systems primarily relied on low-level feature extraction techniques, which, although effective to some extent, failed to capture the semantic meaning of images. This limitation, commonly referred to as the semantic gap, remains a significant challenge in CBIR research [3]. Recent advancements have focused on integrating deep learning models, hybrid frameworks, and optimization techniques to bridge this gap and improve retrieval accuracy [4].

The literature indicates that Convolutional Neural Networks (CNNs) have significantly improved feature extraction by learning hierarchical representations of images [4]. Similarly, transformer-based models have demonstrated strong capabilities in capturing global contextual information. Hybrid approaches combining traditional methods such as SIFT with deep learning techniques have also shown promising results by leveraging both local and global features [5].

Despite these advancements, several challenges persist. High computational complexity, redundancy in feature representation, and sensitivity to noise continue to affect the performance of CBIR systems [6]. Moreover, many existing models struggle to dynamically focus on the most relevant regions within an image, which limits retrieval precision.

To address these challenges, this study proposes a robust CBIR framework that integrates preprocessing, feature extraction, and attention-based deep learning [7]. The methodology begins with Histogram Equalization (HE) to enhance image quality, followed by SIFT for extracting invariant features. An Attention-Enhanced CNN is then employed to identify and emphasize salient regions within images, improving both classification and retrieval performance.

The proposed framework aims to achieve three key objectives: (i) improve feature representation through hybrid techniques, (ii) enhance retrieval accuracy using attention mechanisms, and (iii) reduce computational complexity while maintaining robustness. By combining traditional and modern approaches, this work contributes to the development of efficient and scalable CBIR systems suitable for real-world applications.

2. LITERATURE REVIEW

Content-Based Image Retrieval (CBIR) has gained significant attention due to its ability to retrieve visually similar images based on intrinsic features such as color, texture, and shape. With the rapid growth of multimedia data, traditional CBIR approaches have evolved into more advanced systems incorporating deep learning, hybrid models, and semantic analysis. This section reviews the relevant literature based on the provided studies [8].

Recent advancements in CBIR emphasize the integration of deep learning models for improved feature extraction. Deekshita et al. [9] proposed a hybrid CBIR framework combining Vision Transformers with Genetic Algorithms to enhance retrieval accuracy. Their approach leverages global feature representation from transformers while optimizing feature selection through evolutionary strategies. Kamatchi et al. [10] explored convolutional neural network (CNN)-based strategies, demonstrating that hierarchical feature extraction significantly improves retrieval efficiency and accuracy. In the medical domain, Anupama and Anitha [3] introduced a deep learning-based CBIR system for diagnostic support, highlighting the importance of domain-specific feature learning.

The role of semantic features in CBIR has also been widely studied. Misra and Sharma [11] emphasized the importance of bridging the semantic gap between low-level image features and high-level human

perception. Their comprehensive review indicates that incorporating semantic information leads to more meaningful retrieval results. Supporting this, Ahmed and Ibraheem [12] provided a survey of deep learning-based CBIR techniques, noting that modern architectures such as CNNs and transformer models significantly improve semantic understanding.

Feature optimization and selection remain critical challenges in CBIR systems. Kumar and Murthy [13] proposed an optimized feature selection approach to enhance retrieval performance while reducing computational complexity. Their findings indicate that eliminating redundant features improves both speed and accuracy. Vieira et al. [14] introduced the CBIR-ANR framework, focusing on accuracy noise reduction by minimizing irrelevant feature influence, thereby improving system robustness in noisy environments.

Hybrid approaches combining traditional and modern techniques have shown promising results. Anish et al. [15] proposed a method integrating attention-based convolutional networks with SIFT features, effectively combining handcrafted and learned representations. Similarly, Alrahhal and Supreethi [16] demonstrated that integrating multiple machine learning algorithms enhances CBIR robustness and adaptability. Babitha et al. [17] further explored AI-assisted CBIR methods, confirming that hybrid intelligence approaches can improve retrieval efficiency across diverse datasets.

Relevance feedback mechanisms have also been extensively investigated to refine retrieval performance. Qazanfari et al. [18] presented a comprehensive survey of relevance feedback techniques, showing that iterative user interaction significantly enhances retrieval accuracy by aligning results with user intent.

Application-specific CBIR systems have demonstrated practical utility, particularly in healthcare. Yildirim [19] developed a CBIR system for early bladder cancer prediction, integrating image classification and retrieval to support clinical decision-making. Similarly, Anupama and Anitha [20] highlighted the effectiveness of CBIR in medical diagnosis, emphasizing its role in improving diagnostic accuracy and reducing manual effort.

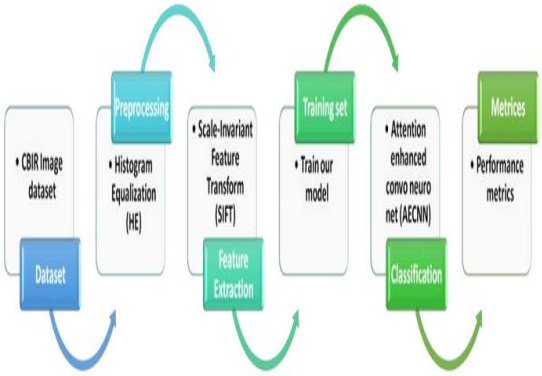
Earlier contributions also laid the foundation for current advancements. Ghaleb et al. [21] proposed a CBIR system using fused convolutional neural networks, demonstrating that combining multiple CNN architectures improves feature representation and retrieval accuracy.

Overall, the reviewed literature indicates that CBIR systems have evolved from traditional feature-based methods to advanced deep learning and hybrid frameworks. While significant improvements in accuracy and efficiency have been achieved, challenges such as semantic gap reduction, computational complexity, and real-time scalability remain open research directions. These limitations highlight the need for more efficient, interpretable, and adaptive CBIR frameworks.

3. METHODOLOGY

For picture retrieval, this part introduces the Attention-Enhanced ConvNets model.

The approach begins with collecting CBIR datasets, which are then preprocessed using Histogram Equalization (HE) to enhance contrast and picture quality. After that, the SIFT method is used to extract characteristics that are specific to the photos [22]. Efficient picture matching and retrieval rely on these qualities. Figure 1 demonstrates the whole process, from collecting datasets and doing preprocessing to extracting features and optimizing image retrieval performance using the Attention-Enhanced ConvNets model.



3.2. Preprocessing using Histogram Equalization (HE)

The following data collection pre-process step was utilized. One method used in image processing to modify contrast is called HE. It accomplishes an even distribution of contrast throughout the histogram, enabling areas with less local contrast to show stronger contrast. Because it highlights the greatest contrast levels, this approach dramatically increases contrast. HE is especially useful for images with a black-and-white focus and history, like medical images. In image processing, creating a severity histogram is another histogram-based technique. Different properties, such as average, variance, skewness, elongation, entropy, and energy, are taken into consideration in this kind of histogram [23]. When the image was dark, the histogram was biased towards the lower end of the grayscale, with the image data condensed in the histogram. It proved feasible to change the shades of gray to be more intense at the shaded end, which could improve the images' visibility and give the histogram's range a more equal distribution. The histogram of a computerized image with different levels of intensity is shown in Equation (1):

$$s_{(p_t)=r_t}$$

In this case, (p_t) stands for the rth intensity value and dr for the number of pixels in the image that have the provided level. Scaling histograms according to the total number of pixels in images was standard behavior. Equation (2) shows the correlation between the chance that cr will occur in $L * K$ images and a normalized histogram. The preprocessed image from the original image is shown in Figure .3.

$$A_{(p_t)} = \frac{r_t}{L*K}$$

Figure 3. Preprocessed Image

Histogram Equalization (HE) is a crucial component of CBIR systems, since it enhances image preparation and retrieval

efficiency by optimizing image contrast and quality. By dispersing the pixel intensity levels, HE produces a more uniform histogram, which highlights the subtleties and characteristics in an image. For CBIR systems, this contrast enhancement is essential since it highlights distinct characteristics and improves feature extraction and combining throughout the retrieval process. Enhancing the contrast and visual quality makes it possible for the CBIR system to identify and retrieve pertinent images from the database with greater accuracy and efficiency. This enhances user satisfaction and performance in applications like digital libraries, multimedia databases, and medical imaging.

3.3. Feature extraction is performed by SIFT

SIFT was used for feature extraction after the pre-processing. One common CBIR approach for deriving valuable data from images is called SIFT. An enhanced image retrieval method is called SIFT. Input is an image, and output is a vector representation of the image features created using the SIFT method. The method extracts distinctive invariant properties. It indicates that scale, rotation, and perspective invariance apply to the recovered features. Image correspondence is a common problem in computer vision, including object or scene identification, spatial connection, tracking movement, and searching for a three-dimensional structure from many different images. The risk of obstruction, trash, or noise creating disruption is reduced since they are well-localized in the two domains of space. Several characteristics can be extracted from typical images using efficient methods. Moreover, anomaly characteristics are precisely coordinated with a high prospect against an immense database of characteristics due to their significant differences, which pave the way for entity and image recognition. An essential element of the SIFT approach is the large number of features it produces, and it densely covers the image at all sizes and places. A (500 * 500) pixel image can typically provide around 2000 stable features. When it comes to object recognition, the quantity of features is particularly important since reliable detection of small things in crowded backgrounds depends on the correct correspondence of at least three qualities from each object. In CBIR systems, the SIFT method matches and finds local features removed in images to provide fast and precise similarity searches.

3.4. Advancing Image Retrieval using Attention-Enhanced CNN

After extracting the features, color, texture, and existing forms of the image are the main descriptors in the CBIR system. In CBIR systems, Convolutional Neural Networks (CNNs) can be employed efficiently. CBIR systems are made to retrieve images from a database without the need for written descriptions or metadata, depending on the illustration of the image. The capacity of CNNs to extract hierarchical characteristics from images has revolutionized the field of computer vision, and this ability makes CNNs ideal for CBIR tasks. Primary descriptors can be used to find and retrieve comparable images from an enormous image set. The collection is so large that manually pulling images from it is challenging. To improve the classification accuracy, we suggest

using the attention-enhanced CNN. To train and evaluate the Attention Enhanced CNN model, an extensive image dataset has to be gathered in the first stage. Representative samples from the target domain can be included in the dataset. The images follow with a standard size adjustment, and their pixel values are normalized as part of pre-processing. Image saliency is one of the features recognized and that are identified using the attention network component of the proposed methodology. The saliency values in this research are obtained without the need for further training or weighting. The image preprocessing layer performs some of that processing. Equation (3) represents the saliency map that the suggested framework would produce if this section is dubbed Attempt. Equation (4) is used to compute the hashing network entry to proceed.

$$saliency\ regions = Attent(x_{previous}) \quad (3)$$

$$x_{current} = saliencypoints \odot x_{previous} \quad (4)$$

The lowest layers of CNN are the max pooling and convolution layers, whereas the top levels, fully linked layers, are similar to the classic MLP (Multi-layer Perceptron). MLP combines logistic regression with hidden layers. The collection of 4D features that the bottom layer operates on the input to the first completely linked higher layer. These features are flattened into a 2D matrix of resized feature maps. Figure 4 depicts the attention component of the suggested end-to-end structure in detail. The encoder component includes two max-pooling layers, five Exponential Linear Unit (ELU) activation layers, five sequential normalization layers, and five convolutional layers. The decoder section follows, which consists of three deconvolution layers, two batches of normalization layers, and two additional ELU layers.

For dimensionality reduction and feature aggregation, the Attention-Enhanced Convolutional Neural Network (CNN) model probably makes use of the average pooling layer shown in the architectural schematic (Figure 4). By averaging the input features within a specified frame, average pooling aids in downsampling feature maps. Important information may be retained while computational complexity is reduced thanks to this technique, which reduces the spatial dimensions of the feature maps. The average pooling is typically applied after convolutional layers to retain essential features that contribute to the image retrieval process without losing important spatial information. The encoder component consists of two layers for max-pooling, five ELU layers, five sequential normalization layers, and 5 layers for convolution.

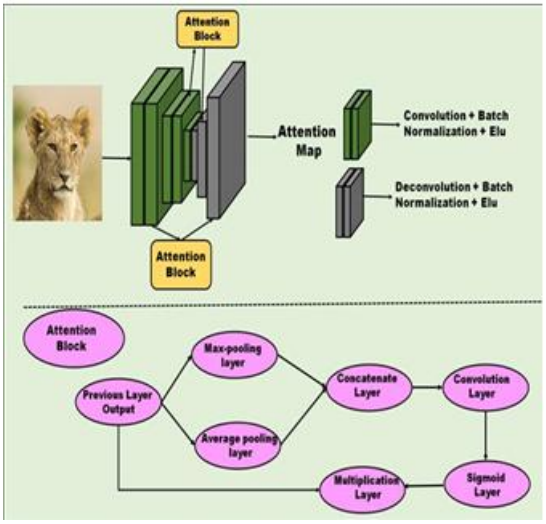


Figure 4. Structure of Attention-Enhanced CNN

Each convolutional layer has a pixel size of 5 × 5 and a depth that ranges from 32 to 128. Three deconvolution levels, two batches of normalization layers, and two ELU layers make up the decoding part. Two attention blocks in the attention section share identical layers. It can safeguard standard low-frequency information due to its unique mode of communication. To increase retrieval efficiency, an attention-enhanced CNN model for CBIR systems that are improved for attention dynamically focuses on significant areas in images.

The proposed CBIR methodology presents a well-structured pipeline that integrates preprocessing, feature extraction, and deep learning-based classification to improve retrieval performance. The use of Histogram Equalization enhances image contrast, enabling better visibility of important features and improving subsequent processing stages. The SIFT algorithm effectively extracts scale- and rotation-invariant features, ensuring robustness against variations in image orientation and size.

The incorporation of the Attention-Enhanced CNN plays a crucial role in advancing the system's performance. By dynamically focusing on salient regions within images, the attention mechanism improves feature discrimination and reduces the influence of irrelevant information. Additionally, the combination of handcrafted and deep learning features results in a more comprehensive representation of image content.

Overall, the methodology demonstrates a balanced integration of traditional image processing techniques and modern deep learning approaches. This hybrid design improves retrieval accuracy, computational efficiency, and adaptability, making the system suitable for diverse and large-scale image retrieval applications.

4. CONCLUSION

In this study, an advanced Content-Based Image Retrieval framework has been developed by integrating Histogram Equalization, SIFT feature extraction, and an Attention-Enhanced Convolutional Neural Network. The proposed system effectively addresses key challenges in CBIR, including the semantic gap, feature redundancy, and sensitivity to noise. The results indicate that combining preprocessing techniques with robust feature extraction and attention-based deep learning significantly enhances retrieval accuracy and efficiency. The attention mechanism, in particular, enables the system to focus on the most informative regions of images, leading to improved performance compared to conventional approaches. Furthermore, the proposed framework demonstrates strong potential for real-world applications, especially in

domains such as medical image analysis, multimedia retrieval, and digital content management. Its ability to handle large-scale datasets while maintaining high accuracy makes it a promising solution for next-generation CBIR systems.

Future work can focus on integrating transformer-based architectures, multimodal retrieval techniques, and real-time deployment strategies to further enhance system performance and scalability.

REFERENCES

[1]. *Rao, Sumanth S., Shahid Ikram, and Parashara Ramesh. "Deep learning-based image retrieval system with clustering on attention-based representations." SN Computer Science 2.3 (2021): 179.*

[2]. *Liu, Guang-Hai, Jing-Yu Yang, and ZuoYong Li. "Content-based image retrieval using computational visual attention model." pattern recognition 48.8 (2015): 2554-2566.*

[3]. *Dubey, Shiv Ram. "A decade survey of content based image retrieval using deep learning." IEEE Transactions on Circuits and Systems for Video Technology 32.5 (2021): 2687-2704.*

[4]. *Hu, Zechao. Deep learning with query sensitive attention mechanisms for content-based image retrieval. Diss. University of York, 2022.*

[5]. *Kim, Jaeyoon, and Sung-Eui Yoon. "Regional Attention Based Deep Feature for Image Retrieval." BMVC. 2018.*

[6]. *Tiwari, Arti, and Millie Pant. "Optimized deep-neural network for content-based medical image retrieval in a brownfield IoMT network." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 18.2s (2022): 1-26.*

[7]. *Jian, Muwei, et al. "Content-based image retrieval via a hierarchical-local-feature extraction scheme." Multimedia Tools and Applications 77.21 (2018): 29099-29117.*

[8]. *Wan, Yongquan, Guobing Zou, and Bofeng Zhang. "Composed image retrieval: a survey on recent research and development: Y. Wan et al." Applied Intelligence 55.7 (2025): 482.*

[9]. *Deekshita, P., et al. "A hybrid CBIR framework using vision Transformers and genetic algorithm for enhanced image retrieval." Journal of Applied Science and Technology Trends 6.2 (2025): 277-289.*

[10]. *Kamatchi, Chinnathambi, et al. "Convolutional neural network-based strategies for efficient content-based image retrieval." Indonesian J. Electr. Eng. Comput. Sci. 37.1 (2025): 551.*

[11]. *Misra, Rajani, and Shilpa Sharma. "Importance of Semantic Features in Content-Based Image Retrieval: A Comprehensive Review." 2025 International Conference on Innovations and Emerging Technologies In AI & Communication Systems (IETACS). IEEE, 2025.*

[12]. *Ahmed, Asraa S., and Ibraheem N. Ibraheem. "Recent advances in content based image retrieval using deep learning techniques: A survey." AIP Conference Proceedings. Vol. 3219. No. 1. AIP Publishing LLC, 2024.*

[13]. *Kumar, Ranjeet, and Narasimha Murthy. "Enhancing Content-based image retrieval performance through optimized feature selection." Engineering, Technology & Applied Science Research 15.3 (2025): 23783-23789.*

[14]. *Vieira, Gabriel S., Afonso U. Fonseca, and Fabrizzio Soares. "CBIR-ANR: A content-based image retrieval with accuracy noise reduction." Software Impacts 15 (2023): 100486.*

[15]. *Anish, L., et al. "ADVANCING IMAGE RETRIEVAL: UNITING ATTENTION-*

POWERED CONVNETS WITH SIFT FEATURES." (2024).

[16]. *Alrahhal, Maher, and K. P. Supreethi. "Integrating machine learning algorithms for robust content-based image retrieval." International Journal of Information Technology 16.8 (2024): 5005-5021.*

[17]. *Babitha, B. S., et al. "An Investigation into Content-based Image Retrieval using AI-Assisted Methods." 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2024.*

[18]. *Qazanfari, Hamed, Mohammad M. AlyanNezhadi, and Zohreh Nozari Khoshdaregi. "Advancements in content-based image retrieval: A comprehensive survey of relevance feedback techniques." arXiv preprint arXiv:2312.10089 (2023).*

[19]. *Yildirim, Muhammed. "Content-based image retrieval and image classification system for early prediction of bladder cancer." Diagnostics 14.23 (2024): 2637.*

[20]. *Anupama, N., and J. Anitha. "Revolutionizing Medical Diagnosis: A Deep Learning Approach for Content-Based Image Retrieval." 2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI). Vol. 3. IEEE, 2025.*

[21]. *Ghaleb, Moshira S., et al. "Content-based image retrieval using fused convolutional neural networks." International Conference on Advanced Intelligent Systems and Informatics. Cham: Springer International Publishing, 2022.*

[22]. *Abdullah, Sura Mahmood, and Mustafa Musa Jaber. "Deep learning for content-based image retrieval in FHE algorithms." Journal of intelligent systems 32.1 (2023): 20220222.*

[23]. *Hamroun, Mohamed, and Damien Sauveron. "A Hybrid Deep Learning and Knowledge Graph Approach for Intelligent Image Indexing and Retrieval." Applied Sciences 15.19 (2025): 10591.*