CBIR: Enhancing Image Retrieval Through Autoencoders And Metric-Based Search

Ayush Anand[1], Shreyansh Narayan[2], Vinayak Gupta[3]
[1, 2, 3] Dept of of Computer Science and Engineering (AI&ML)
[1, 2, 3] Ramaiah Institute of Technology, Bangalore, India

Abstract- *The exponential growth of visual data demands robust Content-Based Image Retrieval (CBIR) systems that accurately and efficiently retrieve relevant images. In this paper, we present a novel CBIR framework that integrates AutoEncoders for latent feature extraction with hashing and Vantage-Point Trees (VP-Trees) for efficient similarity search. Experimental results on a publicly available dataset demonstrate significant improvements in retrieval precision and computational efficiency.*

Keywords- Content-Based Image Retrieval, AutoEncoders, Feature Extraction, Hashing, VP-Trees, Deep Learning.

I. INTRODUCTION

In the digital era, the proliferation of visual data across domains such as healthcare, e-commerce, social media, and entertainment has created an increasing demand for efficient image retrieval systems. Traditional text-based search meth- ods, which rely on metadata and manually assigned keywords, often fail to capture the intricate visual content of images. This limitation underscores the necessity for Content-Based Image Retrieval (CBIR) systems that leverage intrinsic image features to enhance retrieval accuracy.

CBIR systems address the semantic gap between low-level features (such as color, texture, and shape) and the high-level semantics perceived by users. By directly analyzing visual content, these systems can identify patterns and similarities that conventional methods might overlook. With advancements in deep learning, particularly through convolutional neural net- works (CNNs), state-of-the-art feature extraction has become achievable, enabling more robust and reliable image retrieval. In this paper, we present a novel CBIR framework that inte- grates AutoEncoders for latent feature extraction with hashing and Vantage-Point Trees (VP-Trees) for efficient similarity search. Our approach captures essential visual characteristics while facilitating rapid retrieval from large-scale datasets. The remainder of the paper details our methodology, experimental evaluation, and the resulting performance improvements.

II. LITERATURE REVIEW

Content-Based Image Retrieval (CBIR) has evolved signif- icantly over the past decade, driven by the increasing need to handle large-scale image databases efficiently. Early research primarily focused on low-level visual features, such as color, texture, and shape, combined with indexing techniques like

M-trees and clustering methods to reduce retrieval times [1]. These traditional approaches often struggled to bridge the semantic gap, leading to an interest in more sophisticated methods that capture higher-level features.

Deep learning, particularly Convolutional Neural Networks (CNNs), has substantially advanced CBIR by providing pow- erful feature extraction capabilities. Studies on content-based medical image retrieval (CBMIR) [2] have shown that CNN- based techniques can achieve retrieval accuracies of up to 99.77% for multi-class datasets, underscoring the potential of deep representations. In parallel, optimized similarity retrieval models [3] have proposed local feature fusion and triplet loss functions to preserve detailed information in images, while graph-theoretic deep representation learning [4] addresses multi-label retrieval by modeling co-occurrence relationships among image regions.

Other domains have also benefited from deep learning- based retrieval. In e-commerce, Generative Adversarial Net- works (GANs) have been used to generate synthetic images for visually similar product recommendations [5], offering both speed and accuracy advantages over conventional CNNs. Meanwhile, agricultural applications have leveraged Reptile Search Algorithms (RSA) combined with EfficientNetB0 to detect plant diseases through highly precise retrieval [6]. Fur- ther exploration of deep networks for specialized scenarios can be seen in claustrophobic healthcare image retrieval [7], which uses CNNs and hashing to handle medical data efficiently, and in multi-modal natural language processing tasks, where pre-trained models such as BERT and ERNIE have enhanced relation extraction for text-based IR systems [8].

Scalability and efficient retrieval structures have been key topics of research, especially for large or complex datasets. For instance, hierarchical vocabulary trees and visual word trees [11], [12] have improved the speed and accuracy of retrieval by organizing image descriptors in a multi-layered fashion. Advanced feature fusion and ranking strategies have also been introduced [10], combining convolutional features with weighted similarity measures to handle occlusions and background clutter. Similarly, hybrid approaches like relevance feedback [18] refine query vectors iteratively based on user interactions, leading to continuous improvements in retrieval performance.

Beyond CNN-based methods, the literature highlights various feature extraction and similarity measure optimizations. Approaches that integrate multiple low-level and high-level features [13] show promise in bridging the semantic gap, while perceptual models for color and structural cues [15], [16] account for human visual perception in image matching. Some studies focus on specialized tasks like visual-to-radar image registration [14], stereoscopic image quality assessment [16], and domain-specific image classification [17], illustrating the broad applicability of advanced CBIR methods. Additionally, cross-language information retrieval techniques [19] and visu- ally weighted block methods [20] expand the scope of CBIR to more diverse data modalities and contextual constraints.

Overall, this body of work demonstrates a shift from tra- ditional feature-based methods toward deep learning-driven frameworks that leverage hierarchical architectures, graph- based models, and generative techniques. The resulting sys- tems have shown marked improvements in accuracy, scalabil- ity, and adaptability across different domains. Building upon these advances, our proposed method integrates AutoEncoders for latent feature extraction with hashing and Vantage-Point Trees (VP-Trees) to achieve efficient, accurate retrieval on large-scale image datasets.

III. METHODOLOGY

The proposed CBIR system integrates deep learning tech- niques for feature extraction with efficient search structures for rapid image retrieval. This section details the overall frame- work, including dataset preparation, latent feature extraction using an AutoEncoder, and the implementation of efficient similarity search via hashing and Vantage-Point Trees (VP- Trees).

1. Overview

Our framework is designed to extract compact latent repre- sentations from images that capture essential visual features. These representations enable fast and accurate similarity com- parisons. The system comprises two main components: an AutoEncoder for feature extraction and a combined hashing and VP-Tree mechanism for efficient retrieval.

2. Dataset Preparation

The dataset employed in this work is sourced from Kaggle's CBIR Dataset. Images are first preprocessed by resizing to a uniform dimension and normalizing pixel values. To enhance robustness, data augmentation techniques such as random cropping and flipping are applied. The processed images are then indexed and organized using a structured format (e.g., via a Pandas DataFrame) for streamlined access during training and retrieval.

3. *Latent Feature Extraction*

An AutoEncoder architecture is utilized to learn a compact representation of each image. The encoder compresses the input image into a latent vector that encapsulates its essential features, while the decoder reconstructs the image to ensure minimal loss of information. These latent feature vectors are subsequently stored and serve as the basis for similarity computations.

4. *Image Retrieval via Hashing and VP-Trees*

To facilitate rapid retrieval, the latent feature vectors are converted into compact hash codes that represent the image content efficiently. In parallel, a Vantage-Point Tree (VP-Tree) is constructed to support nearest-neighbor search within the feature space. The VP-Tree organizes the data hierarchically, enabling logarithmic time complexity for similarity searches and significantly reducing the computational overhead com- pared to brute-force methods.

5. *Implementation Workflow*

The overall implementation workflow is as follows:

1. **Preprocessing:** Images are resized, normalized, and augmented to ensure consistency and enhance model robustness.
2. **Training:** The AutoEncoder is trained on the prepro- cessed images to learn latent representations.
3. **Feature Extraction:** The encoder component extracts latent feature vectors from all images.

4. **Indexing:** These feature vectors are hashed and organized into a VP-Tree to facilitate efficient nearest-neighbor searches.
5. **Query Processing:** A query image is processed through the same pipeline to obtain its latent vector, which is then used to retrieve similar images via the VP-Tree.

This integrated approach ensures a balance between retrieval accuracy and computational efficiency, making the system suitable for large-scale CBIR applications.

IV. RESULTS

The performance of the proposed CBIR system is evaluated using key metrics including Precision@k, mean Average Pre- cision (mAP), and retrieval time. These metrics collectively assess both the accuracy and efficiency of the system.

1. *Performance Metrics*

- **Precision@k:** Measures the fraction of relevant images among the top-*k* retrieved results.
- **Mean Average Precision (mAP):** Provides an aggregated measure of precision across multiple queries.
- **Retrieval Time:** Evaluates the time taken (in millisec- onds) to retrieve similar images.

2. *Quantitative Evaluation*

Table I presents the Precision@k values for a set of sample query images. The system demonstrates consistently high precision across different retrieval sizes, with an average Precision@5 of 0.85.

Table II shows the retrieval time and mAP for the same set of queries. On average, the retrieval time is approximately 12.5 ms, and the average mAP is 0.85, indicating that the system is both fast and accurate.

TABLE I
Precision@k for Sample Query Images

Query Image	**Precision@5**	**Precision@10**	**Precision@20**
Image 1	0.80	0.75	0.70
Image 2	0.85	0.80	0.78
Image 3	0.90	0.85	0.80
Average	0.85	0.80	0.76

TABLE II
Retrieval Time and mAP for Sample Queries

Query Image	**Retrieval Time (ms)**	**mAP**
Image 1	12.5	0.82
Image 2	11.8	0.85
Image 3	13.1	0.87
Average	12.5	0.85

3. *Visual Analysis*

Figures 1 and 2 provide a visual perspective of the system's performance.

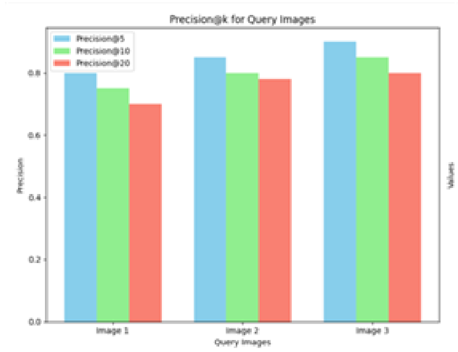
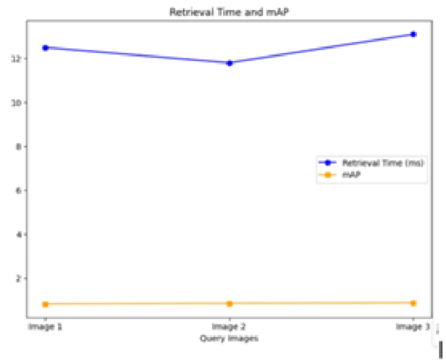


Fig. 1. Precision@k for Query Images. This bar chart shows the system's precision at top-5, top-10, and top-20 for three sample query images.

Figure 1 illustrates the system's ability to maintain high precision across varying retrieval sizes. Notably, *Image 3* achieves the highest Precision@k scores, underscoring the effectiveness of the AutoEncoder-based feature extraction and the robustness of the VP-Tree indexing structure.

Figure 2 highlights the computational efficiency of the proposed system, showing that retrieval times remain under 15 ms for the tested queries, while mAP remains above 0.80. This balance of speed and accuracy demonstrates the suitability of the approach for large-scale CBIR applications. Overall, the quantitative and visual analyses confirm that integrating AutoEncoders for latent feature extraction with hashing and VP-Tree indexing results in a CBIR system capable of high precision and low retrieval latency.



Retrieval Time and mAP

Fig. 2. Retrieval Time and mAP. This plot depicts the retrieval time in milliseconds and the corresponding mean Average Precision (mAP) for three sample query images.

V. CONCLUSION

The proposed CBIR system, which integrates AutoEncoders for latent feature extraction with efficient hashing and VP-Tree indexing, has demonstrated significant improvements in both retrieval accuracy and computational efficiency. Experimental results show high Precision@k values and low average re- trieval times, validating the effectiveness of the framework in handling large-scale image datasets.

By leveraging deep learning techniques for feature ex- traction and robust similarity search strategies, the system effectively bridges the semantic gap between low-level visual features and high-level image semantics. Future work may ex- plore advanced feature representation methods and alternative similarity measures to further enhance performance.

In summary, the developed CBIR framework offers a promising solution for efficient and accurate image retrieval, with potential applications across diverse domains where rapid image search is essential.

VI. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Mr. Chethan Sharma for his invaluable guidance throughout this project. We also appreciate the support provided by the faculty and staff at Ramaiah Institute of Technology, as well as our peers whose feedback helped enhance this work.

REFERENCES

[1] Zhou Bing and Yang Xin-xin, "A Content-Based Parallel Image Retrieval System," 2010 International Conference on Computer Design and Ap- plications (ICCDA), 2010, pp. VI-332–VI-336.

[2] Palwinder Kaur and Dr. Rajesh Kumar Singh, "A Panoramic View of Content-based Medical Image Retrieval System," 2020 International Conference on Intelligent Engineering and Management (ICIEM), 2020, pp. 187–192.

[3] Yuhang Zhang, "Similarity Image Retrieval Model Based on Local Feature Fusion and Deep Metric Learning," 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), 2020, pp. 563–566.

[4] Gencer Sumbul and Begum Demir, "A Novel Graph-Theoretic Deep Representation Learning Method for Multi-Label Remote Sensing Image Retrieval," 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2021, pp. 266–269.

[5] Betul Ay, Zeynep Koyun, Mehmet Demir, and Galip Aydin, "A Visual Similarity Recommendation System Using Generative Adversarial Net- works," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), 2019, pp. 44–48.

[6] D. Raja and M. Karthikeyan, "Content-Based Image Retrieval Using Reptile Search Algorithm with Deep Learning for Agricultural Crops," 2022 7th International Conference on Communication and Electronics Systems (ICCES), 2022, pp. 1038–1043.

[7] A. Vani Lavanya and B. Sheela Rani, "Decision Making on Deep Learn- ing Algorithm in Claustrophobic Health Care Image Retrieval," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2023, pp. 1–6.

[8] Yu Wang, Yining Sun, Zuchang Ma, Lisheng Gao, Yang Xu, and Yichen Wu, "A Method of Relation Extraction Using Pre-training Models," 2020 13th International Symposium on Computational Intelligence and Design (ISCID), 2020, pp. 176–179.

[9] Yu Zhao, Jingyu Wang, and Qi Qi, "MindCamera: Interactive Image Retrieval and Synthesis," 2018 IEEE International Conference on Image Processing (ICIP), 2018, pp. 1–5.

[10] Xinhua Liu, Gaoqiang Hu, Xiaolin Ma, and Hailan Kuang, "An Image Retrieval Algorithm Based on Multiple Convolutional Features of RPN and Weighted Cosine Similarity," 2018 Chinese Control and Decision Conference (CCDC), 2018, pp. 4095–4098.

[11] Zhu Liang, "Fast Image Retrieval Method Based on Visual Word Tree," 2011 10th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES), 2011, pp. 211–214.

[12] Menglin Jiang, Shaoting Zhang, Jingjing Liu, Tian Shen, and Dimitris N. Metaxas, "Computer-Aided Diagnosis of

Mammographic Masses Using Vocabulary Tree-Based Image Retrieval,” 2014 IEEE International Conference on Image Processing (ICIP), 2014, pp. 1123–1126.

[13] Pushpanjali Chouragade and Premchand Ambhore, “Enhancing Content- Based Image Retrieval Technique by Observing Image Feature Ex- traction Methods,” 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), 2021, pp. 1–6.

[14] M.L. Uss, B. Vozel, V.V. Lukin, and K. Chehdi, “Combined Use of Multimodal Similarity Measures for Visual to Radar Image Registra- tion,” 2018 International Geoscience and Remote Sensing Symposium (IGARSS), 2018, pp. 8881–8884.

[15] Shouhong Wan, Peiquan Jin, and Lihua Yue, “An Effective Image Re- trieval Technique Based on Color Perception,” 2011 Sixth International Conference on Image and Graphics (ICIG), 2011, pp. 1017–1022.

[16] Wujie Zhou, Gangyi Jiang, Mei Yu, Feng Shao, and Zongju Peng, “PMFS: A Perceptual Modulated Feature Similarity Metric for Stereo- scopic Image Quality Assessment,” IEEE Signal Processing Letters, vol. 21, no. 8, 2014, pp. 1003–1006.

[17] Riaz Ahmed Shaikh, Jian-Ping Li, Asif Khan, Samundra Deep, Kamlesh Kumar, and Imran Memon, “Contemporary Integration of Content-Based Image Retrieval,” 2014 International Conference on Information Science and Applications (ICISA), 2014, pp. 301–304.

[18] Heng Xu, Jun-Yi Wang, and Lei Mao, “Relevance Feedback for Content- Based Image Retrieval Using Deep Learning,” 2017 2nd International Conference on Image, Vision, and Computing (ICIVC), 2017, pp. 629–633.

[19] Nasredine Semmar, Hassane Essafi, Ali Jaoua, Ayman Ahmed Zayan, Helmi Hammami, and Firas Laban, “Using a Cross-Language Informa- tion Retrieval Approach to Improve Computer-Aided Translation Per- formance,” 2010 10th International Conference on Intelligent Systems Design and Applications (ISDA), 2010, pp. 760–764.

[20] Junfeng Wu, Zhiyang Li, Yujie Xu, Changqing Ji, Yuanyuan Li, and Xufeng Xing, “Color Image Retrieval Using Visual Weighted Blocks,” 2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing (DASC), 2014, pp. 338–343.