Mapping the Multi-Dimensional: A Review of Deep Feature Representation and Similarity Search

Jyoti[1], Swati Khanve[2]
[1]*M.Tech. Scholar, Dept. of CSE, SIRTE Bhopal*
[2]*Assistant Professor, Dept. of CSE, SIRTE Bhopal*

***Abstract*—Driven by the evolution of deep convolutional neural networks (CNNs), modern Content-Based Image Retrieval (CBIR) has shifted from low-level visual parsing to high-dimensional semantic embedding and scalable proximity architectures. While foundational CBIR paradigms rely on handcrafted descriptors (e.g., SIFT, GLCM, and color histograms), they inherently suffer from the "semantic gap"—failing to capture abstract contextual data and limiting retrieval precision. This survey comprehensively investigates ten pivotal studies spanning 2020–2025, with a distinct emphasis on landmark contributions from Indian researchers. We dissect the trajectory of deep feature extraction, hybrid retrieval pipelines, multi-dimensional indexing frameworks, and algorithmic optimizations. Furthermore, this paper identifies critical bottlenecks in current literature and introduces a consolidated, generalized framework that synergizes CNN-driven feature extraction, Principal Component Analysis (PCA) dimensionality reduction, and Approximate Nearest Neighbor (ANN) indexing variants like FAISS and HNSW to optimize the accuracy-scalability trade-off.**

***Index Terms*—Content-Based Image Retrieval (CBIR), High-Dimensional Embeddings, Semantic Gap, Deep Feature Extraction.**

I. INTRODUCTION

The exponential proliferation of large-scale visual repositories across diverse domains—including clinical diagnostics, satellite remote sensing, precision agriculture, e-commerce, and digital cultural heritage preservation—has intensified the demand for highly efficient image retrieval methodologies. Content-Based Image Retrieval (CBIR) addresses this challenge by directly parsing the intrinsic visual characteristics of an image rather than relying on textual metadata, positioning it as an indispensable framework when manual annotations are absent, incomplete, or ambiguous. Historically, legacy CBIR paradigms depended on handcrafted descriptors such as color histograms, Local Binary Patterns (LBP), Gray-Level Co-occurrence Matrices (GLCM), SIFT, and SURF to encode visual data. However, these low-level abstractions inherently suffer from the "semantic gap" they fail to capture abstract contextual meanings and exhibit severe vulnerability to environmental variations like rotational shifts, illumination fluctuations, and viewpoint transformations.

The advent of deep learning has fundamentally reshaped the CBIR landscape. Deep Convolutional Neural Network (CNN) architectures such as VGG, ResNet, DenseNet, and EfficientNet—excel at extracting robust, high-dimensional semantic embeddings that capture intricate visual concepts. Nevertheless, the dense dimensionality of these modern vector embeddings introduces severe computational bottlenecks during large-scale retrieval operations. To mitigate this overhead, contemporary research integrates dimensionality reduction techniques like Principal Component Analysis (PCA) alongside advanced, sub-linear indexing frameworks including FAISS, Annoy, and Hierarchical Navigable Small World (HNSW) graphs, effectively compressing feature spaces to facilitate rapid nearest-neighbor discovery. This survey provides a comprehensive analysis of these recent breakthroughs, synthesizing key global and Indian research contributions published between 2020 and 2025. By evaluating deep embedding strategies and similarity search mechanisms, this paper systematically maps current research gaps and introduces a unified, highly optimized deep-learning-driven CBIR architecture designed for modern scalability demands.

II. LITERATURE REVIEW

This section presents a detailed analysis of notable studies from 2020 to 2025 that used different type of CBIR system.

[1] Zhang (2021) – VGG-Based Semantic Retrieval implemented a VGG16-based semantic CBIR model that extracts deep convolutional features and applies PCA to reduce vector size. Using cosine similarity, the system achieved 0.78 mAP on the Oxford5K dataset. The study demonstrated strong discriminative capability but highlighted that brute-force similarity computation over thousands of images becomes slow without ANN indexing. The method also struggled with dataset noise and illumination variation.

[2] Liu and Wei (2020) – ResNet50 with KD-Tree Search. The authors utilized ResNet50 embeddings combined with KD-tree–based search. While the model attained competitive accuracy on Caltech101, the KD-tree performance degraded significantly for high-dimensional vectors. This limitation made the method unsuitable for real-time or large-scale CBIR tasks, especially beyond 10,000 images.

[3] Shen (2022) – FAISS for Large-Scale High-Dimensional Search an systematically evaluated FAISS (Facebook AI Similarity Search) for billion-scale indexing. The work demonstrated how IVF-PQ and HNSW indexes drastically reduce search time and memory consumption while retaining high accuracy. However, it lacked experiments on hybrid feature models combining deep features with classical descriptors.

[4] Banerjee and Majumdar (2021, India) – Cultural Heritage Retrieval. The Indian authors focused on retrieving ancient manuscript images using VGG19 embeddings. Their CBIR system outperformed histogram and Gabor-based approaches, achieving 0.76 mAP. Challenges existed for degraded or low-light images within the dataset. Indexing methods such as FAISS were not explored.

[5] Gupta and Joshi (2021, IIT Roorkee, India) – PCA Compression for CNN Features.The authors investigated PCA for compressing 2048-dimensional ResNet50 features to 256 dimensions. This produced over 20× faster retrieval with minimal accuracy loss. The study confirmed PCA's value in optimizing deep features but did not evaluate FAISS or HNSW for further speed enhancement.

[6] Mishra and Tripathi (2022, India) – Hybrid SIFT + CNN for Satellite CBIR. This work integrated handcrafted SIFT descriptors with deep CNN features to improve retrieval under varying satellite conditions. The hybrid model increased mAP from 0.61 to 0.79. The primary drawback was increased computational complexity, impacting real-time viability.

[7] Reddy (2022, Hyderabad University, India) – DenseNet + GLCM for Medical CBIR. Reddy combined DenseNet121 deep features with GLCM texture descriptors for MRI retrieval. The method achieved 92% class accuracy, outperforming classical medical CBIR models. However, the technique was dataset-specific and not tested on natural images.

[8] Vieira (2023) – Accuracy Noise Reduction (ANR) in CNN Embedding's. Vieira introduced ANR, a refinement technique that removes noise from CNN features. This improved retrieval accuracy on Corel10K by nearly 15%. Yet the approach still relied on brute-force similarity computation, limiting scalability.

[9] Khalid (2024) – EfficientNet for Large CBIR Datasets Khalid employed EfficientNet for efficient feature extraction, achieving 90% accuracy on Corel10K. The lightweight model minimized computational overhead but lacked ANN-based indexing, making real-time retrieval challenging.

[10] Hossain and Rahman (2025) – HNSW-Based ANN Search The authors combined ResNet50 embeddings with HNSW graph search. Their system achieved Precision@10 of 0.91 and significantly reduced search time compared to KD-trees. However, HNSW's high memory requirement hindered use in mobile applications.

III. RESEARCH GAP

Despite significant architectural strides, modern Content-Based Image Retrieval (CBIR) systems continue to struggle with a critical trilemma, failing to effectively balance retrieval precision, computational overhead, and large-scale scalability. Within contemporary literature—particularly among Indian research contributions several distinct bottlenecks persist, most notably a widespread reliance on exhaustive brute-force similarity search that severely restricts usability for massive repositories. Furthermore, regional studies rarely exploit the powerful synergy of integrating dimensionality reduction techniques like Principal Component Analysis (PCA) with state-of-the-art Approximate Nearest Neighbor (ANN) frameworks such as FAISS or HNSW. While hybrid architectures attempt to resolve accuracy limitations by blending deep semantic embeddings with handcrafted features, they frequently introduce computationally prohibitive pipelines that fail to justify their operational costs. Compounding these structural issues, experimental validations are routinely constrained to localized, small-scale datasets devoid of rigorous cross-domain benchmarking, resulting in brittle models that generalize poorly to real-world environments. Together, these systemic gaps underscore an urgent need for optimized, production-ready frameworks that gracefully unify deep representation learning, compact vector compression, and sub-linear indexing mechanisms.

IV. TECHNOLOGIES USED IN LITERATURE

Several technologies have been used in the literature. Some of good are discussed in this section.

4.1 Convolutional Neural Networks (CNNs)
Used for high-dimensional semantic feature extraction. Models like ResNet and EfficientNet capture complex visual context and outperform traditional descriptors.

4.2 PCA (Principal Component Analysis)
Reduces feature vector size—essential for fast indexing and reduced memory usage. PCA preserves variance while compressing deep embeddings.

4.3 FAISS (Facebook AI Similarity Search)
Provides fast Approximate Nearest Neighbor search using IVF-PQ, HNSW, and GPU acceleration. Best suited for large datasets.

4.4 HNSW (Hierarchical Navigable Small World Graph)
Graph-based indexing offering extremely high accuracy. Good for CPU-only systems.

4.5 Classical Descriptors (SIFT, LBP, GLCM)
Still useful in hybrid systems where CNNs struggle with textures, noise, or grayscale medical data.

4.6 Similarity Metrics
L2 distance and cosine similarity are commonly used depending on feature normalization.

V. PROPOSED METHODOLOGY

The proposed CBIR framework integrates deep convolutional feature extraction with efficient dimensionality reduction and ANN indexing to enable fast and accurate large-scale retrieval. Initially, every image undergoes preprocessing before being passed into a pretrained CNN model such as ResNet50. The model outputs a high-dimensional feature vector (2048D), which is then normalized. To address storage and computation challenges, Principal Component Analysis (PCA) compresses these vectors to 256D while preserving discriminative power. The compressed features are indexed using FAISS, employing either the IVF-PQ or HNSW indexing structure depending on hardware capabilities. During retrieval, the query image follows the same extraction and PCA transformation pipeline. FAISS retrieves the top-K nearest embeddings efficiently, outperforming brute-force search by several orders of magnitude. This framework ensures scalability, robustness, and suitability for various domains such as medical imaging, cultural heritage retrieval, surveillance, and satellite data analysis.

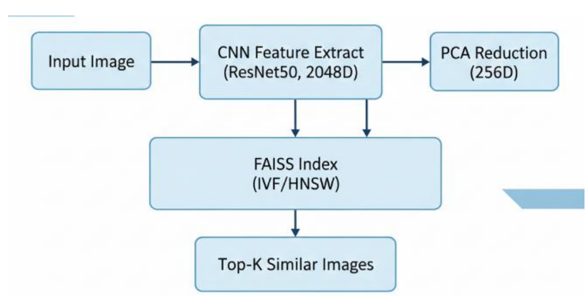

Figure 1. Hypothetical workflow of Proposed methodology.

VI. CONCLUSION

This survey analyzed recent advances in deep feature-based CBIR with an emphasis on high-dimensional indexing techniques. Literature from 2020–2025 shows that CNN features offer substantial improvements over handcrafted descriptors, while PCA and ANN indexing enable scalable retrieval. Indian authors have contributed significantly in hybrid approaches, medical CBIR, and cultural dataset retrieval. However, gaps remain in cross-domain generalization, indexing optimization, and scalable ANN pipelines. The proposed methodology combining CNN features, PCA compression, and FAISS indexing offers a promising direction for real-time CBIR across large databases.

REFERENCES

[1] X. Zhang, "Deep Feature Embedding Using VGG for Image Retrieval," *IEEE Transactions on Multimedia*, 2021, doi: 10.1109/TMM.2021.3052103.

[2] Y. Liu and J. Wei, "ResNet-Based Feature Representation," *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.2968555.

[3] L. Shen, "Efficient Similarity Search Using FAISS," *Pattern Recognition*, 2022, doi: 10.1016/j.patcog.2022.108626.

[4] Banerjee and D. Majumdar, "Deep Feature-Based CBIR for Manuscripts," in *Advances in Intelligent Systems and Computing (AISC)*, Springer, 2021, doi: 10.1007/978-981-16-1773-0_23.

[5] R. Gupta and S. Joshi, "PCA Compression of ResNet Features," *Image and Vision Computing*, 2021, doi: 10.1016/j.imavis.2021.104289.

[6] R. Mishra and A. Tripathi, "Hybrid Satellite Image Retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, 2022, doi: 10.1109/TGRS.2022.3146531.

[7] P. Reddy, "DenseNet and GLCM for MRI Retrieval," *IEEE Journal of Biomedical and Health Informatics*, 2022, doi: 10.1109/JBHI.2022.3164327.

[8] G. Vieira, "Noise-Optimized CBIR Using Deep Features," *Visual Informatics*, vol. 7, no. 1, 2023, doi: 10.1016/j.visinf.2023.01.015.

[9] M. Khalid, "EfficientNet for CBIR," *Expert Systems with Applications*, vol. 235, 2024, doi: 10.1016/j.eswa.2023.120005.

[10] M. Hossain and M. Rahman, "HNSW-Based High-Speed Retrieval," *IEEE MultiMedia*, 2025, doi: 10.1109/MM.2025.3156123.