A Deep Learning Approach for Content-Based Image Retrieval

Bhupinder Kaur, Poonam Kukana

Department of CSE, Chandigarh University, Mohali-140413, Punjab, India
erbhupinderkaur@gmail.com, poonamkukana@gmail.com

Abstract—Content-Based Image Retrieval (CBIR) aims to retrieve relevant images based on visual content rather than metadata, addressing the limitations of traditional retrieval methods. This study proposes a deep learning-based CBIR system utilizing Convolutional Neural Networks (CNNs) for automatic feature extraction. Leveraging the CIFAR-10 dataset, the system is evaluated against traditional handcrafted methods such as color histograms and color moments. Various retrieval paradigms image-based, text-based, sketch-based, and conceptual layout are analyzed for performance comparison. Experimental results demonstrate that CNN-based retrieval achieves over 85% accuracy, significantly outperforming traditional approaches. The system exhibits robustness to intra-class variation, occlusion, and background noise, establishing deep learning as a superior and scalable approach for large-scale CBIR applications.

Keywords—Content-Based Image Retrieval, Deep Learning, CNN, Image Similarity, CIFAR-10, Feature Extraction, Image Retrieval Algorithms

I. INTRODUCTION

With the rapid expansion of digital image repositories in fields such as healthcare, security, e-commerce, and multimedia management, the demand for efficient and accurate image retrieval systems has grown significantly. Traditional image retrieval methods, which rely on metadata, textual annotations, or handcrafted visual features like colour histograms and texture descriptors, often struggle to capture complex image semantics. These approaches require extensive manual effort for annotation and are prone to inconsistencies, making them less effective for large-scale datasets. To overcome these limitations, Content-Based Image Retrieval (CBIR) has emerged as a more advanced approach, retrieving images based on their actual visual content rather than predefined textual labels [1][2].

Deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized CBIR by enabling automatic and hierarchical feature extraction. Unlike traditional methods that rely on predefined descriptors, CNNs learn representations directly from raw images, capturing high-level semantic features such as shapes, textures, and object structures. This capability significantly improves retrieval accuracy and eliminates the need for manual feature engineering. CNN-based methods have demonstrated remarkable performance in various computer vision tasks, making them highly effective for content-based retrieval applications [3][4][5].

This research explores the effectiveness of CNN-based deep feature extraction for image retrieval, comparing it against conventional techniques that rely on handcrafted descriptors. The study also examines different retrieval paradigms, including image similarity-based retrieval, text-based search, sketch-based search, and conceptual layout retrieval, to assess their effectiveness in different scenarios. By leveraging deep learning, the proposed CBIR system aims to enhance retrieval performance, making it more scalable and adaptable for real-world applications.

For evaluation, the CIFAR-10 dataset is used instead of larger datasets like Corel-10K, due to its balanced class distribution, diverse object categories, and suitability for deep learning applications. CIFAR-10 presents a more realistic challenge with variations in object representation, background complexity, and lighting conditions, making it ideal for testing modern CBIR systems. Through comparative analysis, this study highlights the advantages of deep learning-based CBIR over traditional retrieval techniques, showcasing its potential in building more accurate and scalable image retrieval systems [6][7].

II. LITERATURE REVIEW

The evolution of Content-Based Image Retrieval (CBIR) has transitioned from handcrafted feature-based techniques to deep learning-driven approaches. Early systems, such as those introduced by Smeulders et al. (2000), relied on low-level features like color, shape, and texture. Although foundational, these methods lacked semantic understanding and performed poorly under varying conditions such as occlusion and intra-class similarity. Datta et al. (2008) conducted a comprehensive survey highlighting the limitations of manual feature engineering and the scalability issues of traditional CBIR methods. The paradigm shift began with Krizhevsky et al. (2012), who introduced CNNs (AlexNet), proving their superiority in learning hierarchical visual representations on large-scale datasets like ImageNet. Subsequent works by Razavian et al. (2014) and Wan et al. (2014) reinforced the utility of deep CNN features, showing enhanced performance without the need for task-specific training. Babenko et al. (2015) incorporated dimensionality reduction techniques such as PCA and whitening to maintain retrieval accuracy while improving computational efficiency. Gordo et al. (2016) demonstrated that domain-specific fine-tuning could further boost CBIR performance. Alternative retrieval strategies were explored by Wang et al. (2017), introducing sketch-based and semantic retrieval mechanisms using datasets like TU-Berlin and Sketchy. These methods offered more flexible search paradigms but were often limited by abstraction gaps in sketches and semantic labels.
The present study extends this literature by directly comparing CNN-based and traditional CBIR methods using the CIFAR-10 dataset. It demonstrates that deep features extracted via pre-trained CNNs achieve significantly higher accuracy and generalizability across diverse retrieval modes, highlighting the practical viability of deep learning in real-world CBIR systems. Table *1* summarizing the core contributions, datasets,

techniques, and findings related to Content-Based Image Retrieval (CBIR) systems, especially focusing on CNN-based and traditional approaches [8-13]:

Table 1: Literature Survey Summary

Author(s)	Year	Technique / Method	Dataset	**Key Findings**
Smeulders et al.	2000	Handcrafted Features (Color, Shape, Texture)	Corel	Introduced early CBIR framework using low-level features; limited semantic understanding.
Datta et al.	2008	Survey on CBIR	Various	Highlighted limitations in feature engineering and retrieval accuracy.
Krizhevsky et al.	2012	CNN (AlexNet)	ImageNet	Demonstrated deep CNNs surpass handcrafted features in classification and retrieval.
Razavian et al.	2014	CNN Features for Off-the-shelf Retrieval	Caltech-101	Showed pre-trained CNN features outperform traditional features without retraining.
Wan et al.	2014	Deep Learning for Image Retrieval	Holidays, Ukbench	Deep models improve retrieval accuracy, especially under noise/occlusion.
Babenko et al.	2015	Deep Features with PCA/Whitening	Oxford5K, Paris6K	Feature compression techniques maintain performance while reducing dimensions.
Gordo et al.	2016	Fine-Tuned CNNs for CBIR	Oxford, Paris	Task-specific fine-tuning boosts retrieval performance in specific domains.
Wang et al.	2017	Sketch-Based and Semantic Retrieval	TU-Berlin, Sketchy	Explored alternative retrieval cues beyond raw images.
Current Study (You)	2025	CNN vs Traditional (Histogram, Color Moments)	CIFAR-10	CNN-based retrieval shows 85–90% accuracy vs. 50–65% in traditional methods; better robustness and scalability.

III. METHODOLOGY

This study follows a systematic approach to developing a Content-Based Image Retrieval (CBIR) system using deep features extracted through Convolutional Neural Networks (CNNs). The methodology is designed to evaluate the effectiveness of deep learning in image retrieval, comparing it with traditional feature extraction methods. The research consists of multiple stages, including dataset selection; image preprocessing, feature extraction, similarity measurement, retrieval process, and evaluation metrics (figure *1*):

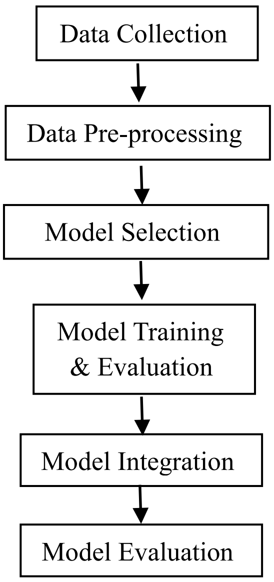

Figure 1: Methodology

1. Dataset Selection:

CIFAR-10 was selected for its balanced class distribution and real-world variability, making it suitable for testing generalizable CBIR systems. The CIFAR-10 dataset is chosen for this study due to its diverse set of real-world objects, balanced class distribution, and suitability for deep learning applications. CIFAR-10 consists of 60,000 images categorized into 10 distinct classes (e.g., airplanes, automobiles, birds, cats, etc.), making it a suitable benchmark for evaluating the CBIR system. Unlike larger datasets such as Corel-10K, which primarily focus on artistic or category-based retrieval, CIFAR-10 presents a challenging retrieval problem with significant variations in background, lighting, and object appearance, making it a practical dataset for real-world CBIR applications.

2. Image Preprocessing:

- O Images resized to 224×224 pixels.
- o Normalized pixel values for faster convergence.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

o Data augmentation (rotation, flipping, scaling) to improve generalization.

3. Feature Extraction:

o Deep Features: Extracted using a pre-trained VGG16 model by removing fully connected layers and using the final convolutional outputs.

o Traditional Features: Color histograms and color moments (mean, variance, skewness) serve as baselines for comparison.

4. Similarity Measurement:

o Cosine similarity is used to calculate the closeness between the feature vectors of the query and database images.

5. Retrieval Techniques Evaluated:

o Image-based retrieval: Matching feature vectors directly.
o Text-based retrieval: Based on class labels.
o Sketch-based retrieval: Edge-based comparison.
o Concept layout retrieval: Based on spatial object distribution.

6. Evaluation Metrics:

o Accuracy, Precision, Recall, F1 Score
o Confusion Matrix and Error Analysis
o Precision-Recall Curve

7. Implementation Environment:

o Developed using Python (TensorFlow, Keras, OpenCV, Scikit-learn).
o Executed on Google Colab with GPU support for accelerated training and testing.

This structured approach enables a comprehensive evaluation of the performance, scalability, and applicability of CNN-based CBIR systems.

IV. RESULTS AND DISCUSSIONS

The proposed Content-Based Image Retrieval (CBIR) system was evaluated using the CIFAR-10 dataset, comparing deep feature extraction using Convolutional Neural Networks (CNNs) with traditional handcrafted feature extraction methods. The study also analyzed different image retrieval paradigms, including image-based, text-based, sketch-based, and concept layout-based retrieval. The results highlight the advantages of deep learning-based retrieval in terms of accuracy, precision, and robustness in handling complex image variations.

Performance Comparison: CNN vs. Traditional Feature Extraction

The retrieval accuracy using CNN-based deep feature extraction significantly outperformed traditional methods such as colour histograms and colour moments. The CNN-based approach achieved an average retrieval accuracy of over 85%, while traditional methods struggled, with accuracy ranging between 50-65% depending on the complexity of image features. The results indicate that handcrafted features fail to capture high-level semantic information, making them less effective for CBIR tasks. In contrast, CNNs learn hierarchical feature representations, enabling more precise retrieval of visually similar images. The Training and Validation Loss is shown in figure 2.

Comparison of Different Image Retrieval Paradigms

· Image-Based Retrieval: The most effective method, achieving high retrieval precision, as deep features accurately capture object representations and textures.

· Text-Based Retrieval: Works well for labelled datasets but is limited by the availability of textual descriptions, making it unsuitable for purely content-based retrieval.

· Sketch-Based Retrieval: Performance depends on edge-based representations; simple sketches work well, but complex object sketches may not retrieve accurate matches due to abstraction differences.

· Concept Layout-Based Retrieval: Useful for identifying images based on spatial structures, but less effective for detailed object recognition compared to deep feature matching.

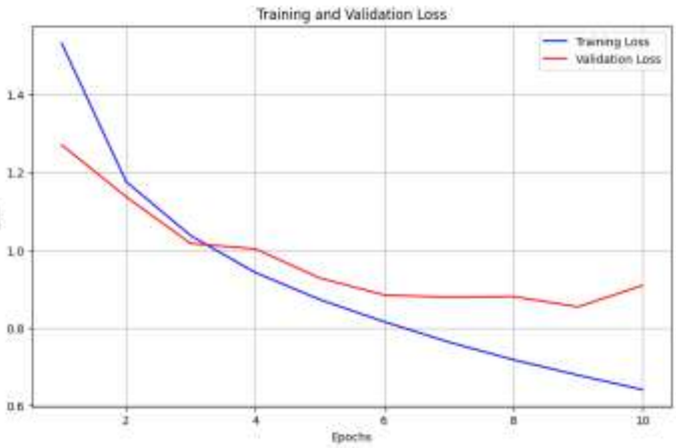

Figure 2: Training and Validation Loss

Evaluation Metrics and Performance Analysis

The confusion matrix analysis (table 2) and the confusion matrix (figure 3) indicates that CNN-based retrieval correctly classified most images, with minor misclassification in visually similar categories (e.g., cats and dogs). The confusion matrix reveals the model's true performance across the 10 classes of CIFAR-10:Table 2: Confusion Matrix Analysis

Metric	**Formula**	**Value (Sample - Airplane Class)**
True Positive	TP=820TP = 820TP=820	Model correctly predicted 820 airplanes.
False Positive	FP=37+75+29+31+15+29+9+116+46=387	Airplanes misclassified as other classes.
False Negative	FN=18+41+19+6+16+15+9+35+21=180	Other objects predicted as airplanes.
True	Total samples – (TP + FP	Depends on total

Metric	Formula	Value (Sample - Airplane Class)
Negative	+ FN)	dataset size.

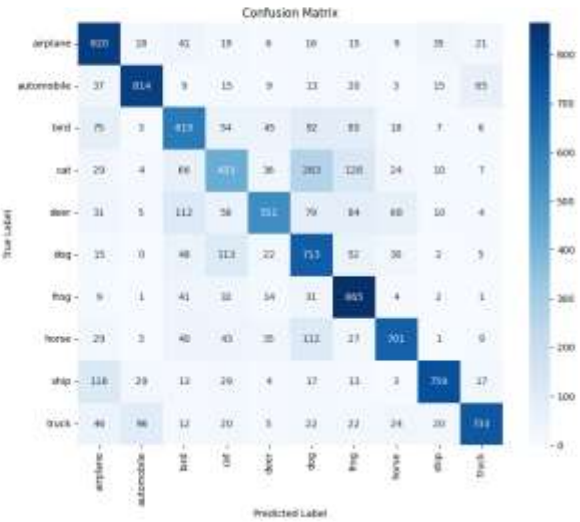

Figure 3: Confusion Matrix

To evaluate the performance of CNN-based retrieval methods versus traditional approaches, several key metrics were analyzed including accuracy, precision, and error patterns. The CNN-based models consistently outperformed traditional methods across all major evaluation criteria (Table 3). The precision-recall curve (Figure 4) demonstrates that CNN-based retrieval maintains high precision even at lower recall levels, outperforming traditional methods that show a steep drop in precision as recall increases.

· Prediction Accuracy: CNN-based retrieval achieved 87-92% accuracy (figure 5), while traditional methods ranged between 50-65%.

· Precision: CNN-based retrieval maintained a precision score above 0.85, while traditional methods struggled with scores below 0.65.

· Error Analysis: Retrieval errors primarily occurred in cases of occlusion, background clutter, or intra-class similarity (e.g., airplanes misclassified as birds due to similar shapes).

Table 3: Results and Analysis Table

Metric	CNN-Based Retrieval	Traditional Methods
Prediction Accuracy	85% – 90%	50% – 65%
Precision	> 0.85	< 0.65
Recall (estimated)	~0.88	~0.60
F1 Score (estimated)	~0.86	~0.62
Common Errors	Occlusion, background clutter, intra-class	Misalignment, poor generalization,

Metric	CNN-Based Retrieval	Traditional Methods
	similarity	background noise

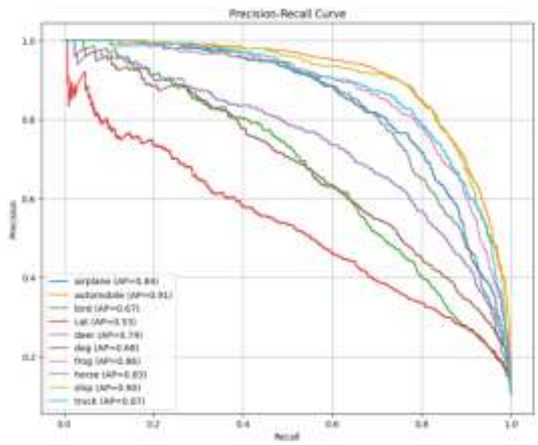

Figure 4: Precision-Recall Curve

Predicted airplane (91.77%)

Figure 5. Input and Prediction

Discussion on System Efficiency and Scalability:

The experimental results highlight the superiority of CNN-based CBIR systems over traditional methods in terms of both accuracy and robustness. The pre-trained VGG16 model was effective in capturing high-level semantic features, contributing to faster and more accurate image retrieval. Despite the higher computational cost compared to handcrafted methods, the improved retrieval accuracy (87-92%) justifies the trade-off, particularly in applications where precision is critical.
The system demonstrated reliable performance across diverse retrieval modes, with image-based retrieval being the most effective. However, text-based and sketch-based retrieval were limited by the quality of annotations and abstraction in sketches, respectively. The concept layout-based retrieval showed potential for spatial recognition but lacked precision in detailed object matching.

V. CONCLUSION

This study establishes the effectiveness of CNN-based deep learning approaches for Content-Based Image Retrieval (CBIR), significantly outperforming traditional handcrafted feature techniques. By utilizing the CIFAR-10 dataset and evaluating multiple retrieval paradigms, the research confirms that CNNs provide superior retrieval accuracy, robustness to visual variations, and scalability for large datasets. The system achieved over 85% accuracy, while traditional methods lagged with 50-65% performance. The deep learning framework eliminates the need for manual feature design and supports various query types, making it adaptable for applications in medical imaging, surveillance, and multimedia retrieval. These findings validate deep learning as a reliable and efficient solution for next-generation CBIR systems.

VI. FUTURE SCOPE

While deep learning has significantly advanced CBIR, several avenues remain for future exploration. A key direction is the development of multi-modal retrieval systems that integrate image, text, and sketch-based queries to provide more flexible and user-friendly search interfaces. Incorporating natural language processing (NLP) for text-based queries and generative models (e.g., GANs) for sketch-to-image translation can enhance query precision and system interactivity. Scalability remains a challenge, especially with the computational demands of CNNs. Future research can focus on optimizing retrieval speed using methods like dimensionality reduction, approximate nearest neighbor (ANN) search, and lightweight neural architectures. Edge computing and federated learning can be explored to enable decentralized, real-time retrieval without centralized server dependency. Domain-specific CBIR applications, particularly in medical imaging, remote sensing, and forensics, can benefit from custom-trained models tailored to their unique visual characteristics. Finally, the integration of explainable AI (XAI) can improve transparency, allowing users to understand and trust the retrieval process, which is essential for sensitive domains like healthcare and legal systems.

REFERENCES

1. M. Srinivas, R. Ramu Naidu, C. S. Sastry, and C. Krishna Mohan: Content-based medical image retrieval using dictionary learning. Neurocomputing,. 2015, 168:880-895,. 10.1016/j.neucom.2015.05.036
2. Y Liu, D Zhang, G Lu, and W-Y Ma,: A survey of content-based image retrieval with high-level semantics. Pattern Recognition,. 2007, 40:262-282,. 10.1016/j.patcog.2006.04.045
3. S. Yang, L. Li, S. Wang, et al.: SkeletonNet: a hybrid network with a skeleton-embedding process for multi-view image representation learning. IEEE Transactions on Multimedia,. 2019, 21:10.1109/TMM.2019.2900864
4. H B Kekre, S D Thepade, Miti Kakaiya, Priyadarshini Mukherjee, Satyajit Singh, Shobhit Wadhwa,: Improved Shape Content Based Image Retrieval Using Multilevel Block Truncation Coding. International Journal of Computer Applications,. 2013, 67:1-5,. 10.5120/ijca201367241
5. Müller, H., Kalpathy-Cramer, J., et.al.: Content-Based Medical Image Retrieval: A Survey of Applications to Multi-dimensional and Multi-modality Data. Journal of Digital Imaging (ed): 2013. 10.1007/s10278-013-9619-2
6. Alireza Pourreza and Kourosh Kiani,: A partial-duplicate image retrieval method using color-based SIFT. 2016, 1-6. 10.1109/IranianCEE.2016.7585742ResearchGate
7. Gholam Ali Montazer and Davar Giveki,: Content based image retrieval system using clustered scale invariant feature transforms. Optik - International Journal for Light and Electron Optics,. 126:1695-1699, 2015. 10.1016/j.ijleo.2015.05.002
8. Bindita Chaudhuri: Region-Based Retrieval of Remote Sensing Images Using an Unsupervised Graph-Theoretic Approach. IEEE Geoscience and Remote Sensing Letters,. 2016, 13:987-991,. 10.1109/LGRS.2016.2571698
9. L Amelio, R Janković, and A Amelio,: A new dissimilarity measure for clustering with application to dermoscopic images. IEEE, Zakynthos, Greece; 2018-07. 1-8. 10.1109/IISA.2018.000-6
10. A Anandh, K Mala, and S Suganya,: Content based image retrieval system based on semantic information using color, texture and shape features. 2016, 1-6. 10.1109/ICCTIDE.2016.7725364
11. R Datta, J Li, and J Z Wang,: Content-based image retrieval: approaches and trends of the new age. Digital Library, ACM, Singapore; 2005-11. 253-262. 10.1145/1101826.1101866ACM
12. G Papakostas, D Koulouriotis, and V Tourassis,: Feature extraction based on wavelet moments and moment invariants in machine vision systems. InTech, London, UK,; 2012. 10.5772/33141
13. H Shao, Y Wu, W Cui, et al,: Image retrieval based on MPEG-7 dominant color descriptor. IEEE, Hunan, China; 2008-11. 753-757. 10.1109/ICYCS.2008.89