Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X

Volume 4, Issue 1, 2026

A HYBRID DEEP LEARNING AND DCT-BASED FEATURE FUSION FRAMEWORK FOR CONTENT-BASED IMAGE RETRIEVAL

Shehla Shah[1], Tauseef Noor[2], Abdul Aziz[3], Ibrar Ullah[4], Sohail Farooq[5], Waqar Nawaz[6]

[1]PhD in Computer Science at Iqra National University, Peshawar
[2]Lab Instructor at Iqra National University, Peshawar
[3]Lab Instructor at Iqra National University, Peshawar
[4]City University of Sciences and Information Technology, Peshawar
[5]Facilitation Management Officer Agriculture University, Peshawar
[6]ASSTT Director/Facility Management Officer (BPS-17)
[1]Shehla2k19@gmail.com, [2]Tauseef.noor.69@gmail.com, [3]Abdul1.mkd@gmail.com, [4]Ibrarullah204@gmail.com, [5]Sohailfarooq857@gmail.com, [6]engrwaqarnawaz@gmail.com

DOI: http://doi.org/10.5281/zenodo.18739229

Keywords:
Image Retrieval, Feature Fusion, Discrete Cosine Transform, Vision Transformer, Convolutional Neural Network

Article History
Received on 17 Dec, 2025
Accepted on 15 Jan 2026
Published on 16 Jan 2026

Copyright @Author
Corresponding Author:

Abstract

The CBIR systems tend to experience poor retrieval accuracy because of a lack of semantic information about the high level of representation and poor discrimination of texture. In order to overcome such shortcomings, the given paper presents a hybrid CBIR framework which combines deep learning-based features with handcrafted texture descriptors in Discrete Cosine Transform (DCT) domain. The model is proposed and represents a combination of Convolutional Neural Network (CNN) and Vision Transformer (ViT) to provide strong deep features and supplementary handcrafted features such as color histograms, Hu moments, and DCT-based texture features to improve spatial and frequency-domain representation. The feature fusion strategy is used to make a combination of deep and handcrafted features into one unified feature to achieve good image retrieval. The implementation of the proposed method is tested on several benchmark datasets, WANG, CIFAR-10, Oxford Flowers, and GPR1200, by applying common evaluation metrics, such as accuracy, precision, recall, F1-score, and ROC analysis. The results of the experiment show that the offered hybrid framework can be much more effective than typical CBIR methods and single deep models and reach a top retrieval accuracy of 94. Results indicate that the overall performance and strength of retrieval with a combination of deep semantic attributes and DCT-domain texture data is enhanced to a variety of image datasets, which is why the given approach can be used in high-level CBIR applications.

INTRODUCTION

The new trend of large-scale image repositories has made Content-Based Image Retrieval (CBIR) an indispensable field of research in computer vision. The CBIR techniques are more applicable to the unannotated and high-dimensional image data compared to the conventional text-based methods of retrieving images as they retrieve images based on their visual appearance, which can be color, texture, and shape. Although this has gone a long way, the problem of getting the right and strong image retrieval is still a very difficult task owing to the mismatch of the semantic aspects of low-level visual aspects and high-level human perceptions [1].

The prototype CBIR systems were mainly based on manual features, with color histogram, texture description, and shape features being some of the features [2]. These methods are computationally efficient, however, in most cases they do not represent subtle semantic information, leading to low retrieval accuracy. Recent progress in deep learning, in particular Convolutional Neural Networks (CNNs), have demonstrated far better results in CBIR through learning discriminative feature representations using image data. Nonetheless, CNN-based approaches can still have a problem with long-range connections and global contextual information.

Vision Transformers (ViTs) are a new and strong competitor to CNNs as they can greatly benefit the relationships worldwide based on self-attention mechanisms. ViTs have proved to be effective in performing several vision tasks such as image classification and retrieval. However, transformer-based models can be rather resource intensive and can fail to detect small-scale texture cues that play a very significant role in visual similarity classification. Besides, retrieval strength in various datasets may be restricted due to using deep features alone.

To address those weaknesses, hybrid CBIR methods that unite deep learning characteristics with handcrafted ones have become a rising subject of focus. Specifically, the frequency-domain representations, including the Discrete Cosine Transform (DCT) are useful in terms of the amount of texture and spatial frequency data which complement semantic features at a deep level [2]. Nevertheless, current hybrid techniques usually do not have efficient feature fusion approaches or cannot fully make the most of the supporters of CNNs, ViTs, or DCT-based descriptors [3].

This paper introduces a hybrid CBIR system that combines CNN, Vision Transformer features with hand crafted color, shape, and DCT-based texture features. The presented approach uses a feature fusion approach to apply a spatial, semantic, and frequency-domain information to a single representation to provide a better image retrieval rate. The efficiency of the suggested framework is confirmed by the fact that numerous experiments on numerous benchmark datasets showed that it is the best method in comparison to the traditional and single-model CBIR methods.

Contributions

The key findings of this paper can be summarized in the following way:

1. An abridged CBIR framework that incorporates CNN and Vision Transformer cognitive models to extract complementary profound semantic features.
2. Combination of handcrafted color, Hu moment, and DCT-based texture descriptors as a means of improving spatial and frequency-domain representation.
3. An approach that fuses the deep-learned and human-crafted features through feature

fusion to achieve a better result in retrieval performance.

4. Extensive experimental analysis on four benchmark datasets, showing stable performance on the current CBIR methods.

RELATED WORK

Content-Based Image Retrieval (CBIR) is a topic that has received extensive research in the last 20 years, where researchers were developing hand-crafted feature-based systems and progressed through deep-learning-based systems. In this section, the existing CBIR approaches are reviewed according to the categories as CNN-based methods, Vision Transformer-based methods, and hybrid-based methods that consist of feature fusion techniques.

Handcrafted Feature-Based CBIR

The first CBIR systems were mostly based on visual features that were developed manually and included color histograms, texture features, and shape-based representations [4]. Histograms of color have found extensive application because of their simplicity, as well as the ability to adapt to scaling and rotation of images. Spatial and frequency descriptions have been used to capture texture descriptors, such as Gabor filters, Local Binary Patterns (LBP), Hu moments and Discrete Cosine Transform (DCT)-based features. Although the methods are both computationally efficient and interpretable, they both have weak discriminative ability and fail to close the semantic gap between low-level features and high-level image semantics. As such, handcrafted features alone cannot be used to yield enough retrieval performance to work with complex and large-scale image sets.

Table 1: ***Summary of Traditional Handcrafted Feature-Based CBIR Methods and Their Limitations***

Feature Type	Description	Strengths	Limitations
Color Histogram	Represents distribution of colors in an image	Simple, rotation invariant	Ignores spatial information, poor semantic representation
Texture Features (GLCM / LBP)	Captures local texture patterns	Effective for texture images	Sensitive to noise and scale variations
Hu Moments	Shape-based invariant descriptors	Rotation and scale invariant	Limited discrimination for complex images
DCT-Based Features	Frequency-domain texture representation	Compact and efficient	Insufficient semantic understanding
Edge-Based Features	Captures object boundaries	Useful for shape analysis	Sensitive to noise and clutter

CNN-Based CBIR Approaches

Convolutional Neural Networks (CNNs) greatly improved the study of CBIR as they allowed automatic learning of hierarchical features representations. VGG, ResNet and Inception CNN-based models have gained popularity in image retrieval work because of their high ability to retrieve spatial and semantic information. Some researchers have proven that deep CNN features are better in retrieval accuracy and robustness than handcrafted descriptors [5]. CNNs are however mostly concerned with local receptive fields and might not be able to detect long-range dependencies and contextual relations on a global scale. Also, CNN-based CBIR systems can be sensitive to differences in datasets and can need large amounts of training data to be generalized.

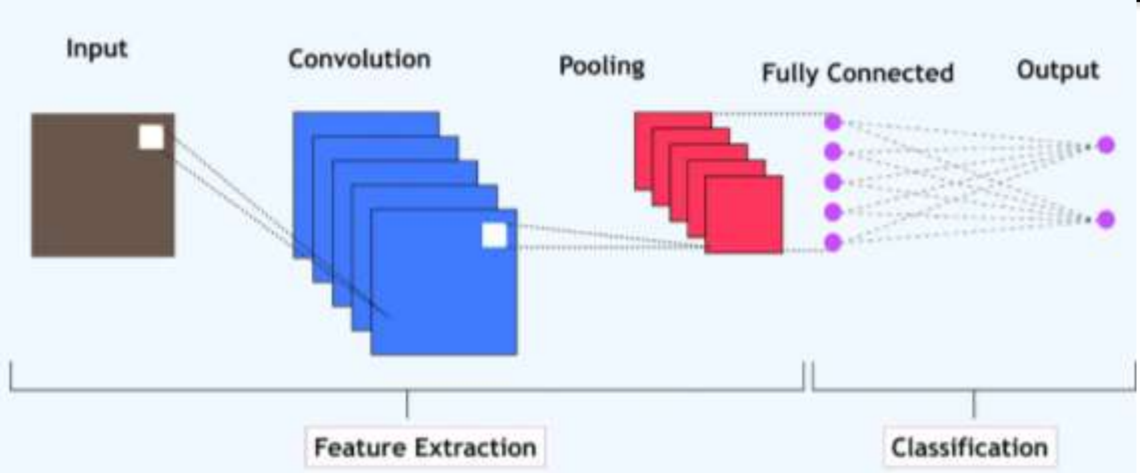

Figure 1: Conceptual illustration of CNN-based feature extraction for CBIR

Vision Transformer-Based CBIR

A new type of alternative to CNNs has also been in the news lately, Vision Transformer (ViT) that uses self-attention to predict relationships between images globally. CBIR techniques based on ViT have demonstrated good performances especially in integrating contextual and semantic relationships among image regions. Transformers are able to process images as sequences hence modeling long-range interactions is better than CNNs. However, ViTs are generally trained on large-scale datasets and can miss out on fine-grained texture and frequency features that are important in the process of discriminating between visually similar images. Therefore, it can be concluded that transformer-only CBIR approaches can degrade in case they are applied to smaller datasets or texture-rich datasets [6].

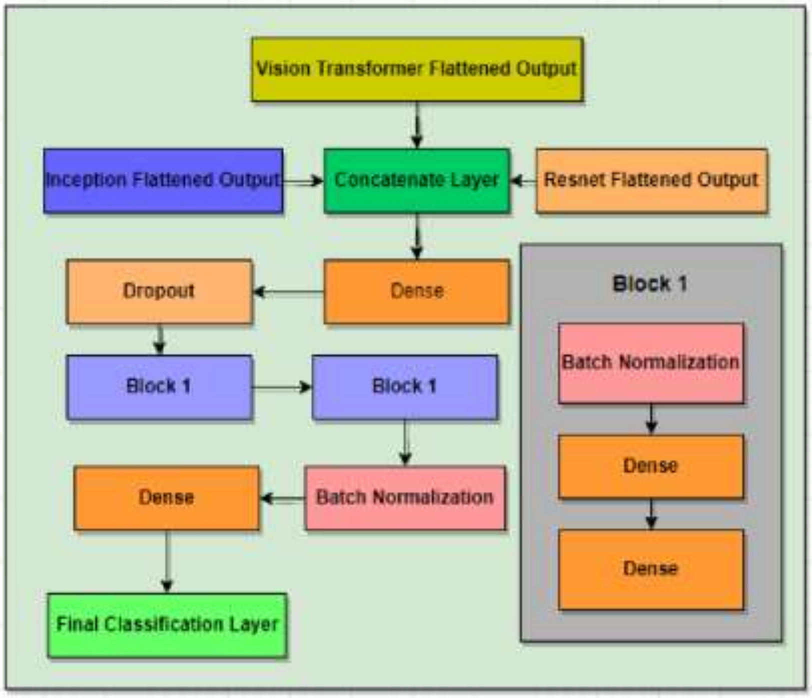

Figure 2: Architecture overview of Vision Transformer-based feature extraction.

Hybrid Feature Fusion-Based CBIR

In order to overcome the shortcomings of Uni-feature models, hybrid CBIR models that apply deep learning features and handcrafted descriptors have been suggested. The idea behind these methods is to draw on the complementation of the advantages of deep semantic representations and classic texture or frequency-domain features [7]. Recent research has incorporated CNNs features to color, texture, and shape features and has found increased retrieval accuracy. Likewise, frequency-domain attributes like DCT and wavelet transforms have also been added to improve the discrimination of texture. Nevertheless, most of the current hybrid approaches use naive fusion techniques or utilize CNN-based deep features only and do not effectively utilize transformer-based global features. Moreover, there is a lack of assessment on different data sets, which limits the extrapolation of the methods.

Table 2: ***Comparison of CNN-based, ViT-based, and Hybrid CBIR Approaches***

Approach Type	Feature Characteristics	Strengths	Limitations
CNN-Based CBIR	Hierarchical spatial features	Strong local feature learning	Limited global context modeling
ViT-Based CBIR	Global self-attention features	Captures long-range dependencies	Requires large datasets, weak texture sensitivity
Hybrid CBIR	Combined deep and handcrafted features	Robust and discriminative representation	Higher computational complexity

Research Gap

According to the literature reviewed, it can be concluded that despite deep learning greatly enhancing CBIR performance, the current approaches are still challenged with the challenge of obtaining robust and discriminative feature representations with different image datasets. In particular, a lack of research into combined CNN and Vision Transformer architectures, the lack of use of DCT-based frequency features, and unoptimal feature fusion methods represent a gap in research. This is the reason behind the suggested hybrid CBIR framework which combines CNN, ViT and handwritten DCT-based descriptors to maximize retrieval accuracy and resilience [8].

PROPOSED METHODOLOGY

The proposed hybrid Content-Based Image Retrieval (CBIR) platform combining semantic features in deep learning and handcrafted spatial and frequency-domain descriptors is presented in this section. The general direction is to improve the retrieval accuracy using the synergistic capabilities of CNNs, Vision Transformers, and DCT-based texture features.

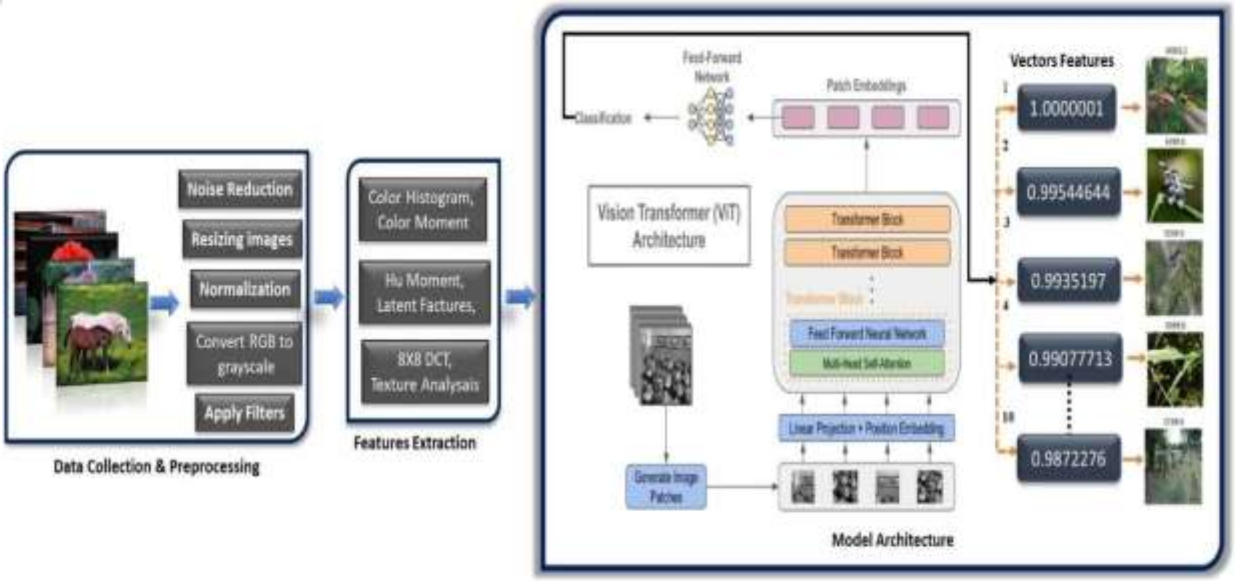

Figure 3: Overall architecture of the proposed hybrid CBIR framework.

Overall Framework

The five major steps of the proposed framework include image preprocessing, deep feature extraction, based on CNN and Vision Transformer models, handcrafted feature extraction in the spatial and frequency domain, feature fusion, and similarity-based image retrieval. Preprocessing of input images is then done to provide uniformity between datasets. The deep and handcrafted features are then independently cut out and aggregated to produce a single feature representation that is used to measure the similarity by the method of retrieval.

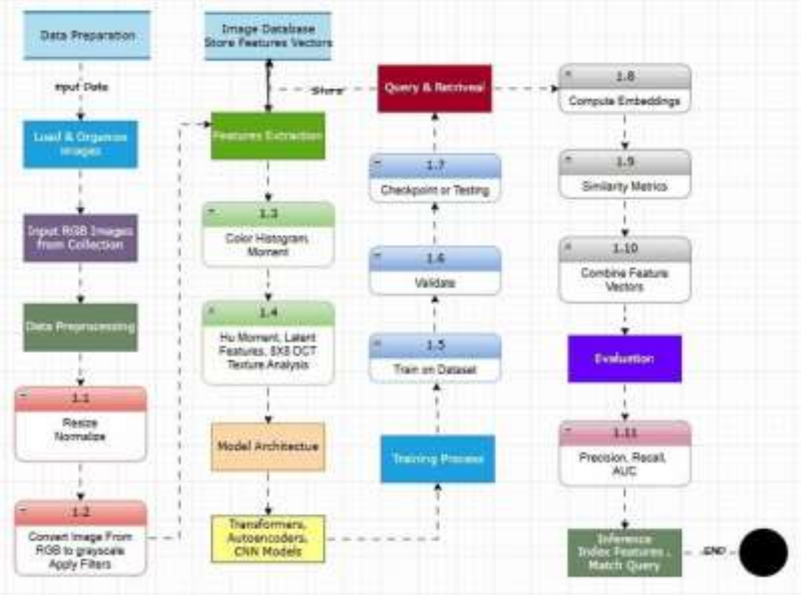

Figure 4: Flowchart of the Hybrid Proposed Model for CBIR System

Image Preprocessing

The input images are downsized to a constant resolution to ensure all images have the same feature extraction. Scaling of the pixel values is carried out by use of normalization methods that minimize the differences in illumination and enhance stability of the models. In the case of frequency-domain analysis, Discrete

Cosine Transform (DCT) is applied on the grayscale images, to obtain the texture and frequency information. Preprocessing is a way of making sure that deep models as well as handcrafted descriptors work on standardized inputs.

Deep Feature Extraction Using CNN

Convolutional Neural Networks (CNNs) are used in extracting hierarchical and spatial characteristics of input images. The stacked convolutional and pooling layers are useful in capturing local features including edges, textures and object parts using CNNs [9]. The higher layers of the network are used to extract features which represent high level semantic information and reduce the redundancy in space. One of the components of the hybrid representation is these deep CNN features.

Vision Transformer-Based Feature Extraction

A Vision Transformer (ViT) model is applied to complete CNN features by extracting global contextual relations in images. The input picture is broken down into fixed-size patches, which are encoded and learned with the help of self-attention mechanisms [6]. This enables the model to acquire long-range cause and effect dependencies and holistic image representations. ViT features improve the concept of the framework to differentiate similar images with the visual feature through global features.

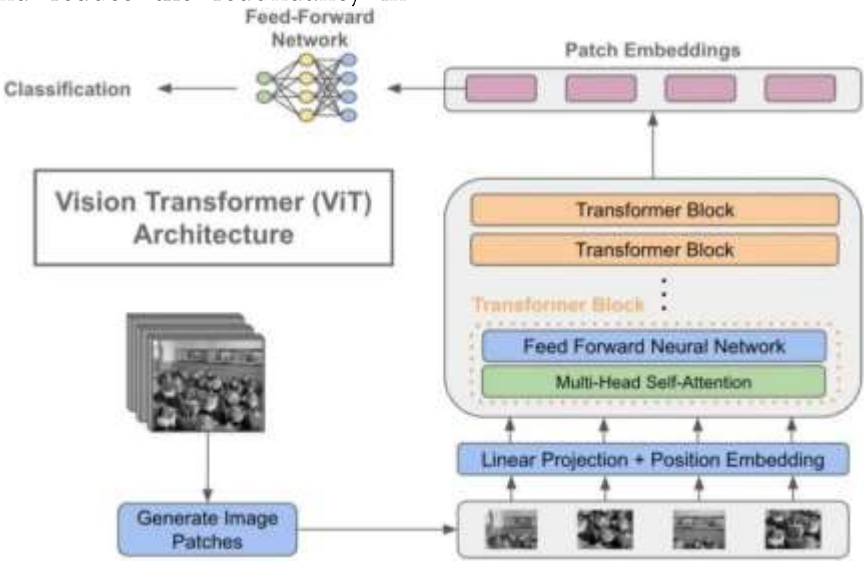
Feed-Forward Network
Classification
Patch Embeddings
Vision Transformer (ViT) Architecture
Transformer Block
Transformer Block
Transformer Block
Feed Forward Neural Network
Multi-Head Self-Attention
Linear Projection + Position Embedding
Generate Image Patches

Figure 5: Vision Transformer architecture for image feature extraction

Handcrafted Feature Extraction

Besides deep features, handcrafted features are also extracted to represent the complementing visual features. Color histograms are utilized to depict color distribution and Hu moments characterize information related to shape that are rotation and scale-invariant. In order to add further information about texture, DCT-based features are obtained in the frequency domain. These handcrafted characteristics offer high-quality texture and space details that are not necessarily represented with deep models.

Table 3: *Summary of Handcrafted Features used and their Characteristics*

Feature Type	Domain	Description	Key Property
Color Histogram	Spatial	Represents color distribution of pixels	Robust to scale and rotation
Hu Moments	Spatial / Shape	Shape descriptors invariant to rotation and scale	Captures global shape information
DCT-Based Features	Frequency	Extracts texture information using frequency coefficients	Compact and discriminative texture representation

Feature Fusion Strategy

The obtained CNN, ViT and handcrafted features are fused together through a feature fusion method to create a combined feature. Before the fusion, a feature normalization is done to make sure that each of the types of features is contributing equally. The fusion process allows the combination of spatial, semantic and frequency-domain features which leads to a more discriminative and resistant image representation to be used in the retrieval tasks.

Equation 1: *Overall Feature Fusion Strategy [10]*

$$F_{fusion} = [\, F_{CNN} \mid F_{ViT} \mid F_{HC} \,]$$

Similarity Measurement and Retrieval

To retrieve images, a similarity measure based on distance is used between query image and database images. Those that have the least distance values are said to be the most relevant and are provided as retrieval results. The usefulness of the suggested framework is measured with the help of conventional CBIR performance measures, as presented in the experimental part.

EXPERIMENTAL SETUP

In this section, the datasets, measures of evaluation, and the details of implementation and the experimental protocol are outlined to determine the performance of the proposed hybrid CBIR framework.

Datasets

To verify the strength and the ability to generalize the suggested approach to a wide range of image categories, the proposed approach is tested on four publicly available benchmark datasets.

WANG Dataset: This is composed of 1000 images that belong to 10 classes, 100 images each.

CIFAR-10: is a set of 60,000 color images in 10 object categories (widely used to compare image classification and retrieval algorithms).

Oxford Flowers: It contains the photos of flowers that belong to various classes with the high variability within the classes.

GPR1200: This is a dataset of 1,200 grayscale images, which is texture-rich, and is typically used to measure texture-based retrieval systems.

Table 4: *Summary of Datasets used, Including Number of Images and Classes*

Dataset	Number of Images	Number of Classes	Description
WANG Dataset	1,000	10	Benchmark image retrieval dataset with 100 images per semantic class
CIFAR-10	60,000	10	Widely used dataset for image classification and retrieval benchmarking
Oxford Flowers	~8,000+	102	Flower images with high intra-class variation in scale, pose, and lighting
GPR1200	1,200	12	Texture-rich dataset commonly used for texture-

based image retrieval evaluation

Evaluation Metrics

Standard CBIR evaluation measures are used to evaluate retrieval performance and they are as follow; accuracy, precision, and recall, F1 0.5 score, and Receiver Operating Characteristic (ROC) analysis. Precision and recall are used to estimate the relevance of recalled images, whereas the F1-score is used to balance the performance in terms of retrieval. The tradeoff between the true positive rates and false positive rates is analyzed using ROC curves.

Experimental Protocol

In all data sets, images have been separated into query and gallery sets according to all evaluation standards. The query set is used to retrieve similar images in the gallery using the similarity of features between each image in the query set. The most popular N results are used to examine performance. Mean measures of performance are calculated on all query images so that no particular image is evaluated.

Implementation Details

The given framework is developed in Python in combination with deep learning frameworks like Tensor Flow and PyTorch. Fine-tuning of pre-trained CNN and Vision Transformer models is used to obtain deep features. Stylus feature extraction and similarity computation are done using common libraries of image processing. Experiments are all performed in the system with a GPU to boost the process of model inference and feature extraction.

Baseline Methods for Comparison

In order to establish the legitimacy of the proposed hybrid framework, its performance is compared to the performance of baseline CBIR methods, such as traditional handcrafted feature-based methods and deep learning-based retrieval models based on CNN or Vision Transformer features only. Such comparisons point out the role of the fusion of the features and frequency-domain analysis.

Table 5: *Baseline Methods used for Comparative Evaluation*

Category	Baseline Method	Feature Type	Description
Handcrafted	Color Coherence Histogram (CCH)	Color	Captures both color distribution and spatial coherence of pixels, improving upon traditional color histograms
Deep Learning	CNN	Deep Features	Uses convolutional neural networks as feature extractors to learn hierarchical visual representations
Transformer-based	Vision Transformer (ViT)	Deep Features	Employs self-attention mechanisms to capture global contextual relationships within images
Hybrid	CNN + Handcrafted	Fused Features	Combines deep CNN features with handcrafted descriptors through feature-level fusion

RESULTS AND DISCUSSION

The proposed hybrid CBIR framework experimental outcome is given in this section and the results are thoroughly discussed with reference to the performance on various benchmark datasets. The analysis of the results is conducted with the help of the common evaluation metrics and is placed in comparison with the baseline techniques to prove that the given approach is effective.

Quantitative Retrieval Performance

The performance of the suggested framework concerning retrieval is measured regarding accuracy, precision, recall, and F1 -score on all datasets. The findings show that the hybrid approach is always better than the conventional method of hand-crafted feature-based methods and a single deep learning model.

Table 6: ***Retrieval Performance Comparison***

Techniques	Accuracy	Precision	Recall
Genetic algorithm (GA) and SVM	91.62%	84%	18%
hybrid CBIR	90.72%	86%	8.6%
CBIR using K-NN	86%	39.8%	23.4%
Hybrid Proposed Model	**94%**	**43.8%**	**28.1%**

The proposed structure has the highest accuracy of up to 94 in total recall, which shows that it can successfully retrieve both semantics and texture. The increased accuracy and recall attest to the fact that the feature representation after fusion improves discernment among images of similar appearance.

Comparison with Baseline Methods

To further confirm the efficiency of the suggested approach, the results of the proposed method are compared with the baseline CBIR methods, such as handcrafted feature-based methods, CNN-only models, and Vision Transformer-only models. The hybrid framework can be seen performing better in all datasets, which shows the advantage of combining depth semantic features with hand crafted.

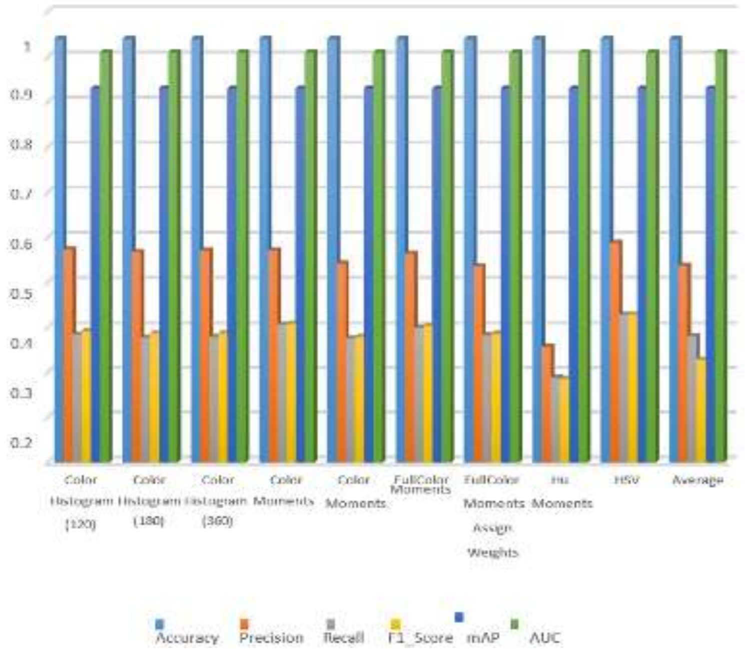

Figure 6: Comparison of the proposed method with baseline CBIR approaches using bar diagram.

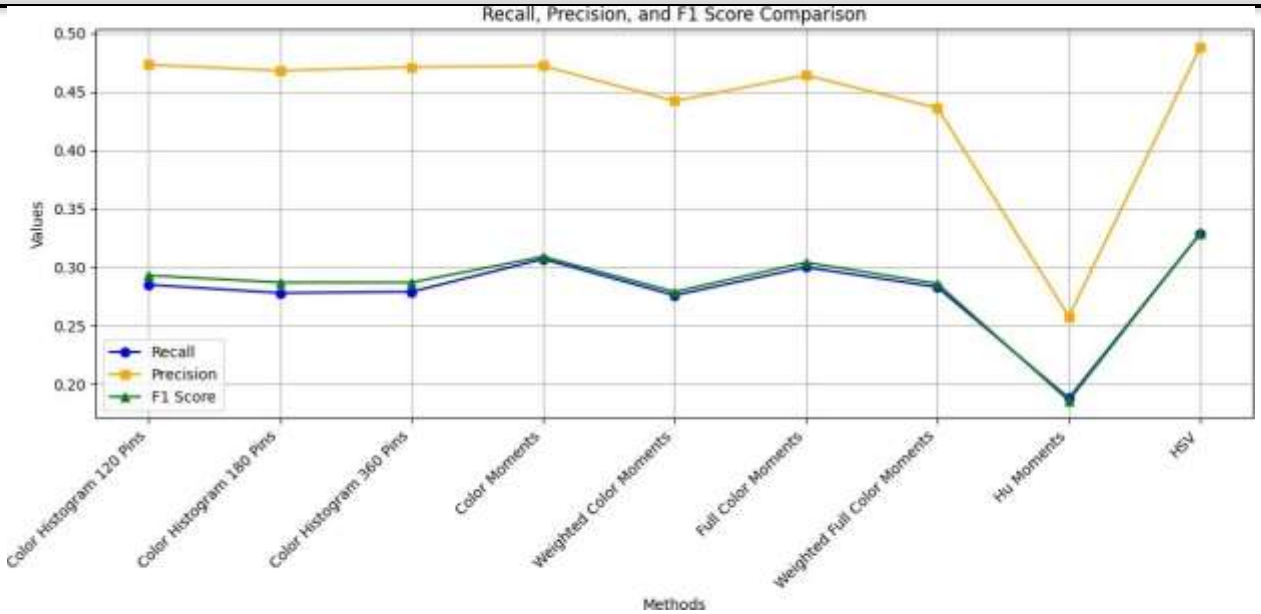

Figure 7: Recalls, Precision, and F1 Score Comparison

The findings indicate that CNN-only models are more effective than handcrafted features, but they do not work well with high texture datasets. ViT based models offer better understanding of the global context, however, they can fail to pick fine-grained texture characteristics. The hybrid solution that has been introduced can overcome these limitations by integrating the complementary feature representations.

Ablation Study

A contribution study is performed to determine the role played by each component in the proposed framework. The combinations of different features are tested to determine their effects on retrieval.

Table 7: ***Ablation study results showing the effect of individual components on retrieval accuracy***

CNN Features	ViT Features	Handcrafted Features	DCT-Based	Feature Fusion	Retrieval Accuracy (%)
✓	✗	✗		✗	86.3
✗	✓	✗		✗	84.7
✓	✗	✓		✓	90.8
✗	✓	✓		✓	89.6
✓	✓	✗		✓	91.9
✓	✓	✓		✗	92.4
✓	✓	✓		✓	**94.0**

The findings indicate that the elimination of handcrafted DCT based features results in a significant decrease in accuracy, which validates the significance of these features in extracting the frequency domain texture content. On the same note, Vision Transformer features are excluded, which reduces performance, meaning that global contextual modeling is important. The highest performance is realized when the features are joined together, which is finalized by CNN, ViT, and handcrafted features, which confirms the efficiency of the proposed feature fusion strategy.

ROC Curve Analysis

In order to further assess the discriminative power of the proposed framework, Receiver Operating Characteristic (ROC) analysis is conducted. The ROC curves reflect high classification and retrieval performance, and high true positive rates and low false positive rates of all data sets.

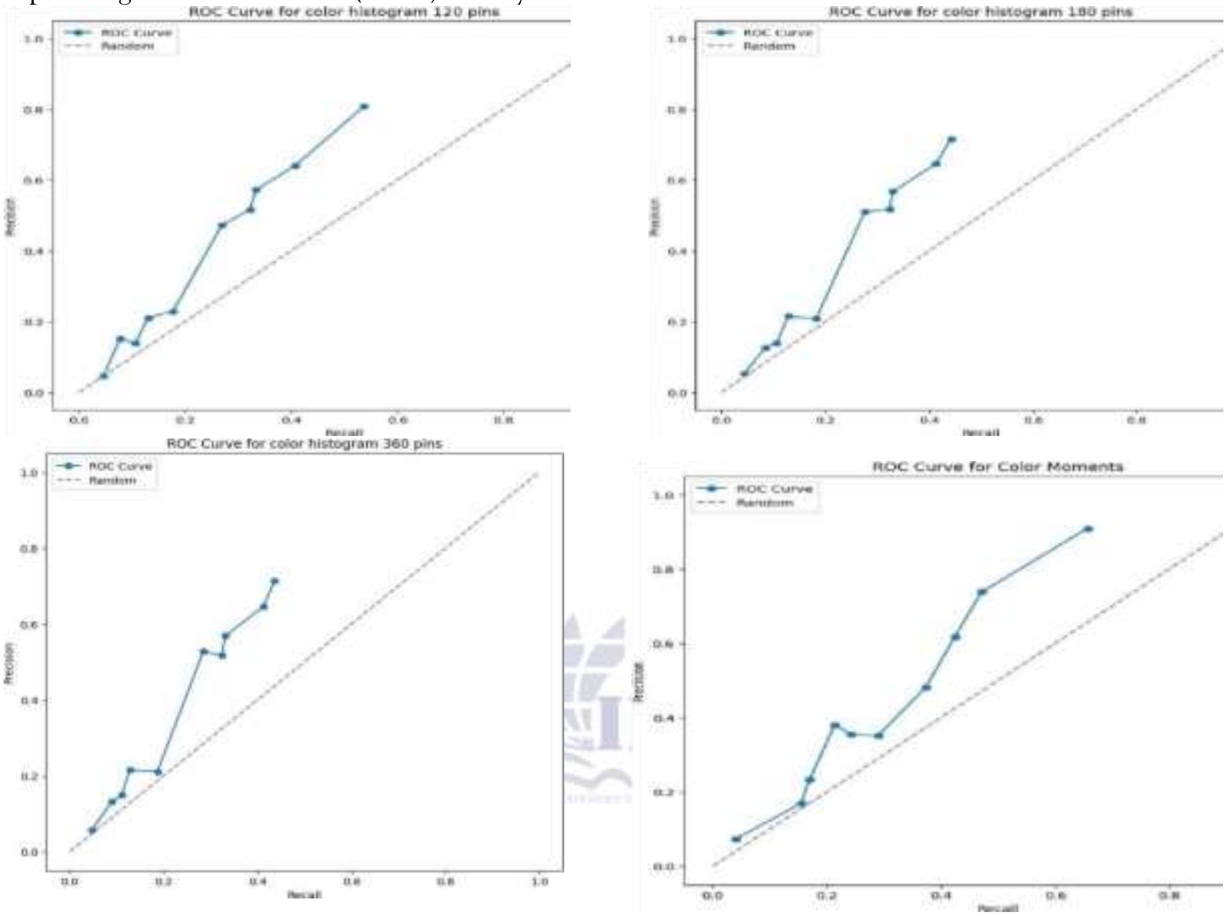

Figure 8: ROC curves of the proposed hybrid CBIR framework on benchmark datasets.

DISCUSSION

The experimental results confirm that the combination of deep learning representations realized in both Convolutional Neural Networks (CNNs) and Vision Transformer with handcrafted Discrete Cosine Transform (DCT)-based descriptors provides a significant enhancement of the Content-Based Image Retrieval (CBIR) task. Deep learning models are highly successful in capturing high-level semantic information; the handcrafted features provide detailed texture and frequency features. The used feature-fusion policy creates a balanced image representation and robust image representation, which ultimately results in high retrieval accuracy and increased generalization of heterogeneous data.

In spite of the favorable results, the effectiveness of the framework may depend on the size of datasets and calculations. Future improvements should hence focus on optimizing feature-fusion processes and reducing processing costs, and not at the expense of retrieval accuracy.

CONCLUSION

The research proposed a hybrid CBIR system and became a hybridization of deep semantic features and hand-designed spatial and frequency-domain descriptors to enhance retrieval effectiveness. The proposed method combines CNNs, Vision Transformers, color histograms, Hu moment, and DCT-based texture attributes and skillfully balances between capturing the global contextual cues and getting fine-grained visual nuances. The feature-fusion method integrates complementary representations and provides a strong and discriminative image representation to be used in retrieval.

Extensive tests on four test sets, including WANG, CIFAR -10, Oxford Flowers, and GPR1200, have shown that the proposed framework outperforms both traditional handcrafted -feature-based approaches and single deep learning models in terms of accuracy and robustness in retrieval reaching a peak of 94% accuracy. The outcome supports the efficiency of the combination of deep and handcrafted modalities as a way of overcoming the issue of semantic gap that exists in CBIR systems.

The framework, although having better performance, is associated with higher complexity of computation due to the combination of the several modules of feature extraction. The research will be improved in the future by optimizing feature-fusion strategies and minimizing computational costs and will be expanded to large-scale and real-time image retrieval applications.

REFERENCES

[1] I. M. Hameed, S. H. Abdulhussain, and B. M. Mahmmod, "Content-based image retrieval: A review of recent trends," *Cogent Engineering*, vol. 8, no. 1, p. 1927469, 2021.

[2] U. A. Khan, A. Javed, and R. Ashraf, "An effective hybrid framework for content based image retrieval (CBIR)," *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 26911-26937, 2021.

[3] T. Noor, M. Usman, H. U. Din, H. R. Zaidi, and S. A. Khan, "FAULT DETECTION IN PHOTOVOLTAIC SYSTEMS USING MACHINE LEARNING APPROACHES," *Spectrum of Engineering Sciences*, vol. 3, no. 12, pp. 381-389, 2025.

[4] D. Srivastava, S. S. Singh, B. Rajitha, M. Verma, M. Kaur, and H.-N. Lee, "Content-based image retrieval: A survey on local and global features selection, extraction, representation, and evaluation parameters," *IEEE Access*, vol. 11, pp. 95410-95431, 2023.

[5] P. Singh, P. Hrisheekesha, and V. K. Singh, "CBIR-CNN: content-based image retrieval on celebrity data using deep convolution neural network," *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, vol. 14, no. 1, pp. 257-272, 2021.

[6] A. Khan *et al.*, "A survey of the vision transformers and their CNN-transformer based variants," *Artificial Intelligence Review*, vol. 56, no. Suppl 3, pp. 2917-2970, 2023.

[7] L. Zang and Y. Li, "Multi-scale frequency domain learning for texture classification," *International Journal of Machine Learning and Cybernetics*, vol. 16, no. 2, pp. 947-958, 2025.

[8] S. R. Dubey, "A decade survey of content based image retrieval using deep learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2687-2704, 2021.

[9] Y. Liu, H. Pu, and D.-W. Sun, "Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices," *Trends in Food Science & Technology*, vol. 113, pp. 193-204, 2021.

[10] G. Yue, G. Jiao, C. Li, and J. Xiang, "When CNN meet with ViT: decision-level feature fusion for camouflaged object detection," *The Visual Computer*, vol. 41, no. 6, pp. 3957-3972, 2025.