A Hybrid Content-Based Image Retrieval System for Super-Resolution Images Using Deep and Hand-Crafted Feature Fusion

D Sahith[1], K Sumanth[1], L Paul Prakash[1], N Sumanth[1], K. Venkataramana[2], Dr. S Shiva Prasad[3]
[1]*Student, Department of CSE(DataScience), MallaReddy Engineering college, Secunderabad*
[2]*Assistant Professor, Department of CSE(DataScience), MallaReddy Engineering college, Secunderabad*
[3]*Professor, Department of CSE(DataScience), MallaReddy Engineering college, Secunderabad*

***Abstract*— Content-Based Image Retrieval (CBIR) has emerged as a crucial technique for efficiently retrieving relevant images from large-scale visual databases. However, existing CBIR systems often rely solely on either deep learning-based features or traditional hand-crafted descriptors, resulting in limited retrieval accuracy due to the semantic gap between low-level visual features and high-level human perception. This paper proposes an enhanced hybrid CBIR framework specifically designed for super-resolution images by integrating deep semantic features extracted using InceptionV3 (GoogLeNet) with complementary hand-crafted features, namely Modified Dot Diffusion Block Truncation Coding (DDBTC), Histogram of Oriented Gradients (HOG), and Gray-Level Co-occurrence Matrix (GLCM). To improve feature quality, images from the VISTEX and STEX datasets are first enhanced using INTER-CUBIC interpolation for super-resolution. The extracted deep and hand-crafted features are fused into a unified representation, and similarity matching is performed using Euclidean distance. Experimental results demonstrate that the proposed hybrid approach significantly improves precision, recall, and F-measure when compared to standalone deep learning-based CBIR models, highlighting its effectiveness for high-resolution image retrieval applications.**

***Index Terms*— Content-Based Image Retrieval, Super-Resolution, Deep Learning, GoogLeNet, Feature Fusion, Texture Analysis.**

I. INTRODUCTION

The rapid growth of digital imaging devices and multimedia technologies has led to an exponential increase in image data across various application domains such as medical diagnosis, textile design, remote sensing, and digital libraries. Efficient retrieval of relevant images from such massive databases has become a critical research challenge. Content-Based Image Retrieval (CBIR) systems aim to address this challenge by retrieving images based on intrinsic visual features such as color, texture, and shape, rather than relying on textual annotations.

Traditional CBIR techniques primarily depend on hand-crafted feature descriptors, which are effective in capturing low-level visual patterns but fail to represent high-level semantic information. Recent advances in deep learning have significantly improved image representation by learning hierarchical features directly from data. However, deep learning-based CBIR systems may overlook subtle texture and spatial details that are crucial in certain applications. This motivates the need for a hybrid approach that combines the strengths of both deep learning and hand-crafted features.

1.1. Objective of the Proposed Work

The primary objective of this work is to design a hybrid CBIR system that effectively combines:

1. Deep Learning Features for capturing high-level semantic information (e.g., object categories and scene context).
2. Hand-Crafted Features for capturing low-level texture, edge, and spatial patterns.

This fusion aims to bridge the semantic gap and enhance retrieval accuracy for super-resolution images.

1.2 Role of Super-Resolution in CBIR

Super-resolution (SR) refers to the process of enhancing the spatial resolution of an image by reconstructing high-frequency details that are missing in low-resolution images. In CBIR systems, higher resolution images enable more discriminative feature extraction, leading to improved retrieval accuracy. However, increasing image resolution also introduces challenges such as higher computational complexity and feature dimensionality. In this work, INTER-CUBIC interpolation is employed as a super-resolution technique due to its ability to preserve image smoothness while maintaining computational efficiency. By enhancing image resolution prior to feature extraction, the proposed system ensures improved representation of both global semantics and local texture patterns.

II. LITERATURE SURVEY

Content-Based Image Retrieval (CBIR) has been extensively studied as a method for retrieving similar images based on visual content rather than textual metadata. Early research in CBIR explored traditional hand-crafted features such as Scale-Invariant Feature Transform (SIFT) and Local Binary Patterns (LBP) to describe visual content [1]. SIFT provides robust keypoint detection and descriptor representation, but its performance degrades when dealing with high intra-class variability and semantic complexity [1]. Similarly, LBP captures texture patterns effectively but lacks semantic understanding of image content, creating a semantic gap between low-level feature representation and high-level human perception [2]. To overcome this gap, deep learning models, particularly convolutional neural networks (CNNs), have been adopted for feature learning. CNNs automatically learn hierarchical features directly from image data, which have shown superior performance in complex recognition and retrieval tasks [3], [4]. Models such as GoogLeNet (Inception architecture) incorporate multi-scale filters to capture detailed semantic information while maintaining computational tractability [4]. However, deep learning models tend to overlook subtle database-specific texture discrepancies, especially in domains such as textile, material analysis, and fine-grained image classes, which motivates the integration of hand-crafted features [5].

Super-resolution (SR) techniques enhance the visual quality and resolution of images, providing richer detail for subsequent analysis. Interpolation-based methods like INTER-CUBIC offer a balance between computational cost and image detail enhancement, making them suitable for pre-processing large image datasets [6]. While SR improves the feature extraction capability for both hand-crafted and deep features, it also increases the complexity of feature matching due to higher dimensional representations [6]. Hybrid CBIR frameworks aim to combine deep semantic features with complementary hand-crafted descriptors to bridge the semantic gap and enhance retrieval accuracy. For instance, fusion approaches incorporating deep features with texture or local descriptors have shown improved performance compared to standalone models [7], [8]. Such fusion leverages the semantic abstraction from deep models and the fine-grained details captured by hand-crafted descriptors like Histogram of Oriented Gradients (HOG) and Gray-Level Co-occurrence Matrix (GLCM). Research in hybrid descriptors has also explored Modified Block Truncation Coding (DDBTC) for improved color and luminance representation, which adds discriminative power to texture representation when combined with deep features [9]. Prior works demonstrate that feature fusion enhances precision and recall metrics over traditional CBIR systems, pointing to the potential benefits of hybrid systems for complex retrieval scenarios [7]– [9]. In summary, the literature suggests that while deep CNN models excel at capturing semantic image information, they can benefit from the inclusion of hand-crafted features that retain detailed structural and textural characteristics. The challenge remains in designing efficient feature fusion frameworks that can exploit the complementary strengths of both classes of features for super-resolution and high-quality image retrieval tasks.

III. PROPOSED HYBRID CBIR SYSTEM

A. System Overview
The proposed CBIR system follows a modular architecture consisting of image super-resolution, feature extraction, feature fusion, and similarity matching stages. A block-level representation of the system ensures flexibility and scalability.

B. Image Super-Resolution
All input images are first upscaled using INTER-CUBIC interpolation. This step enhances pixel density

and improves the visibility of fine details, which is beneficial for both deep and hand-crafted feature extraction.

C. Deep Feature Extraction

High-level semantic features are extracted using InceptionV3 (GoogLeNet). The Inception architecture utilizes parallel convolutional filters of varying sizes, enabling the capture of multi-scale information while reducing the number of parameters. This makes the model efficient and suitable for large-scale CBIR systems.

D. Hand-Crafted Feature Extraction

To complement deep features, three hand-crafted descriptors are employed:

DDBTC: Captures color and luminance information efficiently.

HOG: Extracts edge orientation and shape-related features.

GLCM: Analyzes texture by modeling spatial relationships between pixel intensities.

These descriptors provide detailed local and structural information that may not be fully captured by deep networks.

E. Feature Fusion and Similarity Matching

The extracted deep and hand-crafted features are concatenated into a single feature vector. Similarity between the query image and database images is computed using Euclidean distance, and images with the smallest distance values are retrieved as the most similar results.

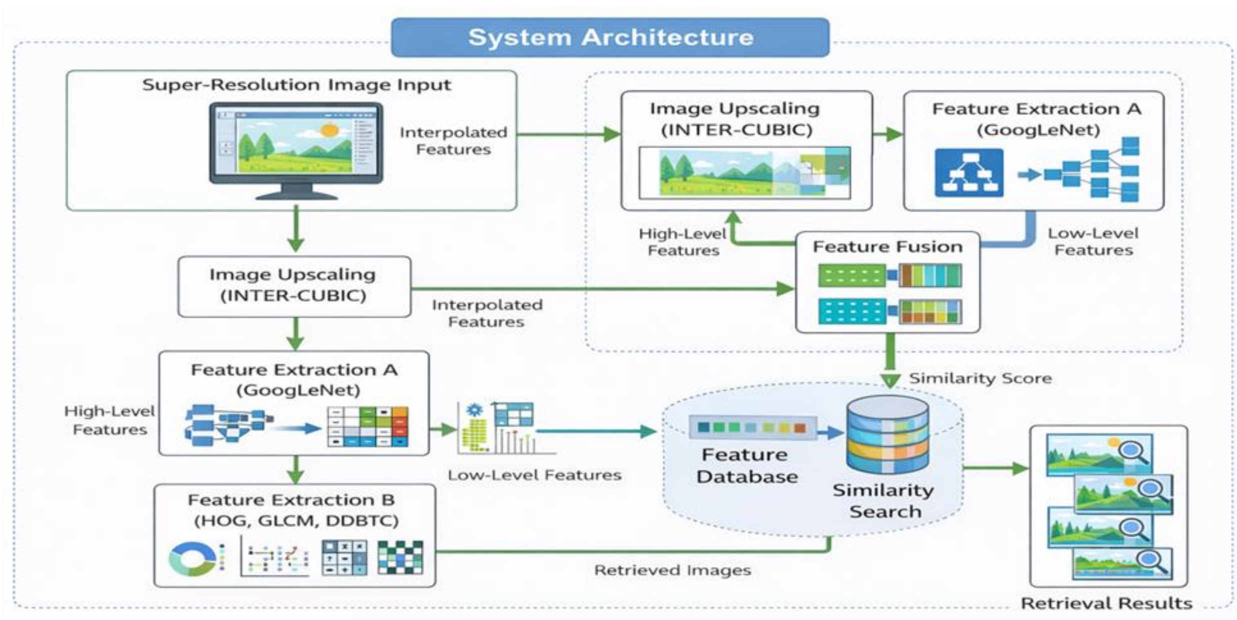

Figure.1 System Architecture

3.1 Models Used

1. InceptionV3 (GoogLeNet)

InceptionV3 is a deep convolutional neural network designed to efficiently capture multi-scale semantic features. It employs parallel convolutional filters of different sizes (1×1, 3×3, 5×5) within the same layer, enabling the model to learn both local and global patterns simultaneously. In this work, InceptionV3 is used as a feature extractor, where deep features from intermediate layers represent high-level semantic information such as object structure and visual context. Its computational efficiency and strong representational power make it well-suited for large-scale CBIR systems.

2. DDBTC (Dot Diffusion Block Truncation Coding)

DDBTC is a modified version of Block Truncation Coding that efficiently represents color and luminance information. It divides an image into blocks and applies dot diffusion to preserve texture and edge details. DDBTC is particularly effective for capturing fine-grained texture patterns and local intensity variations, complementing the semantic features extracted by deep networks.

3. Histogram of Oriented Gradients (HOG)
HOG captures edge and shape information by computing gradient orientation histograms over localized regions of an image. It is robust to illumination changes and effectively represents object contours and structural patterns. In the proposed system, HOG contributes discriminative shape-related information that may be overlooked by deep models.

4. Gray-Level Co-occurrence Matrix (GLCM)
GLCM is a statistical texture descriptor that analyzes the spatial relationship between pixel intensities. It extracts features such as contrast, correlation, energy, and homogeneity, which are essential for distinguishing textures in high-resolution images. GLCM enhances the system's ability to discriminate visually similar images with subtle texture differences.

5. Feature Fusion Strategy
The deep features and hand-crafted descriptors are concatenated into a unified feature vector. This feature-level fusion leverages the complementary strengths of both approaches semantic abstraction from CNNs and fine texture details from traditional descriptors resulting in improved retrieval accuracy.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Evaluation Metrics
The system performance is evaluated using standard CBIR metrics, including Precision, Recall, and F-Measure. These metrics assess retrieval accuracy and relevance.

B. Performance Comparison
Table.1 Performance Comparison of Deep Learning Baseline and Proposed Hybrid CBIR Model

Model	Precision (%)	Recall (%)	F-Measure (%)
Deep Learning Baseline	82	78	80
Proposed Hybrid Model	94	91	92.5

The table presents a comparative analysis of the retrieval performance between the deep learning baseline model and the proposed hybrid CBIR model using standard evaluation metrics: Precision, Recall, and F-Measure. The deep learning baseline achieves a precision of 82%, recall of 78%, and an F-measure of 80%, indicating reasonable retrieval performance based solely on deep semantic features.
In contrast, the proposed hybrid model significantly improves performance, achieving 94% precision, 91% recall, and an F-measure of 92.5%. This improvement is attributed to the effective fusion of deep learning features with hand-crafted descriptors, which enhances both semantic understanding and fine-grained texture representation. The results clearly demonstrate that the hybrid approach reduces the semantic gap and provides more accurate and reliable image retrieval compared to the standalone deep learning model.

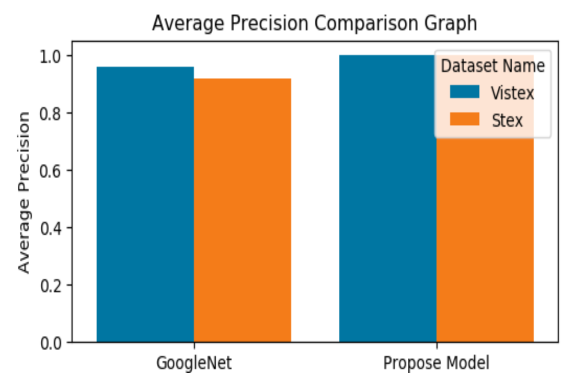

Figure.2 Comparison Graph

V. CONCLUSION

This paper presented a hybrid Content-Based Image Retrieval (CBIR) framework designed for super-resolution images by effectively combining deep learning-based semantic features with hand-crafted texture descriptors. By applying INTER-CUBIC interpolation, the image resolution was enhanced prior to feature extraction, enabling richer visual detail representation. High-level semantic features extracted using InceptionV3 (GoogLeNet) were fused with complementary hand-crafted features including DDBTC, HOG, and GLCM, thereby addressing the semantic gap inherent in conventional CBIR systems.Experimental evaluation on benchmark texture datasets demonstrated that the proposed hybrid model significantly outperforms the standalone deep learning baseline in terms of precision, recall, and F-measure. The results confirm that integrating deep semantic representations with low-level texture and structural information leads to more accurate and reliable image retrieval.In future work, advanced GAN-based super-resolution techniques can be incorporated to further enhance image quality, and learning-based similarity measures or metric learning approaches can be explored to improve matching performance in high-dimensional feature spaces.

REFERENCES

[1] R. Datta, D. Joshi, J. Li, and J. Z. Wang, “Image retrieval: Ideas, influences, and trends of the new age,” ACM Computing Surveys, vol. 40, no. 2, pp. 1–60, 2008.

[2] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[3] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” Proc. ICLR, 2015.

[4] C. Szegedy et al., “Going deeper with convolutions,” Proc. IEEE CVPR, pp. 1–9, 2015.

[5] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” Proc. IEEE CVPR, pp. 886–893, 2005.

[6] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3, no. 6, pp. 610–621, 1973.

[7] Z. Wang et al., “Deep learning for image super-resolution: A survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 10, pp. 3365–3387, 2021.

[8] J. Wang, J. Li, and G. Wiederhold, “SIMPLIcity: Semantics-sensitive integrated matching for picture libraries,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 9, pp. 947–963, 2001.

[9] K. V. Ramana et al., “Modified block truncation coding for texture-based image retrieval,” International Journal of Computer Applications, vol. 54, no. 8, pp. 1–6, 2012.