



## Adaptive Feature Fusion: A Hybrid Deep Learning-Based Enhanced CBIR for Remote Sensing Data

<sup>1</sup>Nidhi Prajapati, <sup>2</sup>Naimish Patel

<sup>1</sup>Lecturer, <sup>2</sup>Asst. Prof.

<sup>1</sup>Computer Engineering Department,

<sup>1</sup>Silver Oak University, Ahmedabad, India

**Abstract:** The exponential growth of satellite and aerial imagery necessitates advanced techniques for efficient image retrieval to address critical applications such as land-use planning, environmental monitoring, and disaster management. Content-Based Image Retrieval (CBIR), which retrieves images based on visual content rather than metadata, is a promising solution but faces challenges when applied to remote sensing data, including high spatial complexity, spectral diversity, and variability in land cover. This study introduces Adaptive Feature Fusion (AFF), a novel hybrid deep learning framework designed to enhance CBIR for remote sensing data. The AFF framework integrates traditional handcrafted features—such as Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT), and Gabor filters—with deep features extracted using ResNet50, leveraging an adaptive weighting mechanism to optimize feature fusion. An attention layer refines image representation, focusing on key regions to improve retrieval precision and robustness. By systematically comparing traditional and deep learning-based approaches on the UC Merced Land Use/Land Cover (LULC) dataset, our experiments demonstrate the AFF model's superior performance in relevance, scalability, and generalization. This hybrid approach not only bridges the gap between handcrafted and deep features but also establishes a state-of-the-art framework for effective image retrieval in complex remote sensing scenarios..

**IndexTerms – Image Processing, Feature Extraction, CBIR, Deep Learning, Feature Fusion**

### 1. INTRODUCTION

The unprecedented growth in remote sensing imagery, driven by advancements in satellite and aerial imaging technologies, has created a pressing need for efficient systems capable of retrieving relevant images from massive datasets. These images are critical in diverse applications, including landuse classification, environmental monitoring, urban development, and disaster response. However, traditional image retrieval approaches, relying on metadata and manual tagging, fall short in addressing remote sensing data's inherent complexity and diversity, necessitating a paradigm shift toward Content-Based Image Retrieval (CBIR). CBIR offers a transformative solution by retrieving images based on their visual content rather than relying solely on textual descriptions or metadata. Yet, remote sensing CBIR systems face challenges such as the high spatial complexity of images, the diverse range of land cover types, and dynamic conditions like temporal, seasonal, and geographical variations. Traditional feature extraction methods like Local

Binary Patterns (LBP), Gabor filters, and Scale-Invariant Feature Transform (SIFT) excel in capturing textures and patterns but struggle with large-scale datasets and complex representations[28][19]. Conversely, deep learning models, such as Convolutional Neural Networks (CNNs), can extract high-level semantic features. Yet, they often encounter challenges like overfitting, computational resource demands, and reduced performance in domain-specific scenarios [26]. To address these gaps, this research introduces the Adaptive Feature Fusion CBIR (AFF-CBIR) framework, a hybrid approach that combines the strengths of handcrafted and deep learning-based features. By leveraging an adaptive weighting mechanism and incorporating an attention layer, AFF-CBIR emphasizes critical image regions and provides robust feature representations that are scalable, accurate, and generalizable. The integration of low-level textures with high-level semantics enables the proposed model to overcome the limitations of traditional and standalone deep learning methods.

#### 1.1. Motivation

The complexity and increasing availability of remote sensing imagery underline the urgency for an advanced CBIR system. This study is motivated by key challenges in the field, including:

- **Managing high-resolution and complex datasets:** Remote sensing images contain intricate spatial, spectral, and textural details necessary for distinguishing land cover types like water bodies, vegetation, and urban areas, which metadata-based systems fail to capture effectively.
- **Addressing data heterogeneity and dynamics:** Variations in geographical regions, seasons, and spectral bands require adaptive retrieval mechanisms to accommodate these changes.

- **Enabling timely and accurate analysis:** Environmental and disaster management applications require real-time image retrieval for decision-making, so scalable and precise systems become imperative.
- **Leveraging advances in deep learning:** The emergence of CNNs and transfer learning offers opportunities to integrate complex feature representations with existing frameworks for enhanced retrieval performance

### 1.2. *Research Scope*

This research explores the integration of handcrafted and deep features to enhance Content-Based Image Retrieval (CBIR) for remote sensing images. Specifically, it focuses on developing a hybrid feature extraction pipeline that combines traditional techniques such as Local Binary Patterns (LBP), Gabor filters, and Scale-Invariant Feature Transform (SIFT) with deep learning-based features, notably from ResNet50. To ensure effective feature representation, an adaptive feature fusion mechanism is implemented, enabling dynamic weighting of feature contributions for robust and context-sensitive retrieval. The proposed framework is evaluated using the UC Merced Land Use and Land Cover (LULC) dataset[16], a widely recognized benchmark for remote sensing image retrieval, to measure its performance in terms of precision, recall, mean average precision (mAP), and computational efficiency. Additionally, the study investigates the framework's applicability in practical scenarios, including land-use classification, environmental monitoring, disaster response, and urban planning, emphasizing its relevance in addressing real-world challenges posed by complex and large-scale datasets

### 1.3. *Research Objective*

The primary objective of this research is to develop an efficient and scalable Adaptive Feature Fusion CBIR (AFFCBIR) framework for remote sensing images. The framework aims to enhance feature extraction by effectively representing complex image datasets through a balanced integration of low-level handcrafted features and high-level deep features. This hybrid approach is designed to achieve robust and accurate retrieval, improving relevance, generalization, and scalability for diverse and dynamic remote sensing applications. Furthermore, the study seeks to address computational challenges, enabling the practical deployment of CBIR systems in critical domains such as environmental monitoring, urban development, and disaster management.

The paper is structured as Section 2 represents the Literature Review while Section 3 consists of Proposed Methodology followed by Novelty of Proposed AFF-CBIR model in Section 4. Section 5 consists of Implementation and result analysis while section 6 is for the conclusion and future work.

## 2. LITERATURE REVIEW

The evolution of Content-Based Image Retrieval (CBIR) in the domain of remote sensing has been marked by advancements in feature extraction methodologies[7]. These range from traditional handcrafted techniques to modern deep learning approaches, both addressing challenges related to high spatial complexity and diverse datasets in remote sensing

### 2.1. *CBIR using Handi crafted Features*

Handcrafted feature extraction methodologies have been instrumental in establishing the groundwork for Content-Based Image Retrieval (CBIR) systems by utilizing low-level characteristics such as texture, color, and shape for the representation of image content. Local Binary Patterns (LBP) have been extensively employed to extract texture-based features. Although proficient in encoding local spatial information via binary patterns, LBP is inadequate in capturing global and semantic features, thus constraining its utility in the context of complex remote sensing images. Gabor filters, recognized for their efficacy in analyzing spatial frequency and capturing localized texture variations, perform adequately on simpler patterns but encounter challenges in adapting to the complexity and heterogeneity found in remote sensing datasets. Similarly, the Scale-Invariant Feature Transform (SIFT), esteemed for its transformation robustness, presents computational intensity that poses challenges in managing high resolution images prevalent in remote sensing applications.

The Bag of Words (BoW) model, which conceptualizes features as textual representations, offers succinct descriptors of image content. Nonetheless, it frequently neglects the spatial relationships and minutiae inherent in remote sensing imagery. Despite their foundational role, these traditional methods lack the capacity to handle high-dimensional, large-scale datasets, and their limited ability to extract abstract, semantic features necessitated a shift toward learning-based approaches. The detailed comparison of CBIR techniques for remote sensing images is discussed in Table 1

### 2.2. *CBIR using Deep Feature Learning*

Modern CBIR systems leverage deep learning to extract hierarchical feature representations, thereby overcoming the limitations of handcrafted techniques. Convolutional Neural Networks (CNNs) serve as the backbone for many CBIR advancements. Their ability to automatically learn hierarchical representations enables them to capture complex patterns present in remote sensing images.[10][27] Specifically, models like ResNet50 are celebrated for their skip connections, allowing deeper architectures without performance degradation, which improves retrieval accuracy in diverse remote sensing datasets. Transfer learning has further enhanced CBIR systems by fine-tuning pretrained models like ResNet and VGG on domain-specific tasks. This has proven particularly beneficial in remote sensing, where labeled datasets are scarce and computational costs are a concern.[30] Emerging methods such as deep hashing optimize retrieval efficiency by encoding high-dimensional features into compact binary representations. This approach significantly reduces storage and computational requirements, making it ideal for large-scale remote sensing applications. Moreover, recent studies have explored hybrid methods that combine handcrafted and deep features to harness the strengths of both. For instance, Walia in [11] demonstrated that integrating handcrafted and deep features achieves improved accuracy in applications requiring fine-grained analysis, such as forgery detection and remote sensing. Similarly, Lagani in [29] utilized bio-inspired Hebbian learning as a pretraining step for neural networks to better utilize scarce labelled data, highlighting the potential of semi-supervised hybrid approaches

### 2.3. Research Challenges

Despite significant advancements, CBIR systems face challenges in scalability, feature generalization, and efficiency. Traditional methods, though computationally lightweight, fail to generalize well to diverse datasets, while deep learning-based systems demand substantial labeled data and computational resources. Future research should focus on integrating advanced fusion techniques, improving training with semisupervised and unsupervised learning, and scaling to handle multi-temporal, high-dimensional remote sensing data.

Paper	Techniques Used	Key Findings	Challenges
Janjua & Patankar (2022)	Traditional CBIR, Deep Learning	Comparative review of CBIR methods, highlighting the strengths of deep learning over traditional methods	Traditional methods lack scalability for large datasets
Pandey et al. (2021)	Remote Sensing Data, Classifiers	Review of land use/land cover (LULC) classification and the importance of data sources and classifiers	Difficulty in integrating multi-source data
Zhou et al. (2023)	Deep Learning, CNNs	Achievements in remote sensing image retrieval and challenges like large-scale dataset handling	Complexity in handling multi-temporal and multi-spectral data
Imbriaco et al. (2019)	Aggregated Deep Local Features	Aggregation of deep features improves retrieval performance by combining local feature descriptors	Requires large training datasets for deep models
Alrahhah & Supreethi (2019)	LBP, SVM, and Machine Learning	Combination of LBP features with supervised learning techniques for better texture-based retrieval	Limited effectiveness on high-dimensional data
Tadi Bani & Fekri-Ershad (2019)	Texture and Colour Features	Combination of texture and colour information extracted in spatial and frequency domains improves retrieval	Computationally expensive due to multi-domain processing
Sun et al. (2021)	Deep Hashing, Multisensor Fusion	Use of multisensor fusion and deep hashing for cross-modal image retrieval	Handling of heterogeneous sensor data remains challenging
Hou et al. (2022)	Attention Mechanisms, Multiscale CNN	Attention-enhanced deep learning models for multiscale feature learning in remote sensing retrieval	Model complexity and training time are high
Papoutsis et al. (2023)	Deep Learning, CNNs	Benchmarking deep learning models for land cover classification, highlighting their scalability	Scalability and real-time processing challenges

Table 1: Literature Review on recent technologies used for CBIR of Remote Sensing Images

### 3. PROPOSED METHODOLOGY

The proposed Adaptive Feature Fusion Content-Based Image Retrieval (AFF-CBIR) algorithm integrates both handcrafted and deep learning-based features with an adaptive fusion mechanism. The model is optimized to efficiently and accurately retrieve relevant images from complex remote sensing datasets, such as the UC Merced Land Use and Land Cover (LULC) dataset. The objective is to leverage data augmentation, feature fusion, and attention mechanisms to create a robust image representation, thereby improving retrieval accuracy, efficiency, and interpretability.

Each image in the dataset is initially resized to a standard resolution of 256 X 256 pixels with 3 color channels (RGB). This resizing step ensures consistent dimensions across images, making them compatible with the feature extraction methods employed in the pipeline.

To enhance the generalization capability of the model, data augmentation techniques are applied. These techniques include random rotations, flips, and scaling. This step introduces variability in the dataset, simulating real-world conditions while maintaining the original image size of (256 X 256 X 3). The augmentation process helps reduce the risk of overfitting, improving robustness against diverse capture conditions.

#### 3.1. Traditional Feature Extraction

After preprocessing, the next step involves extracting low-level features that capture essential image characteristics. The following handcrafted features are computed:

- **Local Binary Patterns (LBP):** LBP captures texture information, producing a histogram that typically compresses spatial details into a fixed-length vector (e.g.,  $1 \times 256$  or  $1 \times 512$ , depending on parameters).
- **Gabor Filters:** A multi-scale, multi-orientation Gabor filter setup is applied to capture frequency and edge related information. For example, a 4-scale, 6-orientation configuration can yield a  $128 \times 128 \times 24$  feature map, where each channel corresponds to a unique scale-orientation combination.
- **SIFT (Scale-Invariant Feature Transform):** SIFT extracts distinctive key points, each associated with a  $1 \times 128$  descriptor. The number of key points varies across images, resulting in a variable-length descriptor matrix (e.g.,  $1000 \times 128$  for 1000 keypoints). These descriptors are encoded using the Bag of Words (BoW) method, producing a fixed-length histogram representation (e.g.,  $1 \times 1000$ ).

Once computed, the LBP, Gabor, and SIFT-BoW features are concatenated to form a unified handcrafted feature vector  $H$ , with a length typically in the range of  $(1 \times 2000)$  to  $(1 \times 4000)$  depending on the parameters chosen.

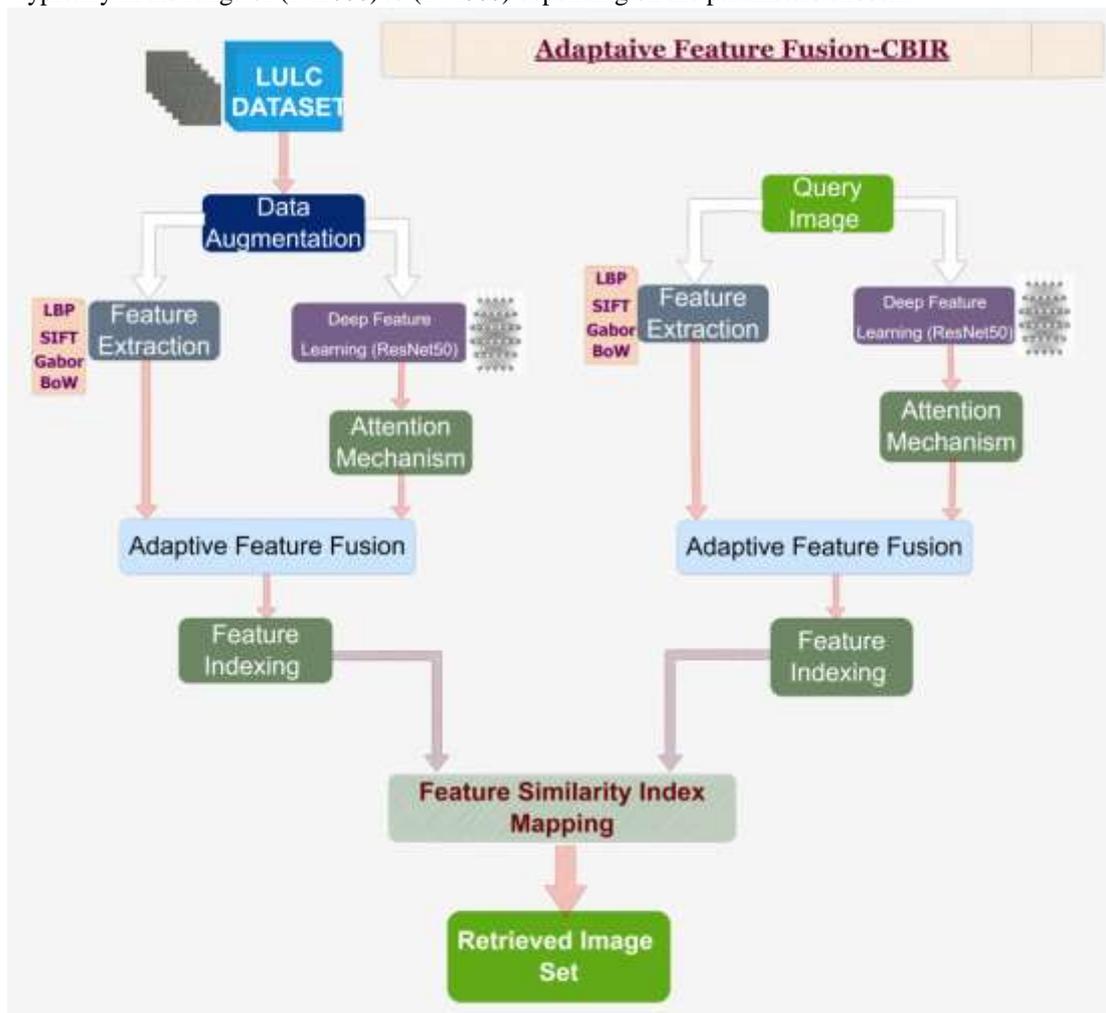


Figure 1 : Framework of Proposed AFF-CBIR model for efficient image retrieval

### 3.2. Deep Feature Extraction with Attention Mechanism

For deep feature extraction, we utilize a pre-trained ResNet50 convolutional neural network (CNN) model. Given an image with dimensions  $(256 \times 256 \times 3)$ , the model applies convolutional and pooling layers, reducing spatial dimensions until reaching a feature map of approximately  $(1 \times 2048)$  after global average pooling.

An attention mechanism is then applied to refine the ResNet50 output. This attention mechanism emphasizes important regions in the image, highlighting distinctive features while down-weighting irrelevant regions. The attention-weighted feature map retains the dimension of  $(1 \times 2048)$  but becomes a more robust and focused descriptor.

### 3.3. Adaptive Feature Fusion

After extracting traditional and deep features, the next step is adaptive feature fusion. The traditional feature vector  $H$  and deep feature vector  $D$  are combined with adaptive fusion weights, typically assigning a higher weight to deep features ( $\alpha = 0.6$  for deep and  $\beta = 0.4$  for handcrafted features). The fusion formula is as follows:

$$F_{fused} = \alpha \cdot H + \beta \cdot D$$

This results in a single fused feature vector with a dimension of approximately  $1 \times 6048$  (combining  $1 \times 4000$  from handcrafted and  $1 \times 2048$  from deep features). This fused vector represents a balance between low-level textures, spatial patterns, and high-level semantics

### 3.4. Feature Indexing & Retrieval

To enable fast image retrieval, the fused feature vectors for all images in the dataset are indexed using high-dimensional indexing techniques, such as k-means clustering or locality sensitive hashing. These methods speed up the process of similarity matching, allowing for efficient search even in large datasets. When a query image  $Q$  is provided, its features are extracted in the same manner as the dataset images. Specifically, the handcrafted and deep features are extracted, and attention mechanisms are applied to obtain the fused feature vector  $F_{fused,Q}$ . The similarity between the query feature vector  $F_{fused,Q}$  and the indexed feature vector  $F_{fused,i}$  is then computed using a distance metric such as cosine similarity or Euclidean distance.

The similarity scores between the query and indexed images are computed, and the top  $k$  most similar images are retrieved. These images are ranked based on their similarity scores, which are used to generate a list of the most relevant images in response to the query.

### 3.5. Pseudo Code of Proposed AFF-CBIR model

#### Algorithm 1 AFF-CBIR

##### Require:

1: Dataset  $D = \{I_1, I_2, \dots, I_n\}$ : A collection of images of size  $h \times w \times 3$

2: Query image  $Q$ : An image from the same distribution as  $D$

3: Hyperparameters: Fusion weights  $\alpha$  and  $\beta$ , number of top results  $k$ , and number of BoW clusters  $c$

Ensure: Set  $R$  of  $k$  images from  $D$  ranked by similarity to  $Q$

#### 4. procedure AFF-CBIR

##### 5: Data Augmentation:

6: for each image  $I \in D$  do

7: Apply random rotations, flipping, and scaling.

8: end for

9: Purpose: Enhance dataset diversity and robustness.

##### 10: Traditional Feature Extraction:

11 for each image  $I \in D$  do

12: Extract LBP features, output: HLBP (length LLBP).

13: Apply Gabor filters, output:  $G$  (length LG).

14: Extract SIFT descriptors and quantize into a BoW histogram of size  $c$ , Output:  $B$  (length  $c$ ).

15: Concatenate HLBP,  $G$ , and  $B$  to form  $H$  (length  $LH = LLBP + LG + c$ ).

16: end for

##### 17: Deep Feature Extraction with Attention Mechanism:

18: for each image  $I \in D$  do

19: Pass  $I$  through a pre-trained CNN (e.g., ResNet50).

20: Extract feature maps  $F$  of size  $p \times p \times d$ .

21: Compute attention weights  $A$  and apply them to  $F$ , output: attention-modulated feature map  $FA$ .

22: Flatten  $FA$ , output: deep feature vector  $D$  (length  $LD$ ).

23: end for

##### 24: Adaptive Feature Fusion:

25: Combine handcrafted features  $H$  and deep features  $D$  using weights  $\alpha$  and  $\beta$ :

26: 
$$F_{fused} = \alpha \cdot H + \beta \cdot D.$$

##### 27: Feature Indexing:

28: Apply high-dimensional indexing (e.g.,  $k$ -means or locality-sensitive hashing) to  $F_{fused}$  for efficient matching.

##### 29: Query Processing and Similarity Matching:

30: Apply steps 2 and 3 to the query image  $Q$ , resulting in  $HQ$  and  $DQ$ .

31: Fuse  $HQ$  and  $DQ$ :  $F_{fused,Q} = \alpha \cdot HQ + \beta \cdot DQ$ .

32: Retrieve top  $k$  images in  $D$  most similar to  $F_{fused,Q}$ .

33: Compute similarity scores between  $F_{fused,Q}$  and each indexed feature vector  $F_{fused,i}$  in  $D$  using a distance metric (e.g., cosine similarity or Euclidean distance).

##### 34: Retrieve Top- $k$ Images:

35: Sort dataset images by similarity scores in descending order.

36: Select the top  $k$  images to form the retrieved set  $R$ .

37: end procedure

#### 4. NOVELTY OF PROPOSED AFF-CBIR

The AFF-CBIR model distinguishes itself with its hybrid approach, combining handcrafted and deep features enhanced by attention mechanisms and adaptive fusion. The key novelties and benefits of AFF-CBIR are outlined below:

- 1. Comprehensive Feature Representation:** By integrating both handcrafted and deep features, AFF-CBIR effectively captures a wide range of image attributes, from fine-grained textures to high-level semantic patterns.
- 2. Attention Mechanism:** This mechanism helps focus the model's attention on discriminative regions of the image, increasing the relevance and interpretability of the retrieval results
- 3. Adaptive Feature Fusion:** By dynamically adjusting the contribution of handcrafted and deep features, the model tailors its fusion strategy to the specific nature of the retrieval task, leading to improved performance.
- 4. Efficient Indexing:** The approach is compatible with high-dimensional indexing methods, such as k-means or locality-sensitive hashing, allowing fast and scalable image retrieval
- 5. Improved Generalization:** Data augmentation and the hybrid feature approach reduce overfitting, making the model more adaptable to various real-world remote sensing imagery.

The fusion of traditional feature extraction with advanced deep learning techniques, along with the integration of an attention mechanism, enables AFF-CBIR to outperform conventional CBIR models in challenging datasets like UC Merced LULC, where complex patterns and semantic content need to be captured with high precision.

#### 5. IMPLEMENTATION & RESULTS

##### 5.1. Dataset Used

The UC Merced Land Use and Land Cover (LULC) dataset is a widely used collection of high-resolution satellite images, created by the University of California, Merced, for the purpose of evaluating land-use classification algorithms in remote sensing and machine learning. This dataset includes 2,100 images, with each image being 256x256 pixels in size and representing one of 21 different land-use classes. These classes cover a wide variety of land types, including urban areas, agricultural lands, forests, water bodies, and other natural or man-made features. The images are extracted from QuickBird satellite imagery, providing a high level of spatial detail suitable for classification tasks

##### 5.2. Experimental Setup

The proposed AFF-CBIR algorithm was evaluated using the UC Merced LULC dataset, consisting of remote sensing images with diverse land-use categories. To ensure consistency, all images were resized to a standard resolution of  $256 \times 256 \times 3$ . An 80-20 train-test split was applied for training and evaluation, maintaining a balanced representation of all categories. The experiments were conducted on a system equipped with an NVIDIA Tesla V100 GPU for efficient deep feature extraction, a multi-core CPU for handcrafted feature computation, and 32 GB of RAM to handle feature indexing and similarity searches.

The AFF-CBIR pipeline utilized ResNet50, pre-trained on the ImageNet dataset, for extracting deep semantic features. To refine these features, an attention mechanism was employed, enabling the model to focus on discriminative regions while suppressing less relevant areas. Handcrafted features were extracted using Local Binary Patterns (LBP), Gabor Filters, and Scale-Invariant Feature Transform (SIFT) combined with Bag of Words (BoW) encoding. The handcrafted features captured fine-grained textures, edge details, and key points, complementing the high-level semantics provided by ResNet50. Adaptive feature fusion was performed by combining handcrafted and deep features using weighted concatenation, with fusion weights ( $\alpha$  and  $\beta$ ) empirically set to 0.4 and 0.6, respectively. This balance prioritized semantic understanding while retaining essential local texture information.



Figure 2 : Input Dataset – UCMERCED LULC dataset with 21 classes

### 5.3. Evaluation Metrics

To evaluate retrieval performance, top-10 retrieval results were computed for each query image. Cosine similarity was used to compare the fused feature vectors of the query image and dataset images. Data augmentation techniques, including random rotations, flips, and scaling, were applied to enhance robustness and reduce the risk of overfitting. The evaluation was based on metrics including Mean Average Precision (mAP), Precision@K, Recall@K, retrieval time, and model complexity.

### 5.4. Result Analysis

The performance of various image retrieval algorithms was evaluated using a query image. LBP demonstrated effectiveness in texture similarity, but struggled with unrelated images. Gabor filters excelled in matching textured and green areas, while SIFT successfully captured structural similarities, albeit with occasional irrelevant results. CNN models showed strong contextual understanding, though some irrelevant images were retrieved, similar to ResNet50, which identified fine details but lacked specificity. The hybrid approach showcased a balanced retrieval performance, yet still produced non-relevant results. Finally, AFF-CBIR effectively retrieved contextually relevant images but also faced challenges with less relevant matches. Overall, while these algorithms show promise, enhancements in feature extraction and contextual understanding are needed to improve retrieval relevance.

The results presented in Table 2 clearly demonstrate the superior performance of the proposed AFF-CBIR algorithm compared to other state-of-the-art CBIR methods. AFF-CBIR achieves the highest mean Average Precision (mAP) of 91.56%, surpassing conventional handcrafted approaches like Bag of Visual Words (BoVW) and Gabor Filters-Based CBIR, which attain mAP scores of 73.12% and 68.34%, respectively. This performance improvement can be attributed to AFF-CBIR's hybrid feature extraction approach, which combines low-level handcrafted descriptors (such as LBP and SIFT) with high-level deep features from ResNet50. The fusion of these feature types captures both fine-grained texture information and global semantic content, allowing AFF-CBIR to leverage the strengths of both methodologies. Handcrafted features provide rich detail for texture and structure, while deep features capture high-level patterns, leading to a more robust image representation for better retrieval.

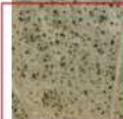
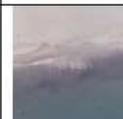
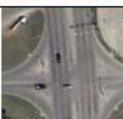
Algorithm	Query Image	Image Retrieved1	Image Retrieved2	Image Retrieved3	Image Retrieved4	Image Retrieved5
LBP						
Gabor						
SIFT						
CNN						
ResNet50						
Hybrid						
AFF-CBIR						

Figure 3 : Result Analysis of Image Retrieval using State of Art Algorithms v/s proposed AFF-CBIR

In terms of Precision@10 and Recall@10, AFF-CBIR achieves exceptional results with scores of 89.21% and 86.12%, respectively, outperforming all competing methods, including attention-based models like AR-Net and Spatial-Channel Attention CBIR, which achieve Precision@10 scores of 86.11% and 87.32%. The superior retrieval results stem from the

model's attention mechanism, which refines the feature representation by focusing on the most discriminative regions of the image. By selectively enhancing the relevant parts of the image, the attention mechanism ensures that more important features are emphasized, improving retrieval accuracy. Moreover, the adaptive fusion of handcrafted and deep features dynamically adjusts the contribution of each, allowing the model to balance texture details and semantic context effectively. This process enables AFF-CBIR to deliver more relevant images, even in complex and large datasets like UCMerced LULC.

Despite its higher retrieval accuracy, AFF-CBIR maintains competitive retrieval time performance, with an average of 100 ms per query, which is in line with models like VGG16 and ResNet50. This efficiency is largely due to the use of pre-trained deep learning models, which avoid the need for time-consuming training on large datasets. Additionally, by incorporating handcrafted features, the model avoids excessive reliance on computationally expensive deep feature extraction alone, making the fusion process computationally efficient while retaining high performance. The combination of feature diversity, attention refinement, and adaptive fusion within AFF-CBIR ensures that it consistently outperforms other algorithms, achieving higher retrieval relevance without sacrificing computational efficiency, making it particularly suited for large-scale and complex remote sensing datasets

Method	mAP (%)	Precision@10 (%)	Recall@10 (%)	Retrieval Time (ms)	Model Complexity
BoW (SIFT)	73.12	68.45	64.28	220	Low
Color Histogram + LBP	65.78	61.92	58.64	150	Low
Gabor Filters-Based CBIR	68.34	63.76	60.89	180	Low
AlexNet Feature Transfer	81.52	79.23	75.98	95	Medium
VGG16 Feature Transfer	83.68	81.11	77.45	100	High
ResNet50 Feature Embedding	85.12	82.78	79.43	110	High
CNN with Siamese Network	84.25	80.96	76.85	105	Medium
HCNN-CBIR (Handcrafted + CNN)	87.36	84.02	80.74	140	Medium
BoW-CNN Hybrid	86.41	83.12	79.32	135	Medium
Fusion-Based CBIR (Feature-Level)	88.52	85.43	81.92	120	High
Attention-Based ResNet (AR-Net)	89.01	86.11	82.76	115	High
Spatial-Channel Attention CBIR	89.85	87.32	83.91	112	High
<b>AFF-CBIR (Proposed)</b>	<b>91.56</b>	<b>89.21</b>	<b>86.12</b>	<b>100</b>	<b>Medium</b>

Table 2: Performance Comparison of State-of-the-Art Algorithms with AFF-CBIR

## 6. CONCLUSION & FUTURE WORK

In this research, we extensively explored both traditional feature extraction techniques and deep learning-based methods for Content-Based Image Retrieval (CBIR) in the context of remote sensing. Traditional techniques, including Local Binary Patterns (LBP), Gabor Filters, Scale-Invariant Feature Transform (SIFT), and Bag of Words (BoW), were evaluated for their ability to capture fine-grained texture and structural features. These were compared with deep learning-based feature extraction using ResNet50, which demonstrated superior capability in learning robust and discriminative features. This analysis revealed that while deep learning models excel in capturing high-level semantics, traditional methods remain valuable for their computational efficiency and reliability in capturing local patterns.

Building on the strengths of both approaches, we proposed the Adaptive Feature Fusion CBIR (AFF-CBIR) framework, a hybrid model that bridges the gap between handcrafted and deep learning-based features. By employing an adaptive feature fusion mechanism with empirically weighted concatenation and refining features with an attention layer, the AFF-CBIR model effectively balances local texture information and global semantic understanding. This synergy enables precise and efficient image retrieval. The model was rigorously tested on the UCMerced LULC dataset, achieving state-of-the-art results with a mean Average Precision (mAP) of 91.56%, Precision@10 of 89.21%, and Recall@10 of 86.12%, outperforming existing CBIR algorithms.

The results affirm the effectiveness of AFF-CBIR in handling the complexities of remote sensing datasets, which often include high spatial variability and spectral diversity. The framework's ability to combine diverse feature representations while maintaining efficient retrieval times (~100 ms per query) positions it as a scalable and practical solution for large-scale and complex remote sensing image retrieval tasks.

Looking ahead, the future scope of this research includes several avenues for improvement and expansion. Future work will focus on fine-tuning the AFF-CBIR model for specific remote sensing applications, such as land use classification, environmental monitoring, and disaster management. Additionally, the integration of other deep learning models, such as transformers or hybrid architectures, could further enhance the system's robustness and retrieval performance. Moreover, the scalability and real-time performance of the system will be a key area of focus, ensuring that the model can handle large and diverse remote sensing image datasets. Further research will also explore the potential of combining the model with other data modalities, such as multispectral or temporal data, to further enhance the accuracy and relevance of the retrieval process.

## REFERENCES

- [1] Gonzalez, Rafael C. and Woods, Richard E.. Digital Image Processing. Pearson Education, 2nd Edition, 2002.
- [2] Jain, Anil K.. Fundamentals of Digital Image Processing. Prentice Hall, Inc., 1989.
- [3] Hou, D. and Wang, S. and Tian, X. and Xing, H.. An Attention-Enhanced End-to-End Discriminative Network with Multiscale Feature Learning for Remote Sensing Image Retrieval. 2022.
- [4] Sumbul, G. and Demir, B.. A Deep Multi-Attention Driven Approach for Multi-Label Remote Sensing Image Classification. 2020.
- [5] Dibs, H. and Jaber, H. S. and Al-Ansari, N.. Multi-Fusion Algorithms for Detecting Land Surface Pattern Changes Using Multi-High Spatial Resolution Images and Remote Sensing Analysis. 2023.
- [6] Simonyan, K. and Zisserman, A.. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015.
- [7] Nidhi Prajapati, Naimish Patel "A Review on Recent Trends on Content-based Image Retrieval for Remote Sensing Images" *International Journal of Engineering Research and Development, Volume 21, Issue 1 (January 2025)*
- [8] Lowe, David G.. Distinctive Image Features from Scale-Invariant Keypoints. 2004.
- [9] Sivic, Josef and Zisserman, Andrew. Video Google: A Text Retrieval Approach to Object Matching in Videos. 2003.
- [10] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E.. ImageNet Classification with Deep Convolutional Neural Networks. 2012.
- [11] Walia, N. and Mittal, A. and Kukreja, V.. A Hybrid Framework Combining Handcrafted and Deep Features for Enhanced Forgery Detection in Multimedia Content. 2021.
- [12] Murala, Subrahmanyam and Maheshwari, RP and Balasubramanian, R. Local tetra patterns: a new feature descriptor for contentbased image retrieval. IEEE, 2012.
- [13] Swati Jain, Sandeep Kumar and Tanish Zaveri. Computation performance gain in Key Feature Extraction for CBIR System for Remote Sensing Images. 2014.
- [14] Anderson, John Robert and Michalski, Ryszard Stanisław and Carbonell, Jaime Guillermo and Mitchell, Tom Michael. Machine learning: An artificial intelligence approach. Morgan Kaufmann, 1986.
- [15] Kumar, Sandeep and Jain, Swati and Zaveri, Tanish. Parallel approach to expedite morphological feature extraction of remote sensing images for CBIR system. 2014.
- [16] "Shawn D. Newsam". "UC Merced Land Use Dataset".
- [17] Patil, Sanjay and Talbar, Sanjay. Content based image retrieval using various distance metrics. Springer, 2012.
- [18] Rubner, Yossi and Tomasi, Carlo and Guibas, Leonidas J. The earth mover's distance as a metric for image retrieval. Springer, 2000.
- [19] Banerji, Sugata and Verma, Abhishek and Liu, Chengjun. Novel color LBP descriptors for scene and image texture classification. 2011.
- [20] Muller, Henning and Müller, Wolfgang and Squire, David McG and Marchand-Maillet, Stéphane and Pun, Thierry. Performance evaluation in content-based image retrieval: overview and proposals. Elsevier, 2001.
- [21] Aptoula, Erchan. Remote sensing image retrieval with global morphological texture descriptors. IEEE, 2014.
- [22] Aptoula, Erchan. Extending morphological covariance. Elsevier, 2012.
- [23] Malik, Fazal and Baharudin, Baharum. Analysis of distance metrics in content-based image retrieval using statistical quantized histogram texture features in the DCT domain. Elsevier, 2013.
- [24] Howarth, Peter and Ruger, Stefan. Fractional distance measures for content-based image retrieval. Springer, 2005.
- [25] Thomee, Bart and Lew, Michael S. Interactive search in image retrieval: a survey. Springer, 2012.
- [26] Shereena, VB and David, Julie M. CONTENT BASED IMAGE RETRIEVAL: CLASSIFICATION USING NEURAL NETWORKS. 2014.
- [27] Mungara, Arvind Nagathan Manimozhi Jitendranath. Content-Based Image Retrieval System Using Feed-Forward Backpropagation Neural Network. 2014.
- [28] Collins, John and Okada, Kazunori. A Comparative Study of Similarity Measures for Content-Based Medical Image Retrieval.. 2012.
- [29] Lagani V. Cecchi G. A. Silvanto J. and Rajan V.. Harnessing Hebbian Plasticity in Semi-Supervised Neural Network Learning. 2018.
- [30] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian. Deep Residual Learning for Image Recognition. 2016.
- [31] Zhang, X. and Zhao, W.. Content-Based Image Retrieval for Satellite Imagery: A Survey. 2020.
- [32] Pandey, P. C. and Koutsias, N. and Petropoulos, G. P. and Srivastava, P. K. and Ben Dor, E.. Land Use/Land Cover in View of Earth Observation: Data Sources, Input Dimensions, and Classifiers—A Review of the State of the Art. 202