

Article

Not peer-reviewed version

Optimizing Remote Sensing Image Retrieval Through a Hybrid Methodology

[Sujata Alegavi](#)^{*} and Raghvendra Sedamkar

Posted Date: 13 May 2025

doi: 10.20944/preprints202505.0920.v1

Keywords: Hyperspectral Images (HSI); Synthetic Aperture Radar images (SAR); Convolutional Neural Networks (CNN); pretrained networks; classification; retrieval; hybrid



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Optimizing Remote Sensing Image Retrieval Through a Hybrid Methodology

Sujata Alegavi ^{1,*} and Raghvendra Sedamkar ²

¹ Associate Professor Internet of Things Department, Thakur College of Engineering and Technology, Mumbai, Maharashtra, India

² Professor Computer Engineering Department, Thakur College of Engineering and Technology, Mumbai, Maharashtra, India

* Correspondence sujata.alegavi@gmail.com.

Abstract: The contemporary challenge in remote sensing lies in the precise retrieval of increasingly abundant and high-resolution remotely sensed images (RS Image) stored in expansive data warehouses. The heightened spatial and spectral resolutions, coupled with accelerated image acquisition rates, necessitate advanced tools for effective data management, retrieval, and exploitation. The classification of large-sized images at the pixel level generates substantial data, escalating the workload and search space for similarity measurement. Semantic-based image retrieval remains an open problem due to limitations in current artificial intelligence techniques. Furthermore, on-board storage constraints compel the application of numerous compression algorithms to reduce storage space, intensifying the difficulty of retrieving substantial, sensitive, and target-specific data. This research proposes an innovative hybrid approach to enhance the retrieval of remotely sensed images. The approach leverages multi-level classification and multi-scale feature extraction strategies to enhance performance. The retrieval system comprises two primary phases: database building and retrieval. Initially, the proposed Multiscale Multiangle Mean-shift with Breaking Ties (MSMA-MSBT) algorithm selects informative unlabelled samples for hyperspectral and synthetic aperture radar images through an active learning strategy. Addressing the scaling and rotation variations in image capture, a flexible and dynamic algorithm, modified Deep Image Registration using Dynamic Inlier (IRDI), is introduced for image registration. Given the complexity of remote sensing images, feature extraction occurs at two levels. Low-level features are extracted using the modified Multiscale Multiangle Completed Local Binary Pattern (MSMA-CLBP) algorithm to capture local contexture features, while high-level features are obtained through a hybrid CNN structure combining pretrained networks (Alexnet, CaffeNet, VGG-S, VGG-M, VGG-F, VGG-VDD-16, VGG-VDD-19) and a fully connected dense network. Fusion of low and high-level features facilitates final class distinction, with soft thresholding mitigating misclassification issues. A region-based similarity measurement enhances matching percentages. Results, evaluated on high-resolution remote sensing datasets, demonstrate the effectiveness of the proposed method, outperforming traditional algorithms with an average accuracy of 86.66%. The hybrid retrieval system exhibits substantial improvements in classification accuracy, similarity measurement, and computational efficiency compared to state-of-the-art scene classification and retrieval methods.

Keywords: Hyperspectral Images (HSI); Synthetic Aperture Radar images (SAR); Convolutional Neural Networks (CNN); pretrained networks; classification; retrieval; hybrid network

I. Introduction

Global change, driven by evolving Earth land cover and use, faces new challenges amid the information revolution. Climate shifts and population demands have altered the global land use/cover map. Advancements in satellite-based sensors enhance geospatial mapping accuracy. Multispectral/Hyperspectral imagery captures spectral signatures of varied materials, enabling

terrain feature identification. Classification, crucial for delineating similar features, uses clustering algorithms and statistical structures. Medium spatial resolution sensor images pose classification accuracy challenges, addressed by advanced computing techniques. Supervised classification utilizes trained samples, while unsupervised classification handles scenarios without training data, fostering improved accuracy in land cover mapping and monitoring.

Remotely Sensed (RS) Image Processing

In the realm of remote sensing, HSI and SAR images are extensively utilized to exploit the wealth of discrete spectral information inherent in an image. The complex nature of HSI/SAR scenes, involving combined pixels, vast data volumes, and limited training samples, presents a formidable challenge necessitating precise recognition methods. Spatial domain classifiers like support vector machines (SVM) and multi-scale breaking ties (MSBT) have historically been employed in the initial phases of HSI/SAR image analysis. However, recent advancements focus on leveraging spectral and temporal data for enhanced image discrimination.

HSI and SAR images play a pivotal role in geo-indexing, target monitoring, area mapping, and environmental monitoring. HSI, possessing two spatial dimensions and one spectral dimension, forms a 3D cube, offering comprehensive information about the physical characteristics of sensed objects. The availability of spatial-spectral information has led to the development of various applications for classification and retrieval, posing challenges due to the high dimensionality of the data.

High-resolution data, collected through sophisticated sensors in Earth observation programs, contributes to the wealth of remote sensing images. HSI/SAR images, with their numerous bands (e.g., 220 bands), provide rich spectral information, facilitating the differentiation of materials. Spectral signatures unique to each material enable effective classification. Advanced sensors capturing fine spatial resolution data exacerbate the dimensionality problem in the spectral domain. The increased spatial resolution allows for the identification of minute spatial structures within the imagery.

II. Literature Survey

In the early stages of this era, extensive research has been conducted on remote sensing image processing, providing a comprehensive overview of diverse methodologies employed for image processing and evaluation in the field. This part offers insights into the depth of research endeavors in this domain.

A. Pre-Processing of HSI/SAR

Jaime Delgado et al. [1] proposed leveraging purist signatures in mixed pixels of hyperspectral images using spectro-unmixing to address spatial resolution limitations. Spatial pre-processing (SPP) techniques were employed before end member identification, but the associated cost and processing time raised concerns. G. Martín, et al. [2] introduced three parallel graphic processing unit (GPU) implementations of the SPP algorithm, marking the first GPU application for spatial pre-processing. Their evaluation showcased reduced execution time, demonstrating real-time feasibility for Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (<https://www.jpl.nasa.gov/missions/airborne-visible-infrared-imaging-spectrometer-aviris/>) sensor data collection.

B. Feature Extraction of HSI/SAR

L. Fang [3] introduced a novel Deep Hashing Neural Network (DHNN) model for HSI classification, employing a trained network to extract similarity-preserving features. The model incorporates a hashing layer to convert real-value features into binary ones, expediting distance computation. A customized loss function is designed to optimize feature distances in Hamming space, and the DHNN output is utilized by a support vector machine (SVM) for HSI classification.

Cui Binge, et al. [4] proposed an Nonlinear Robust Principal Component Analysis (NRPCA)-based feature extraction method combined with Minimum Noise Fraction (MNF) to address challenges in hyperspectral image analysis. The approach involves reducing image size through average image fusion, applying NRPCA to each band, and utilizing MNF for low-level function feature extraction. The SVM method is then employed for identification, considering higher signal-to-noise ratios in the transformed results based on linear quality transformation and signal-noise ratio considerations.

C. Classification of Hyperspectral Images / Synthetic Aperture Radar Images

The limited availability of extensive training data poses a challenge for Deep Learning Neural Network (DLNN) classification in hyperspectral imaging. Addressing this, tailored frameworks for specific tasks or regularization phases with a finite number of samples can enhance generalization. The MugNet network simplifies DLNN architecture for improved performance with few training samples. Additionally, employing transfer learning proves effective in enhancing DLNN accuracy across various scenarios.[5]

In hyperspectral image classification, some studies focus on Active Learning (AL) frameworks that generate regions to constrain sampling. Traditional AL heuristics often overlook spatial uncertainty and batch mode approaches may neglect spatial homogeneity. Zhaohui Xue introduces an Enhanced Uncertainty Measure (EUM) considering community knowledge, utilizing super pixels generated by simple linear iterative clustering (SLIC) to enhance sample diversity. Experimental results demonstrate significant classification precision improvement compared to conventional methods. [6]

D. Deep Learning Methods of Classification

Researchers have explored various multi-scale spatial features to propose spectral-spatial fusion methods, leveraging regulated CNN for hyperspectral image classification. They reconstruct the spectral function as a 2-D image and apply guided filters to remove spatial functions, addressing finite training sampling challenges. [7] The presented spectral-spatial fusion techniques demonstrate significant performance improvements and precise allocation. Additionally, a study discusses integrating multi-scale representation in CNN to enhance object identification, visual recognition, and time series classification, particularly for space IR-point objects. The proposed multi-scale convolution neural network (MCNN) aims to efficiently capture both fast and slow variables in IR signature features, considering motion effects and noise resilience for artifact classification. [8]

E. HSI/SAR Retrieval System

Yansheng Li, et al. [9] introduce deep-hazardous neural networks Scale-Invariant Deep Hybrid Convolutional Neural Networks (SIDHCNNs) optimized with Compressed Sensing - Low-rank and Sparse Representation based Structured Image Registration (CS-LSRSIR) constraints for invariant source learning, aiming for maximum capacity utilization. The proposed SIDHCNNs undergo quantitative evaluation using a dual-source remote sensing dataset comprising 8 standard terrain cover categories with 10 thousand dual samples for each category.

Osman Emre Dai, et al. [10] present a novel content-based image recovery method for remote sensing (RS), involving an image definition approach to distinguish spatial and spectral details. They employ a supervised recovery approach, utilizing descriptors of RS image content, and introduce three spectral descriptors. The method utilizes sparse reconstruction-based classifiers for single-label and multi-label RS image recuperation, enhancing recovery efficiency through improved sensitivity to different dictionary terms.

III. Implementation

The depicted architecture of the proposed HSI/SAR hybrid retrieval system exhibits two distinct modules: the "Database Building" section, responsible for training, labeling, and storing training samples, and the "Retrieval Process" section, focused on testing and retrieving query images. The recovery process for a given query image patch initiates with feature extraction, utilizing pre-calculated target patch features stored in the "building database" section. Leveraging these features, the semantic category of the query image patch is determined, and similarity with target category patches is computed. The retrieval results are subsequently presented in an ordered list. Pre-processing techniques are employed to enhance retrieval precision by mitigating image noise and reducing corresponding database space. Below is the discussion of every component of the architecture in detail.

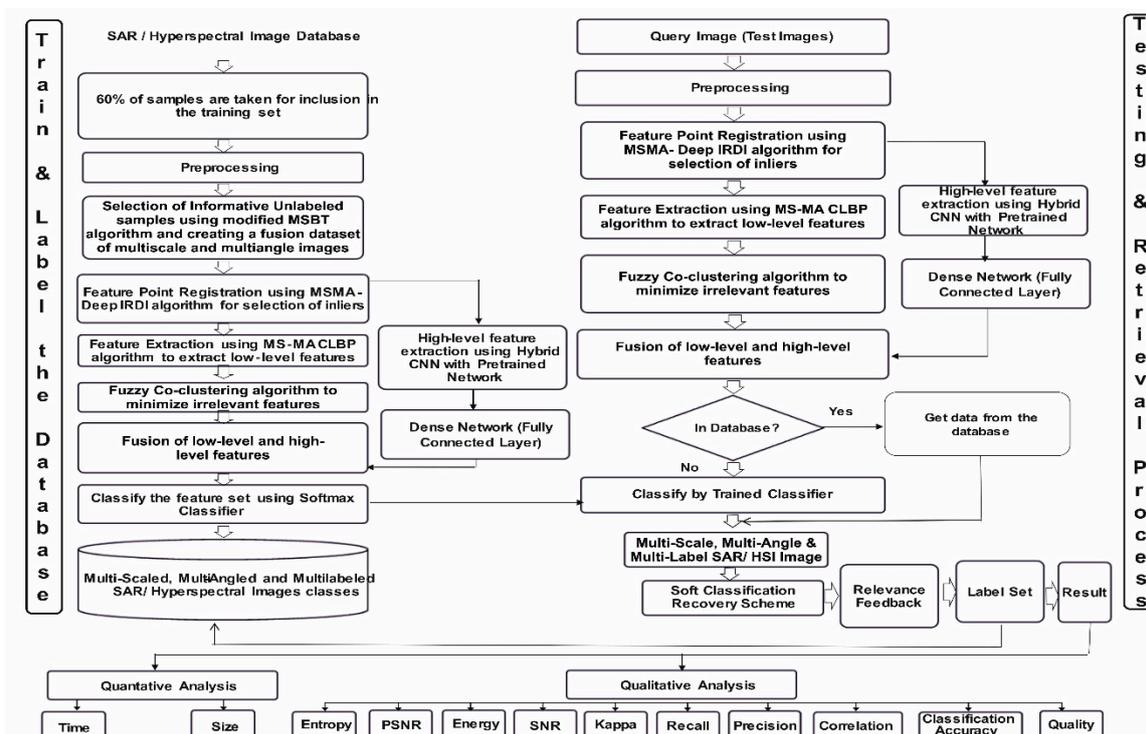


Figure 1. Proposed Hybrid Architecture for HSI/ SAR Image retrieval.

A. Train and Label Database (Database Building)

The primary aim of this architectural phase is to conduct training, labeling of selected samples, and establish a labeled database. This involves sequential stages of pre-processing, sample selection for the training set, image registration, feature extraction, classification, and database categorization. For the creation of a labeled dataset, 60% of diverse category samples are chosen as training samples for raw HSI/SAR images. These images undergo pre-processing using Gabor and independent component analysis (ICA) filters for noise removal and oriental feature capture. Employing scaling, rotation, and fusion, a Multiscale-Multiangle dataset (MSMA) of HSI/SAR images is generated.

To streamline sample selection, the MSMA-MSBT algorithm is introduced, focusing on efficient and informative unlabeled sample inclusion in the training set. Further, feature point registration is performed using the modified DeepIRDI technique, enhancing the robustness of feature point enrollment. These inliers are then fed into the modified MSMA-CLBP module and the hybrid pre-trained-CNN classifier module. Low-level features are extracted using the modified MSMA-CLBP algorithm, addressing the efficiency gap in CNN for low-level feature capture in remote sensing images.

High-level features are obtained through a CNN structure, leveraging initial layers of pretrained networks such as Alexnet [11], CaffeNet [12], VGG-F, VGG-M, VGG-S, VGG-VDD-16, VGG-VDD-19 [13]. These features are subsequently input into the dense network to derive high-level features. Fuzzy co-clustering is applied to reduce irrelevant local features extracted by the modified MSMA-CLBP algorithm, addressing the curse of dimensionality in HSI/SAR images. The Fuzzy Co-clustering algorithm automatically calculates feature weights, optimizing data processing. The fused local features from the Fuzzy co-clustering module and high-level features from the hybrid Pretrained CNN module are input into the classifier, forming the proposed hybrid retrieval network. A SoftMax classifier is employed for final classification, generating multi-labeled HSI/SAR image classes. To tackle post-classification challenges, particularly in mixed pixels prevalent in HSI/SAR images, a soft classification recovery process is proposed. This process adaptively adjusts threshold values for effective pixel assignment in regions of mixed pixels, addressing difficulties in class distinction through normal classification methods.

B. Testing and Retrieval Process (Retrieval Process)

This section is dedicated to querying the labeled database for the most suitable images based on a given query. Upon input of a query image, its features are automatically extracted following the database creation process, including pre-processing. Two scenarios are considered: whether the query image features exist in the designated image database or not. If the features are present, relevant information is directly retrieved from the database. In the absence of features, the query image undergoes classification by the trained hybrid classifier, and subsequent determination of similarities between the query and target images is conducted.

Considering the characteristics of HSI/SAR images, a region-based similarity measurement is introduced, utilizing patch similarity. To address potential classification errors, the query image is not only matched with the most relevant image in the database but also with other relevant images belonging to the same category. To mitigate the inevitable classification errors, a soft thresholding technique is proposed, derived from a confusion matrix. The utilization of various Compute Unified Device Architecture (CUDA) libraries and GPU processing is emphasized for enhanced speed and processing efficiency compared to central processing unit (CPU)-based operations.

C. Multiscale, Multiangle and Multiscale-Multiangle (Fused) Dataset

Scaling:

In the offline datasets HSI/ SAR images are first scaled at different levels to get the scaled dataset and further these scaled images are rotated to get the angle dataset. Finally, these scaled and rotated images are fused together to form a fusion dataset. The different resolutions used for training and testing are 512x512, 256x256 and 64x64. These offline datasets of HSI/SAR images are further used for training and testing purpose. For scaling Gaussian pyramid decomposition is used in which the original HSI/ SAR image is taken as HSI/ SAR-1. Further this HSI/ SAR-1 image is convolved with a gaussian kernel and down sampled to get HSI/ SAR-2 image. c refers to the different scales applied for generating multiscale images. The grey scale value in the HSI/ SAR-1 layer is obtained by,

$$HSI/ SAR - 1 (i, j) = \sum_{m=-c}^c \sum_{n=-c}^c G(m, n) * I_{l-1} (2i-l-m, 2j-l-n) \quad (1)$$

where $*$ is the convolution operation, $HSI/ SAR-1 \in \{2, 3, \dots, l\}$, I_{l-1} represents the previous image reconstructed with gaussian layer $l-1$, l represents the number of layers of Gaussian pyramid, the gaussian window is represented as follows,

$$G(m, n) = \frac{1}{2\pi\sigma^2} e^{-\frac{(m^2+n^2)}{2\sigma^2}} \quad (2)$$

where σ refers to the variance that relates to the Gaussian filter and HSI/ SAR-1, HSI/ SAR-2..., HSI/ SAR-N are obtained by continuous convolution and down sampling operation. Multiple images of 64x64, 256x256 and 512x512 which will yield multi-scaled image dataset are selected.

Rotating:

To understand the effect of rotation on HSI/ SAR images and to analyse the performance of various algorithms on such images, HSI/SAR images are rotated at various angles such as 90°, 180° and 270° which yields multi-angled database which are further used for training and testing purpose.

Fused (MS-MA):

As HSI/SAR images are captured using different sensors which have different resolutions and they are also captured with different instantaneous field of view which distorts the image in scale and angle. To study this effect, scaled and rotated HSI/SAR images are fused together to form a database consisting of both the variations of scaled and rotated images. Different variations of fused dataset such as 512x512-90°, 512x512-180°, 512x512-270°, 256x256-90°, 256x256-180°, 256x256-270°, 64x64-90°, 64x64-180° and 64x64-270° are formed. These images are further used for training and testing purposes.

D. Proposed Modified MSMA-MSBT

The algorithm blends Breaking Ties (BT) technique and Mean Shift (MS) cluster to choose the most explanatory training sampling. To more accurately perform the MS-MA CNN algorithm, we describe various variables. p provides the no of specimens for each iteration to be labelled. We have used Algorithm to construct a collection of N vectors. We take 1 part of N as NV training package with u basic training sampling and we take NY package with unlabelled $A-w$ samples to verify the other part of the N . NV is named after the training set, and stands as $B_S = \{(ai_r, ci_u)\}_{r=1}^d$. If the validation range is unmarked, $NY = \{ci_u\}_{r=1}^{A-w}$, in we select a suggested algorithm, p new samples, such as $D_{MS-MA} = \{ci_u\}_{r=1}^t$ from the validated set, we name them as $D_{MS-MA} = \{ai_r, ci_u\}_{r=1}^p$ and then add D_{MS-MA} into BT. [14]

The parameter " t " represents the number of new samples that are selected by the MS-MA (Multi-Scale Multi-Angle) CNN algorithm from the validation set to be added to the current training set in each iteration.

Algorithm Overview:

Key Variables and Parameters:

Variable " p " denotes the number of specimens labelled for each iteration.

Algorithm constructs a collection of N vectors using a specific technique.

NV represents the training set, comprised of u basic training samples, while NY is the set for validation with unmarked $A-w$ samples.

Algorithm selects p new samples from the validated set, denoted as $D_{MS-MA} = \{ci_u\}_{r=1}^t$ from the validated set, and name them as $D_{MS-MA} = \{ai_r, ci_u\}_{r=1}^p$ and then add D_{MS-MA} into BT.

Algorithm Steps:

$A-w$ samples to verify the other part of the N . NV is named after the training set, and stands as $B_S = \{(ai_r, ci_u)\}_{r=1}^d$. If the validation range is unmarked, $NY = \{ci_u\}_{r=1}^{A-w}$, in we select a suggested algorithm, p new samples, such as $D_{MS-MA} = \{ci_u\}_{r=1}^t$ from the validated set, we name them as $D_{MS-MA} = \{ai_r, ci_u\}_{r=1}^p$ and then add D_{MS-MA} into BT

Input: Basic classifier parameter from multi-scale, e training set, $BT = (ai_u, ci_u)_{r=1}^d$, validating set, $N_y = \{Xi_u\}_{r=1}^{A-w}$ mean-shift parameters, f , candidate samples parameter, β .

$\hat{\theta}$, \hat{e} , or \hat{a} , signifies an estimated value, derived from the algorithm.

e , a , θ refers to the actual values.

1. Train the MS-MA CNN classifier and receive the classifier parameter \hat{e}
2. Forecast and get NY samples using the Classifier $(\hat{ai}_r, \hat{ci}_u)_{r=1}^{A-w}$
3. For each sample in the sub-set $(\hat{ai}_r, \hat{ci}_u)_{r=1}^{A-w}$, compute its BT value as,

$$\hat{\theta}_{i_r} = \min t(\hat{ai}_r = j | ci_r, \hat{e}) - \min p(ai_r = j | ci_r, \hat{e}) \quad (3)$$

4. Arrange $\{\hat{\theta}\}_{r=1}^{A-w}$ in ascending order as $\{\hat{\theta}_{vr}\}_{r=1}^{A-w}$, and depending on the multi-label $\{v_r\}_{r=1}^{A-w}$ take out the first $\beta \cdot p$ samples from NY to be $NBT = \{N_{vr}\}_{r=1}^{\beta \cdot p}$.

5. Find out the p set of samples for $N_{BT} = \{N_{vr}\}_{r=1}^{B,p}$ by mean shift parameter f . Choose a random sample for every sample sub-set and position all selected p sampling pN_{MS-MA} .

$N_{MS-MA} = \{Ni_g\}_{g=1}^p$ Multi-Label N_{MS-MA} to obtain,

$$N_{MS-MA} = \{\hat{a}i_g, ci_g\}_{g=1}^p \quad (4)$$

6. Update BT and NY.

$$B_T = B_T \cup B_{MS-MA}; N_y = N_y - N_{MS-MA} \quad (5)$$

7. Refer to step (1) if a condition is not met; otherwise prevent repetition.

8. Initialization of multi-scale, multi-angle CNN for semantic segmentation: setting size and angle of 3 image patches as 512 X 512, 128 X 128 and 64 X 64 and image angles as 90°, 180° and 270°;

9. Fused dataset is obtained by combining resolution and angle datasets. That is, integrating angle and resolution which is 90°, 180°, 270° and 512 × 512, 128 × 128, 64 × 64 resolution;

10. Extract patches in three dimensions and angles, and normalize image patches;

11. Branch the image patches collected as research specimens and train samples;

12. Train network using the decent gradient algorithm in an end-to-end way;

13. Usage of the trained network to achieve labels for the test samples.

E. Inliers and Outliers in Feature Matching

I. Understanding Inliers and Outliers in Feature Matching

When we compare two images—whether they come from satellites, drones, or any other remote sensing source—we rely on feature matching to align them correctly. In this process, the terms **inliers** and **outliers** become very important.

- **Inliers** are the points in an image that correctly match between the two images. They follow the actual transformation (like rotation, scaling, or translation) between the images, ensuring accurate alignment.
- **Outliers**, on the other hand, are incorrect matches. These could be caused by noise, differences in lighting, or distortions in the images. Too many outliers can lead to errors in the final image alignment.

II. Challenges in Feature Matching

1. Sensor Noise

Different sensors capture images in different ways. Hyperspectral images may be affected by atmospheric noise, while SAR images often suffer from speckle noise, making it harder to detect correct feature points.

2. Geometric Distortions

When two images are taken from different angles or sensors, they may appear distorted. SAR images, for example, often have a foreshortening effect, which changes the apparent shape of objects.

3. Errors in Image Alignment

Traditional methods (like RANSAC [15]) rely on fixed rules to filter inliers and outliers, but these don't always work well for all types of images. If the filtering process isn't adaptive, the method may struggle with different image conditions.

Our Proposed **Modified Deep IRDI Algorithm** improves feature matching by making the process smarter and more adaptive by:

1. Better Selection of Inliers

Instead of using a strict rule to decide which matches are correct, our approach **assigns a probability score to each match**. This means the system can adjust dynamically, depending on the quality of the image.

2. Handles Images at Different Scales

Some features may appear bigger or smaller in different images. Our approach works at **multiple scales**, ensuring that important details are captured correctly even when images differ in size or resolution.

3. Deals with Geometric Distortions

To prevent errors in alignment, our method **checks the spatial consistency of feature points**. If a match doesn't fit well within the overall transformation of the image, it is more likely to be discarded as an outlier.

4. Faster and More Efficient

Using **GPU acceleration**, our method processes images much faster, making it suitable for handling large datasets, such as those used in land cover mapping and disaster monitoring.

F. Proposed Modified Deep IRDI Algorithm

In the process of inlier selection denoted as $M \times N$, a probability matrix PR is generated. This matrix is then subjected to a transformation solver based on the Gaussian mixture model (GMM). The entry PR [m, n] in this matrix represents the probability of matching x_n and y_m . Assuming that x_n corresponds to y_m , a higher probability is reflected in a larger PR [m; n]. This elevated probability indicates a significant correspondence between x_n and y_m , allowing for a visible alignment of characteristics in the respective pair.

The determination of probabilities involves the utilization of both convolution features and geometric structural information. The PR matrix is obtained through a defined operation. [17]

Prepare the $M \times N$ convolutional cost matrix c_{θ}^{conv} by

$$c_{\theta}^{conv}[m, n] = \begin{cases} \frac{d(y_m, x_n)}{d_{\theta}^{max}}, & \text{condition 1} \\ 1, & \text{otherwise} \end{cases} \quad (6)$$

where x_n is point at index n in point set X and y_m are point at index m in point set Y

d_{θ}^{max} is the maximum distance of all matched feature points pairs under threshold θ .

Eq., 6., shows how to extract and match features of HSI/ SAR images and detect objects in a different location captured with a different angle and a different viewpoint. c^{geo} which is acquired by performing a χ^2 test. Both c_{θ}^{conv} and c^{geo} are valued in [0, 1]. The following equations elaborate on how the geometric transformation is estimated, where $h_m^y(b)$ and $h_n^x(b)$ denotes the no. of points decreased in the b^{th} bin surrounding y_m and x_n , respectively.

$$c^{geo}[m, n] = \frac{1}{2} \sum_{b=1}^B \frac{[h_m^y(b) - h_n^x(b)]^2}{h_m^y(b) + h_n^x(b)} \quad (7)$$

Using an element-wise Hadamard product, we calculate an integrated array matrix C (denoted by \odot):

$$C = C_{\theta}^{conv} \odot C^{geo} \quad (8)$$

Finally, the prior probability matrix is calculated using:

$$P_R[m, n] = \begin{cases} 1, & \text{if } y_m \text{ and } x_n \text{ are corresponding} \\ \frac{1-\epsilon}{N}, & \text{otherwise.} \end{cases} \quad (\epsilon \text{ is hyperparameter valued in } [0 \text{ and } 1]) \quad (9)$$

Further, optimal parameters of transformation like (W, σ^2 , ω) are used. The aim of such an approach is to maximize a likelihood function or decrease the negative log-likelihood function equitably: $pold(m|x_n)$ denotes a posterior likelihood. The equation may be rewritten after expanding this equation and omitting derivative redundant terms as:

$$Q(W, \sigma^2, \omega) = \frac{1}{2\sigma^2} \sum_{n=1}^N \sum_{m=1}^M pold(m|x_n) \|x_n - \tau(y_m, W)\|^2 - \frac{1}{2} N_p \log \frac{(\alpha^2 \omega)}{(1-\omega)} - N \log(\omega) \quad (10)$$

W – Transformation parameter and ω – Weighting parameter

Modified Deep IRDI Algorithm implementation:

Input: A and B are the two images where 'A' is the reference image and 'B' is the captured image.

- 1 Initialize parameters from multi-scale and multiangle HSI/ SAR images;
- 2 Pre-match and choose the multi-label feature point sets F and E from A and B ;
- 3 Initialize $\rho = \hat{\rho}$;
- 4 do while
- 5 For each i iterations:
- 6 Calculate the array matrix for convolutional features
 c_{θ}^{conv} as per Equation 6.
- 7 Calculate the array matrix for the geometric structure
 c^{geo} as per Equation 7;
- 8 Calculate the array matrix $C = C_{\theta}^{conv} \odot C^{geo}$; (11)
 Solve the array matrix C linear assignment.
- 9 Calculate the prior probability matrix P_R using Equation 9;
- 10 Update the threshold $\rho \leftarrow \rho - \alpha$;
- 11 end
- 12 E-Steps:
- 13 Compute posterior probability matrix P_O by

$$P_O[m, n] = \frac{P_R[m, n]g_m(x_n)}{P(x_n)}$$
 (12)
- 14 end
- 15 While there is no convergence of equation 10;
- 16 Calculate the transformed picture IZ using interpolation with a thin-plate spline.
- 17 Output: IZ

G. Proposed Modified MSMA-CLBP Algorithm

The modified MSMA-CLBP algorithm is proposed for low-level feature extraction, to improve the efficiency of the system. In the proposed system, HSI/ SAR image dataset is taken as input. For pre-processing various filters like Gabor filter and ICA filter are used for removing the noise from the image. Pre-processing not only clears data of noise but also the data is uniformly modified to be accepted by the network. The modified MSMA-CLBP algorithm is used for low-level feature extraction. Spatial features as well as Speeded-Up Robust Features (SURF) are extracted. Exhaustive feature matches feature set 1 to the nearest neighbours in feature set 2 by computing the pair-wise distance between feature vectors. In next step, all results are fused according to the variance. [16]

Require: Preparation of training dataset and testing dataset

- 1: Capture Gabor operator's characteristics by converging input images
- 2: Adjust the MS-CLBP operator parameters (m, r) and pick the optimal criterion
- 3: Achieve MS-CLBP characteristics depending on the Gabor feature space for every image
- 4: Build the low-level features by the MS-CLBP characteristics

Assure: Works at the lowest level

The Gabor function consists of two parts 1. Real part 2. Imaginary part. By filtering, the actual part, the image and the imaginary part is used mainly for edge detection. The mathematical expression is:

$$g(x, y; \lambda, \theta, \sigma, \gamma) = \exp\left(-\frac{x_0^2 + \gamma^2 y_0^2}{2\sigma^2}\right) \cdot \exp(i(2\pi \frac{x_0}{\lambda} + \varphi)) \quad (13)$$

Where;

$$\begin{cases} x_0 = x \cos \theta + y \sin \theta \\ y_0 = -x \sin \theta + y \cos \theta \end{cases}$$

where x and y are the region of the pixels in the area. The LBP is utilized as a local texture characteristics description. The centre pixel Z LBP value is described as;

$$LBP_{q, e}(Z) = s(g_i - g_c)2^i \quad (14)$$

Where e is the length from the centre point Z and q is the no of adjacent neighbours, g_i is the neighbour's grey value, and $x = (g_i - g_c)$ is the variation among the centre pixel and every neighbour. The proposed CLBP algorithm is based on an LBP algorithm.

$$CLBP_M m, r = \sum_{i=0}^{m-1} s(D_i - T) \quad (15)$$

$$2^i CLBP_S m, r = \sum_{i=0}^{m-1} s(g_i - g_c) 2^i \quad (16)$$

$$CLBP_C m, r = s(g_c - g_n) \quad (17)$$

$$\text{where } D_i = g_i - g_c, g_n = \frac{1}{N} \sum_{j=0}^{N-1} g_j, T = \frac{1}{N} \sum_{j=0}^{N-1} \frac{1}{M} \sum_{i=0}^{m-1} (g_i - g_c)$$

Where, g_n helps in capturing fine-grained texture variations around a pixel.

g_n helps in capturing overall brightness contrast of a pixel against its neighbourhood.

One downside of the CLBP algorithm is that it defines only native single-scale and oriented characteristics. Based on the CLBP, the modified CLBP algorithm is proposed.

IV. Results & Discussions

The experiments were conducted on a MATLAB 2018b platform, utilizing a system with an X64-based PC, Intel Core i5-7400 CPU @ 3.00 GHz, 8 GB RAM, NVIDIA GeForce GTX 1050 Ti GPU, and Windows 10 Pro OS. The proposed hybrid retrieval system underwent testing using both Offline and Online datasets. The Offline dataset, comprising four classes with 100 images each, was meticulously prepared by applying scaling, rotation, and fusion to instill diversity for robust network training. Thirty distinct datasets, totalling over 10,000 images across different categories, were employed for training and testing. Benchmark hyperspectral datasets, including AID, UCMerced_Landuse, WHU-RS19, Botswana, and others, were incorporated. The training set was constructed using 60% of images from various categories. Qualitative and quantitative experiments affirmed the effectiveness of the proposed hybrid retrieval model, demonstrating superior performance compared to other state-of-the-art algorithms.

A. Results of Proposed MSMA-MSBT Algorithm for Online Datasets

The AID dataset results are tabulated, comprising 10,000 images. The MSMA-MSBT algorithm, proposed for training set enrichment, selects 7,090 highly informative unlabeled samples based on uncertainty, leading to increased feature diversity. The selection of samples varies across datasets, contingent on the level of uncertainty associated with each sample. The AID dataset results depicted in Table 1 showcase the selection of 7,090 highly informative, uncertain, and unlabeled samples out of a total of 10,000 images by the MSMA-MSBT algorithm. This approach emphasizes selecting samples with more uncertainty to enhance feature diversity. The method involves choosing the cluster mean and adding the most uncertain samples to the training set, thereby strengthening the classification model. With 30 different classes in the AID dataset, the selection process prioritizes classes with distinctive features. Notably, more samples are chosen from classes like Dense Residential, contributing to efficient classifier training by reducing dimensionality.

Table 1. Selected Samples using MSMA-MSBT algorithm from AID dataset.

Label	Count (selected samples/ total number of samples)
Airport	287/360
Base Land	228/310
Baseball Field	170/220
Beach	293/400
Bridge	209/360

Centre	166/260
Church	180/240
Commercial	200/350
Dense Residential	349/410
Desert	158/300
Farmland	273/370
Forest	153/250
Industrial	293/390
Meadow	183/280
Medium Residential	193/290
Mountain	243/340
Park	253/350
Parking	293/390
Playground	273/370
Pond	323/420
Port	283/380
Railway Station	163/260
Resort	193/290
River	313/410
School	203/300
Sparse Residential	203/300
Square	233/330
Stadium	193/290
Storage Tanks	263/360
Viaduct	323/420

Comparison of performance of MSMA-MSBT with state-of-art algorithms

The proposed MSMA-MSBT method is compared with the SIFT algorithm using satellite and UAV datasets. The table clearly illustrates SIFT's poor performance, particularly in the case of SAR images captured by UAV. The multi-temporal nature of these images, along with diverse sensors, introduces atmospheric noise, significantly impacting the images. Lower-altitude captures amplify the influence of atmospheric noise, making clear categorization and sample selection challenging. While results for satellite-captured images are comparable, the proposed method consistently outperforms the SIFT-based technique in terms of sample selection and classification accuracy. The classification accuracy for both satellite and UAV datasets is superior when employing the proposed method as opposed to the SIFT method. The selection of highly informative and unlabeled samples contributes to the effective training of the classifier using the proposed technique

Table 2. Comparison of selected samples for training using the proposed MSMA-MSBT algorithm for HSI/ SAR datasets.

Dataset	Proposed MSMA-MSBT Accuracy (%)	SIFT Accuracy (%) [6]
Satellite	84.14	71.71
UAV	79.25	42.94

B. Results of Proposed Modified Deep IRDI Algorithm for Online Datasets:

Phase – I & II: Training and Testing Phase

In Figure 2 (A), the feature registration in this stage utilizes the proposed modified Deep IRDI technique for enhanced matching of selected inliers. In this figure, 'a' represents the original image selected after pre-processing and sample selection for training using the proposed MSMA-MSBT

technique. Figure 'b' shows a rotated and scaled image, common in HSI/SAR images captured by different sensors with varying instantaneous field of view (IFOV). Figure 'c' depicts all inliers and outliers, with inliers representing correctly displaced feature points and outliers being incorrectly mapped points in the captured image compared to the reference image. Figure 'd' showcases matching feature points using only inliers, resulting in a robust transformation parameter that aligns the orientation of the captured image with the reference image.

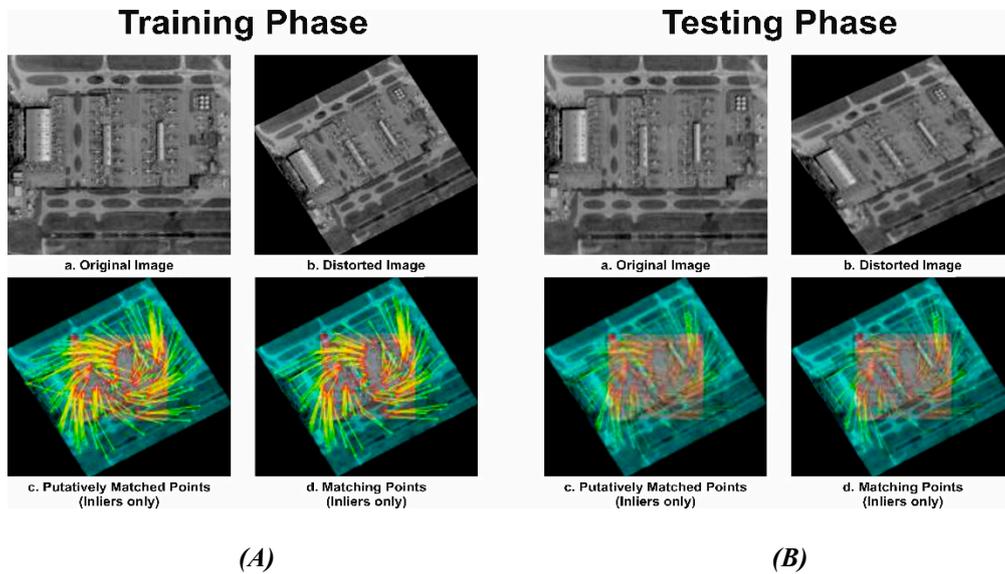


Figure 2. (A) Training and (B) Testing of HSI/ SAR image registration using proposed modified Deep IRDI algorithm for AID/SAR dataset.

In Figure 2 (B), 'a' represents the satellite-captured image used as the query image for registration. The feature registration in this stage employs the proposed modified Deep IRDI technique for improved matching of selected inliers. Figure 'b' presents a distorted HSI/SAR image, common when captured by various sensors with different IFOVs. Figure 'c' illustrates all inliers and outliers, with inliers being correctly displaced feature points and outliers representing inaccurately mapped points in the captured image compared to the reference image. Figure 'd' exhibits matching feature points using only inliers, resulting in a robust transformation parameter that aligns the orientation of the captured image with the reference image.

Comparison of performance of Modified Deep IRDI with state-of-art algorithms

Table 3 presents a comparison between the proposed modified Deep IRDI algorithm and existing state-of-the-art algorithms for online datasets, specifically satellite and UAV datasets. Experimental results demonstrate that the proposed method consistently outperforms the existing algorithms. For satellite images, the proposed modified Deep IRDI achieves the highest accuracy of 98.65%, utilizing inliers for feature point registration. In contrast, RANSAC achieves 84.14%, using both inliers and outliers for registration. Simple Image Registration (IR) algorithm attains an accuracy of 95.65%, while SIFT lags behind at 71.71%. In the case of UAV images, the proposed method achieves an accuracy of 95.43%, outperforming RANSAC (79.25%), IR (93.37%), and SIFT (42.94%). The dynamic selection of inlier feature points contributes to the superior accuracy of the proposed modified Deep IRDI algorithm, establishing it as the most accurate among the compared state-of-the-art algorithms. The average accuracy falls within the range of 80% to 87%. The proposed technique exhibits superior performance, enhancing the overall accuracy of the system. The existing IR system achieves an accuracy of 95.65%, while the proposed Modified DeepIRDI attains 98.65%, indicating a 4% increase in accuracy. The dynamic updating of the 'k' value in the algorithm enables a greater alignment of inliers compared to other techniques, fostering coherence between the captured image and the reference image, ultimately leading to improved image registration accuracy.

Table 3. Comparison of proposed modified DeepIRDI with the existing state-of-art algorithms.

Dataset	Proposed Modified DeepIRDI	RANSAC[15]	IR[18]	SIFT[22]
Satellite	98.65%	84.14%	95.65%	71.71%
UAV	95.43%	79.25%	93.37%	42.94%

The goal of this comparison is to evaluate the effectiveness, flexibility, and robustness of the proposed hybrid framework under varied image acquisition conditions. The figure highlights that existing algorithms tend to perform reasonably well on satellite data but often struggle with UAV-acquired datasets. This discrepancy is due to several factors. Satellite data generally offers broader, more stable coverage and consistent acquisition angles, whereas UAVs introduce challenges like variations in altitude, changes in lighting, more intricate textures, and smaller, highly dynamic scenes. These complexities introduce noise and inconsistencies that many conventional algorithms cannot efficiently handle.

The proposed method, however, demonstrates a superior and more balanced performance across both satellite and UAV images. This suggests a few critical strengths of the technique:

1. **Cross-platform generalization:** The proposed technique maintains high accuracy and classification reliability, regardless of whether the data source is a satellite or a UAV. This reflects its strong generalization capability.
2. **Enhanced feature learning:** The two-level feature extraction mechanism—combining handcrafted (modified MSMA-CLBP) and deep features (from pretrained CNNs)—provides a rich, hierarchical understanding of remote sensing scenes. The integration of these feature types helps to preserve both local textures and global semantic structures.
3. **Mixed pixel handling:** One of the most impressive results is the method's resilience in handling mixed pixels, which are common in HSI and SAR images. UAV datasets, in particular, contain many such mixed pixels due to high scene complexity. The soft classification recovery strategy employed by the model enables finer pixel-wise decision-making, leading to better classification results in these areas.
4. **Consistent performance improvement:** The performance gap between the proposed method and traditional approaches widens in UAV datasets, reinforcing the idea that the proposed technique is more robust under challenging conditions. This means the method doesn't just outperform others in controlled conditions but also excels in more realistic and variable environments.
5. **Real-world applicability:** The consistent and reliable performance on both satellite and UAV images positions this method as a highly viable solution for real-world applications like urban planning, agriculture monitoring, disaster management, and environmental studies, where data may come from multiple platforms and under varied conditions.

Published results are available in [20].

Results of proposed modified Deep IRDI algorithm for Offline (Multiscale (MS), Multiangle (MA), Multiscale-Multiangle (MS-MA) also known as fused) datasets:

Table 4 provides a detailed accuracy analysis for the proposed method compared to existing image registration algorithms, showcasing minimum, maximum, and average accuracy metrics with respect to an offline dataset. The average accuracy ranges from 80% to 87%, with a maximum accuracy of 98.65% observed in the angle variant dataset and a minimum accuracy of 77.25% in the same dataset using the proposed modified Deep IRDI technique. A comparison with the RANSAC algorithm, commonly used for HSI/SAR image registration, reveals average results ranging from 80% to 83%, with a maximum accuracy of 90.81% in the fusion dataset and a minimum accuracy of 73.83% in the same dataset. The proposed algorithm consistently outperforms the existing method, exhibiting an average 4% increase in accuracy attributed to its dynamic inlier selection technique.

Table 4. Represents the comparison of proposed modified Deep IRDI with the RANSAC algorithm with respect to angle, resolution and fusion dataset.

Dataset	Parameter	Modified Deep IRDI (Proposed)	RANSAC [15]
Satellite dataset with angle variant Dataset A	Avg	87.64	83.30
	Min	77.25	82.57
	Max	98.65	84.71
Satellite dataset with resolution variant Dataset B	Avg	82.75	80.54
	Min	81.00	76.19
	Max	84.14	85.10
Satellite dataset with Angle and resolution Fusion Dataset C	Avg	86.38	81.98
	Min	84.49	73.83
	Max	88.99	90.81

C. Results of Training and Testing of Multiscale Multiangle Model for SAR/Hyperspectral Images on Various Pretrained Network for High-Level Feature Extraction

Table 5 displays experimental results on SAR/Hyperspectral datasets using the proposed approach. VGG-S exhibits the highest accuracy among different pretrained CNN models, outperforming others. The models VGG-M and VGG-VDD-16 show similar results, while CaffeNet, Alexnet, and VGG-VDD-19 achieve the same average accuracy, and VGG-F performs the least among these models. These comparisons are based on both original and multiscaled images. For scene classification, seven pretrained networks (Alexnet, CaffeNet, VGG-S, VGG-M, VGG-F, VGG-VDD-16, VGG-VDD-19) are utilized, fine-tuning them in the proposed approach. The analysis reveals that multiscale images outperform original-scale images, as they provide more informative features for class distinction. Among the pretrained models, VGG-S stands out, achieving higher accuracy due to specific characteristics in its convolutional layers. The proposed method attains accuracies around 97% for multiscaled images across all considered pretrained models.

Table 5. Comparison of SAR/Hyperspectral Images using various Pretrained networks.

Accuracy	Original Image	Features of Original Image	Multiscaled Image	Features of Multiscaled Image
Alexnet	83.213	70.504	98.321	75.540
CaffeNet	78.657	79.616	98.321	82.254
VGG-F	87.530	78.177	96.643	79.616
VGG-M	87.290	69.305	98.801	70.743
VGG-S	88.010	73.621	99.041	78.657
VGG-VDD-16	76.499	71.463	97.602	75.779
VGG-VDD-19	78.897	72.902	98.321	73.861

Table 6 presents a comparison of the time required for training the network using GPU and CPU. The results indicate a significant reduction in the time per epoch when utilizing a dedicated GPU, owing to parallel processing capabilities, minimizing both time and computational complexity. The construction of mid-level features, convolutional layer count, and the number of epochs influence the training time. Fine-tuning HSI/SAR data in the initial layers of pretrained networks highlights the impact of different networks on varying data amounts. For a limited dataset, pretrained networks outperform new CNNs built from scratch, but with a larger dataset, constructing a new CNN becomes a viable option. The dedicated use of GPU substantially reduces computational costs.

Table 6. Comparison of GPU and CPU Training time in seconds of SAR/Hyperspectral Images using various Pretrained networks.

	Training Time per Epoch (seconds) with CPU	Training Time per Epoch (seconds) with GPU
Alexnet	312.000	156.00
Caffenet	315.000	157.500
VGG-F	702.000	351.000
VGG-M	798.000	399.000
VGG-S	815.000	407.500
VGG-VDD-16	1500.000	750.000
VGG-VDD-19	1800.000	900.000

D. Comparison of Classification Performance of the Traditional Methods with the Deep Learning Methods Including Our Proposed MSMA-CLBP CNN Method

Table 7 illustrates a comparison between traditional methods, deep learning models, and the proposed MSMA-CNN built from scratch. When a substantial number of training samples are available, building a CNN model from scratch is recommended, allowing for a dedicated design tailored to the input images. The graph clearly demonstrates that the proposed hybrid MSMA-CNN outperforms other traditional and deep learning techniques, achieving a 1% convergence. In comparison to traditional classification algorithms, MSMA-CNN excels in capturing deeper features, resulting in a nearly accurate classification map. The inherent structure of the CNN contributes to a significant increase in classification accuracy compared to existing traditional algorithms. Detailed results can be found in [24–26].

Table 7. Comparison of the proposed Hybrid modified MSMA-CLBP-CNN with existing state-of-art algorithms for online datasets (highest accuracy reported).

Parameters	Classification Method	Classification Accuracy (%)
Traditional Methods	Bag of Visual Words (BoVW) [21]	73.93
	Spatial Pyramid Matching using SIFT (SPM-SIFT) [21]	80.26
	Locality-constrained Linear Coding (LLC) [21]	70.89
	LDA [21]	66.85
Deep Learning Methods	S-UFL [22]	74.84
	LPCNN [23]	89.88
	Hybrid modified MSMA-CLBP-CNN (Proposed)	92.25

E. Comparison of Classification Performance of the Proposed Hybrid System with State-of-Art Algorithms for Online Datasets

The table above illustrates a comparison between the proposed multiscale pretrained CNN classification technique and existing state-of-the-art algorithms for online datasets. The results indicate a clear superiority of multiscale pretrained CNN networks, showing an overall 3% increase in average accuracy compared to traditional algorithms. The advantage lies in the initial training of pretrained networks with millions of parameters and fine-tuning the input images accordingly. Additionally, the consideration of multiscale characteristics allows for the extraction of more precise features, contributing to increased overall efficiency. Among the models, Multiscale VGG-S stands out, attributed to its smaller filter size and varied pooling layers, outperforming both traditional and

other pretrained networks. In scenarios with a limited number of training samples, pretrained models demonstrate superior performance over traditional classification algorithms.

Table 8. Comparison of the proposed Multiscale pretrained network classification with existing state-of-art algorithms for online datasets.

Convolutional Neural Networks	Classification Accuracy %
Spatial BOW [27]	81.19
SIFT + BOW [28]	75.11
SC + Pooling [29]	81.67
SPM [30]	86.8
PSR [31]	89.1
Caffenet [32]	95.02
Google net [33]	94.31
GBRCN [34]	94.53
Caffenet + fine tuning [32]	95.48
Google net + fine tuning [33]	97.1
Multiscale + Alexnet	98.32
Multiscale + Caffenet	98.32
Multiscale + VGG-F	96.64
Multiscale + VGG-M	98.8
Multiscale + VGG-S	99.04
Multiscale + VGG-VDD-16	78.9
Multiscale + VGG-VDD-19	98.32

F. Comparison of Performance Parameters of Proposed Hybrid Architecture for HSI/ SAR Image Retrieval on Offline Datasets

Table 9 presents a comprehensive set of performance metrics for evaluating the proposed hybrid retrieval system on offline datasets of Hyperspectral Imaging (HSI) and Synthetic Aperture Radar (SAR) images. The table includes various image configurations, such as different rotation angles (0°, 90°, 180°, 270°), resolutions (64×64, 256×256, 512×512), and fused datasets that combine both rotation and scaling. The evaluation metrics used are both quantitative (Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Signal to Noise Ratio (SNR) and qualitative (Entropy, Energy, Correlation, Quality Index, Kappa coefficient).

Table 9. Quantative and Qualitative parametric values for offline datasets with proposed Hybrid Architecture for HSI/ SAR Image retrieval.

Dataset	Entropy	Energy	Co-relation	Queue	PSNR	MSE	RSME	Quality index	Kappa	SNR	MAE
0°	6.87	2.08	0.72	0.32	15.67	11.0 1	4.00	0.31	1.00	7.18	7.26
90°	7.38	2.57	0.77	0.35	15.12	11.5 7	4.01	0.38	0.97	7.30	7.50
180°	6.99	2.59	0.75	0.32	15.99	11.4 6	4.06	0.39	1.00	7.24	7.31
270°	6.79	2.04	0.78	0.33	15.79	11.6 9	4.07	0.41	1.00	7.36	7.44
64x64	7.25	3.00	0.79	0.40	15.88	12.0 0	4.19	0.44	0.80	7.87	8.47
256x256	6.90	2.22	0.73	0.34	15.37	11.3 5	4.20	0.44	1.00	7.11	7.27
512x512	7.13	2.20	0.78	0.30	15.92	11.4 7	4.23	0.51	1.00	7.26	7.39
90°-64x64	7.11	3.00	0.73	0.40	16.02	11.9 7	4.30	0.60	0.75	7.19	7.06

90°-256x256	6.81	2.42	0.74	0.38	15.99	11.87	4.40	0.60	0.74	7.27	7.31
90°-512x512	7.36	2.92	0.80	0.35	16.00	11.57	4.41	0.62	1.00	7.22	7.61
180°-64x64	7.59	2.99	0.80	0.38	15.76	12.00	4.48	0.67	0.82	7.60	7.87
180°-256x256	6.90	2.13	0.79	0.39	16.00	11.99	4.50	0.70	1.00	7.17	7.25
180°-512x512	7.46	3.02	0.73	0.31	15.79	11.67	4.49	0.70	1.00	7.24	7.41
270°-64x64	7.21	2.99	0.76	0.40	16.00	12.00	4.42	0.69	0.91	7.45	7.67
270°-256x256	7.15	2.72	0.71	0.32	15.68	11.35	4.50	0.70	1.00	7.26	7.26
270°-512x512	7.65	2.94	0.80	0.32	16.00	11.54	4.49	0.70	1.00	7.25	7.39

The entropy values, ranging from 6.79 to 7.65, represent the amount of information or detail in the images. Higher entropy values, such as 7.65 for 270°-512x512, indicate richer information content. Energy, indicative of pixel intensity uniformity, peaks at 3.02 for 180°-512x512, suggesting strong textural consistency in that configuration. Correlation values measure texture similarity between the reference and processed image, with most configurations achieving values around 0.73–0.80, the highest being 0.80 in configurations like 90°-512x512, 180°-64x64, and 270°-512x512.

In terms of PSNR, which assesses image fidelity, the highest value is 16.02 for 90°-64x64, indicating excellent image quality. Complementary to this, MSE and RMSE are minimized in configurations like 0° and 90°, pointing to better reconstruction accuracy. Specifically, 0° has the lowest MSE (11.01) and RMSE (4.00), making it highly effective for preserving image detail during retrieval.

The Quality Index, representing structural similarity, increases in fused datasets. It peaks at 0.70 in configurations such as 180°-256x256, 180°-512x512, 270°-256x256, and 270°-512x512, demonstrating the advantage of multi-resolution and multi-angle data fusion in capturing structural integrity. Kappa coefficients, which measure inter-rater reliability and classification consistency, remain high (0.74 to 1.00) across most configurations, indicating stable and reliable classification results.

Signal-to-Noise Ratio (SNR) values, which reflect the strength of the image signal relative to background noise, lie between 7.11 and 7.87. Notably, 64x64 achieves the highest SNR (7.87), indicating clearer signal distinction. Lastly, Mean Absolute Error (MAE)—a measure of the average deviation between the predicted and actual values—is lowest in 90°-64x64 (7.06), underscoring this configuration's effectiveness in precise feature reconstruction.

Table 9 demonstrates that incorporating multiscale and multiangle image variants significantly improves retrieval performance. Configurations such as 90°-64x64, 270°-512x512, and 180°-256x256 consistently show superior performance across multiple parameters, validating the efficacy of the proposed hybrid retrieval architecture in handling complex HSI/SAR datasets.

G. Comparison of Performance Parameters of Proposed Hybrid Architecture for HSI/ SAR Image Retrieval on Online AID Dataset

Table 10 provides a detailed evaluation of the proposed hybrid retrieval system using the AID (Aerial Image Dataset), which comprises approximately 10,000 images across 30 diverse classes, offering a rich benchmark for assessing both quantitative and qualitative parameters. The average values for each metric are computed across all classes to ensure consistency in performance analysis. Entropy values, reflecting image complexity, range from 4.30 to 7.57 with an average of 6.70, while energy values, indicating texture uniformity, vary from 1.24 to 9.33 with a mean of 3.68, favoring configurations with smoother spatial patterns. Correlation values remain high, between 0.88 and 1.00 (average 0.98), signifying strong similarity between query and reference image patches, which is critical in remote sensing. PSNR values, indicative of image reconstruction quality, span 17.64 to 18.91 dB (average 18.29), and are inversely related to MSE, which ranges from 8.42 to 18.53 (average 10.05),

suggesting efficient image fidelity. RMSE values are consistently low, between 2.90 and 3.36 (average 3.12), highlighting minimal pixel-level error. The Quality Index varies widely from 0.09 to 3.33, averaging at 0.37, integrating luminance, contrast, and correlation into a single structural similarity score. Kappa coefficients, assessing classification agreement, range from 0.39 to 0.99 with an average of 0.76, indicating strong reliability across classes. SNR values (5.19 to 6.46, average 5.82) suggest the system's effectiveness in preserving signal clarity amid noise, while MAE values (6.30 to 6.46, average 6.39) confirm low average deviation between predicted and original image values. Collectively, the table demonstrates that the proposed hybrid approach maintains high retrieval accuracy, consistent structural similarity, and reliable classification performance across a variety of remote sensing image classes.

Table 10. Quantative and Qualitative parametric values for offline datasets with proposed Hybrid Architecture for HSI/ SAR Image retrieval.

Datasets	Entropy	Energy	Co-relation	PSN R	MSE	RMSE	Quality Index	Kappa	SN R	MAE
Airport	7.15	3.52	0.99	18.24	9.82	3.13	0.27	0.64	5.86	6.46
Base Land	6.50	2.30	1.00	18.91	8.42	2.90	0.17	0.92	5.19	6.39
Baseball Field	7.15	3.52	1.00	18.24	9.82	3.13	0.27	0.64	5.85	6.46
Beach	7.09	2.79	0.99	18.54	9.17	3.03	0.16	0.64	5.64	6.33
Bridge	7.15	3.52	0.99	18.24	9.81	3.13	0.27	0.64	5.85	6.46
Centre	7.15	3.52	0.99	18.24	9.82	3.13	0.27	0.64	5.85	6.46
Church	7.15	3.52	0.99	18.24	9.82	3.13	0.27	0.64	5.85	6.46
Commercial	7.15	3.52	0.99	18.24	9.82	3.13	0.27	0.64	5.85	6.46
Dense Residential	7.15	3.52	1.00	18.24	9.82	3.13	0.27	0.64	5.85	6.46
Desert	4.30	5.45	0.98	17.72	11.06	3.33	3.33	0.93	6.37	6.34
Farmland	5.13	3.97	0.96	17.79	10.89	3.29	0.15	0.92	6.30	6.36
Forest	6.87	1.75	0.99	18.54	9.16	3.03	0.34	0.64	5.56	6.30
Industrial	7.12	1.86	0.99	18.70	8.83	2.97	0.29	0.91	5.39	6.33
Meadow	4.67	7.51	0.88	18.38	9.52	3.08	0.09	0.95	5.72	6.39
Medium Residential	6.78	6.31	1.00	17.92	10.58	3.25	0.25	0.98	6.35	6.36
Mountain	5.86	3.56	0.96	17.64	11.29	3.36	0.26	0.96	6.46	6.41
Park	7.04	2.18	0.98	18.73	8.78	2.96	0.39	0.93	5.37	6.34
Parking	6.37	1.24	0.99	18.52	9.22	3.04	0.32	0.92	5.58	6.37
Playground	6.80	7.55	1.00	18.04	10.29	3.21	0.21	0.91	6.09	6.37
Pond	6.61	1.27	0.99	18.53	18.53	3.03	0.27	0.90	5.57	6.37
Port	6.59	8.89	1.00	17.88	10.68	3.27	0.16	0.59	6.22	6.45
Railway Station	6.55	1.99	0.97	18.90	8.45	2.91	0.30	0.95	5.20	6.36
Resort	7.57	2.00	0.97	18.32	9.64	3.11	0.39	0.52	5.78	6.39
River	7.15	3.52	1.00	18.24	9.82	3.13	0.27	0.64	5.85	6.46
School	7.04	1.32	0.98	18.09	10.17	3.19	0.45	0.52	6.01	6.36
Sparse Residential	6.39	1.29	0.92	18.69	8.86	2.98	0.29	0.99	5.42	6.39
Square	7.23	3.96	1.00	17.98	10.44	3.23	0.22	0.39	6.12	6.32
Stadium	7.15	3.52	1.00	18.24	9.82	3.13	0.27	0.64	5.85	6.46
Storage Tanks	7.26	2.18	0.99	18.45	9.38	3.06	0.38	0.61	5.66	6.40
Viaduct	6.73	9.33	0.99	18.28	9.74	3.12	0.32	0.96	5.84	6.34

H. Comparison of Performance of Region Based Similarity Measure Algorithm with State-of-Art Algorithms

The graph above depicts the recall-precision curve generated for various datasets using the proposed hybrid retrieval technique. Notably, positive results are considered only when the outcome belongs to the same semantic category as the query. The statistical analysis is conducted on 300 images across 30 distinct categories, with the first 100 retrieved image patches taken into account.

SAR and Pavia images exhibit suboptimal performance, primarily due to skew noise, which, even after denoising, diminishes essential features. Conversely, AID and Indian Pines outperform others, as high-resolution hyperspectral images are less affected by environmental noise, enabling effective global region similarity measurements. Detailed results are available in [35].

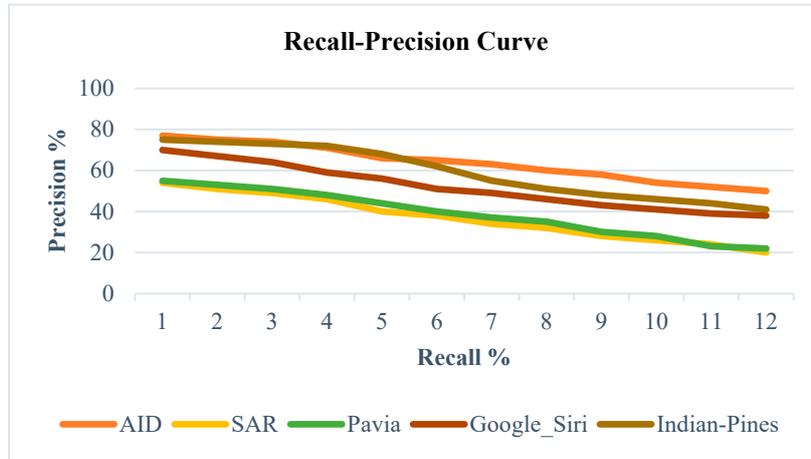


Figure 3. Comparison of Recall-Precision curve of the proposed technique for different datasets.

G. Comparison of Performance of Region Based Similarity Measure Algorithm with State-of-Art Algorithms

The table above presents a comparison of precision values between existing methods and our proposed hybrid retrieval algorithm across various online datasets. Precision, in the context of HSI/SAR images, is determined by the relevance of retrieved images compared to the overall retrieved set using the hybrid retrieval approach. The existing precision is calculated based on the IRM algorithm for image retrieval. Higher precision values indicate a greater number of relevant images retrieved relative to irrelevant ones. Notably, in the case of the AID dataset, our proposed hybrid retrieval algorithm demonstrates enhanced precision.

Table 11. Comparison of average existing precision value with average precision value by proposed method.

Datasets	Average Existing Precision value [36]	Average Precision value using proposed hybrid retrieval algorithm based on Region similarity
AID	0.33	0.79
UC_MERCED	0.51	0.83
WHU_RS	0.66	0.83

The IRM method, relying on patch-based image segmentation followed by SVM classification, struggles to effectively capture features in remote sensed images, especially in areas with mixed regions falling into multiple classes. In contrast, our proposed method utilizes multilabeling, allowing a single patch to be associated with multiple classes, leading to more efficient retrieval and superior precision values. The precision values obtained with our proposed method consistently hover around 0.80 for all three challenging online datasets, outperforming IRM, which yields precision values ranging from 0.30 to 0.69.

Table 12. Comparison of average existing recall value with average recall value by proposed method.

Datasets	Average Existing Recall value [36]	Average Recall value using proposed hybrid retrieval algorithm based on Region similarity
AID	0.49	0.78
UC_MERCED	0.57	0.82

WHU_RS	0.64	0.82
--------	------	------

The provided table presents a contrast between existing recall values and those achieved by our proposed hybrid retrieval algorithm across different online datasets. For HSI/SAR images, the recall is determined by the portion of the total number of relevant instances effectively retrieved through the hybrid retrieval approach, while the existing recall is derived from the IRM algorithm for image retrieval. A higher recall value indicates that a significant proportion of relevant images is successfully retrieved in comparison to irrelevant ones. The IRM method relies on patch-based image segmentation followed by SVM classification, which falls short in effectively capturing the features of remote sensed images. Given that remote sensed images often contain mixed regions concentrated in multiple classes, making clear-cut classification challenging, our proposed method utilizes multilabeling. This approach allows a single patch to be associated with multiple classes, resulting in a more efficient retrieval outcome and consequently, superior recall values. The recall values achieved using the proposed method consistently hover around 0.80 across all three challenging online datasets, surpassing the results obtained through IRM, which range between 0.33 to 0.66.

Table 13. Comparison of average existing F1 score value with average F1 score value by proposed method.

Datasets	Average Existing F1 Score value [36]	Average F1 Score value using proposed hybrid retrieval algorithm based on Region similarity
AID	0.37	0.79
UC_MERCED	0.51	0.82
WHU_RS	0.64	0.82

Precision and recall are quantitative and qualitative measures, respectively. To comprehensively assess the algorithm's performance, the F1 score is considered, which combines precision and recall using a mathematical model to gauge the overall accuracy. A confusion matrix with low false positives and false negatives contributes to a higher F1 score. The table above provides a comparison of existing F1 score values with our proposed hybrid retrieval algorithm for various online datasets. The F1 score for HSI/SAR images is calculated based on precision and recall values obtained through the hybrid retrieval approach, while the existing F1 score is determined using precision and recall values from the IRM algorithm for image retrieval. Achieving low false positives and false negatives in retrieval results leads to a higher F1 score, indicating the system's accurate retrieval of queried images. The IRM method relies on patch-based image segmentation followed by SVM classification, which struggles to effectively capture the features of remote sensed images. Due to the higher concentration of mixed regions in remote sensed images—areas that can fall into multiple classes and cannot be distinctly classified into one—the proposed method incorporates multilabeling. This approach allows a single patch to be assigned to multiple classes, resulting in a more efficient retrieval outcome and yielding improved precision, recall, and subsequently, F1 score values. The F1 score values using the proposed method are around 0.80 for all three challenging online datasets, surpassing the results obtained using IRM, which provides F1 score values between 0.37 and 0.64.

H. Comparison of Retrieval Performance of the Hybrid System with State-of-Art Algorithms

To assess the efficacy of the proposed hybrid retrieval method in comparison to existing retrieval algorithms, a precision-recall graph is plotted, evaluating various techniques on online datasets. The compared retrieval techniques include the FCD method [36], which calculates similarity using the fast compression distance algorithm; SIMPLIcity-I [36], classifying image patches into semantic groups and computing similarity using the IRM distance measure; SIMPLIcity-U [36], employing the UFM distance measure; and Local Gradient Coding (LGC) with different distance measures such as IRM, UFM, and D2.

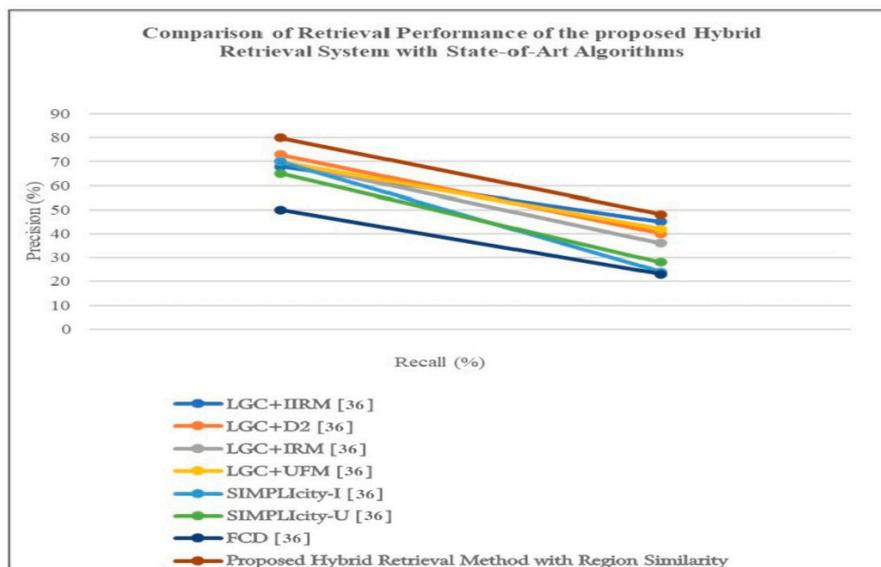


Figure 4. Comparison of the retrieval performance of the proposed hybrid retrieval method with the existing state-of-art algorithms.

In the evaluation, a dataset comprising 1500 samples from diverse categories like Mountain, Ocean, Port, High-density residential, Medium-density residential, Low-density residential, Farm, Plant, Mixed forest, and Water bodies is employed. The precision in percentage vs. recall in percentage is plotted for different retrieval algorithms. Notably, the proposed hybrid retrieval method consistently outperforms the seven existing techniques in terms of precision and recall values.

LGC+D2 exhibits better performance, although its effectiveness diminishes with an increasing number of samples, leading to higher computational complexity. The proposed method, leveraging high-level features derived from CNN and low-level features from the modified MSMA-CLBP algorithm, significantly enhances classification accuracy compared to existing methods. Additionally, region-based global mapping is employed, and similarity is computed between query and target patches.

The proposed system demonstrates stable performance and low computational complexity, particularly when utilizing GPU capabilities. Across various categories, the proposed system exhibits improvements over existing techniques, with a dedicated focus on HSI/SAR images due to their intricate nature. Incorporating soft thresholding enhances post-classification accuracy, resulting in an overall improvement of 0.08%.

Comparison with State-of-the-Art Algorithms:

A. To evaluate the effectiveness of the proposed hybrid system, extensive comparisons were conducted against several state-of-the-art traditional and deep learning-based methods across three critical modules: sample selection, image registration, and classification/retrieval. In the sample selection phase, conventional methods like the Bootstrap Transfer algorithm combined with SIFT achieved an accuracy of 63.71%. In contrast, the proposed method, MSMA-MSBT, which integrates multiscale feature fusion with active learning for unlabeled data selection, demonstrated a significant improvement with an accuracy of approximately 83.36%. This ~20% gain highlights the advantages of employing multiscale and multi-angle image features along with intelligent sample selection, reducing manual labeling effort while enhancing representational richness. In the image registration module, the proposed Deep IRDI algorithm also proved superior. While traditional RANSAC-based methods yielded an accuracy of 84.14% and SIFT achieved only 71.71%, the proposed method reached a remarkable accuracy of 98.65%. Even advanced registration algorithms like the baseline IR model achieved 95.65%, falling short of the dynamic inlier strategy employed in Deep IRDI. This dynamic approach effectively adapts to local feature variations and overcomes the rigid constraints

of fixed inlier definitions used in other methods, resulting in more accurate and consistent registration across multi-temporal and multi-source remote sensing images.

B. In the classification and retrieval stage, traditional approaches such as Bag of Visual Words, Locality-constrained Linear Coding, and Spatial Pyramid Matching using SIFT performed suboptimally, with classification accuracies typically below or around 80%. Deep learning models such as LPCNN have shown promising improvements, reaching classification accuracies close to 89.88%. However, the proposed hybrid approach, which combines MSMA-CLBP handcrafted features with high-level CNN-based features in a purpose-built architecture, surpassed these models with a classification accuracy of 92.25%. The improvement stems from the synergy of low-level texture descriptors and high-level abstract features learned by the CNN, allowing the system to better capture complex spatial and spectral patterns present in remote sensing data.

C. These comparative results clearly demonstrate that the proposed hybrid system consistently outperforms existing state-of-the-art methods across the board. Its modular architecture enables each component—whether sample selection, registration, or classification—to utilize the most effective techniques tailored to remote sensing imagery, addressing specific limitations of individual approaches. By intelligently integrating handcrafted and learned features, leveraging active learning strategies, and implementing dynamic registration techniques, the system delivers improved accuracy, robustness, and adaptability, thereby setting a new benchmark for land use and land cover mapping in remote sensing.

V. Conclusion

Increasing technological advancements in the field of remote sensing have made it easier to capture objects located at remote locations through satellites and aerial platforms. This has expanded the scope of numerous applications across environmental monitoring, urban planning, agriculture, and more. However, processing such complex imagery introduces significant technical challenges. Remote sensing images, particularly Hyperspectral and Synthetic Aperture Radar data, differ from traditional computer vision images in terms of spectral richness, noise characteristics, and spatial complexity.

With the increasing availability of large-scale remote sensing datasets, researchers are actively engaged in acquiring, storing, processing, and analyzing such data. However, despite the advancements in data collection, many systems still fall short in efficient manipulation, indexing, sorting, filtering, summarizing, or searching through such vast databases.

Content-Based Image Retrieval techniques aim to address this by retrieving images based on their visual content, extracting features such as color, texture, and shape. However, Content-Based Image Retrieval systems face the “semantic gap” challenge—a disconnect between low-level features and high-level semantic concepts. This challenge is further amplified in remote sensing images due to mixed pixels, indistinct boundaries, and complex spatial patterns.

To address these issues, the present work proposes a comprehensive hybrid retrieval system. This system integrates multiple modules, starting from the preprocessing of HSI and SAR images, followed by dynamic inlier-based image registration, active learning-based sample selection using a proposed MSMA-MSBT technique, deep learning-based classification, error correction via soft thresholding, and finally, a retrieval module based on region-based similarity matching.

The performance of this system is validated on both offline and online datasets, including the widely recognized AID dataset, which consists of 10,000 images across 30 classes. The focus is particularly on HSI/SAR images due to their richer information content compared to panchromatic or multispectral imagery.

Sample Selection using proposed modified MSMA-MSBT

Classifier performance is significantly influenced by the selection of training samples. The proposed MSMA-MSBT approach selects informative unlabelled samples that enhance the training set quality. HSI/SAR images are augmented at multiple scales and rotated to create a fused dataset, improving performance over traditional single-scale methods.

Experiments show that the fused dataset achieves higher accuracy, with the proposed MSMA-MSBT algorithm yielding a classification accuracy of approximately 83.36%, compared to 63.71% using the standard BT-SIFT combination—a nearly 20% improvement. This demonstrates the effectiveness of multiscale augmentation and selective sampling in boosting classifier performance.

Image Registration using proposed modified Deep IRDI

Image registration is crucial for HSI/SAR images, which are often captured at different times with varying resolutions and viewing angles. Traditional registration techniques using static inlier-outlier selection often fail in the presence of high variability.

The proposed modified Deep IRDI method adopts a dynamic inlier selection approach that reduces misregistration and enhances alignment. The proposed method achieves registration accuracy as high as 98.65% for satellite images and 95.43% for UAV images, significantly outperforming traditional techniques such as RANSAC, IR, and SIFT, whose accuracies range from 42% to 95%. The improvement is attributed to the adaptive and iterative inlier enhancement strategy.

Classification and Retrieval via Hybrid MSMA-CLBP-CNN

Image classification forms the foundation of retrieval accuracy. The hybrid classification approach integrates low-level features extracted via the modified MSMA-CLBP technique and high-level features obtained from a CNN architecture. These features are fused and classified using a softmax layer.

Additionally, a post-classification soft thresholding mechanism corrects misclassified pixels, particularly useful in mixed-pixel scenarios typical in HSI/SAR imagery. Retrieval is performed through a region-based patch similarity mechanism using the L2 norm, which identifies matching image regions with enhanced precision.

This approach achieves a classification accuracy of 92.25%, outperforming traditional methods like BoVW (73.93%), LLC (70.89%), and even recent deep learning models such as LPCNN (89.88%). The hybrid strategy benefits from a dedicated CNN built from scratch tailored to the domain-specific characteristics of remote sensing imagery.

Future Scope: 1. More Features can be tested

The feature selection process is crucial in the proposed hybrid retrieval system. While three features are currently considered, evaluating the system's performance with a broader array of features could provide valuable insights.

2. Application Building for constrained environment

The suggested hybrid retrieval system has the potential for optimization and adaptation to a constrained environment. It could be reengineered to create a mobile application focused on effective online retrieval of HSI/ SAR images, with the potential to support various other applications.

References

1. J. Delgado, G. Martin, J. Plaza, L. I. Jimenez, and A. Plaza, "GPU implementation of spatial preprocessing for spectral unmixing of hyperspectral data," in 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2015, pp. 5043–5046.
2. G. Martin, L. I. Jimenez and A. Plaza, "A new tool for evaluating spectral unmixing applications for remotely sensed hyperspectral image analysis," Proc. Int. Conf. GEOBIA, pp. 1–5, 2012.
3. L. Fang, Z. Liu, and W. Song, "Deep Hashing Neural Networks for Hyperspectral Image Feature Extraction," IEEE Geosci. Remote Sens. Lett., vol. 16, no. 9, pp. 1412–1416, 2019.
4. C. Binge, F. Zongqi, X. Xiaoyun, Z. Yong, and Z. Liwei, "A Novel Feature Extraction Method for Hyperspectral Image Classification," in 2016 International Conference on Intelligent Transportation, Big Data Smart City (ICITBS), 2016, pp. 51–54.
5. W. Masarczyk, P. Głomb, and B. Grabowski, "Effective Training of Deep Convolutional Neural Networks for Hyperspectral Image Classification through Artificial Labeling," 2020.
6. Z. Xue, S. Zhou, and P. Zhao, "Active Learning Improved by Neighborhoods and Superpixels for Hyperspectral Image Classification," IEEE Geosci. Remote Sens. Lett., vol. 15, no. 3, pp. 469–473, Mar. 2018.

7. Y. Guo, H. Cao, J. Bai, and Y. Bai, "High Efficient Deep Feature Extraction and Classification of Spectral-Spatial Hyperspectral Image Using Cross Domain Convolutional Neural Networks," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 1, pp. 345–356, Jan. 2019.
8. Q. Deng, H. Lu, H. Tao, M. Hu, and F. Zhao, "Multi-Scale Convolutional Neural Networks for Space Infrared Point Objects Discrimination," *IEEE Access*, vol. 7, pp. 28113–28123, 2019.
9. Y. Li, Y. Zhang, X. Huang, and J. Ma, "Learning Source-Invariant Deep Hashing Convolutional Neural Networks for Cross-Source Remote Sensing Image Retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6521–6536, Nov. 2018.
10. O. E. Dai, B. Demir, B. Sankur, and L. Bruzzone, "A novel system for content based retrieval of multi-label remote sensing images," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 1744–1747.
11. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", <https://proceedings.neurips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.
12. Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, Trevor Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding", *Computer Vision and Pattern Recognition (cs.CV); Machine Learning (cs.LG); Neural and Evolutionary Computing (cs.NE)*, <https://doi.org/10.48550/arXiv.1408.5093>
13. Karen Simonyan* & Andrew Zisserman+, "Very Deep Convolutional Networks for Large-Scale Image Recognition", Published as a conference paper at ICLR 2015, <https://doi.org/10.48550/arXiv.1409.1556>.
14. Wen Shu 1,2, Peng Liu 1, Guojin He 1, and Guizhou Wang 1, "Hyperspectral Image Classification Using Spectral-Spatial Features With Informative Samples", *2019 IEEE Access*, Volume 7, 2019, DOI: 10.1109/ACCESS.2019.2894766.
15. L.-R. Dung, C.-M. Huang, and Y.-Y. Wu, "Implementation of RANSAC Algorithm for Feature-Based Image Registration," *J. Comput. Commun.*, vol. 01, no. 06, pp. 46–50, 2013.
16. Zhuoqian Yang 1, Tingting Dan 2, and Yang Yang 2,3, "Multi-Temporal Remote Sensing Image Registration Using Deep Convolutional Features", *2018 IEEE Access*, Volume 6, 2018, DOI: 10.1109/ACCESS.2018.2853100.
17. Qiaoqiao Shi 1, Wei Li 1, Fan Zhang 1, Wei Hu 1, Xu Sun 2, and Lianru Gao 2, "Deep CNN With Multi-Scale Rotation Invariance Features for Ship Classification", *2018 IEEE Access*, Volume 6, 2018, DOI: 10.1109/ACCESS.2018.2853620.
18. Q. Huang, J. Yang, C. Wang, J. Chen, and Y. Meng, "Improved registration method for infrared and visible remote sensing image using NSCT and SIFT," in *2012 IEEE International Geoscience and Remote Sensing Symposium*, 2012, pp. 2360–2363.
19. Mahesh and M. V Subramanyam, "Automatic feature based image registration using SIFT algorithm," in *2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12)*, 2012, pp. 1–5.
20. S. S. Alegavi and R. R. Sedamkar, "Using deep convolutional features for multitemporal remote sensing image registration," *Int. J. Adv. Sci. Technol.*, vol. 28, no. 1, pp. 230–240, 2019.
21. B. Zhao, Y. Zhong, and L. Zhang, "A spectral–structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 73–85, 2016.
22. F. Zhang, B. Du, and L. Zhang, "Saliency-Guided Unsupervised Feature Learning for Scene Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2175–2184, Apr. 2015.
23. Y. Zhong, F. Fei, and L. Zhang, "Large patch convolutional neural networks for the scene classification of high spatial resolution imagery," *J. Appl. Remote Sens.*, vol. 10, no. 2, p. 025006, 2016.
24. S. S. Alegavi and R. R. Sedamkar, "Implementation of deep convolutional neural network for classification of multiscaled and multiangled remote sensing scene," *Intell. Decis. Technol.*, vol. 14, pp. 21–34, 2020.
25. S. S. Alegavi, "Transfer Learning for Multiscaled Classification of Hyperspectral Images using VGG-VDD Pretrained Networks," vol. 28, no. 20, pp. 1290–1303, 2019.

26. S. Alegavi and R. Sedamkar, "Classification of Hybrid Multiscaled Remote Sensing Scene Using Pretrained Convolutional Neural Networks," in *Computational Vision and Bio-Inspired Computing*, 2020, pp. 148–160.
27. B. Sirisha, B. Sandhya, C. S. Paidimarry, and A. S. C. Sastry, "Bag-of-Spatial Words(BoSW)Framework for Predicting SAR Image Registration in Real Time Applications," *Procedia Comput. Sci.*, vol. 115, pp. 431–439, 2017.
28. M. Issolah, D. Lingrand, and F. Precioso, "SIFT, BoW architecture and one-against-all Support vector machine," *CEUR Workshop Proc.*, vol. 1179, pp. 0–5, 2013.
29. J. Li et al., "Normalization and dropout for stochastic computing-based deep convolutional neural networks," *Integration*, vol. 65, pp. 395–403, 2019.
30. S. Wu, P. Yang, J. Ren, Z. Chen, and C. Liu, "A Novel Sub-Pixel Mapping Model Based on Pixel Aggregation Degree for Small-Sized Land-Cover," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 3073–3076.
31. J. Wang, C. Luo, H. Huang, H. Zhao, and S. Wang, "Transferring pre-trained deep CNNs for remote scene classification with general features learned from linear PCA network," *Remote Sens.*, vol. 9, no. 3, 2017.
32. Y. Jia et al., "Caffe: Convolutional architecture for fast feature embedding," *MM 2014 - Proc. 2014 ACM Conf. Multimed.*, pp. 675–678, 2014.
33. P. Aswathy, Siddhartha, and D. Mishra, "Deep GoogLeNet Features for Visual Object Tracking," in *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*, 2018, pp. 60–66.
34. F. Zhang, B. Du, and L. Zhang, "Scene Classification via a Gradient Boosting Random Convolutional Network Framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1793–1802, Mar. 2016.
35. S. Alegavi and R. Sedamkar, "Improving Classification Error for Mixed Pixels in Satellite Images Using Soft Thresholding Technique," in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2018, pp. 955–961.
36. L. Jiao, X. Tang, B. Hou, and S. Wang, "SAR Images Retrieval Based on Semantic Classification and Region-Based Similarity Measure for Earth Observation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 8, pp. 3876–3891, 2015.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.