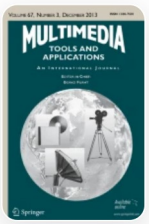
Fusion Based Feature Extraction and Optimal Feature Selection in Remote Sensing Image Retrieval

Published: 11 April 2022

Volume 81, pages 31787–31814, (2022) Cite this article



Multimedia Tools and Applications

Aims and scope

Submit manuscript

Minakshi N. Vharkate & Vijaya B. Musande

Abstract

In remote sensing (RS) community, RSIR (Remote Sensing Image Retrieval) is considered as a tough topic and gained more attention because the data is collected via EO (Earth Observation) satellites. As huge numbers of RS images are available, the lack of labelled samples, complex contents obstructs the understanding of RS images. Therefore, accurate and effective image retrieval (IR) system named fusion based feature extraction and meta-heuristic algorithm

based feature selection is presented in this work for performing RSIR. Pre-processing is done using Kernel PCA (KPCA). Next, fusion of 3 CNN (Fused CNN) architectures namely Visual Geometry Group (VGG 16, VGG 19) and ResNet (Residual Network) is used for feature extraction. The selection of features is performed using Joint MI (Joint Mutual Information) optimized using RFO (Rain-Fall Optimization) algorithm. Next, similarity is measured using Weighted Euclidean Distance (WED) metric. Finally, Relevance Feedback Model (RFM) verifies whether the search results have met the user query. The implementation tool is PYTHON and the three online databases used for testing are WHU-RS19, AID, and UCM. Hence, the simulation outcomes reveal that the presented Fused CNN model achieved improved mAP performances such as 93.693%, 94.716%, and 95.067% on the datasets than the baseline architectures.

1 Introduction

The process of monitoring and detecting the physical appearance of a particular area by

evaluating its emitted and reflected radiation at specific distance using aircraft or satellite is termed as remote sensing (RS). Information is collected from an object/area using distinctive cameras, airborne or satellite sensors. RS images helps the researchers to sense the Earth activities in an intelligent manner [28, 33]. Some of the benefits of RS include characterizing physical items or natural features on the ground, ability to gather data over huge spatial areas, observing the objects as well as surface areas on an orderly manner and monitoring its deviations over time, integrating the RS data with other info which helps in the execution of decisions. Over the past decades, the quality and quantity of RS data is emerging with the rapid advance of earth observation (EO) technologies [13, 25].

Identifying the environmental impacts and its changes is very critical and has turn out to be more effective to social and economic development, as detailed RS imagery is being available progressively [7, 10]. Digital and photographic images are used in examining the changes induced by humans and the natural characteristics on global environment [1]. RS or satellite images are very complex in nature that includes mixed representation of contents which can be utilized as reference data in various applications such as deforestation detection, urban planning, agricltural monitoring etc. [2, 14, 15]. The databases of RS images are high-dimensional, massive and quickly-emerging archives of spatial information. As these images hold substantial quantity of valuable information, manual identification is very difficult [34].

Advancements in the field of computer technology offer newer chances to enhance remote sensing and image detection and enable the growth of newer methods [27]. As RS imagery resolution increases, it is tough to process the images on RS scenes [17]. In large image databanks, the user-defined classes consists of various image groups and distance at different scales in low-level (LL) visual descriptors space [26]. Also, the interactive retrieval process of such image class is very difficult. Hence, RSIR aims at retrieving the data efficiently from the huge RS data collection. In RS, the fundamental task is the process of RSIR. The main objective is to retrieve the RS images efficiently and automatically satisfying the requirements of the user [18, 31]. The categorization of land-use from RS images is essential for managing and monitoring the kinds of human activities involved.

RSIR is considered as a tough topic because of the complicated contents and specific characteristics of RS images. Deep Learning (DL) techniques have advanced the baseline

architecture performance in object detection, speech recognition and image classification. Inspired by the excessive success of DL, the RS community applied this DL technique in effective retrieval of RS images in order to extract semantic high-level features. DL is now emerging as a capable subdivision of Machine Learning which illustrates the high-level abstractions using effective deep network architecture [24, 32]. In large datasets, DL is proficient of identifying intricate structures which is different from the hand-crafted features in learning the useful feature representations.

Recently, DL and CNN (Convolutional Neural Network) models have observed a highest performance in object detection, image classification. The dual CNN models namely VGG baseline architecture and ResNet pre-trained on ImageNet have gained significant performance in RSIR. DL techniques for RSIR, can be divided as unsupervised and supervised feature learning models. CNN is a supervised model in which GoogleNet, VGG-16, SqueezeNet, ResNet, VGG-19 shown superior performance compared to unsupervised methods like auto-encoder and K-Means method. DL networks performs like human brain and learns concepts similar to human perception. Using its domain knowledge, DL learns image features automatically and maps the images directly to the output and overcomes the limitations of handcrafted features [3, 5].

Images have occupied an essential place in all areas of life. The image capturing rate in the domains like medical science and remote sensing has increased due to the advancements in technologies. This lead to the necessity of evolving a smarter image retrieval system and has attracted much attention from researchers around the globe. Existing researches on image retrieval mainly focus on low-level (LL) features like texture, shape and color etc. However, these LL features be likely to achieve inadequate retrieval outcomes in RS image community due to the presence of semantic gap. Also, the RS images consist of non-patterned arrays of pixels, from which the extraction of visual features is required for the understanding of semantic contents. This is due to the fact that identical scenes may represent different complex textures and structures in RS images. The process of RSIR aims to search for the matched areas by examining the similarities among the given image query with that of the databank images. But the retrieval process is very challenging because of the view point, high resolution, spectral nature, coverage area, time difference and numerous image contents. Moreover, these images comprise various semantic objects that complicate the task of

retrieval clearly. Over the ancient times, vivid changes has experienced in RS due to the increased acquisition rate and spatial resolution of RS imagery. Moreover, these changes show significant effects in managing and utilizing the RS images. An essential feature for managing RS data is attaining success in retrieving large volume RS images. Traditional methods on image analysis and feature extraction were directed by means of hand-crafted (HC) features which is benefitted using approaches like BoVW (Bag-of-Visual Words), HOG (Histogram of Oriented Gradients), SIFT (Scale-Invariant Feature Transformation). However, with these HC features on various RS tasks, the local features (LF) illustrate superior performance compared to global features (GF) which are not jointly exclusive. In RSIR, the performance of retrieval depends greatly on the efficacy of the feature representations. To overcome the issues related with RSIR, a novel retrieval approach is developed in this work with the aim of attaining good retrieval accuracy. This motivated the researchers to develop a fused feature extraction model combined with feature selection principle for proper feature representations and to increase the performance of retrieval. The main contributions of this work are as follows:

- To design the novel fusion based feature extraction method using the Fused CNN architectures (VGG16, VGG19, ResNet) for extracting accurate and efficient features.
- To introduce an optimal feature selection model to select the best feature set using Joint MI principle optimized RFO meta-heuristic algorithm model.
- To minimize the time complexity and to enhance the image retrieval accuracy for the RS image dataset.
- To minimize the retrieval of non-relevant images and to improve the retrieval precision with the weighted distance metric (WED) using this fused feature extraction and optimal feature selection approaches.
- To evaluate the system performance based on the relevance feedback model which satisfies the user needs perfectly.

The rest of this paper is systematized as: Section 2 reviews the related works of this paper. In Section 3, the presented methodology is described in detail and the Section 4 examines the simulation results of the presented system. Lastly, Section 5 concludes the manuscript.

2 Related works

Some of the recent works related to the retrieval of remote sensing images are as follows:

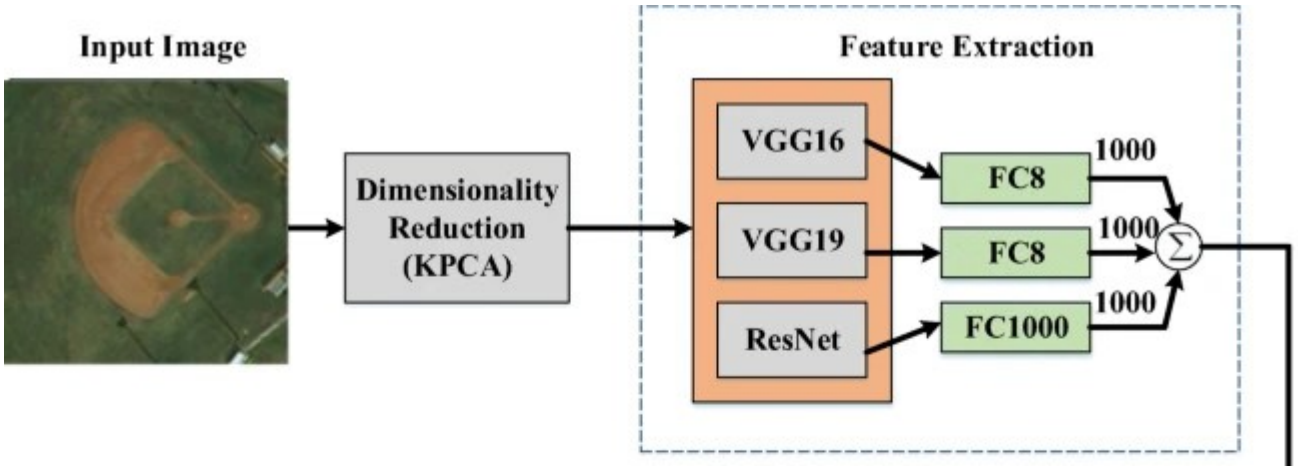
Imbriaco et al. [9] presented an aggregated IR (Image Retrieval) framework named VLAD (Vector of Locally Aggregated Descriptors) which has used both local and attentive convolutional features. Here the feature selection was done with the help of dual attention mechanisms which offers improved performance. The efficacy of retrieval performance was evaluated on 4 various datasets. The proposed VLAD model generates a solitary vector which encodes the VI (visual information) and outperforms the FK (Fisher kernels) and BoW (Bag-of-Words) techniques. Paolo Napoletano [22] introduced the assessment of visual descriptors in an extensive manner for CBIR (content based IR) in RS images. The evaluation procedure involves CNNs, local, global hand-crafted features incorporated with different CBIR models. The 4 various approaches were manual RF (Relevance Feedback), basic CBIR, RF based on active-learning and pseudo RF. The dual datasets used for the evaluation was SceneSat (Satellite Scene) and UCM dataset. Outcomes prove that CNN-based features perform superior to local and global hand-crafted features. Rui Cao et al. [4] proposed a novel approach called TDML-CNN (Triplet Deep Metric Learning-CNN) for CBRSIR. The RS images representative features were constructed with TDML based objective function. To compare the image similarity, ED (Euclidean Distance) metric was effectively employed which retrieves the RS images of similar class. The proposed TDML-CNN approach was tested on dual datasets. The extracted features dimension was reduced using supervised CNN fully connected (FC) layers and unsupervised PCA methods. These methods demonstrate the effectiveness of RSIR. Wang et al. [30] presented an effective RSIR scheme called CBP (compact bilinear pooling) and it was combined with CNNs. The 3 different stages of CBP were pre-training, tuning as well as retrieval. In the first step, the classic architectures of CNN such as ResNet34, VGG16 were pre-trained with the structure of ImageNet. In both networks, CBP layer was adopted before FC layers. Dual mechanisms such as spatial unified attention and channel were used to extract the consistent representations. Next, the network was fine-tuned. Finally, PCA (principal component analysis) was replaced at FC layers to perform the retrieval process. Caihong et al. [20] presented a RSIR model based on ENN (Ensemble Neural Networks). This network model used the texture and color features and also acquired better retrieval outcomes than the usage of only one NN. The AI (Aerial Image) outcomes obtained with ENN was compared with 2

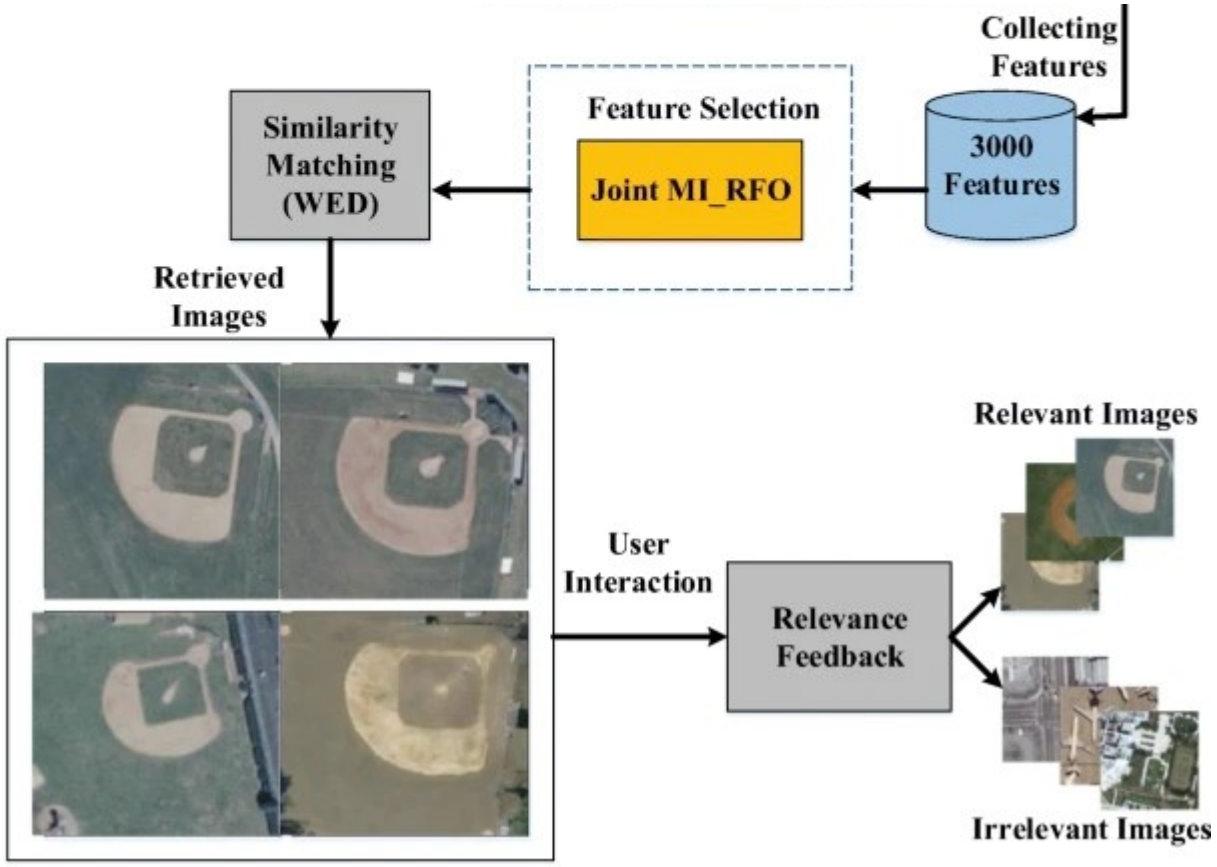
ENNs and 10 individual NNs and the proposed ENN approach has a high degree of precision. Also, the mean precision and coverage rate has achieved dramatic progress of >40% compared with existing approaches. Varsha and Musande [8] introduced a solution to identify the disease in wheat using image processing (IP). The suggested work using IP was termed as a software solution to categorize the plant disease automatically. This work was the integration of image acquisition, IP, image segmentation, feature extraction and neural network, SVM based classification. Musande et al. [21] presented the evaluation via fuzzy-based classifiers for the identification of single crop using temporal MSSI (multi-spectral satellite images). The crop considered was named as a cotton crop. Totally, 5 spectral indices were considered to recognize the cotton crop. The overall accuracy obtained was 96.02% using Noise Classifier. Sandip and Musande [12] introduced the discrimination of cotton crop using multispectral (Landsat-8) data. In order to differentiate the land cover units, the classification approach used was unsupervised K-Means classification. The efficiency and accuracy of the classifier were examined by comparing the outcomes with confusion matrix and kappa statistics.

3 Proposed method

In RS community, there is a great demand for efficient and effective image retrieval. The significant effort on RSIR has been made to improve learning-based features from the perspective of the network structure. In this work, an effective fusion based feature extraction and meta-heuristic algorithm based feature selection is presented for performing RSIR. Figure 1 illustrates the schematic of proposed RSIR framework.

Fig. 1





Schematic of Proposed RSIR Framework

Initially, pre-processing is presented as a dimensionality reduction procedure which is done using KPCA. Next, fusion based feature extraction is processed using the fusion of 3 CNN architectures namely VGG 16, VGG 19 and ResNet. Here, the most prominent features are obtained by merging the features from the last fully connected layers of the 3 fused architectures. After the extraction of features, Feature Section is performed using Joint MI optimized using RFO algorithm. Next is Similarity Matching, the degree of similarity between the query images to images inside each group is measured using WED metric. Finally, RFM is applied to verify that whether the search results met the user query. It can also be utilized as a feedback model to improve the retrieval outcome.

3.1 KPCA based dimensionality reduction

Dimensionality refers to the total number of features (input variables) of a given dataset. The process of reducing the total features in the dataset is known as dimensionality reduction. It can be of linear and non-linear type. Some of the common techniques used for dimensionality

reduction are PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis), KPCA, UMAP (Uniform Manifold Approximation and Projection) etc. However, these models have significant advantages to data visualization. The following dimensionality reduction methods used in machine learning can make the dataset well fit for the model and simplify the classification process. In this work, KPCA based dimensionality reduction method is used to reduce the dataset attributes. KPCA have the ability to solve the problem of non-linearity. The observed data can be converted into possibly infinite (high) dimensional feature space by means of the kernel function.

The non-linear form of PCA is named as KPCA and the information from original data can be represented in reduced form. KPCA make use of ϕ function which converts z (input vector) into a newer vector form$\phi(z)$. In KPCA, using the non-linear map the H_F high dimensional FS (feature space) is mapped from the input data is expressed as

$$ \phi :{\mathfrak{R}}^m\to {\mathrm{H}}_F $$

(1)

$$ z\to \phi (z) $$

(2)

As the FS dimensionality H_F is extremely high it is not feasible to process PCA. Therefore, kernel function is used in this work to avoid this difficulty. The samples in FS is assumed as,

$$ {\sum}_{i=1}^v\phi \left({z}_i\right)=0 $$

(3)

Co-variance matrix function is represented as,

$$ CV=\frac{1}{v}{\phi \phi}^T $$

(4)

where, $\phi = [\phi(z_1), ...\phi(z_v)]$. The representation of eigen values ($\lambda \geq 0$), eigen vectors $m \in H_F$is expressed as,

$$ \lambda s= mCV $$

(5)

$$ m=\sum \limits_{i=1}^v\phi \left({z}_i\right){s}^i $$

(6)

Substitute Eq. (4), (6) in Eq. (5) and signifying $K = \phi\phi^T$ (Gram matrix), then.

$$ v\lambda s=\mathrm{K}s $$

(7)

Here, the term K is the gram matrix. Next, the non-linear principle components are extracted with the test sample z using the kernel function represented as,

$$\left(m\cdot \phi (z)\right)=\sum_{i=1}^v{s}_i\left(\phi \left({z}_i\right)\cdot \phi (z)\right)=\sum_{i=1}^v{s}_ik\left({z}_i,z\right)$$

(8)

Some of the widely used kernels are sigmoid, gaussian and polynomial functions. Here, the gaussian kernel is used as the kernel function which is described as,

$$ {K}_{gau}\left(i,j\right)=\mathit{\exp}\frac{-{\left\Vert {z}_i-{z}_j\right\Vert}^2}{2{\sigma}^2},1\le i\le n;1\le j\le n $$

(9)

where, the feature representation of both i, j^{th} data sample is denoted as z_i and z_j, n signifies

the set of labelled images. Let the representation of feature vector is expressed as,

$$ {z}_i=\frac{z_i}{{\left\Vert {z}_i\right\Vert}^2} $$

(10)

The representation of squared Euclidean norm among the two vectors can be given as,

$$ {\left\Vert {z}_i-{z}_j\right\Vert}_2^2={\left\Vert {z}_i\right\Vert}_2^2+{\left\Vert {z}_j\right\Vert}_2^2-2{z}_i^T{z}_j $$

(11)

From the above Eq. (9–11), the similarity value among the 2 vectors obtained is illustrated as,

$$ {K}_{gau}\left(i,j\right)=\mathit{\exp}\frac{-2\left(1-{z}_i^T{z}_j\right)}{2{\sigma}^2}=\mathit{\exp}\frac{-\left(1-{Z}^TZ\right)}{\sigma^2} $$

(12)

Thus, the input dataset dimensionality is reduced from the selected features subset by the KPCA technique.

3.2 Fusion based feature extraction

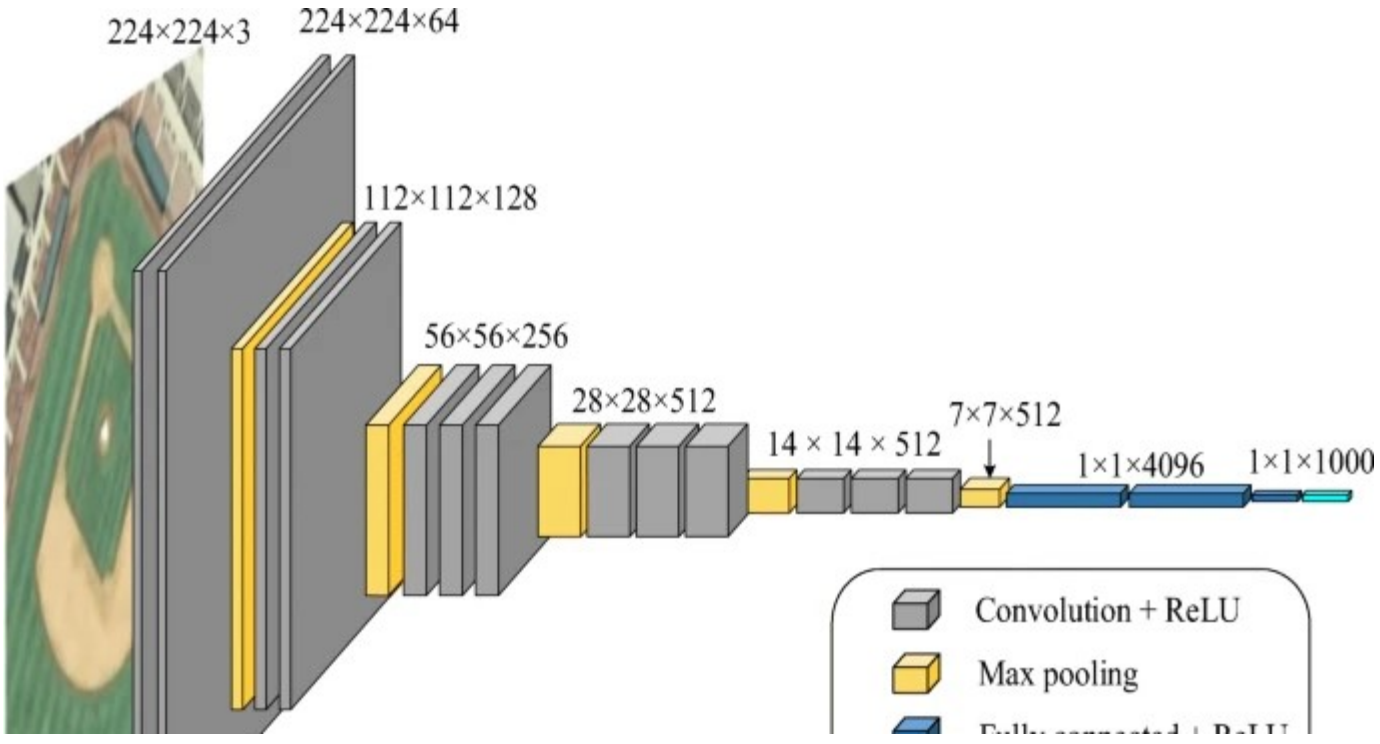
The method of converting the original (raw) data into numerical features and preserving the original dataset information is termed as Feature Extraction. This can also been applied as a dimensionality reduction process. The most important image parts are extracted from the pre-processed data as a feature vector. The processing of input images require number of resources to be calculated. However, this feature extraction process aims to attain the best feature from the dataset. Feature extraction minimizes the volume of redundant data without any information loss. Here, a fusion based feature extraction is presented by means of 3 CNN architecture models such as VGG 16, VGG 19 and ResNet [16].

Generally, the CNN is a similar form of ANN (Artificial Neural Network) which can be used in applications like pattern detection, object tracking, image classification and analysis. CNN architecture have gained more popularity in computer vision tasks after it has won the competition (ImageNet) in the year 2012. The structure of CNN includes the arrangement of layers and each one has its own unique function once the data is fed into it. The layers of CNN are input layer (IL), convolutional layer (CL), activation layer (AL), pooling layer (PL) and fully connected layer (FCL). The IL holds the dimensionality reduced input data. The dot product among image patch and each filter is performed using CL and calculates the output dimensions. The activation function is applied using AL on the output of each element of CL. Next, the PL minimizes the cost of computation and makes the AL output to be memory efficient. The last FCL, takes the input from the former layers and outputs the class-scores computed. Thus, the features extracted with CNNs is termed as learned features.

3.2.1 VGG 16

VGG 16 is a CNN architecture developed in 2014 by Karen Simonyan and Andrew Zisserman in VGG lab of University called Oxford. The term VGG stands for Visual Geometry Group. It is also named as OxfordNet. Fig represents the architecture of VGG 16 model. Figure 2 signifies the VGG 16 network architecture.

Fig. 2



VGG 16 Network Architecture

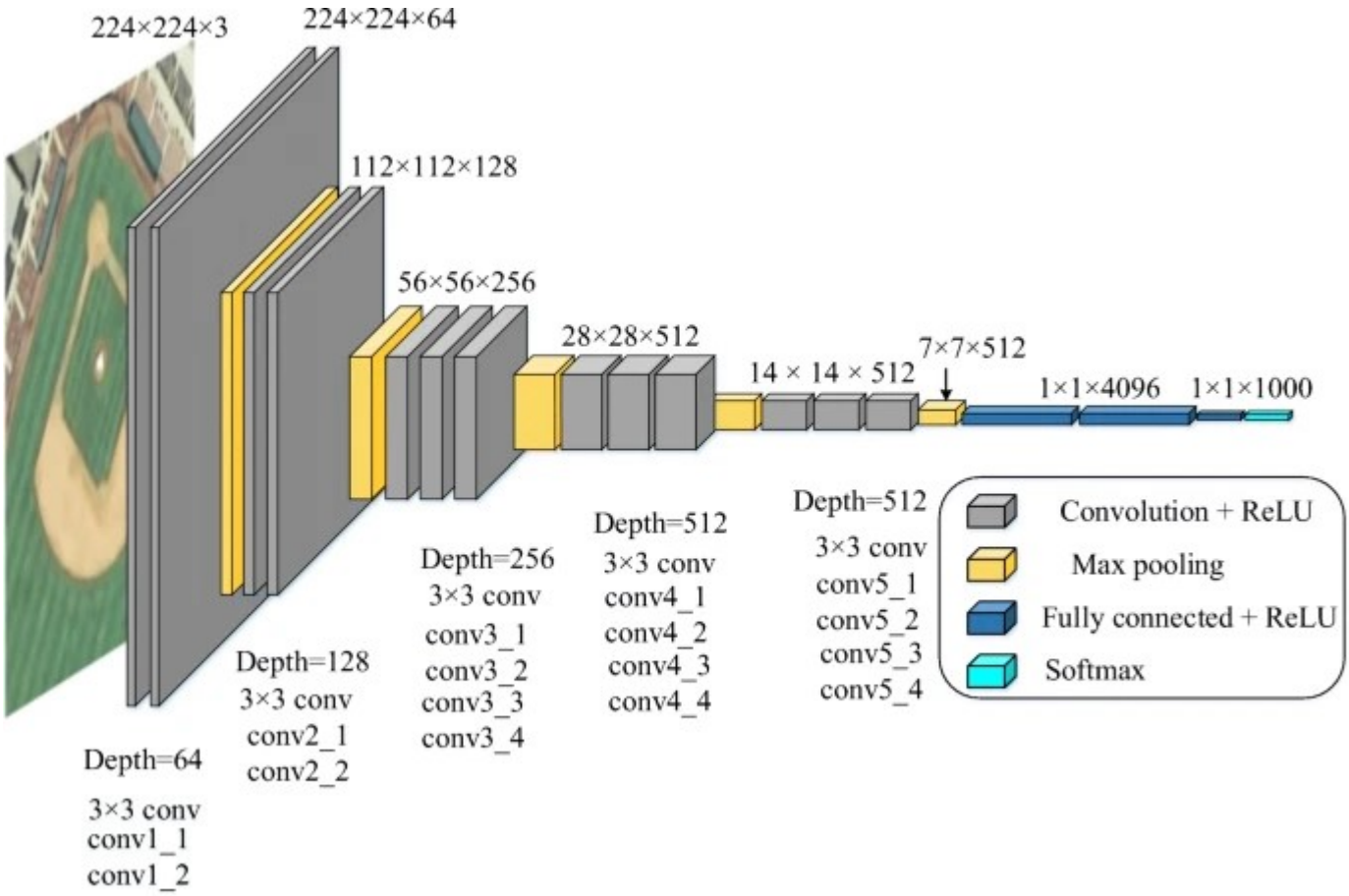
The input dimension of VGG 16 model is 224 × 224 × 3. In VGG 16, the number 16 represents the total 16 layers which is the combination of 3 FCLs and 13 CLs. The VGG 16 model includes 138 M (M = million) parameters. The first 2 layers of VGG 16 have the filter size 3 × 3 and include 64 channels. As an alternative of having more hyper-parameters, VGG 16 model used similar padding which comprises CLs of filter size 3 × 3(stride 1) and Max-PL of 2 × 2(stride 2). This arrangement is followed consistently by both CL and Max-PL throughout the entire architecture. Then the architecture is followed by dual CLs of size 3 × 3and another CL includes 256 filter. Next, there exists dual sets of 3 CLs and one pooling layer. The last layer is the linear classifier (Softmax) which is used only for classification purpose. In VGG 16 architecture, the final FCL (FC8) act as a feature extractor from which the features are acquired.

3.2.2 VGG 19

The architecture of VGG 19 is related to VGG 16 which is also developed by Simonyan and Zisserman. It consists of total 19 layers which is the combination of 16 CLs and 3 FCLs. VGG 19 make use of 3 × 3filters (stride 1) and Max-PL of filters 2 × 2(stride 2). This CNN model includes more layers when compared to AlexNet architecture. The VGG 19 model includes 143 M (M = million) total parameters and placed first position in localization and second position in classification at ILSVR Challenge in 2014.

Figure 3 illustrates the VGG 19 network architecture. The structure of VGG 19 includes 41 layers which is the integration of ReLu layer, FCL, Max-PL, Dropout and Softmax layer. The size of the input image is similar to VGG 16 model with some additional layers. The last layer is the linear classifier (Softmax) which is used only for classification purpose. In VGG 19 architecture, the final FCL (FC8) act as a feature extractor from which the features are acquired.

Fig. 3

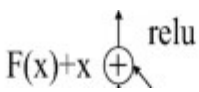


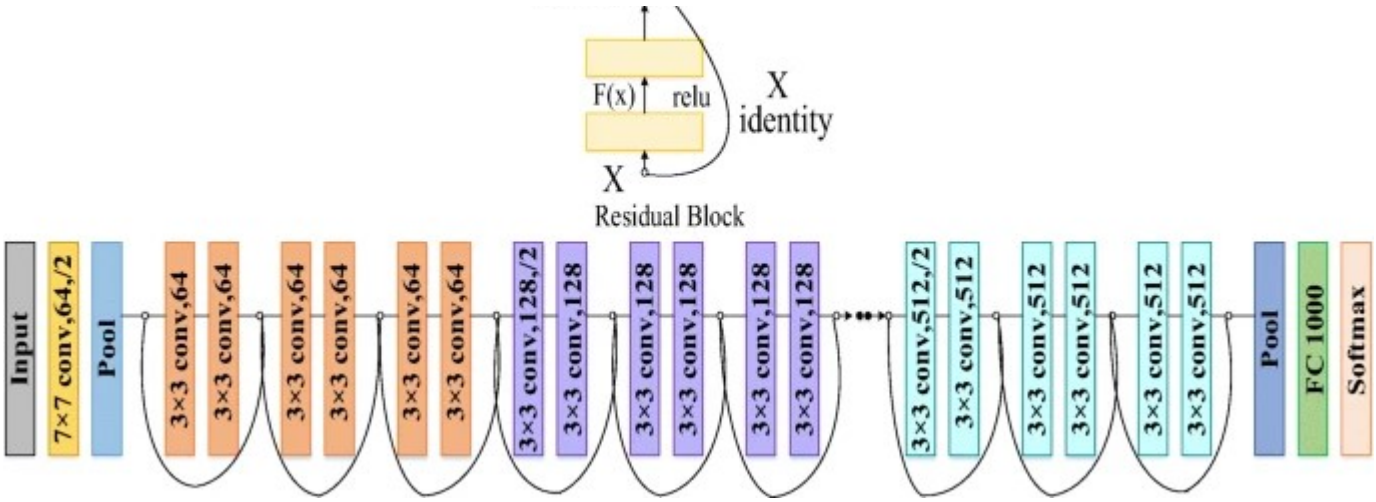
VGG 19 Network Architecture

3.2.3 ResNet

The term ResNet refers to Residual Network introduced by the researchers in 2015 at Microsoft Research. Compared to other network architectures, a new technique named skip connections was introduced in ResNet to overcome the problem of exploding gradient. However, this new approach skips the training from some layers and directly connects to the output. If the architecture performance is degraded due to any of the layers then that layer is skipped using the skip connection technique. Figure 4 displays the Architecture of ResNet-50.

Fig. 4





Architecture of ResNet-50

The residual blocks in ResNet -50 model uses the following design rules such as (a) the layers comprise similar number of filters, for the similar feature map size (b) as the size of feature map splits then filter number becomes doubled. ResNet-50 model includes total 50 weighted layers. The CLs prefers down sampling directly which involves the stride 2. Also, after every convolution as well as before the ReLU activation, batch normalization is done. Identity (I) shortcut is utilized, if the input as well as the output are of similar dimensions. Next, the projection (P) shortcut is utilized if the dimensions increments. However, this shortcut matches the dimensions via 1 × 1convolutions. FCL and a softmax layer is present at the end of the network. In this network, FC 1000 act as the feature extractor layer.

3.3 Joint MI_RFO based feature selection

The process of selecting the optimal number of features is termed as Feature Selection. It is essential to effectively select the features because the retrieval system performance always be governed by the extracted features. Here, the feature selection process is performed using Joint MI (Joint Mutual Information) optimized using RFO (Rain-Fall Optimization) algorithm. The features are selected based on the characteristics such as cost of computation, feature lessening ability, and accuracy. Here, the features are ranked and selected using Joint MI_RFO which offers better accuracy and stability compared to other feature selection approaches.

The term MI (Mutual Information) evaluates the dependencies among dual random variables

Z and K. As stated by Shannon's entropy definition the parameter $H$is used to evaluate the uncertainty existing in Z which is expressed as,

$$ H(Z)=\sum \limits_{z\in Z}p(z) logp(z) $$

(13)

The conditional entropy is expressed as,

$$ H\left(Z|K\right)=\sum \limits_{z\in Z}\sum \limits_{z\in Z}p\left(z,k\right) logp\left(k|z\right) $$

(14)

Therefore, in MI the amount of uncertainty being minimized is expressed as,

$$ I\left(Z|K\right)=H(Z)-H\left(Z|K\right) $$

(15)

However, Joint MI [11] ranks the input variables relevance by means of considering both MI and conditional MI (CMI). The major objective is to remove the features which depends on each other and generate a feature set which has jointly higher power of prediction. This is because the features are jointly relevant and separately redundant. The definition of Joint MI is given as,

$$ I\left({Z}_i,{Z}_j\right)=I\left({Z}_i,K\right)+I\left({Z}_j;K;{Z}_i\right) $$

(16)

However, to obtain the set of feature pairs for maximum Joint MI, both the parameters on the right side (MI and CMI) in Eq. (16) are maximized by means of Eq. (14) and (15). Let the n dimension feature vector can be expressed as Z_i, ..., Z_n and $v < n$be the selected features. The redundant features are removed using CMI by recognizing the inputs which are dependent on

each other by means of the conditional independence stated as,

$$ p\left({Z}_i{Z}_j\left|K\right.\right)=p\left({Z}_iK\right)p\left({Z}_j\left|K\right.\right) $$

(17)

CMI value is zero if the feature Z_j does not contribute to the existing information $[Z_i, \ldots, n-1]$Here, Joint MI is optimized using RFO algorithm. RFO is a nature-inspired algorithm which imitates the behaviour of raindrops trickling nature. The algorithm begins with initial population (i.e. raindrops). Here, the raindrops are considered as the features. The initialization function is expressed as,

$$ {DN}_i=\left[{z}_{i,1}{z}_{i,2}{z}_{i,3}\dots {z}_{i,n}\right]i\in \left\{1,2,3,\cdots, m\right\} $$

(18)

Where, m represents the size of population, n signifies the optimization variables, DN_idenotes the drop number and k^{th} optimization variable is indicated as $z_{i,\,k}$. Raindrops are the important parameter which is generated based on the random uniform distribution as,

$$ {z}_{i,k}= UF\left({l}_k,{u}_k\right) $$

(19)

Where, u_k, l_k be the upper and lower limits, the uniform distribution function is expressed as *UF*. As the raindrop include 'N' elements and indicates a point with radius vector r nearby the point is known as neighbourhood. Next a point \({NeP}_j^i\) is randomly generated during optimization which is expressed as,

$$ \left\Vert \left({DN}^i-{NeP}_j^i\right)\cdot {\hat{u}}_k\right\Vert \le \left\Vert {r}_{ini}\times f\left(\boldsymbol{iteration}\right)\cdot {\hat{u}}_k\right\Vert $$

(20)

Where, $i = \{1, 2, 3. . m\}$, $j = \{1, 2, 3..., np\}$, $k = \{1, 2, 3. . n\}$, r be the real positive vector, r_{ini} represents initial neighbourhood size, *np* signifies the neighbour points and k^{th} dimension unit vector is indicated as \({\hat{u}}_k\). The raindrop can be characterized as active drop, dominant drop and inactive drop.

The drop which include dominant *NeP*(neighbour point) is called as active drop, the drop which include no dominant *NeP*(neighbour point) is called as inactive drop whereas the dominant drop states that among every neighbour points of DN^i, the dominant *NeP* is a point that satisfies the condition in Eq. (21).

$$ {\displaystyle \begin{array}{c}F\left({NeP}_d^i\right)<F\left({DN}^i\right)\\ {} F\left({NeP}_d^i\right)<F\left({NeP}_j^i\right),j\in \left\{1,2,3,\cdots, np\right\}-\left\{d\right\}\end{array}} $$

(21)

Where, F denotes the objective function. Finally, the raindrops are ranked based on the following condition expressed as,

$$ {R}_t^i=\boldsymbol{order}\left({C_1}_t^i\right)\times {\omega}_1+\boldsymbol{order}\left({C_2}_t^i\right)\times {\omega}_2 $$

(22)

$$ {C_1}_t^i=F\left({DN}^i\right){\left|{}_{at{t}^{th}\boldsymbol{iteration}}-F\left({DN}^i\right)\right|}_{at{1}^{st}\boldsymbol{iteration}} $$

(23)

$$ {\left.{C_2}_t^i=F\left({DN}^i\right)\right|}_{at{t}^{th}\boldsymbol{iteration}} $$

(24)

Where, ω_1, ω_2 signifies the weighting coefficients equivalent to 0.5, \({R}_t^i\) states the raindrop rank at t^{th} iteration. By sorting all the scores obtained from raindrops (feature), the features are ranked in ascending order to obtain the set (v) of optimal features.

3.4 Similarity matching

The procedure of evaluating the degree of similarity between the query images to the images inside each group is termed as similarity matching. The technique used to measure the similarity is Weighted Euclidean Distance (WED). The selection of similarity measure play a significant role in enhancing the retrieval system performance. The weighted distance (WD) factor provides additional preference to the recovered images in maximum number of similar categories when comparing with image query. Let Q_u be the query image given to the fine-tuned model and the probability \({pr}^{Q_u}\)for the each image class is obtained which is then used to evaluate the weight of recovered $r_i$image. The weight function is expressed as,

$$ \boldsymbol{We}=1-{pr}_k^{Q_u} $$

(25)

Where, $k$signifies the image class. However, the weighted distance among the Q_u(image query) and r_i(image recovered) can be defined as,

$$ {D}_{\boldsymbol{We}}\left({Q}_u,{r}_i\right)=\boldsymbol{We}\times D\left({Q}_u,{r}_i\right) $$

(26)

Where, $D(Q_u, r_i)$ is measured using the normal Euclidean distance metric.

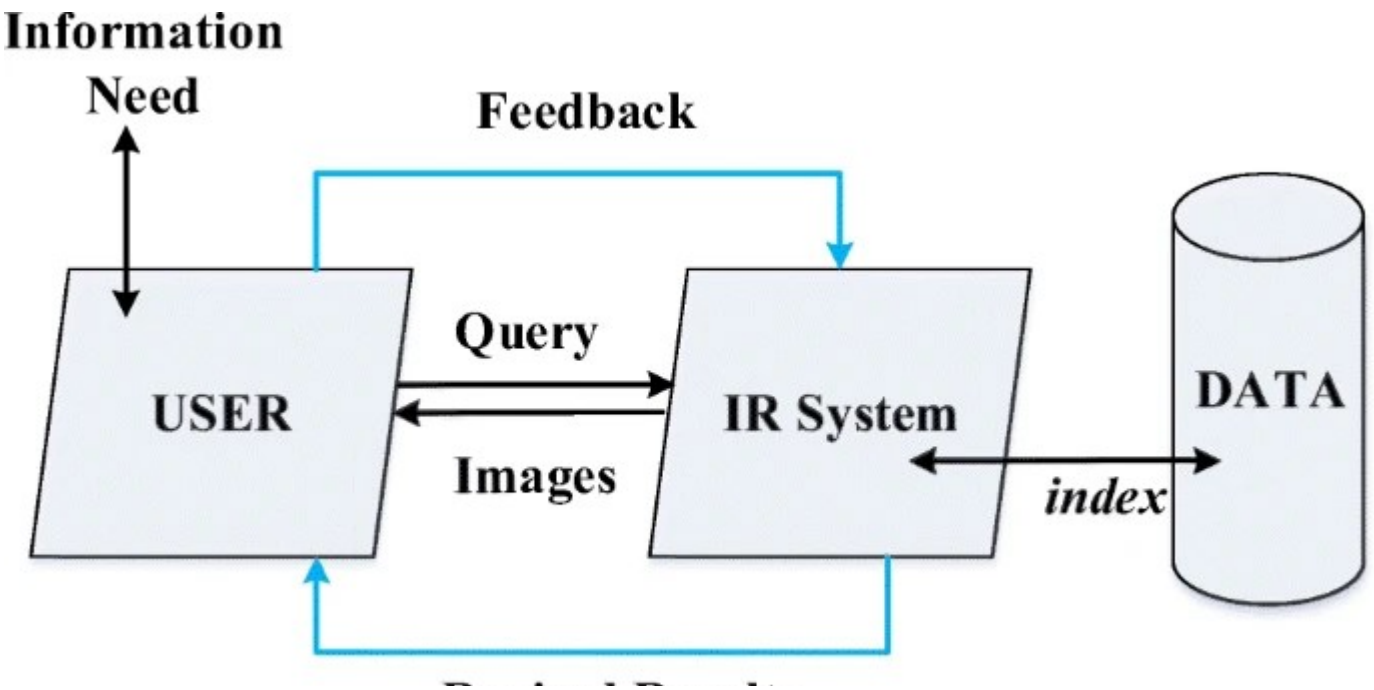
3.5 Relevance feedback model (RFM)

In the field of information technology, relevance retrieval indicates how much the information

retrieved meets the necessity of user. RFM is one of the major features of the information retrieval system (IRS). RFM is defined as the feedback given by the user mainly about the relevance of documents in the first set of results obtained. The basic idea of this model is to include the user in the process of image retrieval in order to enhance the final set of results. The steps involved in RFM are: Initially, a simple query is issued by the user. Then, a set of retrieval results are returned by the system. Next, the user need to mark the results as relevant or non-relevant. Based on the feedback obtained from user, the system calculates better illustration of information. Finally, a set of revised retrieval results are displayed.

Figure 5 displays the RFM structure. RFM can be processed via single or multiple iterations. In IRS, the mechanism of RF is termed as a powerful approach as it automatically adjusts the existing user query with the information given back by the user regarding the relevance of images that are retrieved previously. The image query used is too short and it may not express the information necessities of user in an exact manner. In order to address these problems, RFM is introduced. The idea of RFM is to make use of the relevance decisions offered by the users to improve the query model and further retrieve more relevant results. This study focus on evolving an effective RFM which improves the performance of retrieval system in single/ multiple iteration, in which the user evaluate the relevance of top K images. Hence, RFM is used for attaining accuracy and efficiency in query based information retrieval.

Fig. 5



Revised Results

Structure of Relevance Feedback Model

4 Simulation results

This section represents the details of the implementation of proposed RSIR system. The proposed retrieval system is named as Fused CNN_RFM which is implemented on PYTHON platform and analyse the image retrieval efficiency on three different RS datasets. The proposed Fused CNN_RFM system is tested on the datasets such as UCM (UCMerced Land Use) and WHU-RS19. These datasets can be characterized into testing as well as training sets.

4.1 Dataset depiction

4.1.1 WHU-RS19

The RS database WHU-RS19 (in short, RS19) includes 19 classes with 1005 total aerial images. These RS images were collected from the GE (Google Earth) imagery. The image size is 600 × 600 pixels and every single class have 50 image samples. This database is quite challenging owing to deviations in GSD (Ground Sampling Distance), resolution, scale, orientation and illumination. The image samples of 19 landform categories are shown in Fig. 6.

Fig. 6

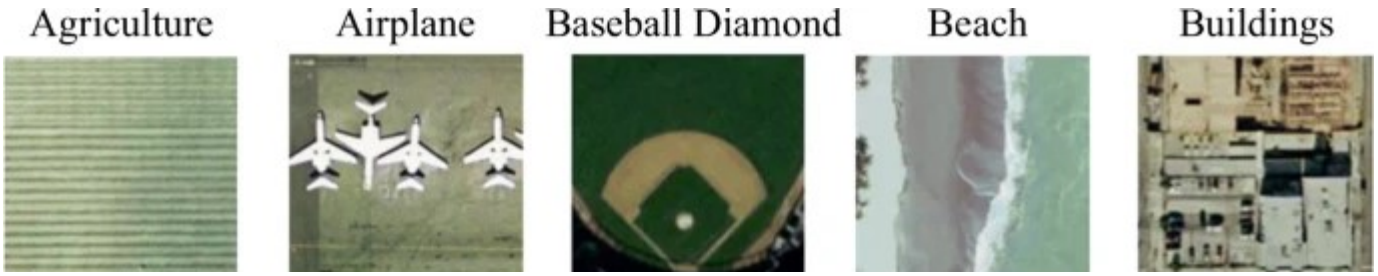
WHU-RS19 Database sample images

The labels of 19 classes from WHU-RS19 database are: airport, viaduct, farmland, meadow, mountain, desert, railway station, river, beach, pond, football field, industrial, park, commercial, residential, parking, forest, and bridge. From 1005 total images, 80% of images are utilized for training the data and remaining 20% is utilized for testing purpose.

4.1.2 UCM database

UCM is a Land Use database with large number of images (2100 RS images) collected from USGS (United States Geological Survey). The total images belong to 21 classes and each single category has 100 images with 256 × 256 pixel and 30 cm spatial resolution. Fig displays the image samples of 21 categories. The image samples of 21 categories are shown in Fig. 7.

Fig. 7



UCM Database sample images

The labels of 21 classes are: agricultural, tennis courts, beach, parking lot, airplane, sparse residential, baseball diamond, runway, mobile home park, runway, chaparral, storage tanks, dense residential, river, overpass, forest, golf course, buildings, freeway, medium density residential, harbour, and intersection. From 2100 (RS images), 80% of images are utilized for training the data and remaining 20% is utilized for testing purpose.

4.1.3 AID database

AID is a large public database that includes 30 categories of aerial scene images. This database holds total 10,000 images and the spatial resolution varies between 0.5 to 8 m. Each class holds 220–420 images with size 600 × 600 pixels. From the total images, 80% of images are utilized for training the data and remaining 20% is utilized for testing purpose.

The images in the dataset are collected from various countries such as USA, France, China, UK, Germany, and Italy. The AID database covers total 30 land-use categories, comprising airport, baseball field, bare land, bridge, beach, center, commercial, church, desert, dense residential, forest, farmland, industrial, medium residential, meadow, mountain, parking, park, playground, port, pond, resort, railway station, river, square, school, storage tanks, sparse residential, stadium, and viaduct. The image samples of 30 categories are shown in Fig. 8.

Fig. 8

AID Database sample images

4.2 Performance measures

The metrics such as AvP (Average Precision), mAP (Mean AP), nMRR (Normalized Modified Retrieval Rank), and R (Recall) in RSIR are considered very important to evaluate the performance of the presented methodology. Recall states the fraction of related/relevant images recovered correctly (CR) to the total database (TD) images. Precision describes the proportion of correctly recovered images (CR) to the total recovered images (TR). The expression for recall, precision is mentioned in Eqs. (27) and (28).

$$ R=\frac{CR}{TD} $$

(27)

$$ P=\frac{CR}{TR} $$

(28)

For a given Q(query image), the formulate to calculate AvP is well-defined in Eq. (29),

$$ AvP=\frac{\sum_{k=1}^N\left(P@k\times rel(k)\right)}{R_N} $$

(29)

$$ P@k=\frac{R_N}{k} $$

(30)

$$ rel(k)=\left\{\begin{array}{c}1,{k}^{th}\mathrm{image}\to \mathrm{relevant}\\ {}0,{k}^{th}\mathrm{image}\to \mathrm{irrelevant}\end{array}\right. $$

(31)

Where, R_N denotes the relevant number of images, k signifies position of retrieval outcome.

The formulate to calculate mAP is stated in Eq. (32),

$$ mAP=\frac{1}{Q}\sum \limits_{q=1}^Q AvP $$

(32)

The formula to evaluate nMRR is specified as,

$$ nMRR=\frac{AR-0.5\left(1+{R}_N\right)}{1.25G-0.5\left(1+{R}_N\right)} $$

(33)

$$ AR=\frac{1}{R_N}{\sum}_{i=1}^R{r}_i^{\ast } $$

(34)

$$ {r}_i^{\ast }=\left\{\begin{array}{c}{\mathit{\operatorname{rank}}}_i,\mathrm{if}

{\mathit{\operatorname{rank}}}_i\le G\\ {}1.25G,\mathrm{otherwise}\end{array}\right. $$

(35)

$$ G=2\ast {R}_N $$

(36)

$$ \boldsymbol{AnMrr}=\frac{1}{Q}\sum \limits_{q=1}^Q{\boldsymbol{nMRR}}_q $$

(37)

Where, the average rank is represented as*AR* and rank of the correctly retrieved *i*-th image is $rank_i$. For several queries, the metric AnMRR (Average nMRR) is expressed in Eq. (37).

4.3 Analysis of fused CNN_RFM performance

This section analyses the performance of Fused CNN_RFM by comparing the simulation outcomes with other similar performance models in order to signify its superiority.

Table 1 indicates the outcomes of proposed model (Fused CNN_RFM) performance. The proposed retrieval performance is tested on the three public sources such as WHU-RS19 and UCM. The outcomes gained from WHU-RS19 and UCM database have verified based on mAP, Recall, nMRR and AnMRR metrics. The proposed model (Fused CNN_RFM) acquired mAP (93.693%), Recall (87.00), nMRR (1.471) and AnMRR (0.077) on WHU-RS19. The proposed performance on AID dataset is mAP (94.716%), Recall (90.20), nMRR (2.243) and AnMRR (0.118) whereas UCM database obtained mAP (95.067%), Recall (90.54), nMRR (1.570) and AnMRR (0.082).

Table 1 Proposed (Fused CNN_RFM) performance

Table 2 represents the comparison of proposed method with few existing related works. The proposed Fused CNN_RFM model is tested on three different datasets such as UCM, WHU-RS19 and AID. Compared to the existing related works on RSIR, the proposed model gained superior mAP performance. From the table it is observed that the presented model outperforms the existing methods for UCM, WHU-RS19 and AID datasets with higher precision. This proves the superiority of the proposed Fused CNN_RFM model when comparing with some existing related works.

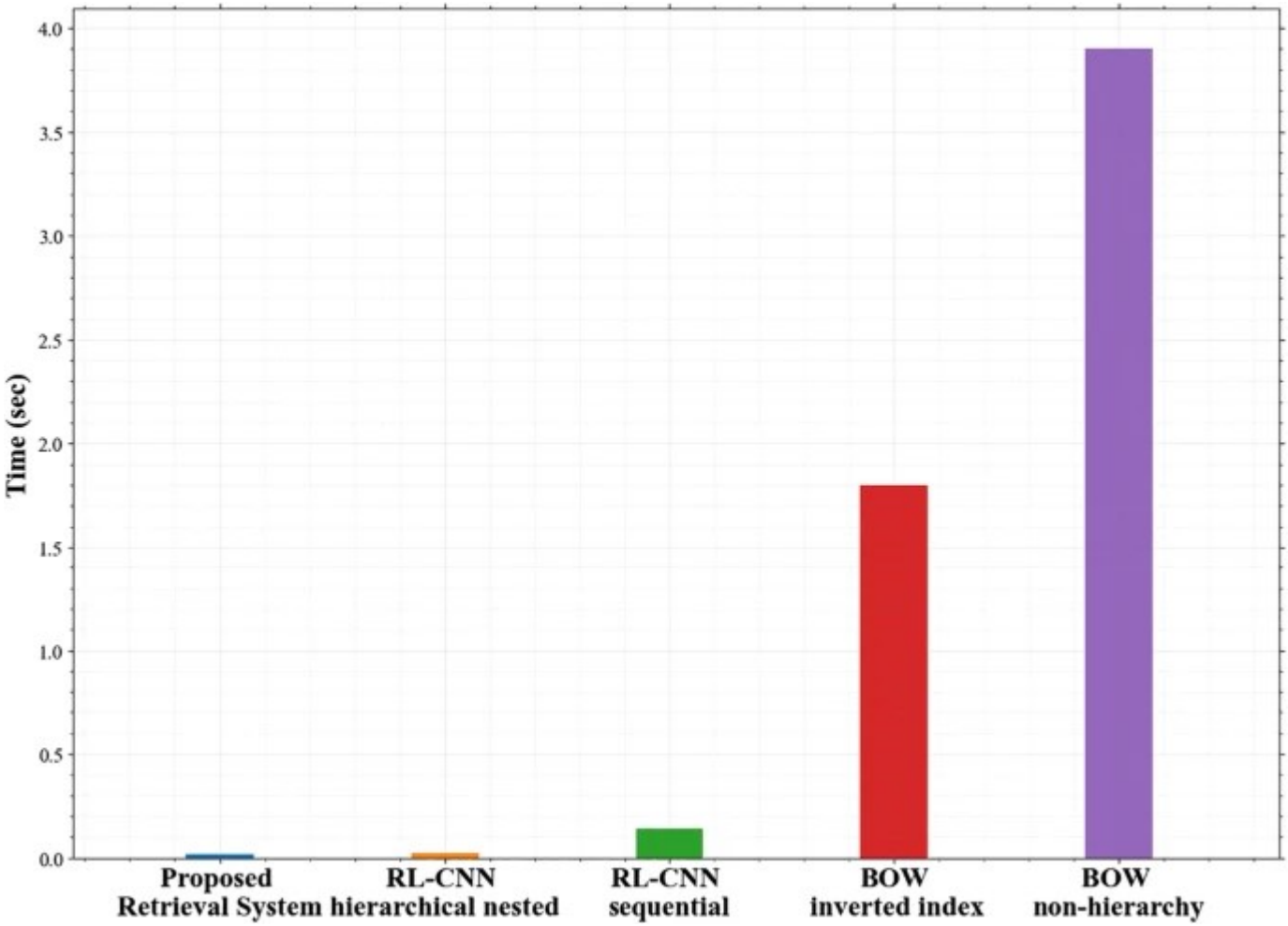
Table 2 Comparison of proposed method with few existing related works

Table 3 shows the proposed model (Fused CNN_RFM) mAP performance for the top 20 retrieved images on UCM database. The baseline architecture models utilized for evaluation are ADLF (Aggregated Deep Local Features), DBOW (Deep BOW) and GOSLm (Global Optimal Structured embedded Learning) [19]. The proposed model (Fused CNN_RFM) estimated mAP (95.067%) which is very high than the baseline architectures ADLF (91.6%), DBOW (83.0%) abd GOSLm (85.8%) for the case of 20 images retrieved.

Table 3 mAP Performance on UCM database

Figure 9 shows the image retrieval time on UCM dataset for proposed (Fused CNN_ RFM) model comparing with other approaches. The baseline approaches named BOW non-hierarchy, RL-CNN hierarchically nested, RL-CNN sequential searching and BOW inverted index are utilized in this work for comparison [23]. The time consumed for image retrieval of proposed (Fused CNN_RFM) method by means of UCM database is (0.019 s) whereas the other models RL-CNN hierarchical nested (0.025 s), BOW inverted index (1.8 s), RL-CNN sequential (0.142 s) and BOW non-hierarchy (3.9 s) respectively. Hence, the proposed (Fused CNN_ RFM) model achieved reduced image retrieval time compared to other models.

Fig. 9

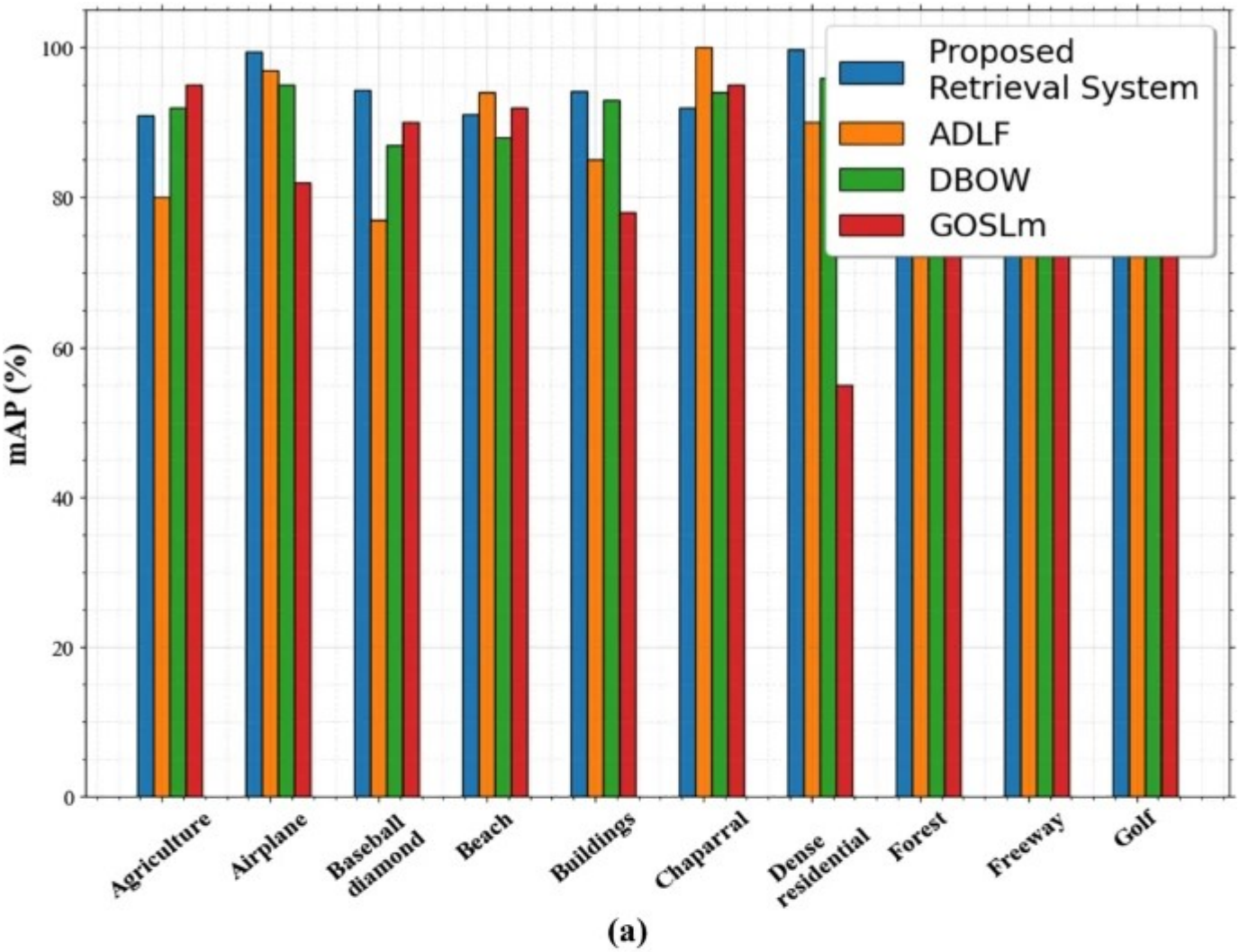


Retrieval Time Comparison (UCM) dataset

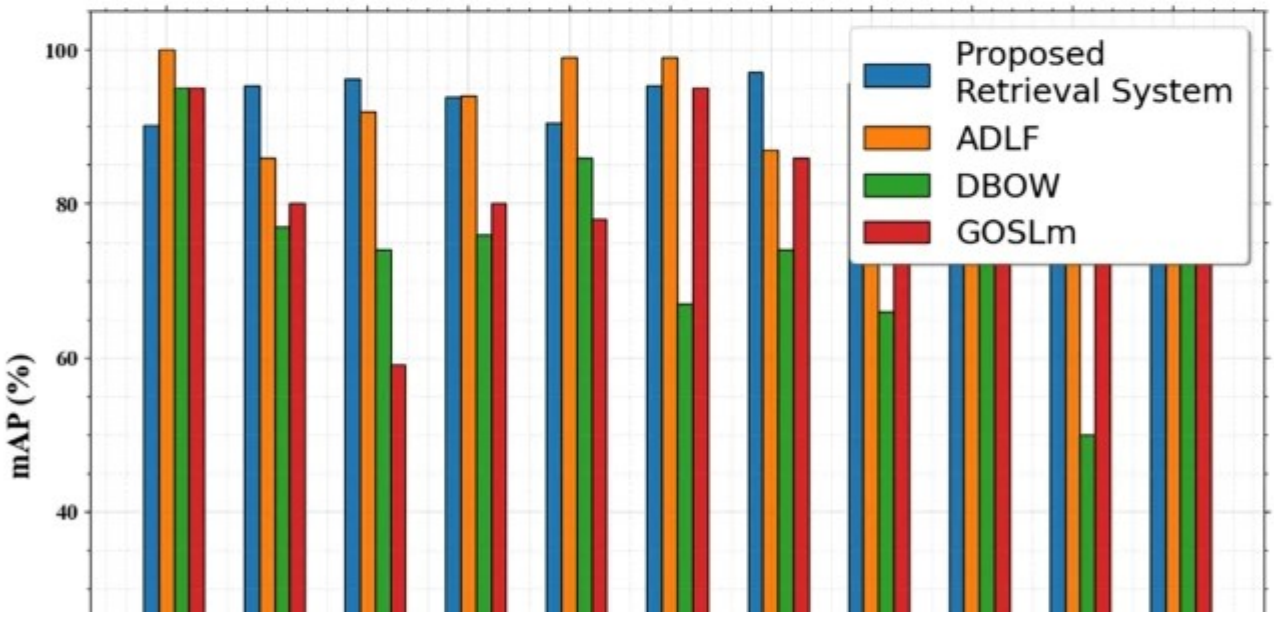
Figure 10 (a) demonstrates the mAP comparison for top N (N = 20) retrieved images for the first 10 categories. Figure 10 (b) demonstrates the mAP comparison for top N (N = 20) retrieved images for the remaining 11 land use categories. The baseline approaches utilized for comparison are ADLF, DBOW and GOSLm. The UCM database 21 categories plotted in terms of precision metric comprises agricultural, tennis courts, baseball diamond, runway, airplane, sparse residential, beach, parking lot, chaparral, storage tanks, buildings, river, forest, dense residential, mobile home park, freeway, overpass, harbour, intersection, golf course, and medium density residential. The top precision is acquired for the dense residential, sparse, freeway, and airplane classes. The proposed Fused CNN_FEM model gained overall mAP (95.067%) for top 20 images retrieved which is greater than the baseline approaches. Hence,

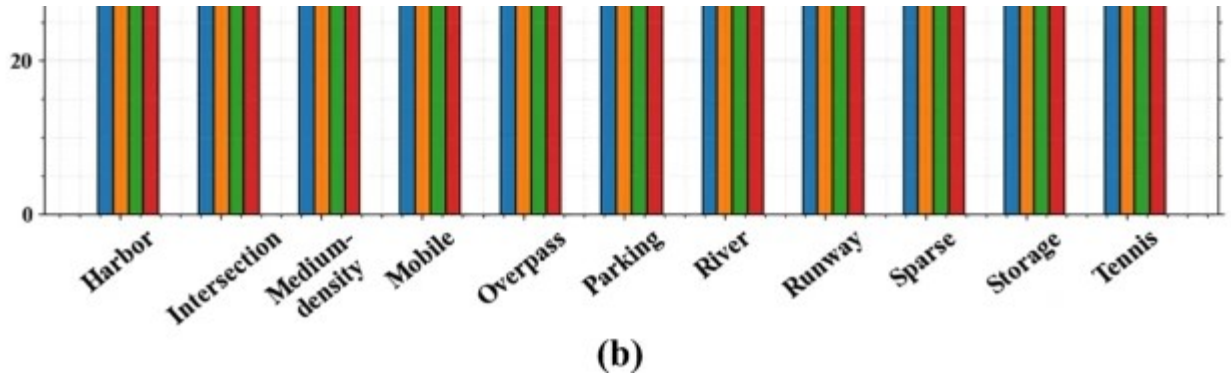
the proposed Fused CNN_RFM acquired better mAP performance.

Fig. 10



(a)

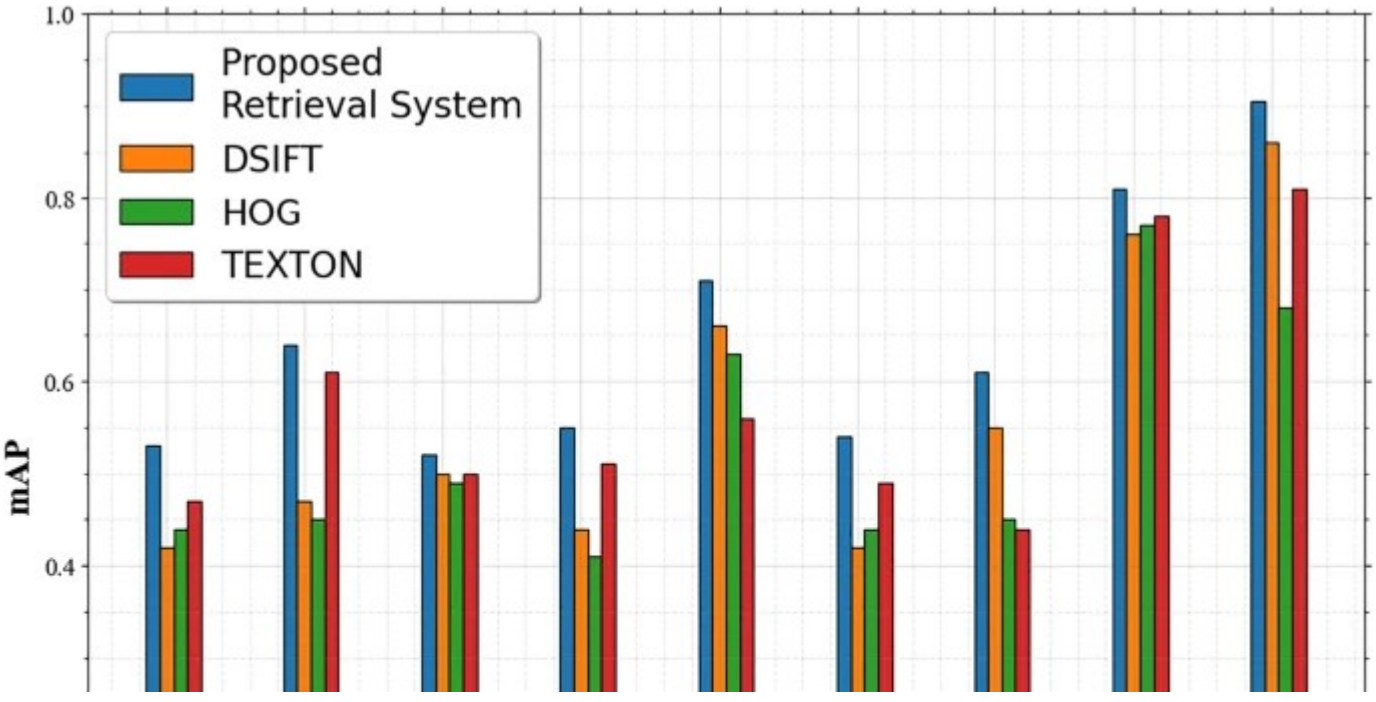


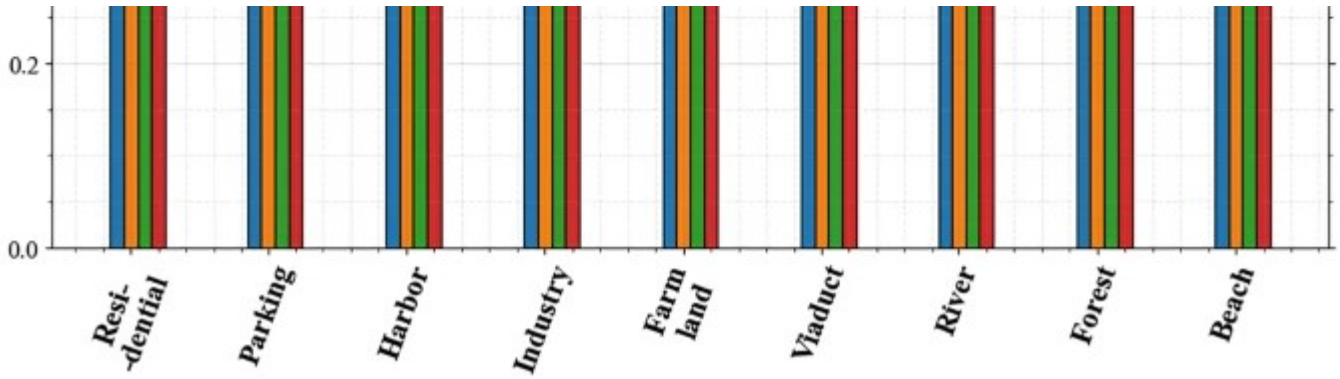


(a) and (b): mAP comparison for 21-classes (UCM database)

Figure 11 shows the WHU-RS19 database mAP comparison for diverse features with reference to nine RS image categories. The following 19 classes in WHU-RS19 database comprises airport, viaduct, farmland, meadow, mountain, desert, railway station, river, beach, pond, football field, industrial, park, commercial, residential, parking, forest, and bridge. Here the mAP value of per-class with diverse features being compared are DSIFT, HOG and TEXTON [6]. The proposed feature model based on Fused CNN gained higher performance for all the categories comparing with baseline features. The beach category attained higher mAP (0.93) than the other classes and the second largest value is obtained by forest category.

Figure 11

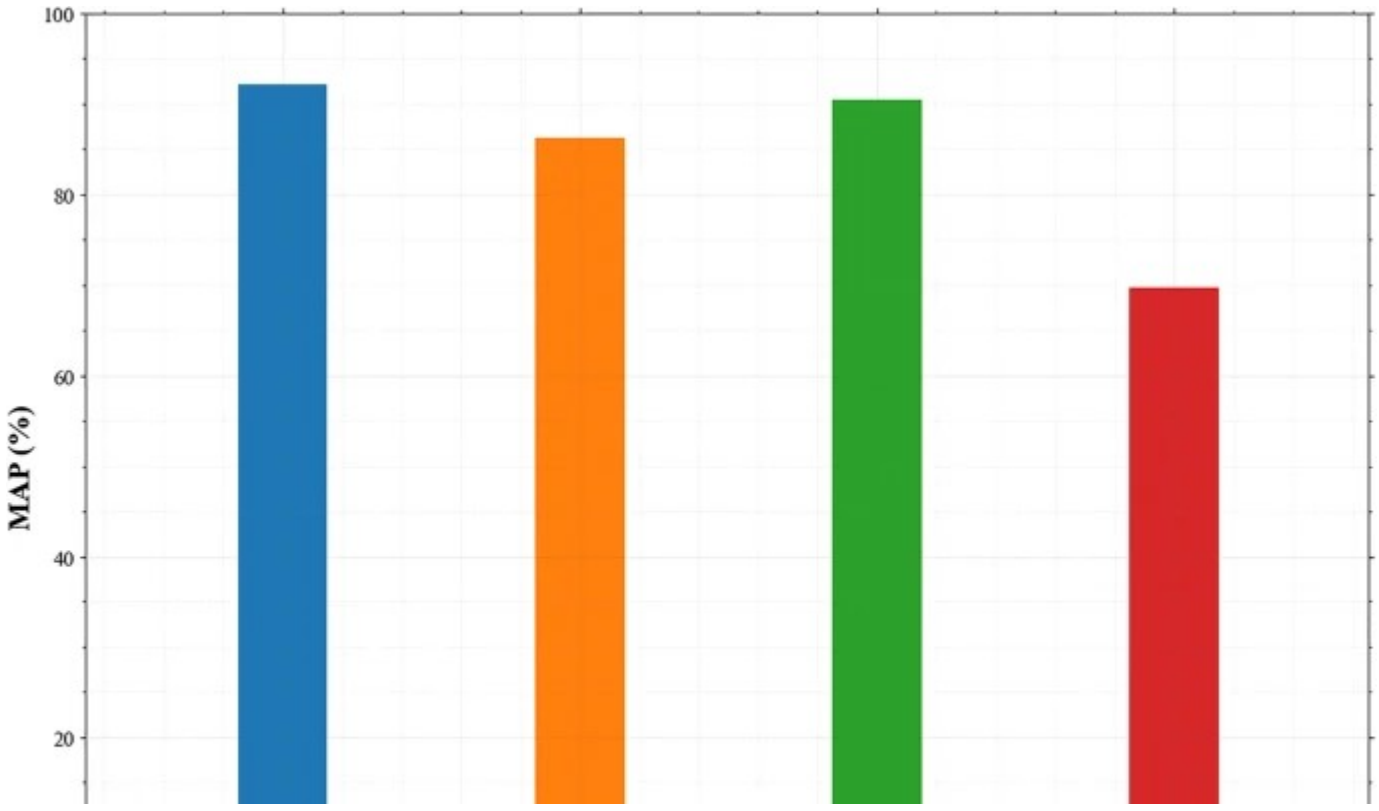


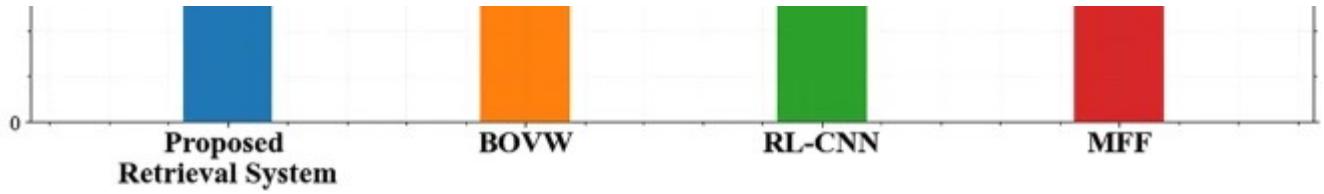


mAP comparison for 9-classes (WHU-RS19 database)

Figure 12 illustrates the retrieval accuracy comparison on the UCM database. The baseline approaches utilized for evaluation are RL-CNN (Representation Learning -CNN), BOVW (Bag of Visual Words), and MFF (Multiple Fused global Features) [23]. The proposed Fused CC_RFM model achieved high MAP accuracy (93.693%) than the other baseline approaches such as RL-CNN(90.5%), BOVW (86.2%), and MFF (69.8%). Hence, the proposed Fused CNN_RFM attained enhanced percentage of retrieval accuracy.

Fig. 12

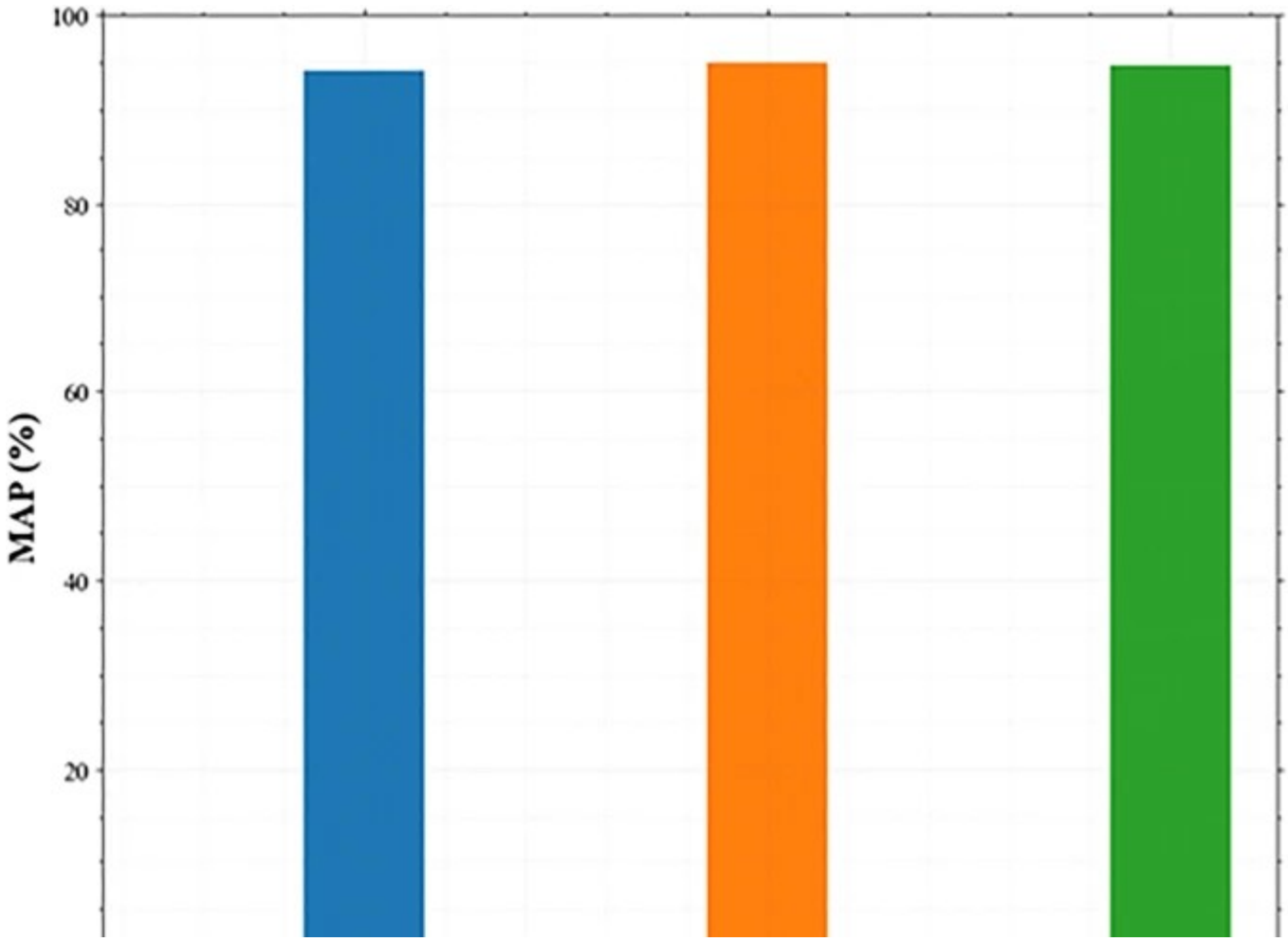


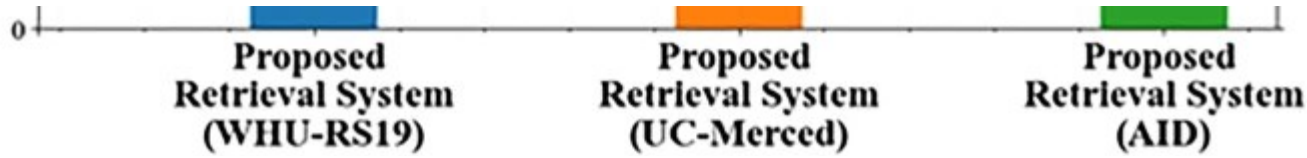


Retrieval Accuracy comparison (UCM database)

Figure 13 displays the proposed method accuracy comparison on the performance of image retrieval using WHU-RS19 and UCM databases. The proposed Fused CNN_RFM model gained MAP accuracy (93.693%) on WHU-RS19. The performance on AID dataset is (94.71%) whereas with the UCM database the proposed Fused CNN_RFM model attained (95.067%). Hence, the proposed Fused CNN_RFM assimilated superior retrieval accuracy measured in terms of percentage.

Fig. 13

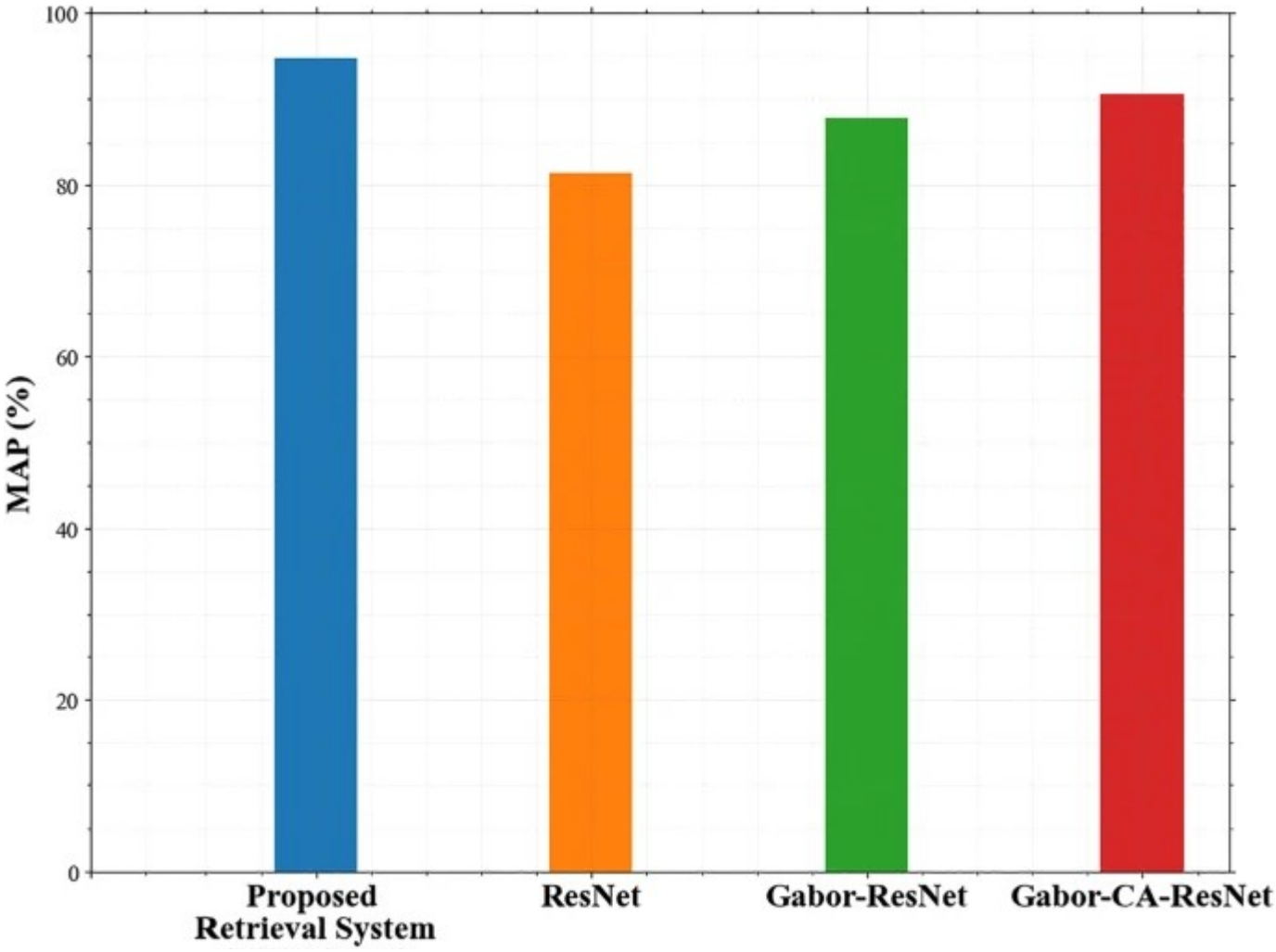




Accuracy comparison on databases

Figure 14 signifies the MAP comparison for AID dataset. The proposed Fused CNN_RFM model is compared with various architectures such as ResNet, Gabor-ResNet and Gabor-Channel Attention-ResNet [35] on AID dataset. The performance of retrieval is 81.37%, 87.69%, and 90.52% with the existing architectures. The proposed model tested on AID dataset have shown superior performance (94.71%) than the other network architectures. This signifies the efficiency of the proposed method.

Fig. 14

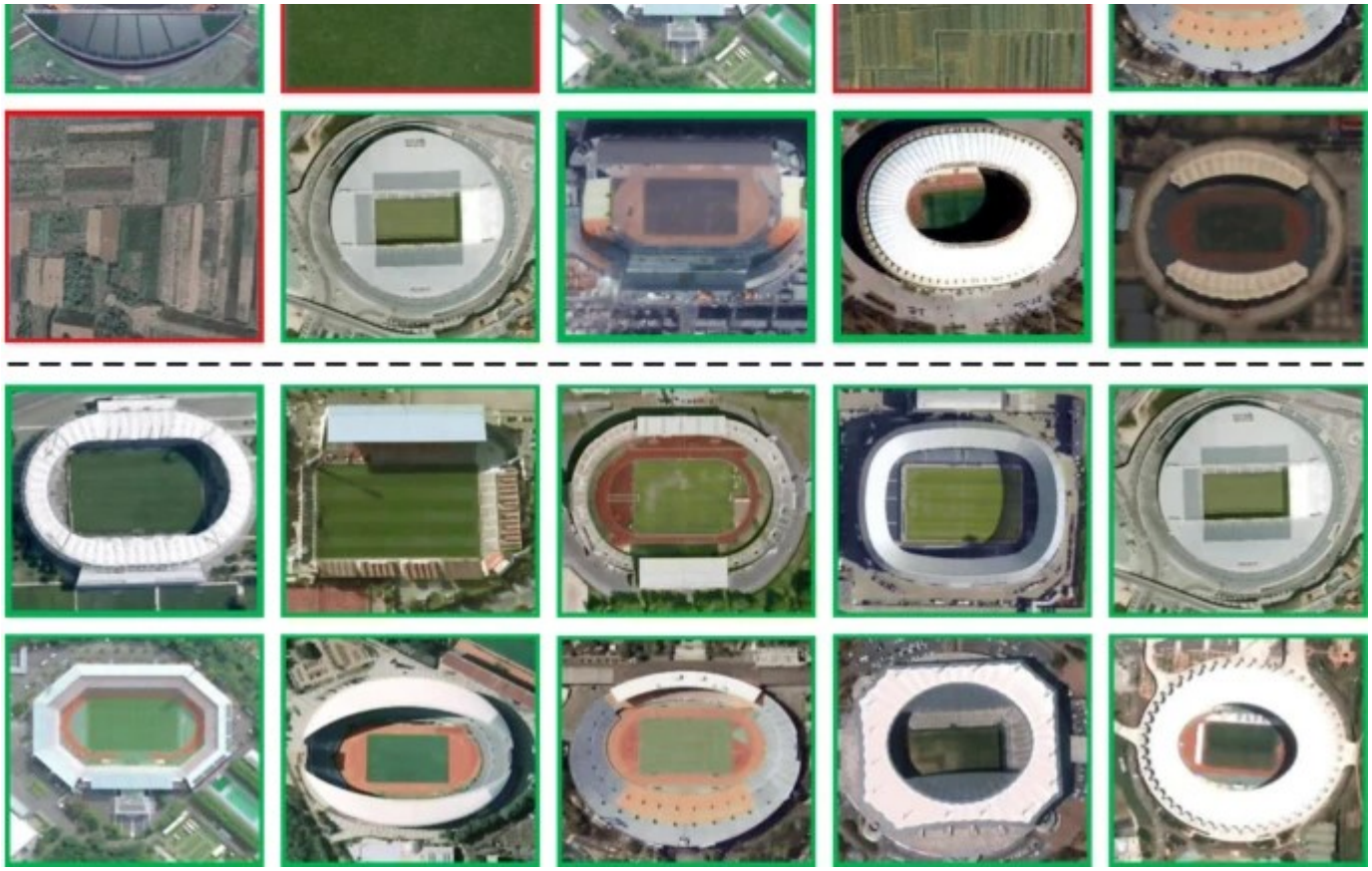


(WHU-RS19)

MAP comparison of AID dataset

Figure 15 displays the retrieval outcomes of WHU-RS 19 database based on the RFM. Here the query image used is Football Field. While querying this image, the retrieval outcome returns some inappropriate images which seem similar to that of the football field image. Some of the inapt images obtained in the first set are mountain, meadow, and farmland category. Figure 16 shows the retrieval outcomes of UCM database based on the RFM. Here the query image used is Dense Residential. While querying this image, the retrieval outcome returns some inappropriate images which seem similar to that of the dense residential image. Some of the inapt images obtained in the first set are parking lot, and medium residential category. Figure 17 shows the retrieval outcomes of AID database based on the RFM. Here the query image used is Beach category. While querying beach image, the retrieval outcome returns some inappropriate images which seem similar to that of the beach image. Some of the inapt images obtained in the first set are pond, and river category. Totally, 10 images are retrieved and the exact images identical to query are notified with a green colour squared structure and the inappropriate images are signified with a red colour square. After the process of fine-tuning, the most relevant images are acquired with dual iterations of relevance feedback model. The first image set is a mixture of relevant as well as non-relevant images. But, in the next iteration, the non-relevant images are fine-tuned and rejected based on user's feedback and final top 10 images are retrieved.

Fig. 15

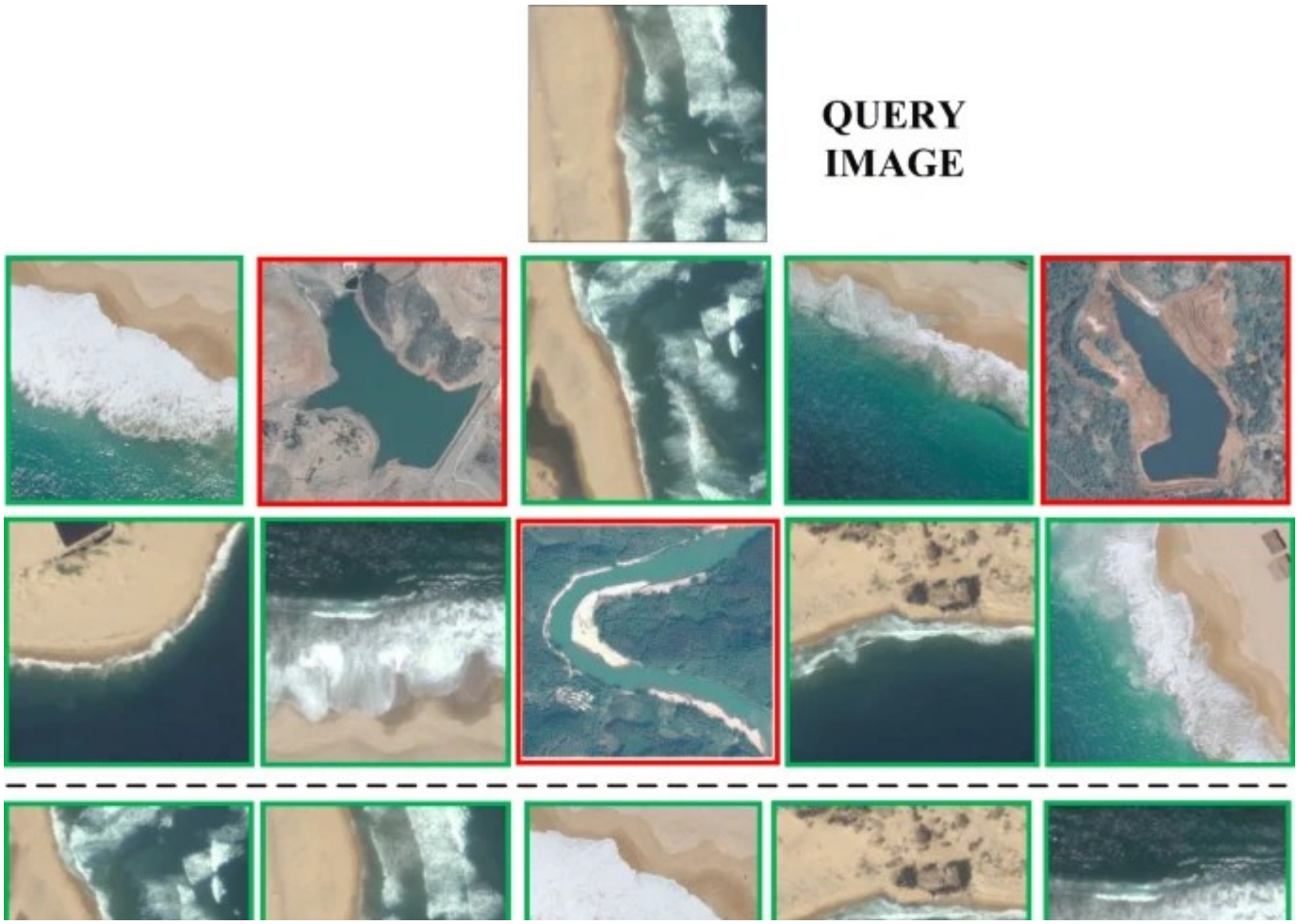


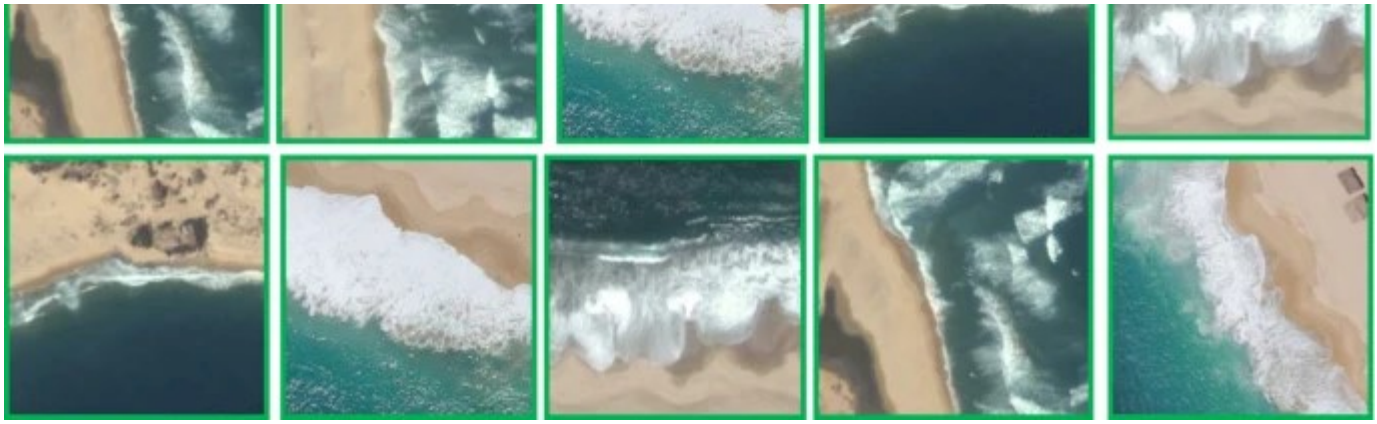
Results on Retrieval of RS images (WHU-RS19 database)

Fig. 16

Results on Retrieval of RS images (UCM database)

Fig. 17





Results on Retrieval of RS images (AID database)

5 Conclusion

In this study, an exact and effective retrieval of RS images is presented with fusion based feature extraction and optimal feature selection framework. With the fused CNN architectures (VGG16, VGG19, ResNet) the last layer act as a feature extractor for all the 3 architectures in which FC8 for VGG16, VGG19 and FC 1000 for ResNet. Then the optimal set of features are selected from the combined set of features using Joint MI_RFO. The proposed Fused CNN_RFM model maintains higher accuracy on image retrieval and preserves multi-dimensional as well as high discriminative image illustrations. Also, the relevance feedback introduced in this proposed framework allows user involvement in image retrieval to enhance the final outcome. The proposed Fused CNN_RFM performance is evaluated on UCM dataset in terms of mAP = 95.067%, R = 90.54, nMRR = 1.570 and AnMRR = 0.082. Also, the remaining two databases named WHU-RS19 and AID are tested and compared for the same. The simulation outcomes on WHU-RS19 database is tested in terms of mAP = 93.693%, R = 87.00, nMRR = 1.471 and AnMRR = 0.077. The proposed performance with AID dataset is mAP = 94.716%, R = 90.20, nMRR = 2.243 and AnMRR = 0.118. Thus, the simulation outcomes reveal that RSIR performance on UCM database by means of Fused CNN_RFM model achieved superior performance than WHU-RS19 and AID databases. The advantages of the proposed work are it offers efficient and effective image retrieval task which retrieves accurate image results based on the user query at minimum time, no need of domain experts, and fast retrieval. The major impact is, with the use of AID dataset many images thus cover a large area and contain a large amount of background which is not appropriate for RSIR. In future, different architectures can

be analyzed, and tested on different RS image datasets to increase the effectiveness of image retrieval in RS images.

References

1. Alberton B, Torres RS, Cancian FL, Borges BD, Almeida J, Mariano GC, dos Santos J, Morellato LPC (2017) Introducing digital cameras to monitor plant phenology in the tropics: applications for conservation. Perspectives Ecol Conserv 15(2):82–90

Article Google Scholar

2. Avtar R, Kumar P, Oono A, Saraswat C, Dorji S, Hlaing Z (2017) Potential application of remote sensing in monitoring ecosystem services of forests, mangroves and urban areas. Geocarto Int 32(8):874–885

Article Google Scholar

3. Boualleg Y, Farah M, and Farah IR (2020) TLDCNN: A Triplet Low Dimensional Convolutional Neural Networks for High-Resolution Remote Sensing Image Retrieval. In 2020 Mediterranean and middle-east geoscience and remote sensing symposium (M2GARSS), 13–16. IEEE.

4. Cao R, Zhang Q, Zhu J, Li Q, Li Q, Liu B, Qiu G (2020) Enhancing remote sensing image retrieval using a triplet deep metric learning network. Int J Remote Sens 41(2):740–751

Article Google Scholar

5. Chaudhuri U, Banerjee B, Bhattacharya A, Datcu M (2020) Cmir-net: a deep learning based model for cross-modal retrieval in remote sensing. Pattern Recogn Lett 131:456–462

Article Google Scholar

6. Deng Z, Sun H, Zhou S (2018) Semi-supervised ground-to-aerial adaptation with heterogeneous features learning for scene classification. ISPRS Int J Geo Inf 7(5):182

Article Google Scholar

7. El Mahrad B, Newton A, Icely JD, Kacimi I, Abalansa S, Snoussi M (2020) Contribution of remote sensing technologies to a holistic coastal and marine environmental management framework: a review. Remote Sens 12(14):2313

Article Google Scholar

8. Gaikwad VP and Musande V (2017) Wheat disease detection using image processing. In 2017 1st international conference on intelligent systems and information management (ICISIM), IEEE 110-112.

9. Imbriaco R, Sebastian C, Bondarev E (2019) Aggregated deep local features for remote sensing image retrieval. Remote Sens 11(5):493

Article Google Scholar

10. Johnson BA, Iizuka K, Bragais MA, Endo I, Damasa B (2017) Magcale-Macandog. Employing crowdsourced geographic data and multi-temporal/multi-sensor satellite imagery to monitor land cover change: a case study in an urbanizing region of the Philippines. Computers. Environ Urban Syst 64:184–193

Article Google Scholar

11. Kashif M, Raja G, Shaukat F (2020) An efficient content-based image retrieval system for the diagnosis of lung diseases. J Digit Imaging 33:971–987

Article Google Scholar

12. Kharat SA, Musande VB (2015) Cotton crop discrimination using landsat-8 data. Int J Comp Sci Inform Technol 6(5):4381–4384

Google Scholar

13. Lang S, Hay GJ, Baraldi A, Tiede D, Blaschke T (2019) GEOBIA achievements and spatial opportunities in the era of big earth observation data. ISPRS Int J Geo Inf 8(11):474

Article Google Scholar

14. Langat PK, Kumar L, Koech R (2019) Monitoring river channel dynamics using remote sensing and GIS techniques. Geomorphology 325:92–102

Article Google Scholar

15. Lausch A, Erasmi S, King DJ, Magdon P, Heurich M (2017) Understanding forest health with remote sensing-part II—A review of approaches and data models. Remote Sens 9(2):129

Article Google Scholar

16. Leonardo MM, Carvalho TJ, Rezende E, Zucchi R, and Faria FA (2018) Deep feature-based classifiers for fruit fly identification (diptera: Tephritidae). In 2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI), 41–47. IEEE.

17. Li Y, Zhang H, Xue X, Jiang Y, Shen Q (2018) Deep learning for remote sensing image classification: a survey. Wiley Interdisciplinary Rev: Data Mining Knowledge Discovery 8(6):1264

Google Scholar

18. Liu Y, Ding L, Chen C, Liu Y (2020) Similarity-based unsupervised deep transfer learning for remote sensing image retrieval. IEEE Trans Geosci Remote Sens

19. Liu P, Gou G, Shan X, Tao D, Zhou Q (2020) Global optimal structured embedding learning for remote sensing image retrieval. Sensors 20(1):291

Article Google Scholar

20. Ma C, Chen F, Yang J, Liu J, Xia W, Li X (2018) A remote-sensing image-retrieval model based on an ensemble neural networks. Big Earth Data 2(4):351–367

Article Google Scholar

21. Musande V, Kumar A, Roy PS, Kale K (2013) Evaluation of fuzzy-based classifiers for cotton crop identification. Geocarto Int 28(3):243–257

Article Google Scholar

22. Napoletano P (2018) Visual descriptors for content-based retrieval of remote-sensing images. Int J Remote Sens 39(5):1343–1376

Article Google Scholar

23. Sadeghi-Tehran P, Angelov P, Virlet N, Hawkesford MJ (2019) Scalable database indexing and fast image retrieval based on deep learning and hierarchically nested structure applied to remote sensing and plant biology. J Imaging 5(3):33

Article Google Scholar

24. Shao Z, Zhou W, Deng X, Zhang M, Cheng Q (2020) Multilabel remote sensing image retrieval based on fully convolutional network. IEEE J Selected Topics Appl Earth Observ Remote Sensing 13:318–328

Article Google Scholar

25. Sudmanns M, Tiede D, Lang S, Bergstedt H, Trost G, Augustin H, Baraldi A, Blaschke T (2020) Big earth data: disruptive changes in earth observation data management and analysis? Inter J Digital Earth 13(7):832–850

Article Google Scholar

26. Tahmasebi P, Kamrava S, Bai T, Sahimi M (2020) Machine learning in geo-and environmental sciences: from small to large scale. Adv Water Resour 103619:103619

Article Google Scholar

27. Temitope YS, Balogun AL (2020) Advances in remote sensing technology, machine learning and deep learning for marine oil spill detection, prediction and vulnerability assessment. Remote Sens 12(20):3416

Article Google Scholar

28. Tiede D, Baraldi A, Sudmanns M, Belgiu M, Lang S (2017) Architecture and prototypical implementation of a semantic querying system for big earth observation image bases. Eur J Remote Sensing 50(1):452–463

Article Google Scholar

29. Vharkate MN, Musande VB (2021) Remote sensing image retrieval using hybrid visual geometry group network with relevance feedback. Int J Remote Sens 42(14):5540–5567

Article Google Scholar

30. Wang Y, Ji S, Lu M, Zhang Y (2020) Attention boosted bilinear pooling for remote sensing image retrieval. Int J Remote Sens 41(7):2704–2724

Article Google Scholar

31. Xia GS, Tong XY, Hu F, Zhong Y, Datcu M, and Zhang L (2017) Exploiting deep features for remote sensing image retrieval: a systematic investigation. arXiv preprint arXiv:1707.07321 2.

32. Xiong W, Xiong Z, Cui Y, Lv Y (2020) A discriminative distillation network for cross-source remote sensing image retrieval. IEEE J Selected Topics Appl Earth Observ Remote Sensing 13:1234–1247

Article Google Scholar

33. Yang C, Yu M, Li Y, Hu F, Jiang Y, Liu Q, Sha D, Xu M, Gu J (2019) Big earth data analytics: a survey. Big Earth Data 3(2):83–107

Article Google Scholar

34. Zhou W, Newsam S, Li C, Shao Z (2018) PatternNet: a benchmark dataset for performance evaluation of remote sensing image retrieval. ISPRS J Photogramm Remote Sens 145:197–209

Article Google Scholar

35. Zhuo Z, Zhou Z (2021) Remote sensing image retrieval with Gabor-CA-ResNet and Split-based deep feature transform network. Remote Sens 13(5):869

Article Google Scholar

Availability of data and materials

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author information

Authors and Affiliations

Department of Computer Science and Information Technology, Dr. B.A.M.U Aurangabad, Maharashtra, India
Minakshi N. Vharkate

Department of Computer Engineering, MIT Academy of Engineering, Alandi(d), Pune, Maharashtra, India
Minakshi N. Vharkate

Department of Computer Science and Engineering, Jawaharlal Nehru Engineering College, Aurangabad MGM University, Aurangabad MGM University, Maharashtra, Maharashtra, India
Vijaya B. Musande

Corresponding author

Correspondence to Minakshi N. Vharkate.

Ethics declarations

Conflict of interest

Authors Minakshi N Vharkate and Dr. Vijaya B. Musande declares that they have no conflict of interest.

Additional information

Publisher's note

Rights and permissions

About this article

Cite this article

Vharkate, M.N., Musande, V.B. Fusion Based Feature Extraction and Optimal Feature Selection in Remote Sensing Image Retrieval. *Multimed Tools Appl* 81, 31787–31814 (2022). https://doi.org/10.1007/s11042-022-11997-y

Received
18 February 2021

Revised
24 June 2021

Accepted
04 January 2022

Published
11 April 2022

Issue Date
September 2022

DOI
https://doi.org/10.1007/s11042-022-11997-y

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Provided by the Springer Nature SharedIt content-sharing initiative

Keywords

Remote sensing
Image information retrieval
Convolutional neural networks
image database
Similarity measurement