iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37321

Remote Sensing Image Retrieval Using Convolutional Neural Network Features and Weighted Distance

Chippy Babu[1], Rose Mariya Abraham[2], Treesa Thomas[3], Dinu Mol Philip[4]
[1, 2, 3]*Dual Degree MCA, Department of Computer Science, De Paul Institute of Science & Technology, Angamaly, MG University*
[4]*Asst. Professor, Department of Computer Science, De Paul Institute of Science & Technology, Angamaly, MG University*

Abstract: Remote sensing image retrieval (RSIR) may be a fundamental task in remote sensing. Most content-based image retrieval (CBRSIR) approaches take an easy distance as similarity criteria. A retrieval method supported weighted distance and basic features of Convolutional Neural Network (CNN) is proposed during this letter. the strategy contains two stages. First, in offline stage, the pretrained CNN will be fine-tuned by some labelled images from our target data set, then accustomed extract CNN features, and labelled the pictures within the retrieval data set. Second, in online stage, we extract features of the query image by using fine-tuned CNN model and calculate the load of every image class and apply them to calculate the space between the query image and also the retrieved images. Experiments and methods are conducted on two Remote Sensing Image Retrieval data sets. Compared with the state-of the-art methods, the proposed method significantly improves retrieval performance.
Keywords: Convolutional Neural Network (CNN), Remote Sensing Image Retrieval (RSIR), Feature Extraction, Weighted distance, Visual Geometric Group (VGG16).

I. INTRODUCTION

Remote sensing images become vital within the fields like in geological analysis, urban planning, natural disaster monitoring, weather prediction, resource investigation, and so on. How to efficiently retrieve the remote sensing images according to users need from large image database becomes one a challenging and emerging research topic within the field of remote sensing. Content Based Image Retrieval (CBIR) could be a useful method to resolve more problems supported the features of image, such as intensity, shape, texture, and structure. Content based RSIR (CBRSIR) is a full of life and challenging research topic that has attracted the eye of researchers round the world. Because of the lack of data, it is difficult to train a novel model from the beginning CNN makes it easier and user-friendly method to retrieve remote sensing images from a large dataset.

During these years it has been proved that, an explained literature has introduced on high resolution remote sensing (HRRS) image retrieval problem. Many of the researchers have applied different styles of descriptors on remote sensing image retrieval purpose, and these descriptors are often classified into two main categories. These include local features extracted at interest point locations, like intensity features, spectral features, shape features, structural features, texture features, etc. and also global features, obtained by encoding local features into one vector representation using bag-of-words (BoW), vector locally aggregated descriptors (VLAD), or their variants.

The common ancient image representations have terribly restricted capability in dealing with the quality of HRRS scenes and multi formats of HRRS classes. To beat the weakness of existing HRRS image retrieval ways, we tend to use deep features derived from Convolutional Neural Networks (CNNs) to displace conventional representations. CNNs have earned superb achievements in computer vision analysis space throughout recent years, apart from original image classification task, CNNs proven to own inter changeableness as terribly powerful feature extractor. Deep features are outstanding in finding varied visual issues, like image classification, object detection, fine-grained recognition, and image instance retrieval.

Remote sensing community has applied deep learning techniques to extract high level semantic features for CBRSIR. to beat the prevailing problems of HRRS image retrieval methods, we implement deep features from Convolutional Neural Networks (CNNs) to import conventional representations. CNNs proved to possess a very powerful feature extractor. Deep features are dominant in solving various visual problems, like image classification, object detection, and image instance retrieval. In this paper for image retrieval process we use a weighted distance and CNN features. Here we fine-tune the pre-trained CNN model to calculate the load of every class within the retrieved data set for the query image. We also give more importance to the retrieved images in additional similar classes with the query image.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

II. EXISTING FRAMEWORK

The standard process of constructing a picture retrieval system consists of two parts, i.e., an initial stage of feature extraction that presumably extracts discriminating information from the photographs and an indexing framework that enables efficient querying to search out similar image sets.

During extraction, we opted to use a multiscale dense sampling strategy during which descriptors are computed for same sized and same-shaped patches over multiple scales, rather than a KEY POINT-based sparse approach, because it provides a more robust representation.

These descriptors are then combined in three different schemes so as to represent the entire image, namely, BoW, VLAD, and a quantized binary version of VLAD denoted by VLAD-PQ.

BoW represents, which is initially introduced to index large numbers of text documents for retrieval purposes. It has been used in a number of computer vision studies for recognition tasks, and it became a standard baseline method as it is easy to implement and it shows good retrieval performance. The idea behind the framework is to represents each image by the occurrence frequency of some discriminating features.

The next strategy represents image representation, which is named VLAD, is a more recent approach that is based on Fisher kernels. It was first introduced for classification problems in the context of DNA and for protein sequence analysis. By treating all the extracted descriptors from an image, the independent samples drawn from the imaging process from a particular scene, VLAD makes use of the Fisher kernel is used to compute a descriptor to represent the image. The VLAD is represented in the form by aggregating the difference vectors of the descriptor codeword pairs.

VLAD-PQ is a product-quantized binary version of the VLAD descriptor, is utilized. Quantization is a mapping technique it is used to represent high-dimensional vectors with small-length bit strings. The input vectors with cluster centroids and it contain a high number of examples are necessary during the learning of the quantizer function. This problem intensifies the input dimension size increases.

Local features are introduced for the latest generations of aerial and satellite imagery whose increased spatial resolution. An extensive evaluation of local invariant features is used for image retrieval of Land-Use/Land-Cover (LULC) classes in high-resolution aerial imagery is performed. It contains different design parameters on a bag-of-visual-words (BOVW), which includes saliency- versus grid-based local feature extraction. The size of the visual codebook, clustering algorithm is used to create the codebook. The dissimilarity measure is used to compare for BOVW representations.

The BOVW approaches do not perform better than the best standard approaches. It represents a robust alternative that is more effective for certain land-use classes. It has shown that BOVW approach with proposed spatial co-occurrence kernel consistently improves the performance. The potential of local standard spectra, calculated with a dense strategy, for describing the content of aerial images, with CBIR as the final target. Furthermore, we combine them with the Locally Added Descriptor Vector (VLAD) to form a visual vocabulary.

More specifically, pattern spectra are one of the earliest and most powerful content descriptors that mathematical morphology offers. They were originally introduced to explain the scale distribution of objects within images using histogram-like representations. within the course of your time they were expanded to explain not only the scale, but also the distribution of shape and texture also because the distribution of arbitrarily selected attributes between the linked image components. In addition, their efficient implementation through tree-based imagery has made them effective, efficient, multi-scale, universal, holistic (i.e., global) content description tools.

Therefore, first contribution to the state of the art is to apply sample spectra to the CBIR of aerial images for the first time. In addition, the theoretical fundamentals were recently introduced in order to calculate them on site. The first reported implementation of local pattern spectra was in the context of universal image classification, where they were calculated from key regions captured by MSER.

After the promising results that have been achieved in this context, in this article we examine for the first time the computation of local pattern spectra not from prominent regions, but from the entire input image, without using a global approach. dense grid with fixed passages. The SIFT was researched together with VLAD, it examined and reported on the results of combining locally calculated sample spectra with the visual vocabulary obtained via VLAD.

Problem Definition: Unlike normal scene images, special characteristics of structures are absent in Remote sensing (RS) images since they are taken from top view. Also, it is difficult to retrieve a remote sense image from a large database.

III. PROPOSED SCHEME

With the speedy development of remote-sensing technology and also the increasing range of Earth observation satellites, the amount of image datasets is growing exponentially. The management of big Earth information is additionally turning into progressively advanced and tough, with the result that it is often onerous for users to access the representational process that they're curious about quickly, expeditiously. to deal with these challenges, we tend to propose a remote-sensing image-retrieval model supported convolutional neural network (CNN).

Remote Sensing Image Retrieval (RSIR) is a fundamental activity in remote sensing. Most content-based RSIR approaches use a simple distance as a criterion for similarity. In this scheme we propose a recovery method based on the weighted distance and the fundamental properties of the Convolutional Neural Network (CNN). The method consists of two phases. First, in the offline phase, the pre-trained CNN is provided with some tagged images from the target data set, then it is used to extract the CNN features and the images are tagged in the retrieval data set. Second, in the on-line phase, we use the adapted CNN model to extract the CNN function from the query image and calculate the weight of each images in the class and use them and compare it with distance between the query image and the retrieved images.

Here two winning CNN models pretrained on ImageNet which are evaluated for remote sensing image retrieval in our recent work, namely, the far-famed baseline model VGG and therefore the ResNet (Residual Network) model. VGG model gains smart performance without a complicated computation. In 2015, ResNet achieved progressive performances in several computer vision tasks. ResNet heavily concentrated on the convolutional network by adapting residual perform and identity mapping.

Unlike normal scene images, special characteristics of structures are absent in Remote sensing (RS) images since they are taken from top view. Also, it is difficult to retrieve a remote sense image from a large database. By using basic features of CNN and by calculating weight of each images in our data set and comparing it with the query image here we can retrieve remote sensing images easily from a large dataset. Here we retrieve images using Convolutional Neural Network features and by calculating Weighted Distance.

Here we use normalization for image classification and labelling every image within the data set. Clustering techniques are accustomed to label images within the data set relating to color, size and form. The complete data set images are labelled based on extracted features and this labelling results used for classifying the image. An optimization is performed on the extracted features employing a changed genetic algorithmic program to accumulate a far better image retrieval method. The k-means clustering technique is employed for classification processes in which activation update is performed. Finally, the resulting images retrieved.

We use optimization algorithm where, optimizers are algorithms or methods used to change the values of our neural network such as weights and learning for loss reduction. In this proposed system we use optimization algorithm to get the accurate or an optimum value. An optimization algorithm is a procedure wher it is executed iteratively by comparing various solutions till an optimum or a satisfactory solution is found. Here we compare the weighted distance of the query image with the images in the data set.

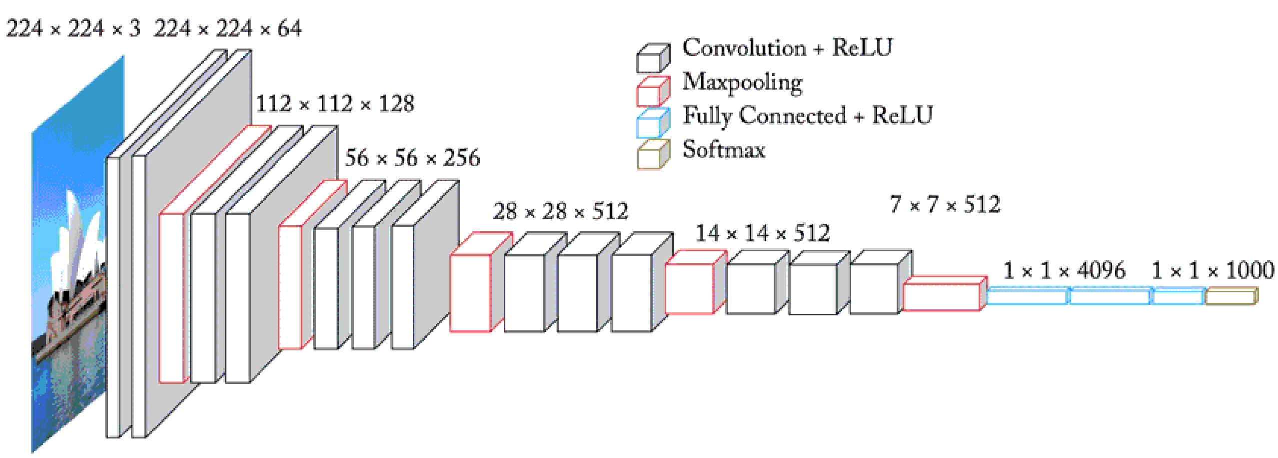


Fig. 1.0 Architecture of VGG16

There are 16 layers in VGG16(Visual Geometric Group) with 13 convolutional layers, and 3 fully connected layers.

The input to this will be an image of dimensions (224, 224, 3), where 3 indicates 3 RGB channels. The first two layers of VGG16 will have 64 channels of 3*3 filter size and same padding.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Then after this comes a max pool layer of stride (2, 2), where two layers which have convolution layers of 256 filter size and filter size (3, 3). This again will be followed by a max pooling layer of stride (2, 2). It is same as the previous layer. Then there again comes 2 convolution layers of filter size (3, 3) and 256 filter. After this, there comes 2 sets of 3 convolution layer and a max pool layer. Each will be having 512 filters of (3, 3) size with same padding. After all this process the image will be passed to a combination of two convolution layers. Filters of size 3*3 will be used here in convolution and maxpooling layers. In some of the layers, it also uses 1*1 pixel it is for manipulating the number of input channels. Padding will be done after each convolution layer to prevent the spatial feature of the image.

We use padding because the information on the borders of images are not preserved as well as the information in the middle. To unravel these issues, we use padding, that is a method of adding layers of zeros to our input pictures so to avoid the issues mentioned before.

After we apply filters to the image its results will captured in feature map. So, after the stack of convolution and max-pooling layer, we got a (7, 7, 512) feature map. We flatten this output to form it a (1, 25088) feature vector. After this there comes 3 fully connected layer. There the first layer takes input from the last feature vector and gives outputs a (1, 4096) vector, second layer also outputs a vector of size (1, 4096) but output of third layer will be 1000 channels for 1000 classes of ILSVRC (ImageNet Large Scale Visual Recognition Challenge). Then after the output is then passed to softmax layer in order to normalize the classification vector. After the output of classification vector top-5 categories for evaluation, all the hidden layers use ReLU (Rectified Linear Unit) as its activation function. ReLU is an activation function, which is used in convolution layer and fully connected layer. ReLU results in faster learning and it also decreases the likelihood of vanishing gradient problem. This is the reason why it become more efficient.

IV. FRAMEWORK DEMONSTRATION

In our approach we first briefly introduce two CNNs that we use and the fine-tuning with the two pretrained CNN models, then describe the CNN features used to retrieve image. Next, the weighted distance is presented. At last, the process of image retrieval method has been described.

Inside these processes it includes, fine tuning CNN models, image segmentation, feature extraction, labeling each image in the dataset, classification, at last process of image retrieval. Here we have two parts in our proposed approach, one part is offline and the other is online stage.

In this approach we use a weighted distance and the simple CNN features to retrieve image. We need to calculate weight of each class in the retrieved data set for query image by fine-tuning the pretrained CNN model. More preference will be given to the retrieved images which is more similar classes with the query image. The framework of our method is shown in Fig. 2.0.

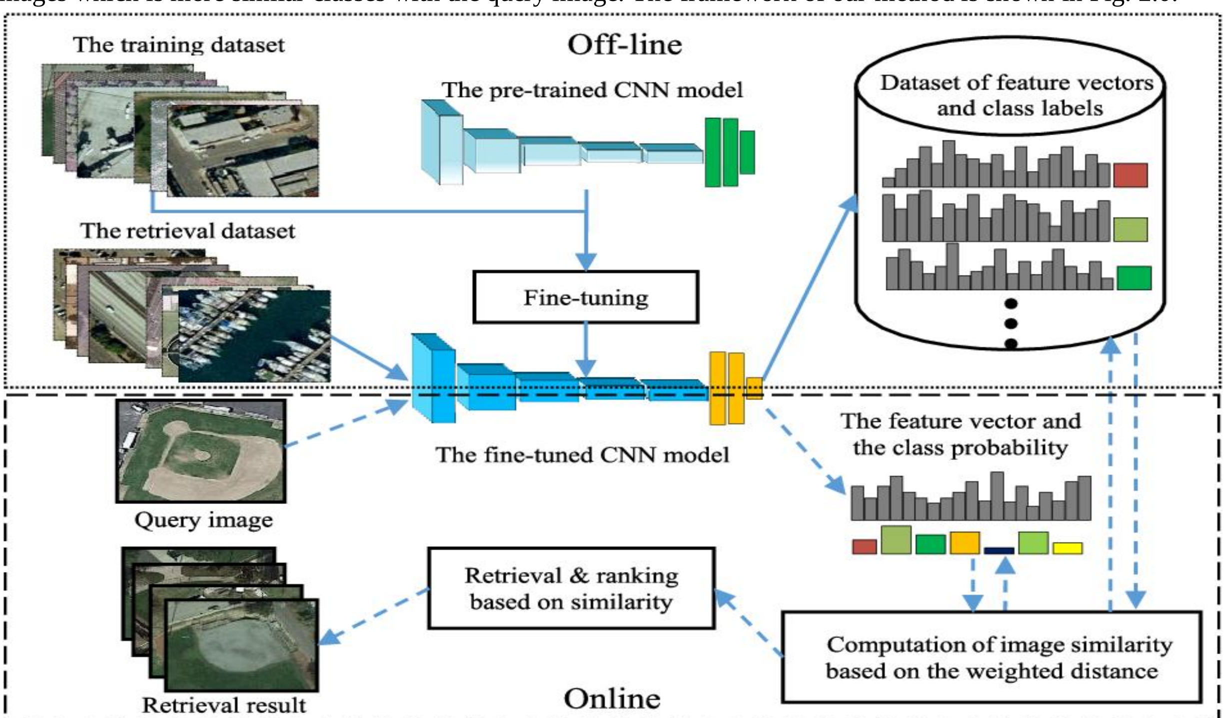


Fig. 2.0 Framework of our Proposed Approach

In the offline stage, we will fine-tune the pretrained CNN model with the labeled images in our data set. Then, label and extract the CNN (Convolutional Neural Network) features for the images inside the retrieved data set with the fine-tuned CNN model. Finally, we will build a data set with feature vectors and its class labels.

In the online stage of our proposed methodology, a query image will be fed into fine-tuned CNN model to calculate the CNN features and so the class of the image likelihood. Then, the weighted distances between the question image and the retrieved images are computed. After that, we sort the retrieved images according to ascending order per the weighted distances. Finally, we will get the retrieval results.

Here, two successful CNN models pretrained on ImageNet are evaluated for remote sensing image retrieval in our proposed scheme. They are VGG (Visual Geometric Group) and ResNet model. Because of lack of data it will be difficult to train a model from beginning. So, fine-tuning is the best way to get a better effect with few iterations by adjusting the pretrained parameters to better suit the target data set. Regarding the fine-tuning process, the softmax layer normalized by the last fully connected or convolutional layer is used for classification. The outputs from the higher-level layers of CNN will be having better performance. These outputs from the fine-tuned VGG16 model and the last convolutional layer will be considered as the features for remote sensing image retrieval. Each image in our dataset should be labelled and it is necessary to mention each of its classes. So, we feed each images in our dataset into the fine-tuned CNN model. Then we apply softmax function to the output for converting the output to the probability for each class. Here, softmax function is used to represent the reliable probability in the network input which is mainly used to normalize output values to fit between zero and one. Then we calculate weighted distance. It is mainly used to give more preference to the retrieved images which are more similar classes with the query image. The more the class probability of the query image, the less is that the distance between the query image and also the retrieved images of that particular class.

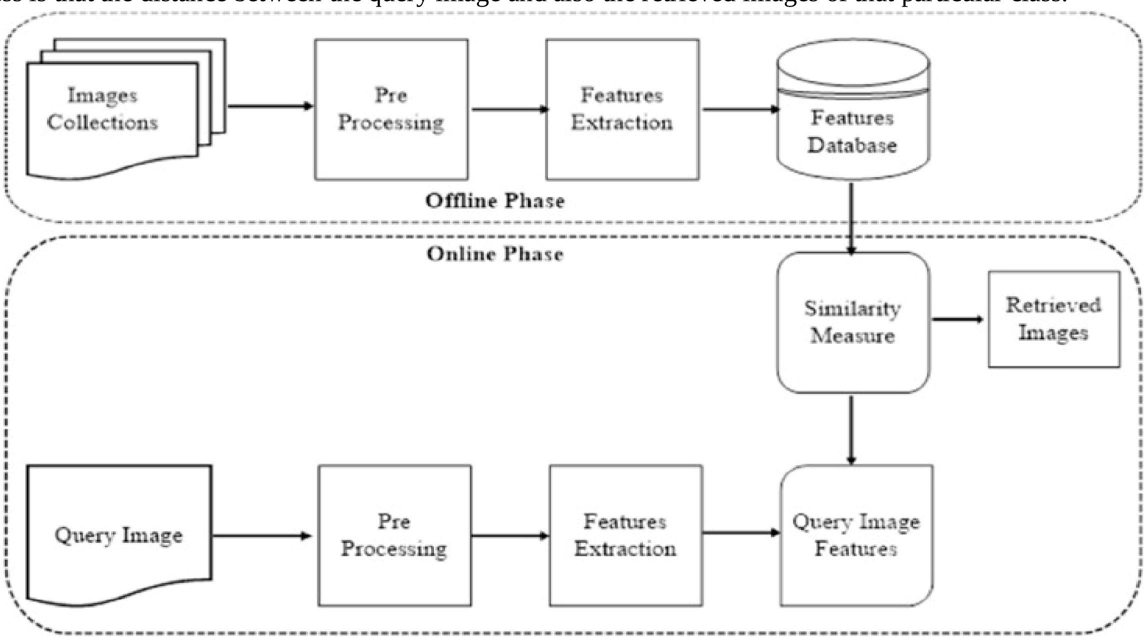


Fig.3.0. Proposed remote sensing image retrieval system

V. FOCAL POINTS

A. Data Collection

Here we collect remote sensing image datasets which are publicly available. Here we have University of California, Merced (UCMD) data set, which consists of 21 land-use categories cropped from the United States Geological Sur_vey National Map. We have four classes of images like airport, playground, bare land and railway station. Each class has 100 images with 256×256 pixels.

B. Image Pre-processing

Pre-processing is done to transform any raw data before it is fed to the machine learning or deep learning algorithm. For example, there will be bad classification performances when training a convolutional neural network on raw images. In order to overcome this, we will do image pre-processing. All images in the UCMD (UC Merced Land use Data set) data set are reduced to a resolution of 224 x 224 in order to reduce the computing time. The images were then saved in RGB format, which are the three channels.

C. Feature Extraction

A deep learning algorithm, namely, the VGG-16(Visual Geometry Group), was employed to extract the features. A CNN model is a combination of two parts: one is feature extraction and the other is classification part. The convolution + pooling layers perform feature extraction. For instance, consider an image, detection of features such as two eyes, long ears, four legs, a short tail and everything present in that image will be done by convolution layers. Here, we consider the outputs of high-level layers from the fine-tuned VGG16 model and the last convolutional layer from the fine-tuned model as the features for remote sensing image retrieval.

D. Classification

Each image in the retrieval data set needs to be labelled which class it belongs to. Image classification may be a supervised learning problem: outline a group of target categories (objects to spot in images), and train a model to acknowledge them victimization labeled example images.Image classification is that the method of categorizing and labeling teams of pixels or vectors at intervals a picture supported specific rules.Classify the images into classes like agricultural, airplane, buildings, beach, baseballdiamond, chaparral, denseresidential, forest, freeway, golfcourse, harbor, intersection, mediumresidential, runway, river, mobilehomepark, overpass, parkinglot, sparseresidential, storagetanks, tenniscourt.

E. Evaluation

We use weighted distance to give more preference to the retrieved images in more similar classes with the query image. Measuring the efficiency & effectiveness of training model. ReLU, activation function has been used to evaluate the CNN model. We observed that ReLU gives better classification performance in comparison to other structures. The Euclidean distance is used to evaluate the performance of the proposed method.

F. Testing

Upload query image to detect the type of images that belongs to classes like agricultural, airplane, baseballdiamond, beach, buildings, chaparral etc. It is by calculating the weighted distance of query image and comparing it with the images in the data set.

G. Image Retrieval

When a query image is fed into the fine-tuned model, we get a probability for each class and use it to calculate the weight of a retrieved image in the retrieval data set. We will calculate weighted distance of the query image in the online stage. Then we will compare it with the weighted distance of the images in the data set. And the images which are of the similar to the query image will be retrieved.

VI. EXPERIMENTs AND ANALYSIS

Here, we evaluate the performance of the proposed method for remote sensing image retrieval on remote sensing data set which are publicly available. It is University of California, Merced (UCMD) Data set. It consists of 21 land-use categories from United States Geological Survey. We have four classes like airport (100 items), playground (98 items), bare land (98 items), railway station (98 items) in our data set.

We evaluate per class mean average precision values for the CNN features on data sets. The average values of the methods using the pretrained feature without the weighted distance will be having poor performance for buildings, storage tanks, and tennis court, but our methods with the weighted distance which achieve better results. Average values of the methods using the pretrained feature are approximately 60%, while our methods are approximately 90%. It can conclude that our methods can get better results than the pretrained features.

The number of the training images using in the fine-tuned process is an important factor of the class prediction precision. Therefore, we have a tendency to build the experiment to investigate the impact of coaching data volume on retrieval performance. We take a series of images per image class to fine-tune the pretrained CNN model, which these images are randomly split into training and testing data sets. The results of pretrained CNN features with weight are similar to those of fine-tuned features with weight. It means that our methods can get good performance by the two kinds of features.

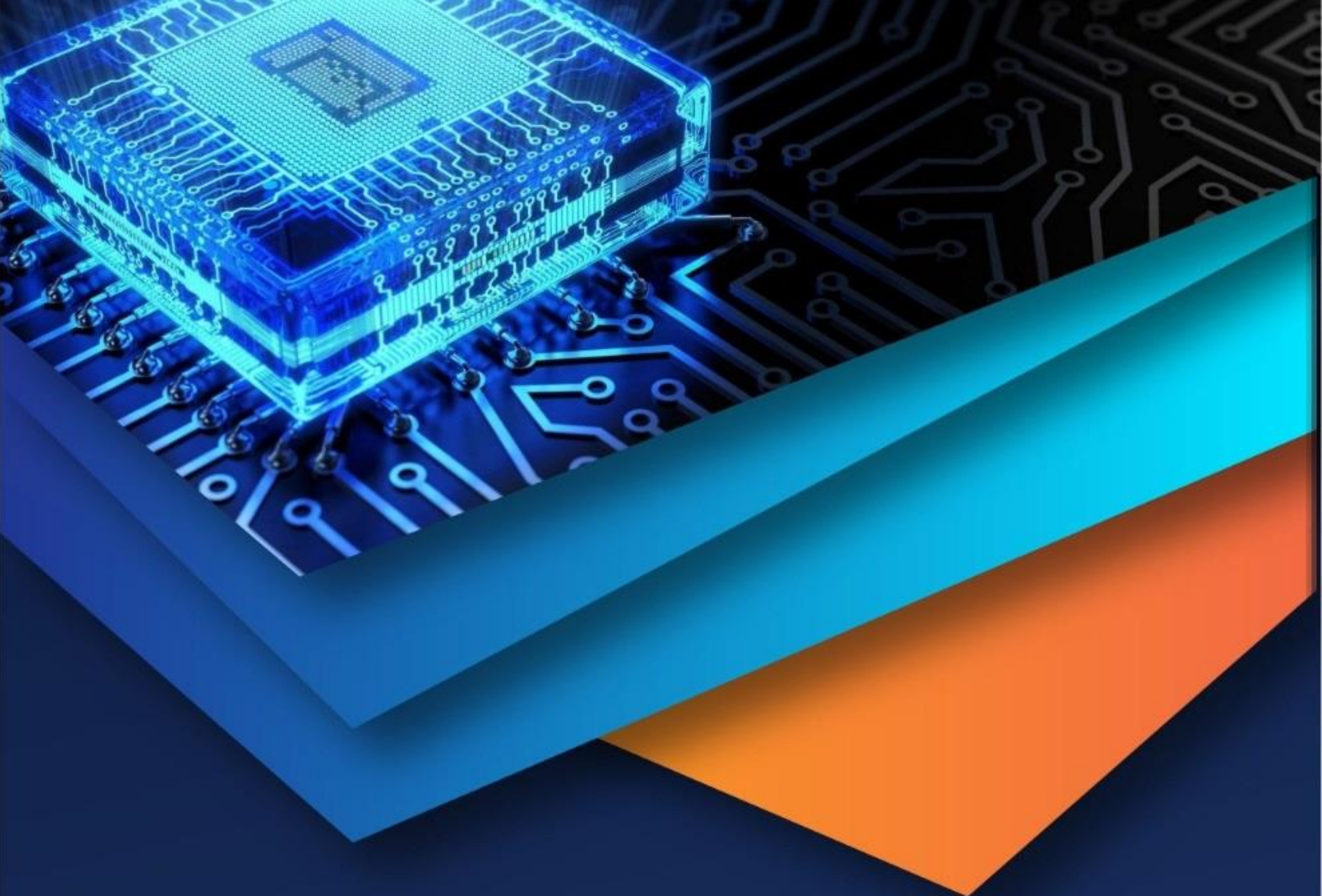
To validate the effectiveness of the proposed method, we compare our method with the state-of-the-art methods. The results prove that our methods are better than those of the other methods when comes in training. In briefly, our method is simplified and can get excellent performance with some training images.

VII. CONCLUSION

Automatically and with efficiency to retrieve the remote sensing images that users would really like from huge image databases becomes one in each of the difficult and rising analysis topics inside the sector of remote sensing. Because of the lack of data, it is difficult to train a novel model from the beginning. CNN makes it easier and user-friendly method to retrieve remote sensing images from a large dataset. Within the offline stage, we use the fine-tuned CNN models to extract image features and label the class of image inside the retrieval data set. In the online stage, we need to calculate the weight of each class in step with the class chance of the question image and used it to regulate the gap between the question image and thus the retrieved photos. The experimental results on the UCMD and PatternNet data sets demonstrate that the planned methodology achieves higher performance compared with that of any of the other progressive ways.

REFERENCES

[1] S. Özkan, T. Ates, E. Tola, M. Soysal, and E. Esen, "Performance analysis of state-of-the-art representation methods for geographical image retrieval and categorization," IEEE Geosci. Remote Sens. Lett., vol. 11, no. 11, pp. 1996–2000, Nov. 2014.
[2] Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in Proc. ACM Int. Conf. Adv. Geogr. Inf.Syst., 2010, pp. 270–279.
[3] Fei Liu, Yong Wang, Fan-Chuan Wang "Intelligent and Secure Content-Based Image Retrieval for Mobile Users",ACCESS.2019.2935222, IEEE Access,in Aug 2019
[4] Russakovsky, O., Deng, J., Su, H., et al. "ImageNet Large Scale Visual Recognition Challenge." International Journal of Computer Vision (IJCV).
[5] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014)
[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learn. Represent., Sep. 2015, pp. 1–14.
[7] Y. Ge, S. Jiang, F. Ye, C. Jiang, Y. Chen, and Y. Tang, "Aggregating CNN features for remote sensing image retrieval," Remote Sens. Land Resour., to be published.
[8] Q. Bao and P. Guo, "Comparative studies on similarity measures for remote sensing image retrieval," in Systems, Man and Cybernetics, 2004 IEEE International Conference on, vol. 1. IEEE, 2004, pp. 1112–1116.
[9] T. Bretschneider, R. Cavet, and O. Kao, "Retrieval of remotely sensed imagery using spectral information content," in Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International, vol. 4. IEEE, 2002, pp. 2253–2255.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY