Remote Sensing Image Retrieval using Convolutional Neural Network with Weighted Distance and Result Re-Ranking

Arjun N.K

M.Tech Student, Dept. of CSE Engineering, AWH Engineering College, Kerala, India

---***---

Abstract - *Remote sensing image retrieval (RSIR) is a fundamental task in remote sensing. This letter proposes a content based image retrieval method based on weighted distance and basic features of Convolutional network. in this method there are two stages. First, in offline stage, the pre-trained CNN is fine-tuned by set of labeled images and extract CNN features to label retrieval data set. Second, in online stage, we use the fine-tuned CNN model to extract the CNN feature of the query image , and calculate the weight of each image class and apply them to calculate the distance between the query image and the retrieved images. HOG and LBP techniques are used with CNN to extract features. a feedback based reranking method is introduced to optimize the retrieved images Experiments are conducted on two RSIR data sets. Compared with the state-of the- art methods, the proposed method is simplified but efficient, significantly improving retrieval performance.*

*Key Words***:** *Convolutional neural network (CNN), remote sensing image retrieval (RSIR), weighted distance.*

1. INTRODUCTION

Remote sensing images become very important in the field of geological analysis, urban planning, natural disaster monitoring and assessment, weather prediction, resource investigation, and so on, with the development of remote sensing technology, both quality and quantity of remote sensing (RS) images have been growing quickly. How to automatically and efficiently retrieve the remote sensing images from large image databases becomes one of the challenging and emerging research topics in the field of remote sensing. Content based image retrieval (CBIR) is a useful method to solve the hard problem, which is based on the features of image, such as intensity, shape, texture, and structure.

Over the years, a rich literature has been developed on high resolution remote sensing (HRRS) image retrieval problem. Researchers have applied different forms of descriptors on RS retrieval purpose, and these descriptors can be classified into two main categories by default. There includes local features extracted at interest point locations, such as intensity features, spectral features, shape features, structural features , texture features, etc. and also global features, obtained by encoding local features (i.e., the scale-invariant feature transform (SIFT)) into a single vector representation using bag-of-words (BoW), vector locally aggregated descriptors (VLAD), or their variants.

Nonetheless, the common traditional image representations have very limited capability in coping with the complexity of HRRS scenes and multi formats of HRRS categories. To overcome the weakness of existing HRRS image retrieval methods, we employ deep features derived from Convolutional Neural Networks (CNNs) to displace conventional representations. CNNs have attained amazing achievements in computer vision research area during recent years, except for original image classification task, CNNs proved to have transferability as very powerful feature extractor. Deep features are prominent in solving various visual problems, such as image classification, object detection, fine-grained recognition, and image instance retrieval.

This letter uses a weighted distance and the simple CNN features to retrieve image. We fine-tune the pretrained CNN model to calculate the weight of each class in the retrieved data set for query image. We give more preference to the retrieved images in more similar classes with the query image. We also use HOG and LBP features for feature extraction and feedback oriented reranking method to filter and sort the result the framework of our method is shown in Fig. 1.

Experiments show that the retrieval method with weighted distance can retrieve more desired images. The remainder of this letter is organized as follows. In Section II, our remote sensing image retrieval method based on CNN features and weighted distance is described. In Section III, experimental results and analyses are presented. We conclude this letter in Section IV.

2. METHODOLOGY

In this section, 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. Finally, the process of the proposed image retrieval method is described.

2.1. CNN Models

CNN architecture is the most effective deep learning approach in various fields such as image classification, object recognition, and image scene analysis, CNN consist of many

layers such as convolution, pooling, and fully connected layers. The network architecture used is the Convolutional neural network model called VGG-16. Fig.1. shows the VGG-16 architecture. This architecture has got 16 layers which include convolution layers followed by pooling layers and finally the fully connected layers. Weight sharing is performed in Convolutional layer and in pooling layers the sub sampling of outputs of the convolution layer for lowering the data rate from the previous convolution layer is done. The main feature of the VGG-16 network is the fixed size of input is 224 x 224..

2.2. Fine-Tuning the CNN Models

CNNs have been dominant in feature extraction, and have gradually replaced traditional methods in the field of computer vision. The achievement of CNNs is mainly due to the fact that deep network structures bring a large number of nonlinear functions, and weight parameters can be automatically learned from the training data. However, remote sensing image datasets cannot provide a large amount of data for CNNs training from scratch. CNNs pre-trained on massive datasets have been used to extract feature embedding, which has been proven to be effective and efficient, even when the training data set has a lot of difference with the remote sensing image. It is difficult to train a novel model from the beginning because of the lack of data and hard adjustment of parameters.
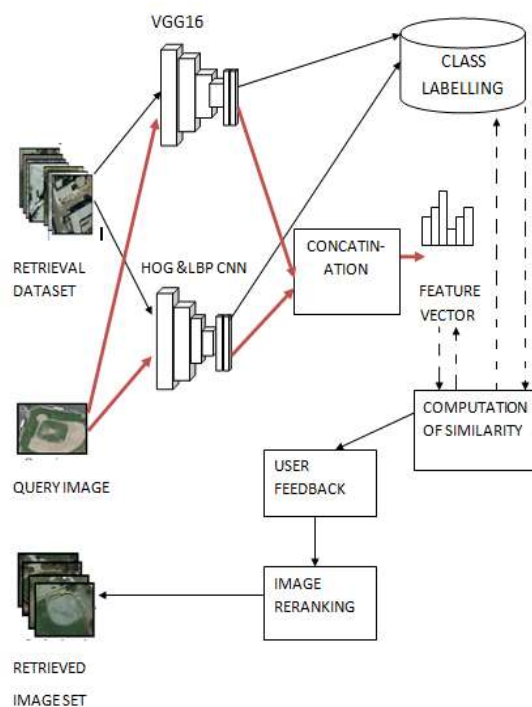

Fig. 1. Framework of our proposed approach.

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. In this letter, the pretrained CNN models, VGG and ResNet pretrained on ImageNet data set, are fine-tuned using the target retrieval data set. The two pretrained CNNs are implemented and fine-tuned using

MatConvNet, which is a MATLAB application of CNNs for computer vision applications. MatConvNet gets advantages in CNN building blocks under MATLAB which is a friendly development environment.

2.3 Feature Extraction

The outputs from the higher level layers of CNN (e.g., Fc6 and Fc7 from VGG Model) have been shown to be very effective generic features which achieve state-of-the-art performance in image retrieval . Here, we consider the outputs of high-level layer (Fc6 and Fc7) from the fine-tuned VGG16 model and the last convolutional layer (Pool5) from the fine-tuned ResNet50 model as the features for remote sensing image retrieval.

There are two methods are introduced with CNN architecture to extract image features are

2.3.1 Local Binary Pattern (LBP):

LBP feature extractor has been applied in many areas including text identification and face recognition Classical LBP feature extractor finds the intensity values of points in a circular neighborhood by considering the values which have small circular neighborhoods around the central pixel .

2.3.2. Histogram of Oriented Gradients (HOG):

HOG captures features by counting the occurrence of gradient orientation. Traditional HOG divides the image into different cells and computes a histogram of gradient orientations over them. HOG is being applied extensively in object recognition areas as facial recognition.

2.4. Labeling Image Retrieval Data Set

Each image in the retrieval data set needs to be labeled which class it belongs to. We feed each image into the fine tuned model and apply the softmax function to the output of the fine-tuned model for converting the output to the probability for each class.

$$p_i = \frac{e^{a_i}}{\sum_{k=1}^{K} e^{a_k}}$$

where a is a K vector of the inputs to the output layer, and

i indexes the output units, $i = 1, 2, \ldots, K$. The class of image I is predicted as the class with the max probability.

2.5. Weighted Distance

CNN is very effective in distinguishing image class, so we use the ability of CNN to improve remote sensing image retrieval. The motivation of weighted distance is to give more preference to the retrieved images in more similar classes with the query image. The more the class probability of the query image, the less is the distance between the query image and the retrieved images of that class. When a query image q is fed into the fine-tuned model, we get a probability pq for each class and use it to calculate the weight of a retrieved image r in the retrieval data set according

$$w = 1 - p_k^q$$

where k is the class of the image r. The weighted distance between the query image q and the retrieved image r is reformulated as

$$dw(q, r) = w \times d(q, r)$$

where $d(q, r)$ is calculated using the conventional distance chosen suitably. The Euclidean distance is used to evaluate the performance of the proposed method, the top row is the retrieval results using the distance $d(q, r)$; there are five irrelevant images in the top nine results. While the bottom row is the results using the weighted distance $dw(q, r)$, the top nine results are all correct. The method with the distance $d(q, r)$ returns six irrelevant images, while the method using the weighted distance $dw(q, r)$ only returns three irrelevant images. It can be seen that the weighted distance can improve the retrieval result
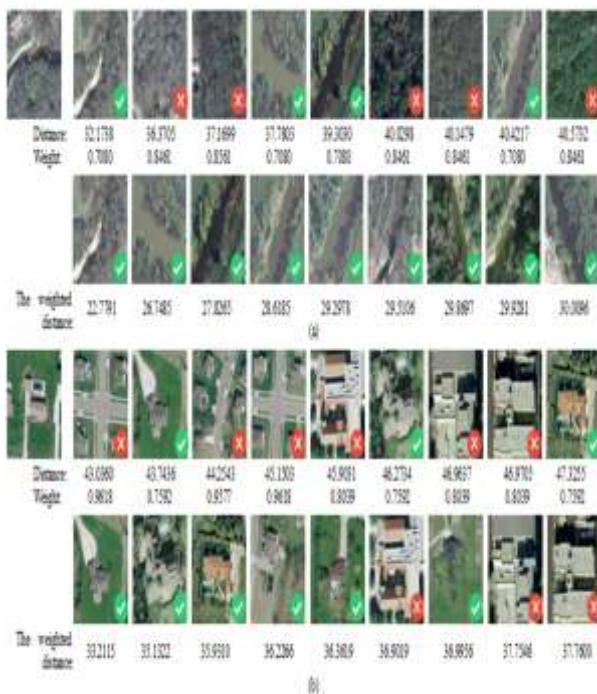

Fig. 2. Comparison of image retrieval results between different distances. (a) and (b) are two retrieval examples. In both (a) and (b), (Top row) Results with the distance $d(q, r)$, (Bottom row) Results of the weighted distance $dw(q, r)$, (First column) Query images.

2.6. Process of Image Retrieval

The framework of the proposed method is shown in Fig. 1. It consists of two parts: the offline part and the online part. In the offline part, we fine-tune the pretrained CNN model with some labeled images. Then, we label and extract the CNN features for the images in the retrieved data set with the fine-tuned CNN model. Finally, we build a data set with feature vectors and class labels. In the online part of our method, a query image is fed into fine-tuned CNN model to calculate the CNN features and the class probability. Then, the weighted distances between the query image and the retrieved images are computed. After that, we sort the retrieved images in

ascending order according to the weighted distances. Finally, we get the retrieval results.

2.7. IMPROVED K -MEANS CLUSTERING (lKC) ALGORITHM

Clustering is widely studied in data mining community. It is used to partition data set into clusters so that intra-cluster data are similar and inter-cluster data are dissimilar.. Kmeans is one of the most widely used clustering algorithms. Its essence is to reach the purpose of stepwise refinement through iteration. It has the following steps:

I) Pick K data points randomly to represent the initial centers of clusters;

2) Repeat steps (3) and (4), until the clusters are unchanged or criterion function has converged;

3) For each data point, find the cluster center that is closest to the data point. Assign the data point to the cluster whose center is closest to it, update K clusters;

4) Recalculate the center of each cluster.

This algorithm is based on the squared-error criterion .In traditional K-means algorithm, the choice of initial centers is random, so the clustering results are usually only locally optimal and fluctuant. If we choose some more appropriate initial centers, clustering results will be more reasonable, and the speed of clustering will also be faster. To remedy this problem, we take the following steps:

1) Calculate the distances of the data points between each other in the data set U;

2) Select two nearest data points to find a subset AI, and remove them from the data set U;

3) Calculate the distances between the subset AI and the remaining data points in data set U, and put the data that is nearest to subset AI into AI. Repeat this process until the data points in Al reach a certain number. The distances between the subset Al and a data point can be calculated by the following formula:

d (x , AI) = mine d (x , y), y E A 1)

4) Repeat steps (2) and (3) until K subsets AI, A 2 , ... , A k are formed;

5) Calculate the average values of the K subsets: ml, m2, ... mk . We choose ml, m2, ... mk as initial centers. Calculate the distances of the data points between each other in the data set U;

2.8. RELEVANCE FEEDBACK

Through applying IKC, the image search result can be partitioned into K clusters and K cluster centers are formed. The image that is nearest to the center is deformed as representative image of the cluster. There are two steps in relevance feedback: In the first step, the user is required to label the relevant representative images of the K clusters .The clusters whose representative images are labeled can form an image set G. For the image set G, we calculate the standard deviation of the ith dimension of the feature vectors, and its inverse is assigned to update the corresponding weight

formula. In the second step, we recalculate the distance between the query image q and the image p in search result using the following formula:

$$dis(Hq,Hp) = aDw(Hq,Hp)+ f/Dw(Hq,mp)$$

where mp is the center of the cluster Cp which the image p belongs to, and Dw(H q, mp) is the weighted Euclidean distance between the query image and the cluster Cp • a and f/ are weighting factors to adjust the importance of Dw(Hq,Hp) and Dw(Hq,mp) The distance between the query and the images in the nearer cluster is adjusted to a smaller value, otherwise the distance is modified to a larger value by formula . According to formula, we re-order the ranks of the images in search result in decreasing order.

3. EXPERIMENTS AND ANALYSIS

Here, we evaluate the performance of the proposed method for CBRSIR on two remote sensing data sets of urban area which are cropped from the Google kerala Map. Each class has 100 images with 256×256 pixels. The other one is rural area dataset cropped from the Google kerala Map. The data set contains 38 classes, each of which has 800 images.

3.1 Per Class mAP

We evaluate per class mean average precision (mAP) values for the CNN features on both data sets average values of the methods using the pretrained feature without the weighted distance are poor performance and methods with the weighted distance achieve noticeably better results. Average values of the methods using the pretrained feature are approximately 70%, while our methods are approximately 92%. It can conclude that our methods can get better results than the pretrained features. Moreover, it can be filter and sorted out the result using feedback based re ranking.

4. CONCLUSION

In this letter, we presented a remote sensing image retrieval method based on weighted distance and CNN. In the offline stage, we used the fine-tuned CNN models to extract image features and label the class of image in the retrieval data set. In the online stage, we calculated the weight of each class according to the class probability of the query image and used it to adjust the distance between the query image and the retrieved images. K-means clustering and relevance feedback to re-rank the search result. The experimental results demonstrate that the proposed method achieves better performance compared with existing methods

REFERENCES

[1] 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.

[2] 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.

[3] G. J. Scott, M. N. Klaric, C. H. Davis, and C.-R. Shyu, "Entropybalanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 5, pp. 1603–1616, 2011.

[4] A. Ma and I. K. Sethi, "Local shape association based retrieval of infrared satellite images," in *Multimedia, Seventh IEEE International Symposium on*. IEEE, 2005, pp. 7–pp.

[5] G.-S. Xia, J. Delon, and Y. Gousseau, "Accurate junction detection and characterization in natural images," *International Journal of Computer Vision*, vol. 106, no. 1, pp. 31–56, July 2014.

[6] D. Dai and W. Yang, "Satellite image classification via two-layer sparse coding with biased image representation," *Geoscience and Remote Sensing Letters, IEEE*, vol. 8, no. 1, pp. 173–176, 2011.

[7] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2010, pp. 270–279.

[8] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 675–678.

[9] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531*, 2014.

[10] Flickr photo sharing service, http://www.flickr.com/.

BIOGRAPHIES

ARJUN N.K, M.Tech student of AWH Engineering college, APJ Abdul Kalam Technological University, Kerala India.