Remote Sensing Image Retrieval using CNN and Weighted Distance as Classifier

Chippy Ravindran E[1], Flower Abraham Mundackal[2]

[1]PG Student, Dept. of Electronics & Communication Engineering, College of Engineering, Poonjar, Kerala, India
[2]Assistant Professor, Dept. of Electronics & Communication Engineering, College of Engineering, Poonjar, Kerala, India

---***---

Abstract - *Remote sensing image retreival (RSIR) is the basics of remote sensing. The quality of remote sensing data consist of its spatial, spectral and temporal resolutions. This image retreival method is based on neural network and weighted distance algorithm. It consists of two phases. In the first phase features of the querry images are extracted by using a pre- trained CNN model. The second phase is the classification of the extracted features. First the weighted distance method is used as the classifier for calculating the mean average precision (mAP) values. Then both CNN and weighted distance are used as a combination to calculate the mAP values. On comparison it can be identified that combination of CNN and weighted distance produces the best results.*

***Key Words*: Remote Sensing Image Retreival (RSIR), Convolutional Neural Network, Weighted Distance, mAP**

1. INTRODUCTION

Detecting and monitoring the physical charecteristics of an area by measuring its reflected and emitted radiation at a distance (typically emitted from a satelite or an aircraft). Special cameras will collect remotely sensed images, which help the researchers to sense things about the earth. With the development of the imaging technique the number of remote sensing (RS) images has increased explosively in the recent years. A fundamental step for most of the RS image processing task is to find the specific content from a ver large amount of RS images. Thus, it is important to develop the most efficient and accurate retreival method. Conventional content based remote sensing image retreival methods mainly use low level features to represent the content of the remote sensing images which includes texture, spectral and shape features. The local, low-level features are often transformed into mid-level representations to improve the retrieval performance by feature encoding techniques, for example, vector of locally aggregated descriptors[2], bag-of-visual-words (BoVW)[3],[4] and bag-of-features[5]. The proposed method is based on CNN which is a special class in neural network. For improving precision weighted distance method is combined with CNN to solve the problem. This combination of convolutional neural network and weighted distance provides better precision values as compared with the precision values produced by either CNN or weighted distance[1] alone as classifier.

1.1 Objectives of the study

The main objective of this paper to obtain a better image by remote sensing image retreival. This mainly includes:

- First the weighted distance method is used as the classifier for calculating the mean average precision (mAP) values.
- Then both CNN and weighted distance method are used as a combination to calculate the mAP values.
- On comparison it can be identified that the combination of CNN and weighted distance produces the best results or the mean average precision values.

2. METHODOLOGY

Convolutional neural network algorithm is a multi-layer perceptron which is specially designed for identification of 2D image information. CNN has three layers namely input layer, hidden layers and output layer.

The main objective of the Convolution Operation is to extract the high-level features such as edges, from the input image. ConvNets need not be limited to only one Convolutional Layer which is used. Conventionally, the first ConvLayer is responsible for capturing the Low-Level features[6] such as edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features as well, giving us a network which has the wholesome understanding of images in the dataset, similar to how we would. The algorithm has two main processes, convolution and sampling. the key technology of CNN is the local receptive feel, sharing of weights, subsampling by time or space, so as to extract features and reduce the size of the training parameters.

Similar to the Convolutional Layer, the Pooling layer is responsible for reducing the spatial size of the Convolved Feature[6]. This is to decrease the computational power required to process the data through dimensionality reduction. Furthermore, it is useful for extracting dominant features which are rotational and positional invariant, thus maintaining the process of effectively training of the model. There are two types of Pooling: Max Pooling and Average Pooling. Max Pooling returns the maximum value from the portion of the image covered by the Kernel. On the other hand, Average Pooling returns the average of all the values from the portion of the image covered by the Kernel. Max Pooling also performs as a Noise Suppressant. It discards the noisy activations altogether and also performs de-noising along with dimensionality reduction. On the other hand, Average Pooling simply performs dimensionality reduction as a noise suppressing mechanism. Hence, we can say that Max Pooling performs a lot better than Average Pooling. Adding a Fully-Connected layer is a (usually) cheap features as represented by the output of the convolutional layer. The Fully-Connected layer is learning a possibly non-linear function in that space.

A batch normalization layer normalizes each input channel across a mini-batch. To speed up training of convolutional neural networks and reduce the sensitivity to network initialization, use batch normalization layers between convolutional layers and nonlinearities, such as ReLU layers.

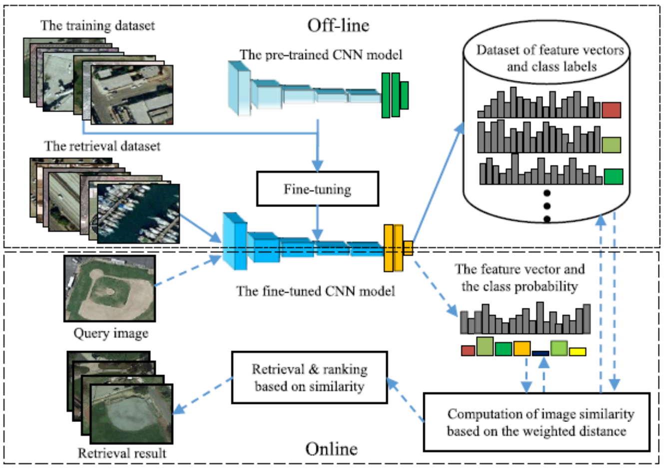

Fig 1 : Proposed Framework

This work mainly comprises of two stages - the training stage and the testing phase. In the first stage, the extracted image features from all 21*100 images of dataset. The dataset used for training is the UCMD dataset. Then, store the trained CNN features (21*100*4096) along with their labels (4096*1). In second phase, select an input querry image and feed to the neural network. The result will be ten most matching images. The CNN layer pattern is given below

conv1 → relu1 → norm1 → pool1 →

conv2 → relu2 → norm2 → pool2 →

conv3 → relu3 →

conv4 → relu4 →

conv5 → relu5 → pool5 →

fc6 → relu6 →

fc7 → relu7 →

fc8 → last layer

The dataset used for the testing phase is Uc Merced Land used dataset. This is a 21 class land image dataset. there are 100 images for each class. Each image measures 256x256 pixels. The images are manually extracted from large images from the USGS national map urban area imaginory collections foer vareious urban areas around the country. The pixel resolution of this public domain imaginory is one foot.

Fig 2 : UCMD dataset 1

Fig 3: UCMD dataset 2

The image is pre-processed before running the network. Pre-processing involves resized imagesto a fixed size as input and removing the mean; also image intensities are normalized in the range [0,255]. The next step is running the CNN which will return a structure with the output of the network layer. This CNN architecture consists of convolution with filter bank, feature pooling, normalisation and so on. CNN architecture contains 5 convolutions with filter bank, 3 feature pooling, 2 normalisation and 3 fully connected layer. The convolution layer computes the convolution of the input map x with a bank of k multi-dimensional filters f and biases b. the activation function for all weight layers is the REctification Linear Unit (RELU 1-7). Local response normalization is applied in first two layers (norm 1 & norm2). The max pooling downsampling factor is applied in first two layers and the fifth layer (pool 1,2 & 5).

Softmax layer computes the probability based on which the weight is identified by the network. The weight, W is given by

$$W = 1 - P^{k}_{q} \qquad (1)$$

Distance is calculated by Eucledian distance. Multiplying Eucledian distance by the weight gives weighted distance

$$dw(q, r) = w \times d(q, r) \qquad (2)$$

Weighted distance for each class is calculated and features are classified into class with less weighted distance.

3. EXPERIMENTAL RESULTS & ANALYSIS

First, a querry image is selected from the databasewhich is given below.

Fig 4 : Querry Image

Then it is passed on to a fine tuned neural network. Then the Convolutional Neural Network extracts its features. The total number of CNN features[7] extracted are 4096. These 4096 features are classified[8] and based on this classification ten most matching images are displayed.

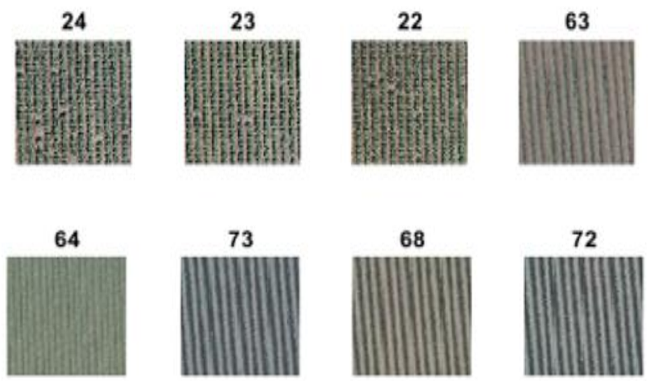

Fig 5 : Retreived Images

Precision of the retreival process can be analysed by using mean average precision (mAP)[9],[10] values. mAP consists of two parameters - precision and recall. Precision measures how accurate is your predictions. i.e. the percentage of your predictions are correct. Recall measures how good you find all the positives. For example, we can find 80% of the possible positive cases in our top K predictions.

Extracted features[7] are classified based on two algorithms- weighted distance algorithm and combining weighted distance with CNN classification algorithm. mAP values for both algorithms are calculated and both mAP values are compared.

As this work is based on CNN when the total number of training dataset increases, the mAP values also increases. There are two graphs shown below of which first one has lesser number of dataset while the second one has more number of dataset.

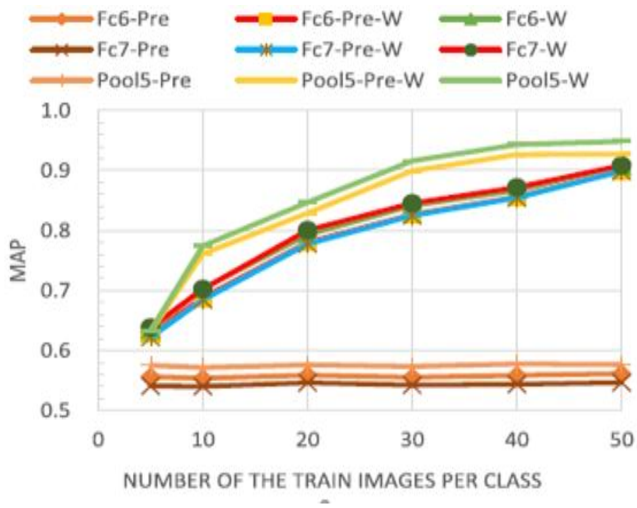

Fig 6 : Performance and training of 50 images

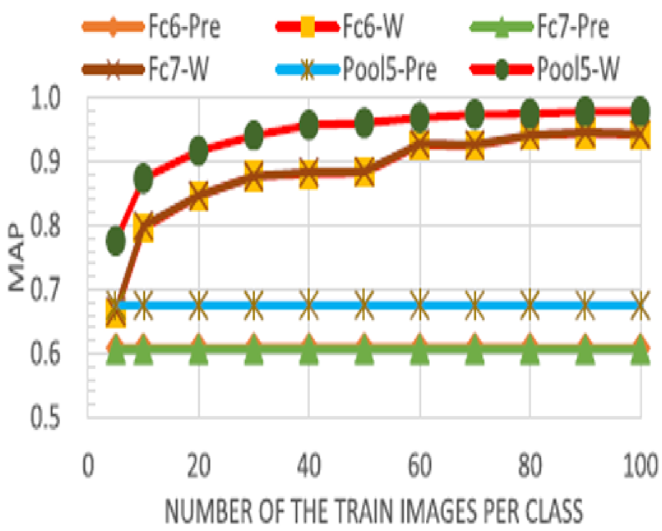

Fig 7: Performance and training of 100 images

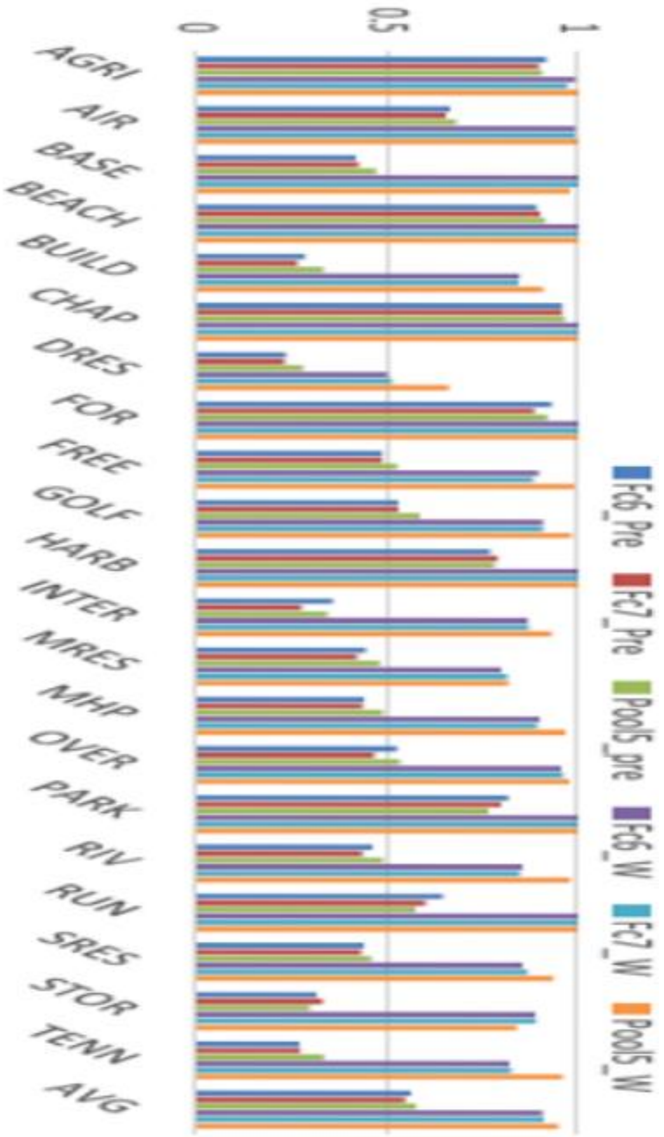

Fig 8 : mAP values for UCMD database

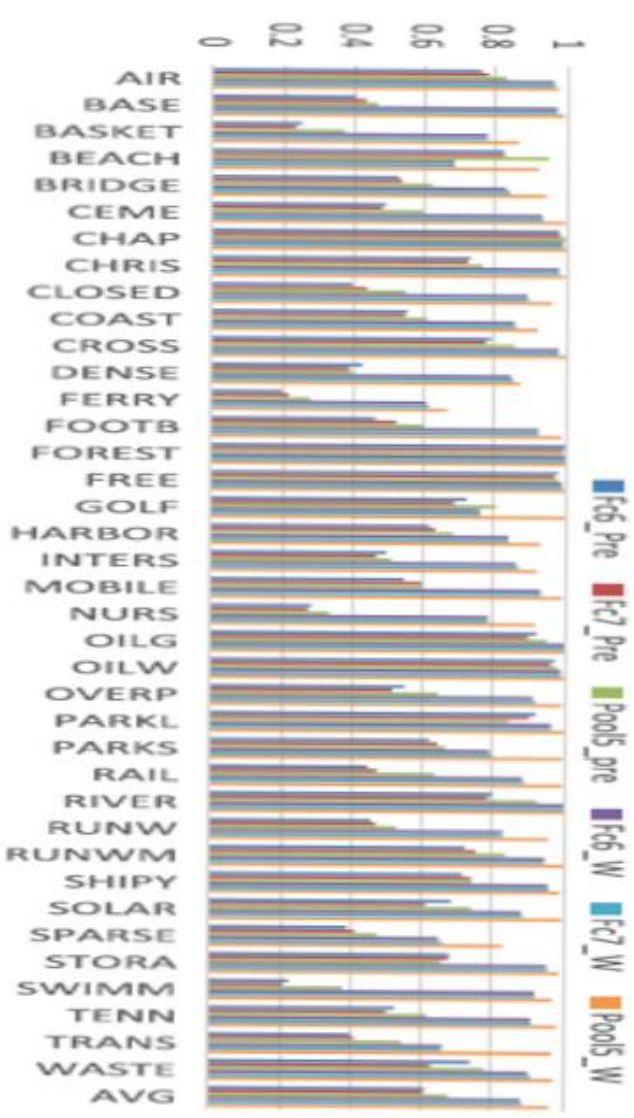

Fig 9 : mAP values for patternNet database

4. CONCLUSION

In this paper we have presented a neural network based system for retreiving remote sensing images. This has two stages of operations. In the first phase the features of querry are extracted. Feature extraction is accomplished by CNN algorithm. For classification we propose a more accurate algorithm rather than using the traditional algorithm. Classification algorithm is a combination of CNN algorithm and weighted distance algorithm. This combination provides a better precision value as compared with both CNN and weighted distance method alone.

REFERENCES

[1] Famao Ye, Hui Xiao, Xuqing Zhao, Meng Dong, Wei Luo, and Weidong Min," Remote Sensing Image Retrieval Using Convolutional Neural Network Features and Weighted Distance",IEEE GEOSCIENCE AND REMOTE SENSING LETTERS

[2] 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.

[3] Y. Yang and S. Newsam, "Geographic image retrieval using local invariant features",IEEE Trans. Geosci. Remote Sens., vol. 51, no. 2, pp. 818–832, Feb. 2013.

[4] Y. 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.

[5] N. Passalis and A. Tefas, "Learning neural bag-of-features for largescale image retreival",IEEE Trans. Syst., Man, Cybern. Syst., vol. 47, no. 10, pp. 2641–2652, Oct. 2017.

[6] Y. Ge, S. Jiang, F. Ye, C. Jiang, Y. Chen, and Y. Tang, "Aggregating CNN features for remote sensing image retrieval",Remote Sensing Land Resources (to be published)

[7] W. Zhou, S. Newsam, C. Li, and Z. Shao, "Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval",Remote Sensing, vol. l9, no. 5, p. 489, May 2017.

[8] W. Min, M. Fan, X. Guo, and Q. Han, "A new approach to track multiple vehicles with the combination of robust detection and two classifiers,"IEEE Trans. Intell. Transp. Syst., vol. 19, no. 1, pp. 174–186, Jan. 2018.

[9] Y. Yu and F. Liu, "Aerial scene classification via multilevel fusion based on deep convolutional neural networks," IEEE Geosci. Remote Sens. Lett.,vol. 15, no. 2, pp. 287–291, Feb. 2018.

[10] https://en.wikipedia.org/wiki/Precision and recall.