Deep Convolutional Neural Network using Local and Global Clustering Feature Extraction for Content Based Image Retrieval System

K Ramanjaneyulu
Research Scholar, ECE Department
JNTUCEK, Kakinada
ramanjaneyulu.k@qisit.edu.in

K Veera Swamy
Professor, ECE Department
Vasavi College of Engineering,
Hyderabad, Telangana
k.veeraswamy@staff.vce.edu.in

CH. Srinivasa Rao
Professor, ECE Department
JNTUKCEV, Vijayanagaram,
Andhra Pradesh
ch_rao@rediffmail.com.

***Abstract*— In this paper, deep convolutional neural network using clustering feature extraction for content based image retrieval system has been implemented. To extract the feature extraction from huge database deep convolutional neural network is required. The input image are converted into various images like RGB Color space, YCbCr color space and Gray scale images. The feature extraction is purely based on the clustering process. Each image is applied to various filters like sharpening, edge contours etc, The deep convolutional neural network has convolutional, pooling and softmax classifier. The experimental results are implemented using CoreL database. The proposed method gives better results as compared to existing methods in terms of recall, precision and F-Score criteria.**

Keywords*— Deep Learning, Content based Image Retrieval, Convolution Neural Network

I. INTRODUCTION

Use of digital image processing based applications has increased many fold with availability of low price disk storages and high speeds processors. Image databases containing millions of images are now cost effective to create and maintain [9]. Earlier image retrieval systems are text-based; in which images are manually annotated and then indexed according to the annotation. However, with the exponential increase in the volume of images database, the task of user-based annotation becomes very cumbersome [10]. Major Limitations of text annotation based image search are:

a. Image annotation is subjective
b. Image annotation may fails to convey the complete andappropriate details of image.

Content based Image Retrieval (CBIR) system extract and represents the visual features of images like color, shape, texture and uses these features to discover visually similar images. Color histogram is one of the important and widely used feature to describe the content of image. It represents the proportion of specific colors in an image [11]. Visual features are broadly categorized as Global and Local features. Global features are extracted from image by treating it as a single entity where as local features are extracted from a specific region of the image [12]. CNN is a Feed-Forward Neural Network that can extract topological properties from an image. It can recognize patternswith extreme variability as a result many researchers [1, 2, 4] have used Convolution Neural Network for image featureextraction & representation for CBIR. Rest of the paper is organized as follows.

Section II describes the related work carried out by other researchers using Neural Network for CBIR. Section III briefly explains theConvolution Neural Network and section IV presents the experimental system architecture. Section V presents the experimental results and section VI concludes the paper.

II. RELATED WORK

Ruigang Fu et. al. [1] have presented a CBIR system using Deep Convolution Neural Network and Linear Support Vector Machine. Deep features are extracted from images using Convolution Neural Network and similar images are identified using deep features and linear Support Vector Machine. Retrieved similar images are ranked based on distance between retrieved image and trained hyperplane. Kun Hu et. al. [2] have proposed Multi-Level Pooling method to extract object-aware deep image features from different layers of Convolution Neural network. Features extracted from different layers are used to create a short representative feature vector for CBIR. Domonkos Varga and Tamás Szirányi [3] have presented a supervised learning framework for CBIR which learns the probability-based semantic-level similarity and feature-level similarity simultaneously. Xinran Liu et. al. [4] have described a novel method for CBIR using combination of Convolution Neural Network and Radon Barcodes. Initially Convolution neural Network is used for Global classification of images; then Radon Barcodes are used for retrieval of similar images.Shun-Chin Chuang et. al. [5] have proposed a CBIR system using Multiple Instance Neural Network. It takes samples of related and unrelated images from user and uses color histogram of these images to train the Neural Network. Then this trained Neural Network is used to retrieve the similar images.

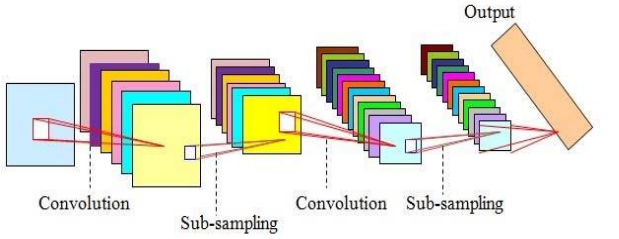
P. Muneesawang and L. Guan [6] have presented an adaptive CBIR system using Neural network. For a CBIR system, there are two major mechanisms or compo- nents which are respectively image representation for image indexing and similarity measure for database search. or image representation is expected to be discriminativeso as to distinguish images. More importantly, it is also expectedto be invariant to certain transformations. Based on image repre- sentation, the similarity measure between two images should reflect the relevance in semantics.

K. -H. Yap and K. Wu [7] have proposed an adaptive system using Fuzzy relevance feedback and Radial Bias Function Network. Users Fuzzy interpretation of image similarity is used to update the parameters of Radian Bias Function network in order to improve the CBIR performance.

Samy Sadek et. al. [8] have described the design of CBIR system using Cubic Splines Neural Network. With the help of Cubic Splines Neural Network, CBIR system is able to learn the non-linear relationship among the images and improvesthe performance retrieval performance.

III. CONVOLUTION NEURAL NETWORK

Figure 1 shows the general structure of Convolution Neural Network (CNN). It is a special type of Multilayer Feed-Forward Neural Network designed to recognize visual patterns from image pixels. It can recognize patterns with extreme variability



Fig, 1.General Structure of Convolution Neural network

Each convolution layer of a CNN is composed of multiplefeature maps. Feature maps are in the form of a plane and all the neurons of feature map are constrained to share the same set of synaptic weights. Each neuron in CNN takes inputs from a receptive field in the previous layer, which enables it toextract local features.

Convolution layer is followed by a sub-sampling layer. This layer performs local averaging and sub-sampling of poolif image pixels, which in-turn reduces the resolution of featuremap. By reducing the spatial resolution of the feature map, certain degree of shift and distortion invariance is achieved.

IV. SYSTEM ARCHITECTURE

Figure 2 shows the process of training the Deep CNN model for CBIR. Images present in the image dataset are clustered into 'K' number groups using K-NN clustering algorithm and then given as a input to the Deep CNN model to train it. After training, features are extracted from trained CNN model to represent the images in training dataset.

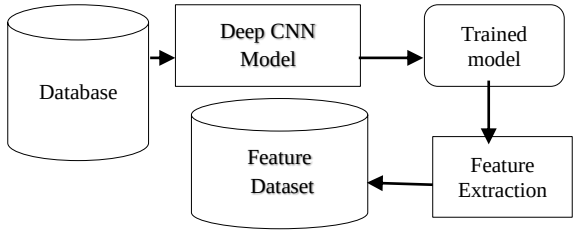


Fig. 2: Training Deep CNN Model for CBIR

Table 1 shows the internal structure of Deep CNN model usedto train the CBIR system.

TABLE I. DETAILS OF CONVOLUTION NEURAL NETWORK

Layer Description	Output Shape
Input	(128, 128, 3)
Convolution (3*3*32)	(128, 128, 32)
Convolution (3*3*32)	(128, 128, 32)
Max Pool (2*2)	(64, 64, 32)
Convolution (3*3*32)	(64, 64, 32)
Convolution (3*3*32)	(64, 64, 32)
Max Pool (2*2)	(32, 32, 32)
AFFINE (512 Units)	(512, 1)
AFFINE (512 Units)	(512, 1)

Features of query image are extracted using trained CNN model. These features are compared with features of available images using Euclidian distance measure to discover similar images.

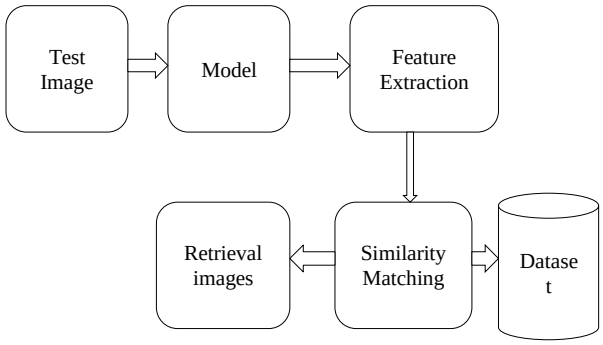


Fig. 3. Process to retrieve Similar Images

V. EXPERIMENTAL RESULTS

Wang 1000 image dataset [13, 14] is used to perform the experiments. It contains images belonging to various categories like Tribal persons, Sea shores, buses, manmade structures, flowers, etc... Experiments are performed using machine having 8 GB RAM and 2.30 GHz Quad-core processor. The images are converted into YCbCr, RGB color space and Gray. Experiments are conducted on three types of images by grouping them in 20 and 30 clusters. The following section describes the results obtained for all the six cases. Feature-based approaches improve retrieval performance by extracting more discriminative and powerful features. Early methods mainly used conventional features. Shao et al. combined color and texture features to improve the performance of RSIR. Color-Texture-Structure-Spectral Speeded up Robust Fea- tures (CTSS-SURF) is a novel local representation for remote sensing image. It is achieved by dividing images into several parts and then designing regional feature vectors. In this way, it can effectively overcome the challenges of RSIR, such as scale, illumination, shift, and rotation variation. Some methods begin to use CNN features through deep learning techniques. Li et al. combined the deep features and conventional features to represent remote sensing images, then use collaborative affinity metric fusion to get

retrieval results.

One critical direction of CBIR in future research is to collect more and larger datasets. Deep learning techniques are data- driven. In general, as long as there emerge new and large scale datasets, we can train the good deep neural network models to refresh the retrieval accuracy and solve the database search prob- lems. However, in the training process, the over-fitting problems may hinder the breakthrough of the learning algorithm. So, more and larger datasets are necessary and valuable. On the other hand, one of the most famous local features is SIFT, which mainly involves two steps: interest point detection and local region description. In recent years, many local feature extraction methods are the extensions of SIFT. For example, Zhou et al. developed the binary signature of the SIFT descriptor with two median thresholds determined by the original descriptor itself. Moreover, a new indexing scheme BSIFT for CBIR is established with this binary SIFT Furthermore, on the basis of SIFT, the edge is also added into the feature descriptor to establish Edge- SIFT and so on. Apart from the feature extraction methods of image key points like SIFT, some local features methods extract the features on the dense grids, possibly at multiple scales inde- pendently of the image content. In fact, a variety of local descriptors have been developed in recent years. Since these method have different kinds of superiority as claimed, it is rather difficult to select the best one for a retrieval task. Nevertheless, Madeo et al. made a comparative analysis among some typical local descriptors from three aspects: speed, compactness, and discrimination.

When the sample image changes greatly with a large set of some kind local features through the dataset, it is often necessary to aggregate these local features into a vector representation with a fixed length for the subsequent database search via the similarity comparison of a query against all the database images. Most of these aggregation schemes need to make a clustering analysis on these local features to obtain a codebook of the centers of the obtained clusters. According to this codebook, the original feature vector can be aggregated in different ways.

Sivic et al. pro- posed the Bag-of-Words (BoW) method which uses the k-means algorithm to create the codebook. Then the clustering center near- est to the feature point is used to replace the feature point. Usually, this aggregation scheme can lose certain detailed information and the generated BoW vector is very sparse. Perronnin et al. fur- ther proposed the Fisher Vector (FV) which aggregates local descriptors using the Gaussian Mixture Model(GMM). Actually, GMM can be used for clustering analysis, and it considers the dis- tance from the feature point to each cluster center directly. In the FV method, each feature point is represented by a linear combina- tion of all cluster centers. And, this aggregation scheme also loses some information in the process of GMM modeling. Based on BOW and FV schemes, Jegou et al. proposed Vector of Locally Aggregated Descriptors (VLAD) scheme. On one hand, like BOW, VLAD only considers the cluster center closest to the feature point, and saves the distance from each feature point to the cluster center closest to it. On the other hand, like FV, VLAD considers the value of each dimension of the feature point that has a more detailed description of the local information of the image. More importantly, the VLAD feature has no loss of information. Some other works [24–28] are the improvement or extension of VLAD, which have been demonstrated by many experiments.

For the general instance retrieval, more and larger instance datasets can make the search applicable to many search purposes. If the CNN in a CBIR system is trained with a larger dataset which combines the large person re-identification dataset and large e-commerce product retrieval dataset, it may be able to efficiently apply to both clothing search and commodity search. For various specialized instance retrieval, more and larger datasets are also crucial to the performance of retrieval. The new state-of-art methods can be only established with the larger scale and richer forms of datasets. To effectively use the dataset, the label of the new dataset should be accurate enough to eliminate some ambiguity problems in the rel- evance of image content, such as commodity icon data.

Case 1: Color Model: YCbCr, Number of Clusters: 20

Figure 4 shows the experimental results obtained for case 1. Red curve indicates average precision and green curve indicates average recall obtained at different threshold values. It is observed that the crossover point is at 0.37.

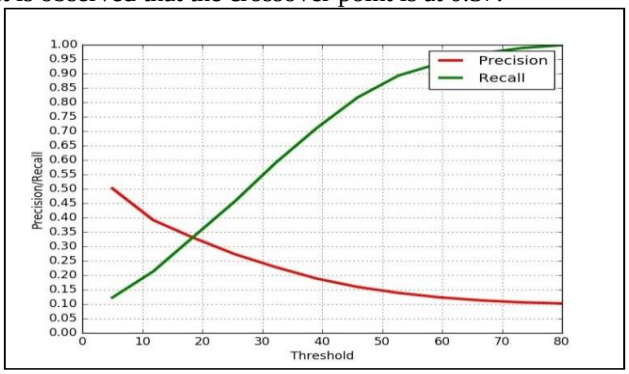


Fig. 4. Precision vs Recall Graph (Color Model: YCbCr, Clusters: 20)

Case 2: Color Model: YCbCr, Number of Clusters: 30

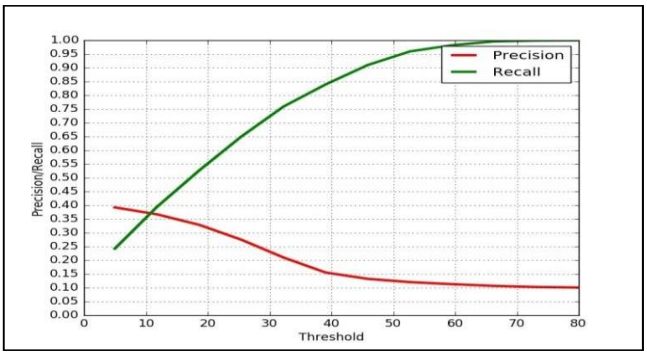


Fig. 5. Precision vs Recall Graph (Color Model: YCbCr, Clusters: 30)

Figure 5 presents the experimental results for case 2. Red and Green curve indicates average precision and average recall obtained at different threshold values. It can be observed that

the crossover point is obtained at 0.365.

Case 3: Color Model: RGB, Number of Clusters: 20 Experimental results obtained for case 3 are shown in Figure 6. Average precision obtained is shown using red curve and average recall obtained is shown using green curve. Crossoverof precision and recall is observed at 0.37.

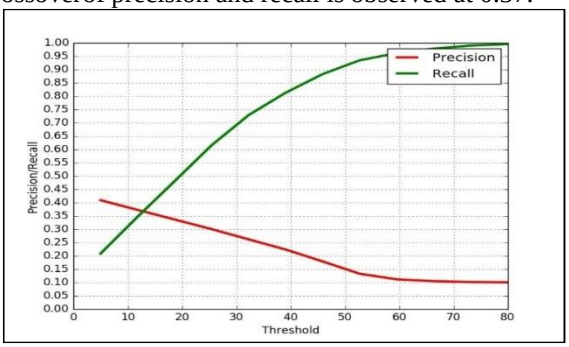


Fig. 6. Precision vs Recall Graph (Color Model: RGB, Clusters: 20)

Case 4: Color Model: RGB, Number of Clusters: 30

Experimental results for case 4 are presented in figure 7. Average precision and average recall obtained at different threshold values is shown using red curve, green curve respectively. Precision-recall crossover is observed at 0.35.

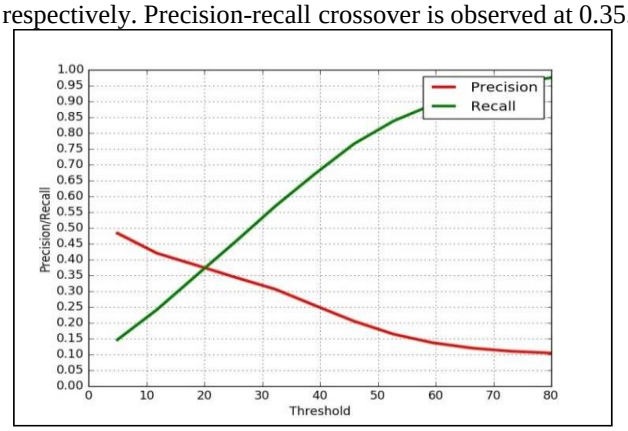


Fig. 7. Precision vs Recall Graph (Color Model: RGB, Clusters: 30)

Case 5: Color Model: Gray, Number of Clusters: 20

Figure 8 shows the experimental results obtained for case 5. Red curve indicates average precision and green curve indicates average recall obtained at different threshold values. It is observed that the crossover point is at0.34.

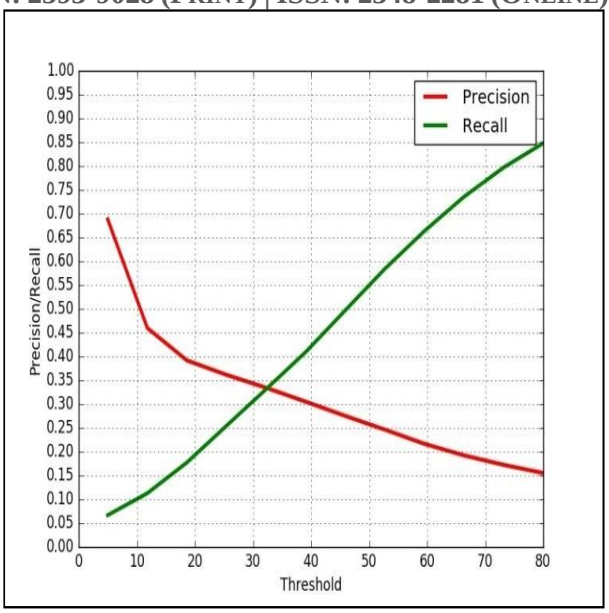


Fig. 8. Precision vs Recall Graph (Color Model: Gray, Clusters: 20)

Case 6: Color Model: Gray, Number of Clusters: 30

Experimental results obtained for case 6 are shown in Figure 9. Red and Green curve indicates average precision and average recall obtained at different threshold values. It can be observed that the crossover point is obtained at 0.37.

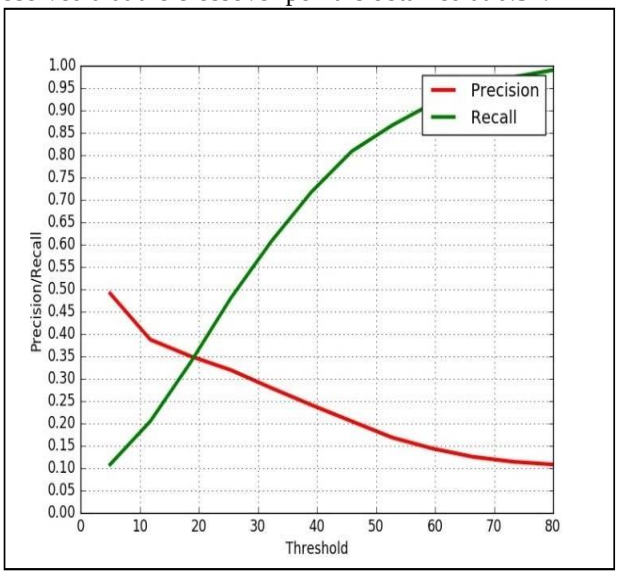


Fig. 9. Precision vs Recall Graph (Color Model: Gray, Clusters:

30)TABLE II. SUMMARY OF EXPERIMENTAL RESULTS

Sr. No.	Color Model	Number of clusters	Crossover point
1	RGB	20	0.37
2	RGB	30	0.35
3	YCbCr	20	0.37
4	YCbCr	30	0.365
5	Gray	20	0.34
6	Gray	30	0.37

Table II summarizes the experimental results obtained for Gray Images, RGB color space, YCbCr color space with different

image clusters. The result shows that, more number of clusters are required for Gray images to obtain the same crossover point as compared to RGB and YCbCr color spaces.

VI. CONCLUSION

This paper presents the CBIR system using Deep CNN. Content-Based Image Retrieval from technological and practical applications in the last decade. First, we review the developments of image representation (or feature extraction) and database search for CBIR. We then present the typical practical applications of CBIR on fashion image retrieval, person re-identification, e- commerce product retrieval, remote sensing image retrieval and trademark label image retrieval, respectively. Finally, we discuss the challenges and potential research directions in the future with the emergence of big data and the utilization of deep learning tech- niques. Wang 1000 image dataset is used to perform the experiments. For RGB, YCbCr color space and Gray images precision-recall crossover point is obtained at 0.37 for different number of clusters. Further improvement in performance can be verified by increase in number of Convolution layers.

REFERENCES

[1] Ruigang Fu, Biao Li, Yinghui Gao, Ping Wang, "Content-based image retrieval based on CNN and SVM", 2nd IEEE International Conference on Computer and Communications (ICCC), 2016, pp. 638-642

[2] Kun Hu, Yuan Dong, Hongliang Bai, "Multi-level convolutional channel features for content-based image retrieval ", Visual Communications and Image Processing (VCIP), 2016, pp. 1-4

[3] Domonkos Varga and Tamás Szirányi, "Fast content-based image retrieval using convolutional neural network and hash function ", IEEE International Conference on Systems, Man, and Cybernetics (SMC),2016 , pp. 26-36

[4] Xinran Liu, H.R. Tizhoosh, J. Kofman, "Generating binary tags for fast medical image retrieval based on convolutional nets and Radon Transform", International Joint Conference on Neural Networks (IJCNN), 2016

[5] Shun-Chin Chuang, Yeong-Yuh Xum Hsin Chia Fu and Hsiang-Cheh Huang, "A Multiple-Instance Neural Networks based Image Content Retrieval System", First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC'06), pp. 412- 415

[6] P. Muneesawang and L. Guan, " A neural network approach for learning image similarity in adaptive CBIR", IEEE Fourth Workshop on Multimedia Signal Processing 2001, pp. 257-262

[7] K.-H. Yap and K. Wu, "Fuzzy relevance feedback in content-based image retrieval systems using radial basis function network", IEEE International Conference on Multimedia and Expo, 20015

[8] Samy Sadek, Ayoub Al-Hamadi, Bernd Michaelis, and Usama Sayed, "Image Retrieval Using Cubic Splines Neural Networks", International Journal of Video & Image Processing and Network Security, Vol: 9 No: 10, pp. 17-22

[9] Syed Hamad Shirazi, Noor ul Amin Khan, Arif Iqbal Umar, Muhammad Imran Razzak, Saeeda Naz, Bandar AlHaqbani, "Content-Based Image Retrieval Using Texture Color Shape and Region", International Journal of Advanced Computer Science and Applications, Vol. 7, No. 1, pp. 418-426

[10] Miss. Aboli W. Hole1, Prof. Prabhakar L. Ramteke, "International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 10", pp. 45-49

[11] Ms. Pragati Ashok Deole,, Prof. Rushi Longadge, "Content Based Image Retrieval using Color Feature Extraction with KNN Classification", , International Journal of Computer Science and Mobile Computing, Vol.3 Issue.5, pp. 1274-1280

[12] Rehan Ashraf Khalid Bashir, Aun Irtaza and Muhammad Tariq Mahmood, "Content Based Image Retrieval Using Embedded Neural

[13] Y. Li, L. Shapiro, J.A. Bilmes, A generative/discriminative learning algorithm for image classification, in: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 2, IEEE, 2005, pp. 1605–1612.

[14] A.P. Berman, L.G. Shapiro, A flexible image database system for content-based retrieval, Comput. Vis. Image Underst. 75 (1–2) (1999) 175–195