Improving CBIR Accuracy using Convolutional Neural Network for Feature Extraction

Amjad Shah[1], Rashid Naseem[1], Sadia[2], Shahid Iqbal[1], and Muhammad Arif Shah[1]

[1] Department Of Computer Science, City University of Science and Information Technology, Peshawar, Pakistan
[2] Lecturer of Computer Science at Higher Education Department, Khyber Pakhtunkhwa, Pakistan
amjads99@gmail.com, rashid@cusit.edu.pk, sadia.rehman71@gmail.com, shahidiqbal930@gmail.com

***Abstract*—Content Based Image Retrieval (CBIR) becomes a very challenging taks due to the rapid growth in multimedia content and its visual complexity. From query by image to retrieval of relevant images, CBIR has different phases. However, features extraction of images is one of the important phases. Recently Convolutional Neural Network (CNN) shows good results in the field of computer vision due to the ability of extraction features from the images. This paper introduces CNN for features extraction from images, in CBIR system. Euclidean distance is used for association among query and stored images using the extracted features. Performance of the proposed work is evaluated using precision. The proposed work shows improved results as compared to the existing works.**

keywords: Content Based Image Retrieval; Convolutional Neural Networks; Feature Extraction

I. Introduction

The multimedia content plays an important role in wide range of area's like investigation, medical care, and social networking etc. This yields an urgent need for developing a very competent retrieval systems to entertain human needs. The multimedia database contains a huge amount of information of different types such as texts, audios, images and videos. CBIR is one of the most challenging research areas in last decade due to the visual complexity of images and large size of the image databases. A human can understand and interpret image content while machine can not. There is a huge gap between human perception and machine description also known as a semantic gap. Because of the semantic gap, it is an exigent task for CBIR to access multimedia databases. Studies have been conducted to reduce the semantic gap. Various techniques have been developed to minimize the semantic gap between human high-level perception and machine low-level description.

Few of the existing methods are used in the extraction of visual features while some focus on detecting objects in an image and interconnection of different objects in the image. Feature extraction is one of the most important phase in CBIR. Features play a vital role in accuracy for retrieval images. CBIR is based on the visual features of i.e color, texture and shape. The recent study in this area focuses on more powerful features to retrieve images according to user need. This paper introduces Convolutional Neural Network (CNN) for features extraction from the images.

The rest of this paper is organized as: Section II discusses the related literature and background study. Section III presents the proposed methodology while experimental results are shown in Section IV. Addition to that, results are analysed in Section V. Section VI concludes this paper.

II. Related Work

CBIR is an approach which employ visual contents query by user, to search relevant images from large scale image databases. Since 1990 it has been an active and fast progressive research field. During the past decade, notable development has been performed among both theoretical research and technical development to highlight and solve the issues. However, there continue deep taxing research problems with appeal to researchers beyond a couple of disciplines. CBIR is a set of phases/methods for retrieving relevant images from a large scale image database based on automatically extracted image features. Feature extraction is a very important phase of CBIR system because it represents an image in a form that computer can manipulate. There are two categories of extracted features global (color, texture and shape) and local (corner and edges) [1].

A. Local Features

There are many local features descriptors that describe local information like region or segments or corners in the image. To extract these local feature -there exist many algorithms like Scale-Invariant Feature Transformation (SIFT). SIFT is a very popular local feature extraction algorithm introduced in last decade by Lowe [2]. SIFT is invariant to scale and rotation but it has high dimensional at matching. To overcome the high dimensionality matching problem Speeded Up Robust Feature (SURF) has been introduced by Herbert [3]. SURF is inspired from SIFT and the authors claim that SURF is faster. However, SURF give poor performance on rotational invariance. Another algorithm for local feature is called Histogram of Oriented Gradients (HOG) introduced by Dalal and Trigg [4] to provide better performance as compared to existing local descriptors. HOG can categorize the object appearance and shape. As from the previous review, we found that no standalone global features or local features are enough to describe the image. We need such features that can describe all visual aspects of the image completely.

B. Global Features

There are many feature extraction algorithms that extract global features from images including color, texture, and shape. Color is the most prominent and attractive feature of an image. It has a close relation to objects, foregrounds, and backgrounds. The common color representations are color histogram, color moments [5], color correlogram [6] and color co-occurrence matrix [7]. There are two categories of color spaces: linear color spaces (CMY, YIQ, RGB, XYZ, and YUV) and Non-Linear color spaces (L*a*b, HSV, Ng) [8]. Color features in CBIR are very popular but are not sufficient to describe an image completely [9]. Limitations with color feature descriptors are: lack of perceptual similarities and spatial information. Texture has no proper definition but according to Alzubi [1], what is left after considering color and shapes is known as Texture. Texture feature is very beneficial in the image for CBIR, however, it has some limitations i.e., complex to compute, accuracy and noise sensitivity. Many texture based CBIR systems have been proposed to improve the accuracy of CBIR. Some common feature texture algorithms are Markov Random Field (MRF) [10], Edge Histogram Descriptor (EHD) [1], Steerable Pyramid Decomposition (SPD) [11], Gray Level Co-occurrence Matrix (GLCM) [12]. Shape is another global feature descriptor, many researchers combine shape with color or texture to improve CBIR System. Many algorithms have been proposed to extract shape features from image such as Multi-Texton Histogram (MTH) [13], Curvature Scale Space (CSS) [14], Fourier Descriptors [15]. Shape feature descriptor is sensitive to translation, scaling, rotation invariance and stability. Many researchers have proposed different CBIR systems by combining features texture, color, and shape to achieve higher accuracy and efficiency [1].

III. Proposed Methodology

In this paper we propose deep learning algorithm called CNN for feature extraction in order to achieve better retrieval results for CBIR system. Figure 1 shows the proposed framework of CBIR using CNN as feature extractor.

A. CNN for Feature Extraction

The proposed work uses CNN as a feature extractor instead of any conventional feature extractor. CNN becomes a valuable research topic in the field of machine learning and computer vision. So the basic purpose of using CNN as a feature extractor is to compare CNN based CBIR system with conventional CBIR system in order to find if it's better in any way. Beside image classification, CNN is also beneficial for object detection task. In this work we are utilizing CNN's Alex Net architecture for feature extraction [16]. The Alex Net has eight trained layers. The first five are convolutional layers while remaining three are the fully connected layers. This work utilizes the 7th layer of the architecture for feature extraction with the 4096 dimensions/features per image. The process of CNN's Alex Net starts over the image dataset for feature extraction. Then stores the extracted features in a features database. A user then query an image in order to get

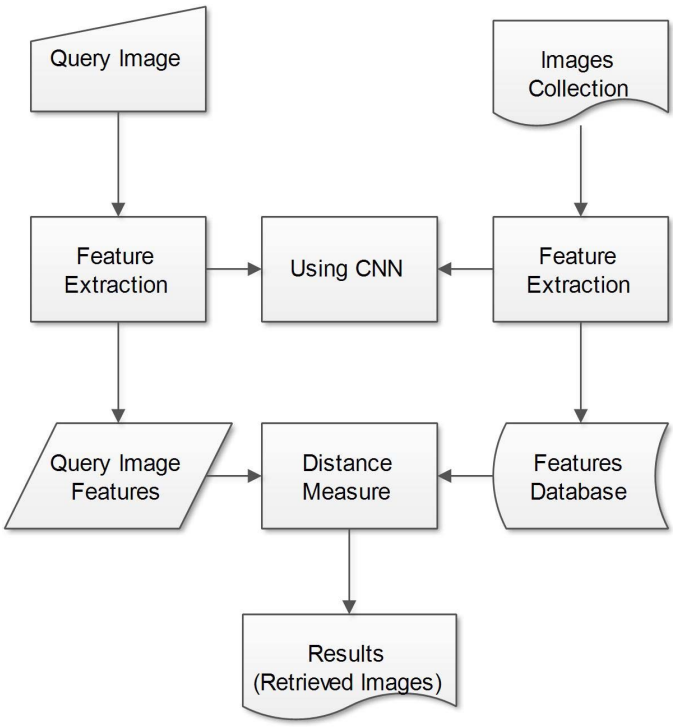

Fig. 1. Proposed CBIR Pipeline View

the results based on the similarities of the extracted features by using a distance formula which ranks the result in descending order to get the top most relevant results.

B. Distance Measure

The distance measure between the query image and the dataset is measured using Euclidean distance. Query image features is compared with dataset features using Euclidean distance (see Equation 1). A set of relevant images is retrieved then arranged them in descending order of their distance score computed by Euclidean distance. In Equation 1 *(x,y)* are the two dimensions of an image while *(a,b)* are the dimensions of another image.

$$dist((x,y),(a,b)) = \sqrt{(x-a)^2 + (y-b)^2} \qquad (1)$$

C. Image Datasets

The performance of the proposed work is examined using Corel [17], Catech-101 [18] and Li Photography [19] datasets. Corel image dataset contains 1000 images with ten different categories like mountains, african people, flowers, beach, food, and dinosaurs. Similarly, Li Photography dataset contains different categories. For the proposed work, we have selected 183 images and categorized into three types i.e., flowers, trees, and brushes. Caltech 101 dataset has 101 categories. For the proposed work, we have selected six categories butterflies, dolphins, crocodiles, water lily, faces and wild cats. The dimensions of all images are converted into 227 x 227 for feature extraction using CNN. Table I shows the summary of

Fig. 2. Example images from each category of Corel dataset

image sets used in this work. Figures 2, 3 and 4 are showing images from each category of corel, caltech and li dataset respectively.

TABLE I
STATISTICS OF DATASETS

Data Set	Number of Categoires	Number of Images
Corel	10	1000
Li Photography	3	183
Caltech	6	712

IV. EXPERIMENTS

Experiments were conducted to evaluate the performance of the proposed work. All the experiments were implemented in MATLAB 2016a running on personal system core i5 with 4 GB RAM while for feature extraction core i7 with Nvidia 4.0 was used. Figure 5 shows Caltech-101 dataset retrieved images with respect to the query image. Figure 6 shows Coral dataset retrieved images with respect to the query image. Figure 7 shows the Li photography most similar images with respect to the query image.

V. PERFORMANCE EVOLUTION

Performance of the proposed work is evaluated and measured against the existing system [21]. The proposed system retrieves the desired set of similar images from the dataset based on the score of Euclidean distance. Performance of the proposed work is measured using precision and recall. Precision shows the effectiveness of the system and recall

Fig. 3. Example images from each category of Caltech dataset

Fig. 4. Example images from each category images of Li dataset

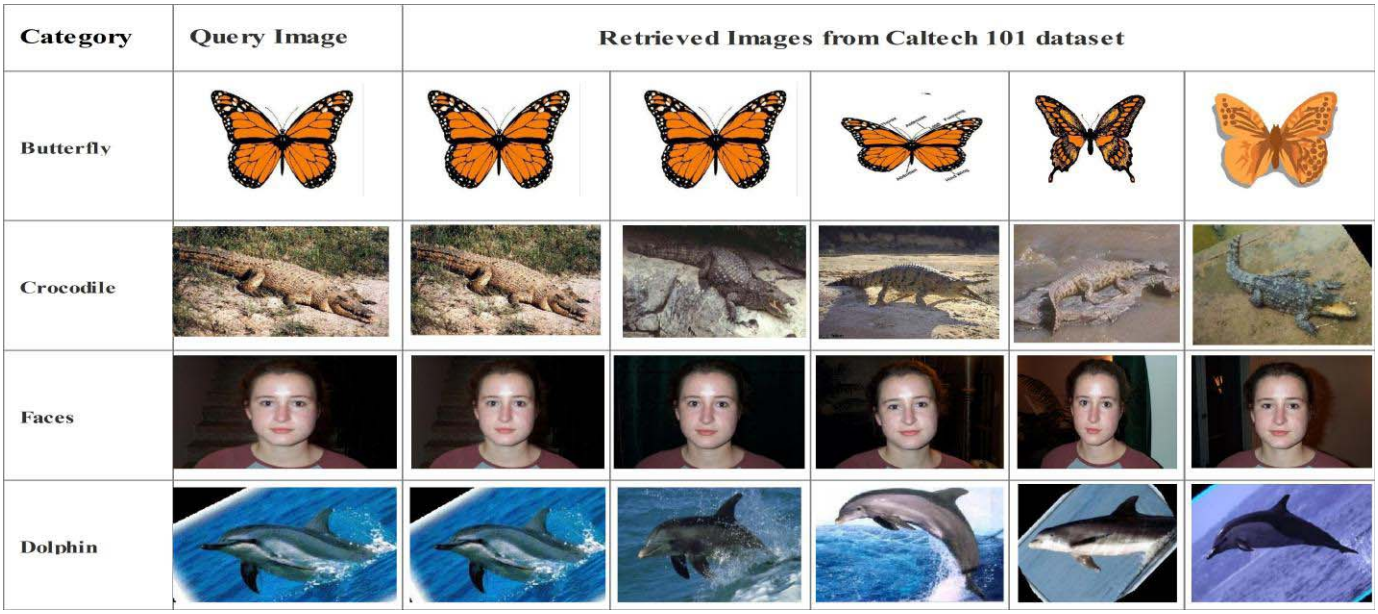

Fig. 5. Query images and the desired set of retrieved images from Caltech dataset

TABLE II
EXPERIMENTAL RESULTS FOR THE PROPOSED WORK AND THE EXISTING WORK USING PRECISION

Method/Cat	People	Beach	Monuments	Buses	Dinosaur	Flowers	Elephant	Horse	Mountain	Food	Ave
Yu [17]	0.849	0.356	0.616	0.818	1.000	0.931	0.591	0.928	0.404	0.682	0.717
Lin [19]	0.683	0.540	0.562	0.888	0.993	0.891	0.658	0.803	0.522	0.733	0.727
Guo [20]	0.847	0.466	0.682	0.885	0.992	0.733	0.964	0.939	0.474	0.806	0.779
Anandh [21]	0.767	0.836	0.752	0.879	1.000	0.944	0.727	0.843	0.645	0.638	0.803
Chiang [22]	0.060	0.930	0.260	0.070	1.000	0.880	0.680	0.260	0.260	0.930	0.533
Proposed Work	**0.882**	**0.865**	**0.959**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.970**

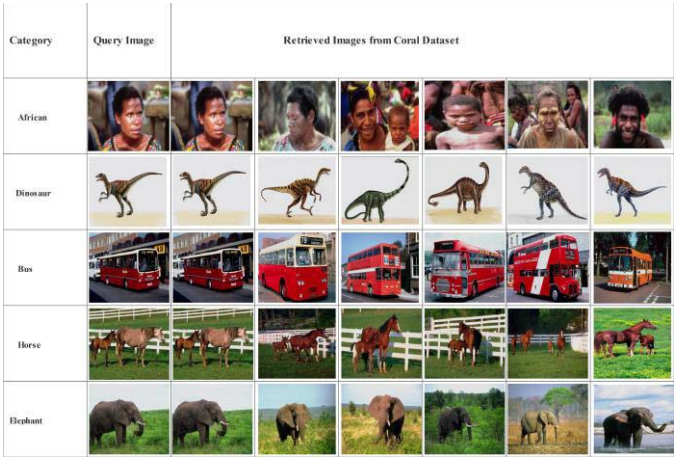

Fig. 6. Query images and the desired set of retrieved images from Corel dataset

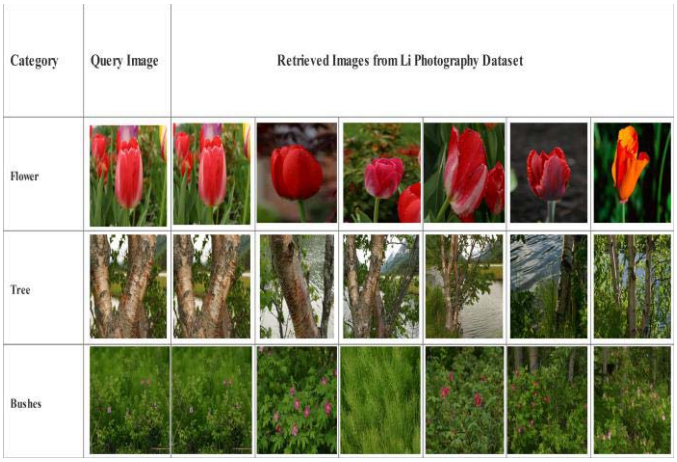

Fig. 7. Query images and the desired set of retrieved images from Li dataset

shows the accuracy of the system. Precision and recall can computed as:

$$Precession = \frac{\text{No of retrieved relevant images}}{\text{No of relevant images}}$$

$$Recall = \frac{\text{No of retrieved relevant images}}{\text{No of retrieved images}}$$

Results present that the proposed work achieves higher accuracy as compared to existing retrieval system [21]. Table II

lists the comparison of proposed work with the existing work in terms of precision using Corel dataset. Table III shows the Accuracy rate of Caltech 101 dataset with a comparison to existing work. Table IV shows the accuracy of Corel dataset with a comparison to existing work. While Table V shows the accuracy of Li Photography dataset and comparison with existing work.

TABLE III
CATEGORY WISE COMPARISON OF PROPOSED WORK WITH EXISTING WORK IN TERMS OF PRECISION USING CALTECH DATASET

Category	Existing Work	Proposed Work
Butterfly	66%	93%
Crocodile	80%	88%
Dolphin	58%	97%
Faces	70%	100%
Water Lilly	94%	100%
Wild Cat	50%	88%

TABLE IV
CATEGORY WISE COMPARISON OF PROPOSED WORK WITH EXISTING WORK IN TERMS OF PRECISION USING COREL DATASET

Category	Existing Work	Proposed Work
Food	70%	94%
Mountain	64%	86%
Horse	94%	98%
Dinosaurs	100%	100%
Buses	90%	100%
Elephant	70%	94%
Flowers	96%	100%
Monuments	82%	86%
Beach	80%	84%
African People	84 %	96%

TABLE V
CATEGORY WISE COMPARISON OF PROPOSED WORK WITH EXISTING WORK IN TERMS OF PRECISION USING LI DATASET

Category	Existing Work	Proposed Work
Flowers	90%	87%
Trees	88%	96%
Bushes	86%	89%

VI. CONCLUSION

This paper presents an improved CBIR system is proposed by using CNN Alax Net architecture for features extraction. CNN extracts 4096 features per image. The experiments conducted on different image datasets and comparison of proposed CBIR system with existing work revealed that proposed work results in higher accuracy in terms of precision and accuracy rate i.e., 95% for Corel, 97% for Caltech 101 and 88% for Li Photography datasets. We believe that this study initiates studies using CNN for feature extraction in CBIR system and can be a basis to extend to advanced CBIR approaches in future; we suggest to reduce the number of features extracts by CNN Alax Net architecture. Using of classification and clustering techniques is the prominent direction for research to improve the accuracy of CBIR system.

REFERENCES

[1] A. Alzu'bi, A. Amira, and N. Ramzan, "Semantic content-based image retrieval: A comprehensive study," *Journal of Visual Communication and Image Representation*, vol. 32, no. July, pp. 20–54, 2015.

[2] D. G. Lowe, "Distinctive image features from scale invariant keypoints," *Int'l Journal of Computer Vision*, vol. 60, pp. 91–11 020 042, 2004.

[3] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3951 LNCS, pp. 404–417, 2006.

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, vol. I, pp. 886–893, 2005.

[5] X. D. X. Duanmu, "Image Retrieval Using Color Moment Invariant," *Information Technology: New Generations (ITNG), 2010 Seventh International Conference on*, pp. 200–203, 2010.

[6] Jing Huang, S. Kumar, M. Mitra, Wei-Jing Zhu, and R. Zabih, "Image indexing using color correlograms," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 191, no. 3-4, pp. 762–768, 1994.

[7] V. Kovalev and M. Petrou, "Multidimensional Co-occurrence Matrices for Object Recognition and Matching," *Graphical Models and Image Processing*, vol. 58, no. 3, pp. 187–197, 1996.

[8] X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, "Content-based image retrieval by integrating color and texture features," *Multimedia Tools and Applications*, vol. 68, no. 3, pp. 545–569, 2012.

[9] J. W. Z. Zhang, "Content-Based Image Retrieval using color and edge direction features," *2010 2nd International Conference on Advanced Computer Control*, pp. 459–462, 2010.

[10] G. R. Cross and A. K. Jain, "Markov Random Field Texture Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 1, pp. 25–39, 1983.

[11] E. P. Simoncelli and W. T. Freeman, "The Steerable Pyramid : A Flexible Architecture For Multi-Scale Derivative Computation Eero P Simoncelli GRASP Laboratory , Room 335C 3401 Walnut St Philadelphia , PA 19104-6228 William T Freeman Mitsubishi Electric Research Laboratories Cambridge , MA 02," *Image (Rochester, N.Y.)*, vol. III, pp. 444–447, 1995.

[12] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[13] G. H. Liu, L. Zhang, Y. K. Hou, Z. Y. Li, and J. Y. Yang, "Image retrieval based on multi-texton histogram," *Pattern Recognition*, vol. 43, no. 7, pp. 2380–2389, 2010.

[14] S. Abbasi, F. Mokhtarian, and J. Kittler, "Curvature scale space image in shape similarity retrieval," *Multimedia Systems*, vol. 7, no. 6, pp. 467–476, 1999.

[15] D. Zhang and G. Lu, "A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures," *International Conference on Intelligent Multimedia and Distance Education*, vol. 1, pp. 1–9, 2001.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.

[17] http://wang.ist.psu.edu/docs/related/.

[18] http://www.vision.caltech.edu/Image_Datasets/Caltech101/.

[19] http://sites.stat.psu.edu/ jiali/index.download.html.

[20] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, ser. CVPR '97. Washington, DC, USA: IEEE Computer Society, 1997, pp. 762–.

[21] A. A. Associate, "Content Based Image Retrieval System based on Semantic Information Using Color , Texture and Shape Features," in *Computing Technologies and Intelligent Data Engineering (ICCTIDE), International Conference on*. Kovilpatti, India: IEEE, 2016.

[22] T.-W. T. T.W Chiang, "Content-base image retrieval using multiresolution color and texture feature," in *J inf Technol Appl (CVPR '97)*, vol. 1. IEEE Computer Society, 2006.