Enhancing Content Based Image Retrieval Systems by using CNN as Feature Extractor

Amjad Shah[1], Yasir ali[1],

[1]*Higher Education Department, Khyber Pakhtunkhwa, Pakistan*

amjads99@gmail.com

Abstract

The Contents Based Image Retrieval (CBIR) has become a daunting task due to significant evolution in multimedia contents and its related visual complexity. CBIR encompasses various phases from query to the retrieval of images but the most significant phase is feature extraction. In a bid to enhance the accuracy of CBIR, numerous studies have been performed but the output of Convolution Neural Networks (CNN) is amazing in the domain of computer vision. CNN has the potential of extracting features and it can be applied as classifier as well. In our proposed work, the CNN has been applied as feature extractor in place of traditional feature extractors. Similarity measurement was calculated with the help of Euclidian distance formula, precision and recall were used for measuring the performance of the system. In this paper it has been discussed that how our proposed work produces better results as compared to the previous works.

Key words: Convolutional Neural Networks, Content Based Image Retrieval, Feature extraction

1. Introduction

The contents are multimedia are important for different fields such as health-care, social networking, investigation etc. For this purpose, it is indispensable to develop a smart retrieval systems in order to address the human needs. The data base of multimedia is composed of a bulk of information such as video, audio, text, images and other graphical data. In last decade CBIR has become the most stimulating research domain due to images complexity and huge size of image databases. A human is able to understand and interpret the contents of image but machine fails to do so. There exists a huge semantic gap between human's perception and machine interpretation. This semantic gap has led to-wards the challenging task of accessing multimedia database via CBIR. In this regard different studies have been conducted to diminish this sematic gap. Similarly, various techniques have been proposed for the reducing semantic gap i.e. between human level perception and machine based perception. Among the existing methods, some are focused to extract the visual features and some of them are intended to focus upon the detection of objects and interconnection of objects in image. The feature extraction is known the important phase of CBIR due to its impacts upon the accuracy of image. CBIR is based on visual features like colour, shape and texture. A recent research in this area is to study more powerful feature to retrieve images based upon the needs of users.

**Corresponding author address: Yasir Ali, email: yasiuop007@gmail.com Mobile: 03459700648*

2. Related Work

CBIR is composed of different phases but features extraction is considered as the most important one due to its representation of image in the form that is can easily be understood by the machine. The extract is divided in to two broad categories such that one is known as global features which is described in terms of colour, texture and shape and local features are presented by edges and corners [1]. These two features are discussed in next section of this research paper.

2.1. Global Features

There are variety of feature extraction algorithms that can be used to extract global features i.e. colour, texture and shape. Among the global features, colour is important and striking component of image due to its close relationship with objects, foreground and background. Colour can be represented by colour histogram, color moments [2], colorcorrelogram [3] and color co-occurrence matrix [4]. The colour spaces can be divided into two categories: linear colour spaces (e.g., CMY, YIQ, RGB, XYZ, and YUV) and Non-Linear color spaces (e.g L*a*b, HSV, Ng) [5].In CBIR, colour features are very famous but it fails to describe the image completely [6]. Problems associated with colour feature descriptor are due to lack of perceptual similarities and spatial information. Ac-cording to Alzubi [1] the texture cannot be defined in proper way but what is left on image after removal of colour and shape. The texture is very important entity for CBIR but still there are some limitations exist in it such that noise sensitivity, complex to compute and accuracy. For enhancing contents based image retrieving, various CBIR systems have been proposed. For extraction of shape features from image, many algorithms have been proposed such as are Markov Random Field (MRF) [7], Edge Histogram Descriptor (EHD) [1], Steerable Pyra-mid Decomposition (SPD) [8], Gray Level Co-occurrence Matrix (GLCM) [9]. Shape is second global feature descriptor and many researchers mix up shape with color or texture in order to enhance CBIR system. Many algorithms have been proposed to extract shape features from image like Multi-Texton Histogram (MTH) [10], Curvature Scale Space (CSS) [11], and Fourier Descriptors [12]. Shape feature descriptor is sensitive to translation, scaling, rotation invariance and stability. Many researchers have presented various CBIR systems by mixing up features such as texture, colour and shape to achieve higher accuracy and efficiency [1].

2.2 Local Features

Like global features descriptors, there exist various local features descriptors that describe the local information like region, segments or corners in the image. The most significant local features extraction algorithms are Scale-Invariant Feature Transformation (SIFT). SIFT is a very popular local feature extraction algorithm in last decade introduced by [13]. SIFT is invariant to scale and rotation. But SIFT suffers from high dimensionality, when it comes to matching. In or-der to address the problem of high dimensionality, Speeded Up Robust Feature (SURF) algorithm was introduced by Herbert [14].

This algorithm was inspired by SIFT and it is faster result than SIFT. But SURF algorithm suffers from limitation of poor performance on rotational invariance. Dalal and Trigh [15] also presented local feature algorithm known as Histogram of Oriented Gradients (HOG), which provides better performance in comparison to current local descriptors. HOG has the ability to categorize the appearance and shape of object. From literature review, we observed that no standalone global features and local features algorithm are sufficient enough describe the image. We are in dire need of such features that can describe the entire visual aspects of the image fully.

3. Proposed Methodology

For achieving better retrieving results in CBIR system, we proposed deep learning algorithm named as Convolutional Neural Networks (CNN) for extraction of features. This was the first time that CNN was used for feature extraction in CBIR system. The CNN can be applied to both feature extraction and classification but in our framework it has been used as feature extractor. Our proposed feature extractor framework for CBIR is depicted in Figure 1. The images were collected from various sources and then these images were stored in feature data base in form of features. Then CNN algorithm was applied for extraction of features from images data set. The working of CNN has been elaborated in next section 3.1 and onward of this paper.

3.1 Using CNN features in CBIR

In the proposed work, the main focus was to use CNN as a feature extractor instead of using any other conventional feature extractor. CNN has been emerged as the most hot research topic in the domain of computer vision and machine learning. The main idea behind using CNN as feature extractor is to compare the CNN based CBIR system with conventional CBIR system and to investigate that how much it is better in all respects and any other way. CBIR not only can be used as image classi cation but it is also useful for detection of object. Our proposed frame work inspired by CNN's Alex Net architecture for feature extraction which is comped of eight trained layers [16]. In this algorithm rst ve layers are convolutional and three layers are fully connected. The proposed algorithm utilizes 7th layer of Alex Net architecture for extraction of features with a dimension of 4096 features per image. The Alex net's CNN is run over the images' data set for feature extraction.

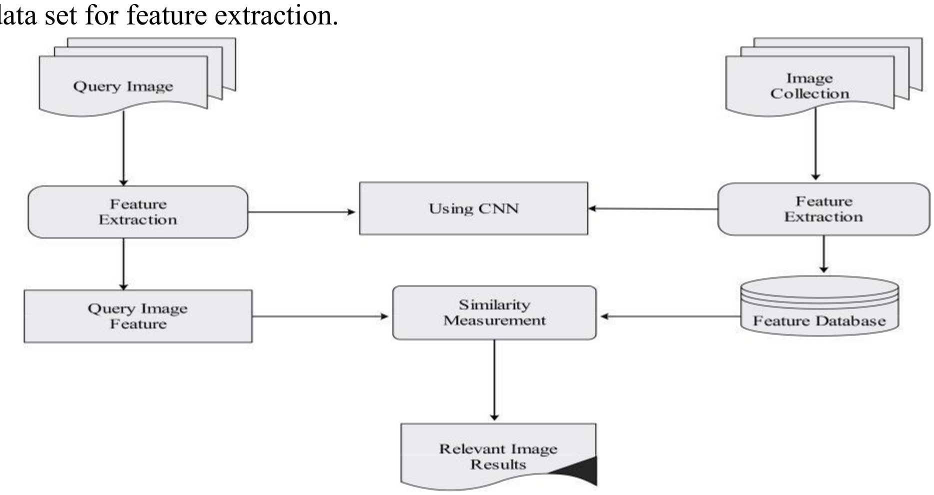

Fig. 1. Proposed CBIR framework using CNN as feature extractor

The extracted features are stored in a features database. Query image is provided for the purpose knowing about the results based upon similarities of extracted features. The similarity was measured by using Euclidian distance formula. This formula ranks result in descending order to get top most relevant results.

3.2 Similarity Measure

Euclidean distance formula was used for finding similarity measure between the query image and the data set. Query image features is compared with dataset features using Euclidean distance (see Equation 1). A set of relevant images were returned and arranged in descending order of their similarity scores computed by using Euclidean distance.

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

3.3 Image Datasets

In our proposed work, three data sets were used such as Corel [20], Catech-101[22] and Li Photography [21] datasets. The performance of proposed work was examined on these data sets. Corel image dataset is composed of 1000 images having 10 different categories such as mountains, African people, flowers, beach, food, and dinosaurs. Similarly, Li Photography dataset contains different categories. In our proposed work, we 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 selected six categories butteries, dolphins, crocodiles, water lily, faces and wild cats. The dimensions of all images were converted into 227 x 227 for feature extraction using CNN. Table 1 shows the summary of image data sets used in our proposed work. Table 1 shows the summary of image data sets used in our proposed work. Figure 2, 3 and 4 are showing images from each category of Corel, Caltech and Li photography dataset respectively.

Table 1. Statistics of datasets

Data Set	No's of Categories	No's of Images
Corel	10	1000
Li Photography	3	183
Caltech	6	712

4. Experiment

The performance of proposed work was evaluated with by experimental work and it was implemented in MATLAB 2016a on personal system possessing the processing abilities such as core i5 and memory capacity of 4 GB. For extraction of features core i7 with Nvidia 4.0 was used. Figure 5 shows Caltech-101 dataset similar images with respect to the query image. Figure 6 shows Coral dataset similar images with respect to the query image. Figure 7 shows the Li photography most similar images with respect to the query image.

5. Performance Evaluation

The performance of proposed work is evaluated and measured in comparison to the existing system [17]. The proposed system retrieved the desired set of similar images from the dataset based upon the score calculated by using Euclidean distance formula. Performance of the proposed work is measured by calculating precession and recall rate. Precision shows the e effectiveness and recall indicates the system accuracy. Precession and recall can computed by the following equations.

$$Precision = \frac{\text{No of relevant images return}}{\text{No of images return}}$$

$$Recall = \frac{\text{No of relevant images return}}{\text{Total No of relevant images in the database}}$$

Our proposed work computes classification accuracy from support vector machine classifier. The results obtained from proposed work achieves higher accuracy as compared to existing retrieval system [17]. The detail about precession of proposed work in comparison to existing works by using Coral dataset is depicted in Table 2. The accuracy rate upon Caltech 101 dataset, Corel dataset and Li Photography in comparison to existing works are displayed in 3, 4 and 5 respectively.

Fig. 2. Each category images of Corel dataset

Fig. 3. Each category images of Caltech dataset

Fig. 4. Each category images of Li dataset

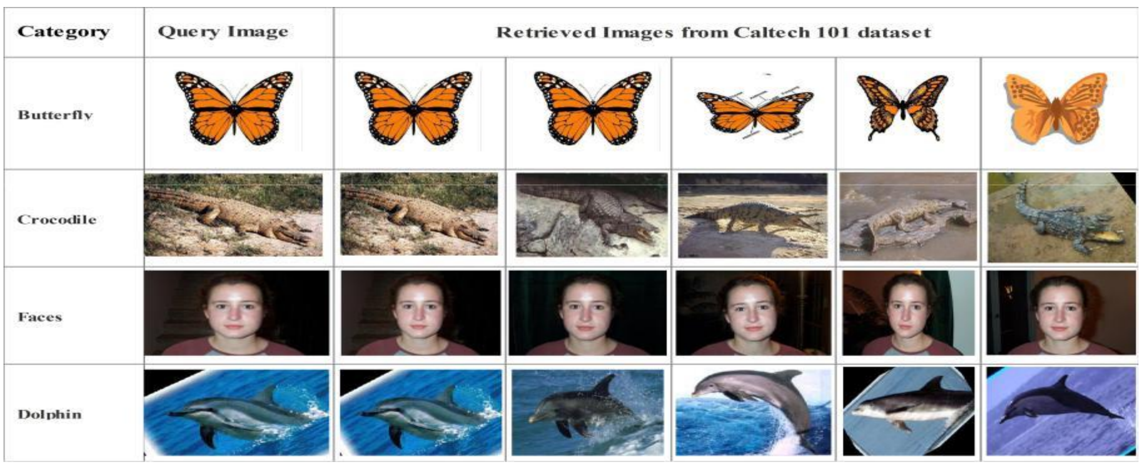

Fig. 5. Query images and the desired set of relevant images from Caltech dataset

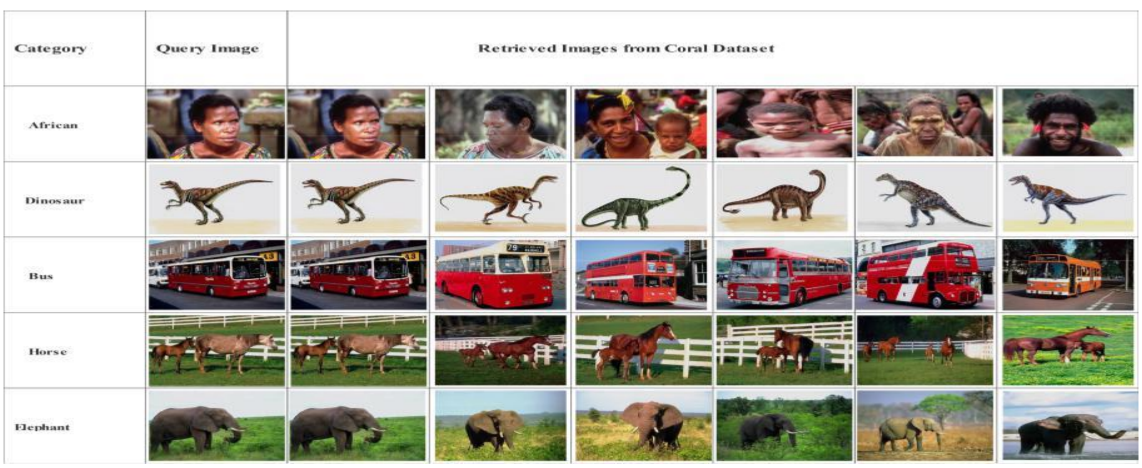

Fig. 6. Query images and the desired set of relevant images from Corel dataset

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

Table 2. Comparison among the proposed work and the existing work in terms of precision

Method/Cat	People	Beach	Monuments	Buses	Dinosaur	Flowers	Elephant	Horse	Mountain	Food	Ave
Guo [18]	0.8	0.466	0.682	0.885	0.992	0.733	0.964	0.939	0.474	0.806	0.779
Anandh [17]	0.7	0.836	0.752	0.879	1.000	0.944	0.727	0.843	0.645	0.638	0.803
Yu [20]	0.8	0.356	0.616	0.818	1.000	0.931	0.591	0.928	0.404	0.682	0.717
Chiang [19]	0.0	0.930	0.260	0.070	1.000	0.880	0.680	0.260	0.260	0.930	0.533
Lin [21]	0.6	0.540	0.562	0.888	0.993	0.891	0.658	0.803	0.522	0.733	0.727
Proposed Work	0.8	0.865	0.959	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.970

Table 3. Category wise comparison of proposed work with existing work in terms of accuracy rate using Caltech dataset

Experimental Results		
Category	**Existing Work**	**Proposed Work**
Butter y	66%	93%
Crocodile	80%	88%
Dolphin	58%	97%
Faces	70%	100%
Water Lilly	94%	100%
Wild Cat	50%	88%

Table 4. Category wise comparison of proposed work with existing work in terms of accuracy rate using Corel dataset

Experimental Results		
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 5. Category wise comparison of proposed work with existing work in terms of accuracy rate using Li dataset

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

6. Conclusion

In this paper, a proposed framework was presented based upon Alex Net architecture, which is an efficient image retrieval system for extraction of features. After performing comparative analysis it is understood that the proposed work in comparison with the existing work yields higher accuracy rates in term of precession and accuracy rate for each dataset such as 95% for Corel, 97% for Caltech 101 and 88% for Li Photography datasets.

7. Acknowledgement

The authors want to extend sincere thanks to the colleagues, who supported and shared valuable advices to complete this research work.

References

1. A. Alzubi, A. Amira, and N. Ramzan, Semantic content-based image retrieval: A comprehensive study, J. Vis. Commun. Image Represent., vol. 32, no. July, pp. 2054, 2015.
2. X. D. X. Duanmu, Image Retrieval Using Color Moment Invariant, Inf. Technol. New Gener. (ITNG), 2010 Seventh Int. Conf., pp. 200203, 2010.

3. Jing Huang, S. R. Kumar, M. Mitra, Wei-Jing Zhu, and R. Zabih, Image indexing using color correlograms, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 191, no. 34, pp. 762768, 1994.

4. V. Kovalev and M. Petrou, Multidimensional Co-occurrence Matrices for Object Recognition and Matching, Graph. Model. Image Process., vol. 58, no. 3, pp. 187197, 1996.

5. X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, Content-based image retrieval by inte-grating color and texture features, Multimed. Tools Appl., vol. 68, no. 3, pp. 545569, 2012.

6. J. W. Z. Zhang, Content-Based Image Retrieval using color and edge direction features, 2010 2nd Int. Conf. Adv. Comput. Control, pp. 459462, 2010.

7. G. R. Cross and A. K. Jain, Markov Random Field Texture Models, IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-5, no. 1, pp. 2539, 1983.

8. E. P. Simoncelli and W. T. Freeman, The Steerable Pyramid : A Flexible Architec-ture 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. 444447, 1995.

9. R. M. Haralick, Statistical and structural approaches to texture, Proc. IEEE, vol. 67, no. 5, pp. 786804, 1979.

10. G. H. Liu, L. Zhang, Y. K. Hou, Z. Y. Li, and J. Y. Yang, Image retrieval based on multi-texton histogram, Pattern Recognit., vol. 43, no. 7, pp. 23802389, 2010.

11. S. Abbasi, F. Mokhtarian, and J. Kittler, Curvature scale space image in shape similarity retrieval, Multimed. Syst., vol. 7, no. 6, pp. 467476, 1999.

12. D. Zhang and G. Lu, A Comparative Study on Shape Retrieval Using Fourier Descriptors with Di erent Shape Signatures, Int. Conf. Intell. Multimed. Distance Educ., vol. 1, pp. 19, 2001.

13. D. G. Lowe, Distinctive image features from scale invariant keypoints, Intl J. Com-put. Vis., vol. 60, pp. 9111020042, 2004.

14. H. Bay, T. Tuytelaars, and L. Van Gool, SURF: Speeded up robust features, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioin-formatics), vol. 3951 LNCS, pp. 404417, 2006.

15. N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, Proc. - 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, CVPR 2005, vol. I, pp. 886893, 2005.

16. A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet Classi cation with Deep Convolutional Neural Networks, Adv. Neural Inf. Process. Syst., pp. 19, 2012.

17. A. A. Associate, Content Based Image Retrieval System based on Semantic Infor-mation Using Color , Texture and Shape Features, in Computing Technologies and Intelligent Data Engineering (ICCTIDE), International Conference on, 2016.

18. Huang, Jing and Kumar, S. Ravi and Mitra, Mandar and Zhu, Wei-Jing and Zabih, Ramin, Image Indexing Using Color Correlograms, Proceedings of the 1997 Con-ference on Computer Vision and Pattern Recognition (CVPR '97), International Conference on, 1997.

19. T.W Chiang, T-W Tsais, Content-base image retrieval using multiresolution color and texture feature, J inf Technol Appl (CVPR '97), International Conference on, 2006.

20. http://wang.ist.psu.edu/docs/related/

21. http://sites.stat.psu.edu/ jiali/index.download.htm

22. http://www.vision.caltech.edu/ImageDatasets/Caltec