

Performance Evaluation of Content-Based Image Retrieval Using Convolutional Neural Network

A.P.Nilawar^{1*}, C.G.Dethe²

¹Department of Electronics and Communication Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India

²Director, UGC HRDC, R.T.M.N.U., Nagpur, Maharashtra, India

¹aa.nilawar@gmail.com

Abstract

In today's world due to multimedia development, there is a huge image database. Content-Based Image retrieval (CBIR) is a widely used method for image retrieval from the large image database. Existing retrieval methods are based on the basic content of an image like color, Shape, and Texture. The system based on basic features requires more time for processing and provides less accuracy. To reduce time and improving accuracy we are proposing CBIR Using CNN in this paper. CNN is used for feature extraction and for similarity measurement ANN is used. In this technique, the user has to provide an image as an input query image. The similar images related to the query image are displayed as a result. The performances of a system are evaluated by precision and mean average precision (MAP). After comparing with existing methods we found encouraging results which leads to improving accuracy.

Keywords: Content-Based Image Retrieval, CNN, DNN, LSH

1. Introduction

In this day and age, there is an expansion in the component of interactive media databases. It comprises of information like image, sound, video, and so forth. The utilization of images in the area of medicinal, individual, news coverage makes a lot of image databases. In this manner, the Content-Based Image Retrieval system created to manage huge image databases. Content-Based Image retrieval system restores a lot of images from a progression of images in the database to fulfill the user's need with closeness assessment. CBIR system is based on features like color, shape, and Texture. CBIR incorporates the extraction of features from image essentially based on its content, estimating likeness between the query image and image in the database, and positioning from generally tantamount to the assorted image. CBIR system has two essential stages in the main stage the pre-handling for image database alluded to as image highlight extraction basically based on the sort of features like color, shape, and Texture. What's more, in the second stage estimation of likeness from database image alluding query image gave by means of the user.

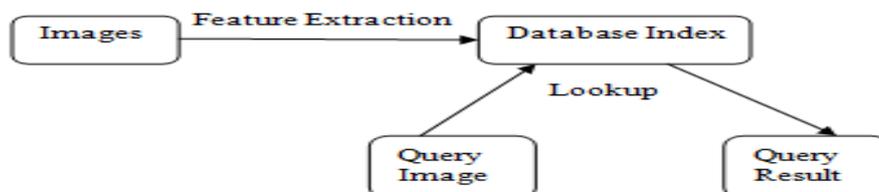


Figure 1. Basic CBIR System

The above figure explains the basic operation of the CBIR System. In the retrieval system, there are essentially three most important fields text-based image retrieval, retrieval based on visual feature, and semantic features. In text-based image retrieval system images are provided through labels and for retrieval cause similarity of these labels are used. A predominant disadvantage of these types of structures is that a lot of human work evolved for image labeling and if the database of images in a massive quantity than it is pretty difficult. And labeling of images is no longer user friendly it relies upon database creators labeling strategies. Then the visual-based image retrieval systems used for image retrieval purposes and low-level features are considered for feature extraction. The systems based on visual features along with semantic features in use. There are three most important features are existing in an image specifically color, texture, and shape [1, 2]. All these methods having disadvantage is that they are time-consuming because for each feature we require to train image database then compute the results. Hence CNN is used for feature extraction. Once the pre-training did it will provide faster results.

2. Literature Survey

2.1 Neural Network

A Neural Network has an independent variable as input then through neurons connected to the output. Basically, the neural network consists of input values, Weights, activation function, and output values [3]. Generally, there are four activation functions are used namely Threshold function, Sigmoid function, Rectifier function, and Hyperbolic Tangent function[4]. Depend on application appropriate activation function needs to be used. The values of weights are estimated by two methods of Gradient Descent and Back Propagation. The inputs are applied to output through a neuron. Inputs are connected to neurons by weights. And inside the neurons activation function present. After validating output weights are estimated. And using Backpropagation weights are adjusted.

2.2 Convolutional Neural Network

The Convolutional Neural Network is a neural network having four steps. The first step is applying convolution to input. Basically, convolution is a mathematical operation in which two functions are used. The values of the functions are changed due to convolution. For the application of CNN in image processing, the feature detector is needed. Generally, the Feature detector is a matrix of size 3x3. In convolution operation, this feature detector applied to an image and after convolution operation, we get a matrix with some values, which are different from original [5]. In this step feature detector applied to input image which results in the generation of the feature map. After feature map generation for removing unwanted information, the ReLU rectifier function is used [6]. In the second stage, the pooling operation performed. There are three types of pooling are used namely max pooling, mean pooling, and sum pooling [6]. Stage three consists of flattening which generates a feature map. And in the last stage full connection done. The softmax and cross-entropy are two functions used for improving CNN.

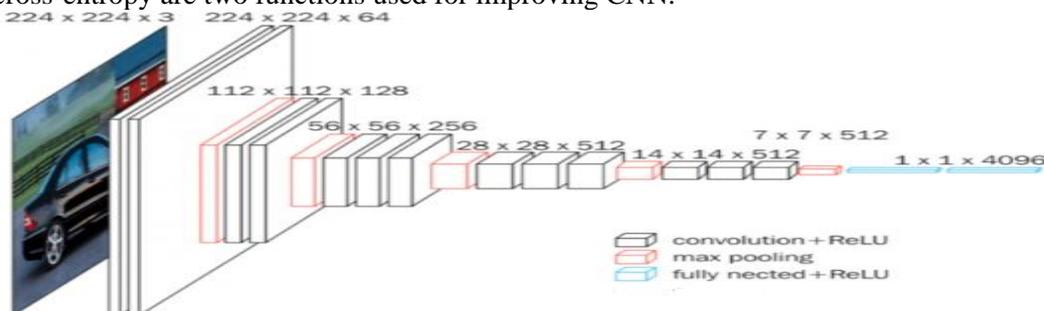


Figure 2. CNN Used for Feature Extraction

The above figure 2 shows Convolutional Neural Network which is used for feature extraction. In this network CNN along with ReLU used along with max pooling. The input for this network is an image. The output provides a feature vector of 4096 elements. This network is also known as the VGG16 network [8].

2.3 Similarity Measurement

The similarity between the query image and images in the database measured by various methods. Euclidean distance, cosine distance, and Hamming distance are popular algorithm for measuring similarity. In this paper, we will use the Approximate Nearest Neighbor method (ANN). ANN also is known as Locality Sensitive Hashing (LSH). Let us consider the database of images having various categories like bus, bird, planes, etc. After applying CNN feature vectors are generated for each image. Each vector represents the feature of an image. In ANN each Feature vector provided to hashing function. Basically hashing function is nothing but thresholding. It will convert feature vector value in either 0 or 1. After applying hashing function the vectors having closer values are put into one common basis vector. In this way various basis vectors created. During retrieval basis vector of the query, image compared with the basis vector of the image database and provides results. In this process since the basis vector consist of various vectors hence retrieval takes less time as compared to other functions.

3. Proposed Method

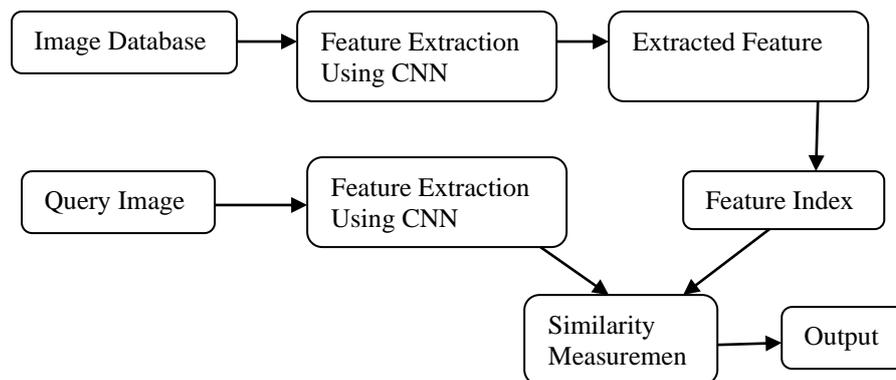


Figure 3. Proposed Method

Figure 3 provides the proposed method of data flow. The algorithm for the proposed method is explained below.

Step 1: In an image database several images with different categories are present.

Step 2: These images are fed to CNN for feature extraction.

Step 3: For each image feature vector generated. Each vector consists of 4096 elements.

Step 4: Features are indexed in the database.

Step 6: The user has to provide an input image considered as a query image.

Step 7: Features of a query image extracted using CNN.

Step 8: Using ANN similarity between features of query image matched with index.

Step 9: Finally top 10 similar images are displayed.

Step 10: Repeat from step 6 for another input.

4. Performance Evaluation

The performance of the system evaluated using two parameters Precision and Mean Average Precision (MAP). Precision is the ratio of the number of relevant images retrieved to the total number of images retrieved.

$$MAP = \sum_k^K \frac{Avg P(k)}{K} \tag{1}$$

Where k = value of precision for query and K is the number of query images.

5. Experimental Results and Discussion

The image database consists of 1000 images with 20 categories [9]. The hyperparameters used to train the image databases are 50 epochs and 32 batch size. The performance of CNN is given in figure 4.

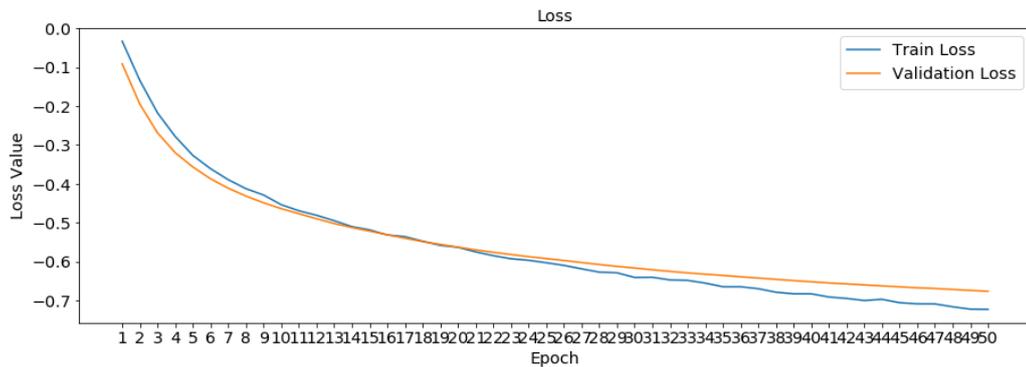


Figure 4. Performance of CNN

After training the model we evaluate the performance of the system by providing a query image. The results show the top 10 matched images from the image database. Figure 5 shows sample results that are tested on a few query images.

Query Image	Results

Figure 5. Sample Results of the proposed system

The database consists of 20 categories and in each category 50 images. Figure 5 shows sample results of some categories like airplanes, cats, and birds. Top 10 matching images displayed as a result.

Table 1. Precision and MAP

Category	Precision %	Category	Precision %
Airplane	96.18	Dining Table	96.28
Bicycle	96.28	Dog	98.40
Bird	96.28	Horse	97.40
Boat	96.70	Motor Bike	96.18
Bottle	95.70	Person	96.50
Bus	97.24	Plant	95.40
Car	98.20	Sheep	96.28
Cat	96.25	Sofa	95.20
Chair	96.90	Train	96.24
Cow	96.85	TV	98.36
MAP		96.64	

Table 2. Comparisons with existing methods

Methods	Images in Database	Precision %
Color in YCBCR, Color Moment [10]	1000 Images	82.70
HSV Color Histogram and Hu moment invariant [11]	1000 Images	84.20
CBIR Using CNN [Proposed]	1000 Images	96.64

Table 1 represents the performance of the proposed system with precision and MAP. We can conclude that in all categories results provides promising results. Table 2 Provides a compression analysis with existing methods.

6. Conclusion and Future Work

We can conclude that the performance of CBIR improves significantly using CNN. Since the existing methods based on image content like color, shape, and texture provides less accuracy as compared to the proposed system. We can use CNN for improving accuracy and retrieval time. In all categories presented in the database, the proposed system provides significant results. In this system, we are using pre-trained CNN. In the future, we can design CNN depends on image size in the database. We can evaluate CNN using different image databases.

References

- [1] A. Latif, "Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review", Journal of Mathematical Problems in Engineering, vol. 2019, (2019), pp. 01-21. DOI: <https://doi.org/10.1155/2019/9658350>
- [2] Ganar A.N, Gode C.S., and Jambhulkar S.M., "Enhancement of Image Retrieval by Using Color, Texture and Shape Features", Proceedings of International Conference on Electronic System, Signal Processing and Computing Technologies, India,(2014),pp.251-255
- [3] Y.LeCun, L.Bottou, G.Orr, and K.Muller, "Efficient BackProp", (1998), <http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf>
- [4] X.Glorot, A.Bordes, and Y.Benjio, "Deep Sparse Rectifier Neural Network", Proceedings of International Conference on Artificial Intelligence and Statistics, FL, USA, Vol.15, (2011), pp.315-323
- [5] J.Wu," Introduction to Convolutional Neural Network", LAMDA Group, China, (2017)
- [6] C.Kuo," Understanding CNN with Mathematical Model", (2016), <https://arxiv.org/pdf/1609.04112.pdf>
- [7] D.Scherer, A.Miller and S.Behanke," Evaluation of pooling operations in convolutional architecture for object recognition", Proceedings of International Conference on Artificial Neural Network, Greece,(2010)

- [8] K.Simonyan and A.Zisserman, "Very Deep Convolutional Networks for Large Scale Image Recognition", Proceedings of ICLR, **(2015)**, pp.1-15
- [9] Cyrus Rashtchian, Peter Young, Micah Hodosh, and Julia Hockenmaier, "Collecting Image Annotations Using Amazon's Mechanical Turk", In Proceedings of the NAACL HLT **(2010)**
- [10] Y.Dandotiya and A.Atre, "Image Retrieval Using Edge Detection, RLBP, Color moment method for YCbCr and HSV Color Spaces", Proceedings of IEEE International Conference on Electronics Communication and Aerospace Technology, **(2017)**, pp.662-668
- [11] R.Rajkumar and Dr.M.V.Sudhamani, "Development of retrieval system using color from multi-segment and shape features from an image as a combination", International Journal of Engineering and Technology, Vol.9, Iss.1 **(2019)** DOI:[10.3590/ijeat.A1291.109119](https://doi.org/10.3590/ijeat.A1291.109119)