

Application of Convolutional Neural Network for Content Based Image Retrieval

Rahul M.Bhutada¹, Prof.V.S.Mahalle²

¹ Department of Computer Science & Engineering, Shri Sant Gajanan Maharaj College of Engineering,
SGBAU University, Shegaon, India,

² Guide, Department of Computer Science & Engineering, Shri Sant Gajanan Maharaj College of Engineering,
SGBAU University, Shegaon, India,

Abstract— The problem is to extract the image features that represent the image contents in a database. Such kind of extraction requires a evaluation of retrieval performance of features of image. Related work is the foundation for understanding and improving more knowledge regarding precise field. This paper proposes Content Based Image Retrieval via Convolutional neural network. Content Based Image Retrieval is a technique for extracting features of images like color, shape, texture, blob detection, edge detection, contour detection and etc. That sort of work of extracting features from image will be done by convolutional neural network here. Once features are extracted from a query image then same procedure will be used for extracting features of database images. For comparing features of both images Euclidian distance can be used to produce the best result. The key part of retrieval system is feature extraction. Until now, the only way of penetrating these collections was based on keyword indexing, text, or just by browsing. In this paper we survey some practical aspect of current CBIR texture, and shape for exact and efficient CBIR after doing the study of related works.

Keywords— CBIR, CNN, Features, Image Retrieval, Classification, Neural Network

I. INTRODUCTION

In the recent past the advancement in computer and multimedia technologies has led to the production of digital images and cheap large image repositories [6]. Size of image set has increased rapidly due to this, including digital libraries, medical images etc. To tackle this rapid growth it is required to develop an image retrieval system which operates on a large scale. The primary aim is to build a robust system that creates, manages and query image databases in an accurate manner. CBIR is the procedure of automatically indexing images by the extraction of their low-level visual features, like shape, Color, and texture, and these indexed features are solely responsible for the retrieval of images [2]. Thus, it can be said that via navigation only, browsing, query-by-example etc. we can calculate the similarity among the low-level image contents which will be used for the retrieval of same images. Images are a representation of points in a high dimensional feature space and a metric is used to measure the similarity or dissimilarity between images on this space. Hence, those images which are closer to the query image are similar to it and are accessed. Representations of the Image feature and similarity measurement are very essential to calculate retrieval performance of a CBIR system and from decades researchers have studied them extensively [4]. A variety of techniques image pixels captured by machines and high level semantic concepts perceived by humans. Such a problem poses fundamental challenge of Artificial Intelligence from a high-level perspective that is how to build and train 2 intelligent machines like human to tackle real-world tasks [3]. One promising technique is Machine Learning that attempts to address this challenge in the long-term. In the recent years there have been important advancements in machine learning techniques [6]. Deep Learning is an important breakthrough technique, which includes a family of machine learning algorithms that attempt to model high-level abstractions in data by employing deep architectures composed of multiple non-linear transformations.

II. LITERATURE SURVEY

Inventors have already proposed different methods to improve the system of content based image retrieval.

- [1]. Surbhi Jain, Joydip Dhar proposed applications of CNN towards solving classification and retrieval problems. For retrieval of similar images, they agreed on using transfer learning to apply the GoogleNet deep architecture to their problem. Extracting the last-but-one fully connected layer from their training of GoogleNet CNN model served as the feature vectors for each image, computing Euclidean distances between these feature vectors and that of their query image to return the closest matches in the dataset in Image Based Search Engine Using Deep Learning.
- [2]. Pakkrudin B, Ravi G. H. proposed Color and texture features based image retrieval system. In this paper author worked on Wang database contains 1000 JPEG images. From result analysis of those proposed system author got Precision of 100% at recall worth is 5% and the quality reduction to 47.57% at 50% recall.
- [3]. Mushtaq Ali, Le Dong, Yan Liang, Zongyi Xu, Ling He, Ning Feng the creators presented one image feature, for example, weighted average of colourful image. This feature portrays the probability of the sum of the color values of any of the relating pixels of the three planes in the given RGB image. There are four steps to find this feature, First

include comparing pixel of the three planes of shading picture, Second figure the sum of the occurrences of all different values in the result array, Third find the probability distribution function of the occurrences of each different value in the result array, Finally calculate the weighted average using different values and their related probabilities. Major goal of this paper is to compare the average retrieval time of proposed system with color averaging technique. Hence, the proposed system retrieves the desired images very quickly. The average precision, recall and F_measure of the results are generated.

- [4]. H. B. Kekre, Kavita Sonawane the authors proposed color histogram based bins approach in RGB, XYZ, Kekre's LXY and L'X'Y' color spaces. The proposed technique contains seven steps as follows: First take the input as RGB image, Second convert the input RGB image to XYZ, LXY and L'X'Y' color spaces, Third divide the image into three planes of individual color spaces, Forth calculate the histograms for each of the three planes of the image from each color space, Fifth dividing each histogram into two parts by computing the Centre of gravity and perform the "8 bins formation" procedure, Sixth bins formation: Seventh eight bins obtained after execution of step sixth.. According to Kekre transform features vectors are calculated by using a combination of row mean and column mean of input image and database images. Similarly HSV color space quantifies the color space into different regions, by calculating the precision and recall parameters of choosy images from the database, contrast among the two algorithms and the efficiency of the mixture of both the algorithms can be calculated.
- [5]. VenuShaha, Pawan Bhatt, Mahesh Dasarath, Shipra Gupta and Jacki Bhavsar the author proposed three image indexing technique for retrieval of images. The query procedure algorithm is used to find the features of the query image. The Indexing procedure is used to find the features of the database images and Finally Semantic Indexing algorithm is used to find the features of the collected images. In this approach medical images are used as a database which is available in google search engine.
- [6]. Ahmed J.Afifi and wesam M.Ashour proposed a content- based image retrieval method using color and texture feature. To extract the color feature from the image the colour moments has calculated where the image will be in the HSV colour space. To extract the texture feature the image to be in gray-scale and ranklet transform is performed on it.

III. ARCHITECTURE OF CBIR

The Hidden layers (Feature extraction layers) part in this part, the network will perform a number of convolutions with RELU and pooling operations in between which the features are detected.

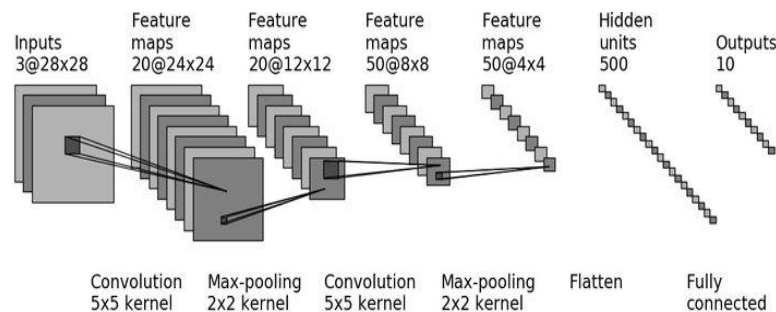


Fig. 1 Architecture of CNN with all layers

A. Convolutional Layer

The 2D convolution is a fairly simple operation, start with a kernel, which is simply a small matrix of weights. Kernel slides over the 2D input data, performing an element wise multiplication with the part of the input it is where on, and then sum up the results into a single output pixel. The kernel repeats this process for every location it slides on, converting a 2D matrix of features into yet another 2D matrix of features. The output features are crucially, the weighted sums (with the weights being the values of the kernel only) of the input features located in the same location of the output pixel on the input layer. Whether or not an input feature falls within this same location, gets determined directly by whether it's in the area of the kernel that produced the output or not. It means that size of the kernel directly checks how many (or few) input features get combined in the production of a new output feature.

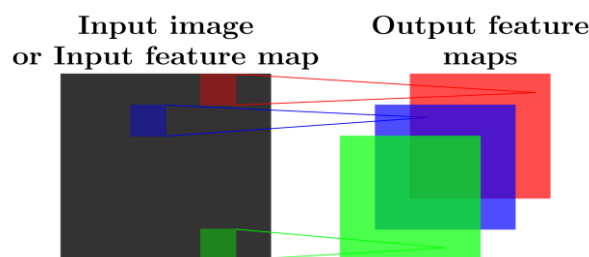


Fig. 2 Convolutional layer

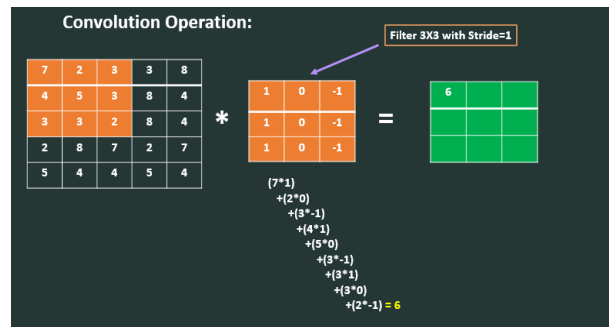


Fig. 3 Convolution step1

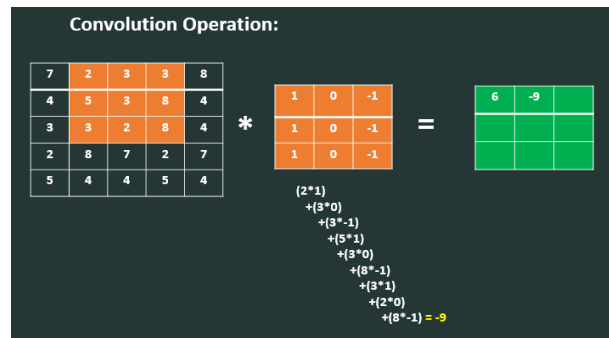


Fig. 4 Convolution step2

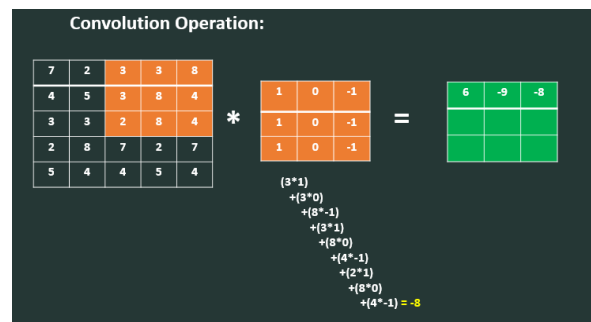


Fig. 5 Convolution step3

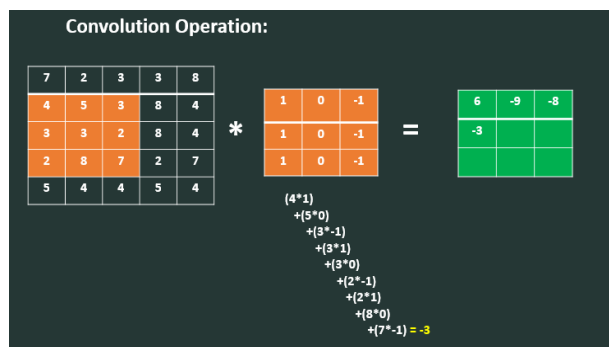


Fig. 6 Convolution step4

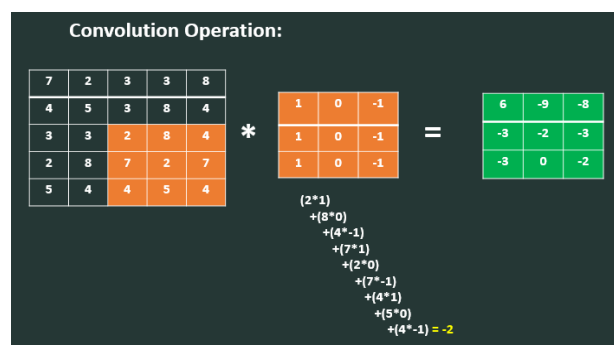


Fig. 7 Convolution step 5

B. RELU Layer

The Rectified Linear Unit is the most commonly used activation function in deep learning models. The function returns 0 if it receives any negative input, but for any positive value it returns that value x back. After every convolution layer, it is convention to apply a nonlinear layer (or activation layer) suddenly afterward. Behind this layer the aim is to introduce nonlinearity into a system that generally has just been calculate linear operations during the convolution layers (just element by element multiplications and summations). Nonlinear functions like tanh and sigmoid were used in past, but researchers found out that RELU layers work more better because the network is able to train a lot faster (as because of the computational efficiency) without making a huge difference to the accuracy. It also helps to reduce the vanishing gradient issue, which is the issue where the lower layers of the network train very slow because the gradient decreases exponentially through the layers. The Rectified linear unit layer applies the function $f(x) = \max(0, x)$ to all of the values in the input area. Basically Relu layer changes all the negative activations to 0. This maximizes the nonlinearity in the model and the overall network without affecting the receptive fields of the convolutional layer.

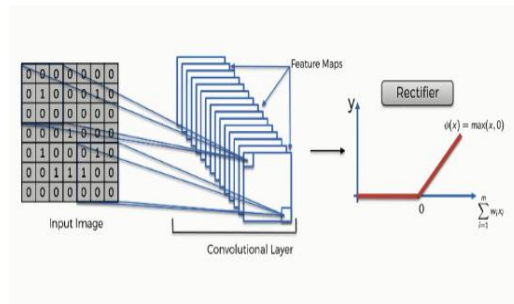


Fig. 8 Pictorial Representation of RELU Layer

C. Pooling Layer

To reduce the computational complexity, a pooling layer or sub sample layer is usually used immediately after a convolutional layer. After rectified linear unit layer's, programmers may apply a pooling layer. It is also called as a down sampling layer. In this category, there are also several layer options, with max pooling being the most popular. This takes a filter (normally with size 2x2) and a stride with the same length. It then applies it to the input volume and outputs the maximum number in every sub region that the filter convolves around.

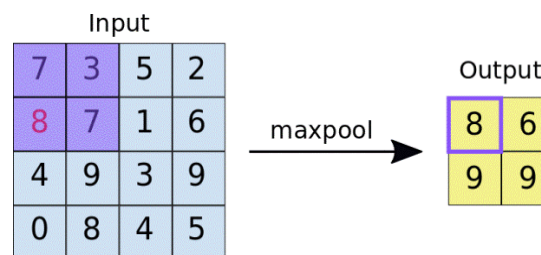


Fig. 9 Pooling Layer

For pooling layers other options are average pooling and L2-norm pooling. The instinctive reason behind this layer is that once we know that a specific feature is in input part (there will be a high activation value), its exact location is not that much important as its relative location to the other features. As you can think, this layer forcefully reduces the spatial dimension (the length and the width change but not the depth) of the input volume.

D. Fully Connected Layer

The Classification part The Fully Connected Layers will serve as a classifier on top of these extracted features. It will allot a probability for the object on the image being what the algorithm predicts it is. This layer basically takes an input volume (whatever the output is of the conv or ReLU or pool layer preceding it) and outputs an N dimensional vector N is the number of classes that the program need to choose. For example, if you wanted a digit classification program, N would be 10 as there are 10 digits. Each number in this N dimensional array represents the probability of a certain class. For example, if the resulting vector or array for a digit classification program is $[0.1 \ 0.1 \ 0.75 \ 0 \ 0 \ 0 \ 0 \ 0.05]$, then this represents a 10% chance that the image is a 1, a 10% chance that the image is a 2, a 75% probability that the image is a 3, and a 5% probability that the image is a 9. The way this fully connected layer works is that it sees at the output of the previous layer and checks which features most correlate to a particular class. For instance, if the program is predicting that some image is a dog, it will have high values in the activation maps that represent high level features like a paw or 4 legs, etc. Like this if the program is predicting that some image is a bird, it will have high values in the activation maps that represent high level features like wings or a beak, etc. Basically, a FC layer looks at what high level features most strongly correlate to a particular class and has particular weights so that when you compute the products between weights and previous layer, you receives the correct probabilities the original for different classes.

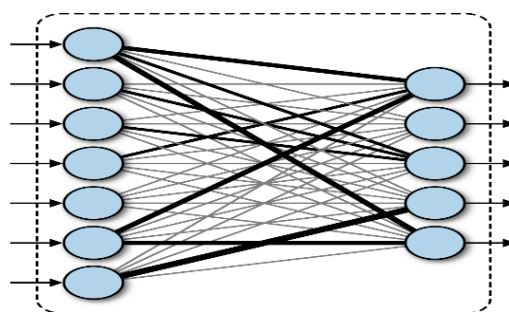


Fig. 10 Fully Connected Layer

IV.APPLICATIONS

There are varieties of applications for content based image retrieval it can be used for architectural and engineering, digital libraries, art collections(museums,..), crime prevention(security filtering,..), intellectual property, medical diagnosis(X-ray, CT, MRI,..), commerce(fashion, catalogue,..), entertainment(personal album,..), military(radar, aerial,..) and etc.

V. ADVANTAGES

1. CBIR has become a very active area research for two major research communities, Database Management and Computer Vision.
2. Feature Extraction methods are easy, effective and less expensive.
3. Time requires is less to find all those related image.
4. More than one related outcomes occur by only one search (If more than one equally likely image present in the database).
5. CNN is very helpful to process on image, text or video
6. It will do feature extraction automatically developer don't need to write code for it.
7. CNN is more efficient in case of memory utilization
8. CNN is super simple means very low complexity in case of feature extraction

VI. RESULT ANALYSIS OF CBIR

The most common measures to evaluate a system are response time and memory space used. A better system if the response time and the memory space are smaller this is the case of our system because we retrieve the query image in the appropriate class and not in the entire images database. But with CBIR systems, in addition to these two measures,we are interested in other measures. Usually in the CBIR system, the user is interested in the relevant responses of the system. So information retrieval systems require the evaluation of the accuracy of the response. This type of evaluation involves the system research performance. The indexation system and Image searching is system used for information retrieval. In this section, we will describe the two most important measures: recall and precision. These measures are interrelated. Therefore this relationship is frequently described by a recall and precision curve. Then we present other measures that are also used to evaluate information retrieval systems.

Recall is defined as the ratio of the number of relevant images retrieved to the total number of relevant images in the dataset.

$$recall = Ra/R$$

Here Ra is the relevant retrieved images, R is the total number of relevant images present in the database.

Precision is the ratio of number of relevant images retrieved to the total number of images retrieved.

$$Precision = Ra/A$$

Where A is the total number of images retrieved.

Table I: Existing and proposed system Accuracy for Top 10 Predictions

Number of top 10 Predictions	Accuracy
Using Feature descriptors	84%
Using CNN	91%

VII. CONCLUSION

This is our proposed concept that focuses on Content Based Image Retrieval by using convolutional neural network. The Problem here is to extract the image features that effectively represent the image contents in a database. Such an extraction requires a detailed evaluation of retrieval performance of image features CBIR is used to search a specific image from a large database. CBIR performs interactive searching of images from the dataset. Up to now CBIR is not the most accurate technique for finding desired outcomes. But by using Convolutional neural network it works good. One promising technique is Machine Learning that attempts to address this challenge in the long-term. In the recent years there have been important advancements in machine learning techniques. Finding more features and extracting that from query image is the very complex term in CBIR but as because of CNN it becomes little bit easy and so we have thought of using CNN for more accurate result. Convolutional neural network is an important breakthrough technique, which includes a family of machine learning algorithms this technique is still in research state for finding more better and accurate output. At present this technique is implemented by Google.

REFERENCES

- [1]. Surbhi Jain and Joydip Dhar. "Image based search engine or content based image retrieval using machine learning", Proceedings of 2017 Tenth International Conference on Contemporary Computing (IC3), 10-12 August 2017.
- [2]. Pakruddin.B, Ravi G.H, "Content-Based Image Retrieval Using Color and Texture Features", Proceedings of the International Journal for Scientific Research & Development| Vol. 4, Issue 03, 2016.
- [3]. Mushtaq Ali, Le Dong, Yan Liang, ZongyiXu, Ling He and Ning Feng , "A COLOR IMAGE RETRIEVAL SYSTEM BASED ON WEIGHTED AVERAGE", Proceedings of the IEEE conf, 2014.
- [4]. H. B. Kekre, Kavita Sonawane, "Comparative Study of Color Histogram Based Bins Approach in RGB, XYZ, Kekre's LXY and L'X'Y' Color Spaces", Proceedings of the IEEE International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA), P 364-36, 2014.
- [5]. Venu Shah, Pavan Bhat, Mahesh Dasarath, Shipra Gupta and Jackie Bhavsar, " Content Based Image Retrieval Using Combination of Kekre Transform and HSV Color Segmentation", Proceedings of the IACSIT International Journal of Engineering and Technology, Vol. 3, No. 5, October 2011.
- [6]. Wesam M.Ashour, Ahmed J. Afifi, "Content-Based Image Retrieval Using Invariant Color and Texture Features", Proceedings of the IEEE, conf, 2012.