Content Based Image Retrieval Using Deep Learning Convolutional Neural Network

Arshiya Simran¹, Shijin Kumar P.S² and Srinivas Bachu³
¹M.Tech Scholar, Department of Electronics and Communication Engineering, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, Telangana, India.
²³Associate Professor, Department of Electronics and Communication Engineering, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, Telangana, India.
¹shijinkumarps@yahoo.com

Abstract. Content-based image retrieval (CBIR) is a widely used method for image retrieval from large and unlabeled image collections. However, users are not satisfied with the traditional methods of retrieving information. Moreover the abundance of online networks for production and distribution, as well as the quantity of images accessible to consumers, continues to expand. Therefore, in many areas, permanent as well as widespread digital image processing takes place. Therefore, the rapid access to these large image databases as well as the extraction of identical images from this large set of images from a given image (Query) pose significant challenges as well as involves efficient techniques. A CBIR system's efficiency depends fundamentally on the calculation of feature representation as well as similarity. For this purpose, they present a basic but powerful deep learning system focused on Convolutional Neural Networks (CNN) and composed of feature extraction and classification for fast image retrieval. We get some promising findings from many detailed observational studies for a number of CBIR tasks using image database, which reveals some valuable lessons for improving the efficiency of CBIR. CBIR systems allow another image dataset to locate related images to such a query image. The search per picture function of Google search has to be the most popular CBIR method.

Keywords: Image Retrieval, Convolutional Neural Network, Content Based Image Retrieval, Deep Learning, Query Image.

1. INTRODUCTION

In recent years, along with Bing photo search, there seems to be a rapid growth in search engines: CBIR engine of Microsoft (Public Company), CBIR machine of Google, Note: Not running on all images (public company), CBIR search engine, Gazopa (private company), Imense Image Search Portal (private company) and the like. Com (Private enterprise), the retrieval of images has also proven to be a challenging mission [1]. Also with support of the present period, writers can scan for textual statistics very quickly, but this scanning approach calls for people to explain each pixel manually inside the database, which is almost difficult for very large datasets or for pictures with the purpose of being created mechanically, e.g. Photographs from surveillance cameras. It has additional disadvantages because within the definition of pictures there might be a potential to skip images that use specific equal terms. "Systems focused on categorizing snap shots in semantic groups such as" tiger "as a" animal "subclass will debar the issue of miss-categorization, however it will entail additional attempt to choose the pix that is possibly" tigers "with the assistance of a usage, but they are all most handy as a" animal [2]. The CBIR technique is opposed to conventional approaches, which are seen as fully concept-based approaches [3].

The According to several common methods introduced in recent years, one of which has several disadvantages, such as the histogram; first this representation leads to the lack of spatial detail necessary to accurately represent the material of an image. Second, in quantification, the use of such a histogram raises the problem of characteristic spaces [4]. CNN is primarily designed to work with the variability of 2D forms, and all other strategies have seen to outperform. Multiple modules, including attribute extraction, classification as well as paradigm learning, are made up of recognition frameworks. They make it possible to train such multimodal systems globally using gradient-based approaches to maximize an overall output assessment [5]. In comparison to previous methods, the binary methods require pairwise inputs for binary code learning, the feature representation has the best CNN output, the generalisation potential of the extracted features, the relationship between dimensional reduction as well as loss of accuracy in CBIRs. A form of artificial neural feed-forward network where the data is located is the Convolutionary Neural Network (CNN). They are biologically inspired Multi-layer perceptron (MLP) invariants that are designed for minimum pre-processing purposes [6]. In image and video recognition,
these models are used extensively. Convolutional neural networks use very little pre-processing compared to other feature extraction as well as classification algorithms. Orthodox neural networks that are very good at classifying images have far more parameters and require a long time to learn on the CPU [7]. The first aspect of CNN is the process of transformation. Authors investigate a deep learning system for content-based image retrieval (CBIR) and perform a comprehensive series of empiric studies for a variety of CBIR tasks by applying a state-of-the-art deep learning process, i.e., convolutional neural networks (CNNs) for the learning of image representation features. Authors derive some promising findings from the observational studies as well as reveal some useful observations to answer the unanswered questions [8]. Through attempting to grasp the overhead of obtaining the complete data collection of original raw images for use in CNNs, writers first begin this work [9]. The authors then clarify that our compression architecture does not negatively affect the efficiency of the CNN model classification [10] [11].

In the ILSVRC-2012 competition, Writers joined the version of this model and earned a winning top-five test error rate of 15.3 percent compared to 26.2 percent for the second-best entry. [12]. Our final network consists of five convolutionary and three fully interconnected layers, and this depth seems to be significant: the author states that the size of the network is primarily constrained by the amount of memory available to current GPUs and the amount of training time that the author schedules. Our network will take between five and six days to train two GTX 580 3 GB GPUs. All our experiments demonstrate that our results can only be improved if we wait for faster GPUs and stronger datasets to be used [13][14].

2. METHODOLOGY
This section explains the suggested framework for the CBIR scheme that employs DConvNet as shown in Fig.1. CNN’s working can be described as follows: Sliding philtres are applied to the input by a 2-D convolution layer. By shifting the philtres vertically as well as horizontally over the input, the layer covers the input and calculates both the weight as well as the input point product, applying the concept of discrimination. The ReLU layer performs a threshold function for each input variable where any value below zero is set to zero. The final pooling layer is sampled by dividing the input into rectangular areas and measuring the boundary of each region. A fully connected layer multiplies the input by a mass matrix and adds it to the vector.

As per the facts, DL-CNN training and testing includes allowing any source image to be classified by artefacts with probabilistic values varying from [0,1] The kernel or philtre, The corrected linear unit (ReLU), the max pooling, the fully linked layer as well as the SoftMax classification layer are used for a sequence of convolution layers. Fig.2 demonstrates the DL-CNN architecture used for improved attribute representation for word images over traditional retrieval systems in the suggested technique for the CBIR scheme [15].

The convolution layer in Fig. 2 is the key layer from which the characteristics are extracted from a source image as well as preserves the relationship between pixels by using small blocks of source data to learn the features of the image. It is a mathematical function that considers two sources, such as the \( I(x, y, d) \) source image where \( x \) as well as \( y \) indicate space coordinates, i.e. row and column count. \( d \) is denoted such as dimension of an image (here \( d = 3 \). Although the source image is 3GB) as well as a related input image philtre or kernel, the image can be referred to as \( F(k_x, k_y, d) \).
Figure 2. Representation of convolution layer process

Output obtained to the input image as well as the filter convolution process seems to have a size of \( f(x-k_x+1), (y-k_y+1) \). This is recognized also as feature map. Fig 3a gives an example of the convolution method. Let us presume that the input image is 5×5 and the filter is 3×3 in dimension. The image function map of the input image is obtained by multiplying the values of the filter as seen in Fig. 3b.

Networks using the hidden layer corrective technique have been referred to as the linear correction unit (ReLU). This function of ReLU \( g(\cdot) \) is a simple calculation that returns the input value directly when it returns zero if the input price increases zero afterwards.

\[
g(x) = \max\{0, x\}
\]  

(1)

The primary component analysis is a machine learning technique used to decrease dimensionality. It uses fundamental mathematical and linear algebra matrix operations to measure a source data projection in identical and smaller dimensions. PCA may be considered a projection technique in which m-column or attribute data is projected by m or even smaller columns onto a subspace while retaining the source data's most important portion. Enable n x m to appear in the source image matrix and result in a J that is a projection of I. Measuring the mean value for each column is the main step. Next it excludes the mean column value; the values in each column are centred. Now, the centred matrix covariance is being computed. Finally, compute each covariance matrix's own value decomposition, that gives a list of own or exclusive principles. These vectors are the paths or elements of the reduced subspace J, while these vectors represent the full path amplitudes. Now by descending their own values, these vectors can be sorted to range the elements or axes of a new subspace to (I,K). In general, patented vectors referred to as the key components or functions are chosen.

Figure 3. Example of convolution layer process (a) an image with size 5×5 is convolving with 3×3 kernel (b) Convolved feature map.
A metric must be established to determine distance from picture word question \( I_q \) and obtained word images \( I_i \). If the question and word images obtained are the same (bit per bit), we need a measurement process. Therefore, we want a measure of resemblance in which the distance value including its images considered has its number of equivalent bits. Fig. 4 provides a detailed explanation on the Euclidean distance.

3. RESULTS AND DISCUSSIONS

We addressed the simulation results of the CBIR method in this section. With a few datasets, the suggested algorithm was checked and the outputs were shown in the figures below. Fig. 5 reveals that photographs are recovered using the recommended CBIR system. We used two commonly used Precision and Recall metrics as a measure of efficiency. Precision tests the CBIR algorithm's ability to retrieve only relevant images, while Recall selects the CBIR algorithm's ability to retrieve all relevant images in accordance with Eq. (2) as well as Eq. (3) respectively. Fig. 6 provides a comparison on the performance of proposed system with existing systems in terms of mean average Precision (mAP) and mean average Recall (mAR).

\[
\text{mAP} = \frac{\text{total number of relevant images retrieved}}{\text{total number of retrieval images}}
\]

\[
\text{mAR} = \frac{\text{number of relevant images retrieved}}{\text{number of relevant images in database}}
\]

| Input Image | mAP  | mAR  |
|-------------|------|------|
| Bird        | 86.14| 90.34|
| Dog         | 82.85| 86.45|
| Car         | 88.91| 91.34|
| Cat         | 83.23| 85.78|
| **Average** | **85.23** | **88.53** |
Figure 5. Retrieved dog images using DConvNet CBIR system

Table 2. Performance comparison of CBIR systems

| Methodology          | mAP  | mAR  |
|----------------------|------|------|
| Color Histogram      | 68.86| 70.85|
| DWT                  | 78.85| 79.56|
| DConvNet-PCA (Proposed) | 85.23| 88.53|

Figure 6. Performance comparison of CBIR systems

4. CONCLUSION
This article proposed an effective CBIR method with pair-wise hamming distance using DConvNet and PCA. Through developing large-scale deep convolutionary neural networks to learn efficient image representation of images, the authors implement a CBIR deep learning system. Authors carry out a systematic sequence of empiric experiments for thorough testing of deep convolutionary neural networks, with the application of a number of CBIR tasks under different conditions, in order to understand the characteristics of representations. Proposed system provides mAP and mAR of 85.23 and 88.53 respectively. The results of the simulation showed that the proposed CBIR method achieved superior efficiency through the acquisition of more appropriate images. Furthermore, using mAP and mAR, the performance assessment of the proposed CBIR system is seen and contrasted with the current CBIR systems discussed in the literature.
5. References

[1] Liu Y, Zhang D, Lu G, and Ma W Y, 2007, A survey of content-based image retrieval with high-level semantics, *Pattern Recognition*, 40(1), pp 262–282.

[2] Le Cun Y, Bengio Y, and Hinton G, 2015, Deep learning, *Nature*, 521(7553), pp. 436–444.

[3] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, and Wojna Z, 2016, Rethinking the inception architecture for computer vision, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826.

[4] Babenko A, Slesarev A, Chigorin A, and Lempitsky V, 2014, Neural codes for image retrieval, *European conference on computer vision*, Springer, pp. 584–599.

[5] Xia R, Pan Y, Lai H, Liu C, and Yan S, 2014, Supervised hashing for image retrieval via image representation learning,” *AAAI*, 1(2), pp. 2-7.

[6] Chen J. C, and Liu C. F, 2015, Visual-based deep learning for clothing from large database, in *Proceedings of the ASE Big Data & Social Informatics*. ACM, pp. 42-48.

[7] Iliukovich Strakovskaia A, Dral A, and Dral E, 2016, Using pre-trained models for fine-grained image classification in fashion field, *Proceedings of the First International Workshop on Fashion and KDD*, pp. 31-40.

[8] Shrivakshan G and Chandrasekar C, 2012, A comparison of various edge detection techniques used in image processing, *International Journal of Computer Science Issues*, 9(5), pp. 272–276.

[9] Maurya N and Tiwari R, 2014, A novel method of image restoration by using different types of filtering techniques, *International Journal of Engineering Science and Innovative Technology*, 3(1), pp.32-40.

[10] Kandwal R, Kumar A, and Bhargava S, 2014, “Review: existing image segmentation techniques,” *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(4), pp. 35-42.

[11] Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, and Le- Cun Y, 2013, Over feat: Integrated recognition, localization and detection using convolutional networks,” *arXiv preprint*.

[12] Wu P, Hoi S, Xia H, Zhao P, Wang D and Miao C, 2013, Online multimodal deep similarity learning with application to image retrieval, *Proceedings of the 21st ACM international conference on Multimedia*, pp. 153–162.

[13] Liu S, Song Z, Liu G, Xu C, Lu H, and Yan S, 2012, Street-to shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3330–3337.

[14] Yamaguchi K, Kiapour M. H, Ortiz L. E, and Berg T. L, 2015 “Retrieving similar styles to parse clothing,” *IEEE transactions on pattern analysis and machine intelligence*, 37(5), pp. 1028–1040.

[15] Wan J, Wu P, Hoi S C, Zhao P, Gao X, Wang D and Li J, 2015, Online learning to rank for content-based image retrieval.