Received: 13 July 2022 | Revised: 9 August 2022 | Accepted: 30 August 2022 | Published online: 13 September 2022

RESEARCH ARTICLE

Artificial Intelligence and Applications
2022, Vol. 00(00) 1–5
DOI: 10.47852/bonviewAIA2202326

BON VIEW PUBLISHING

Deep Learning-Based Image Extraction

K. S. Krupa[1,*], Y. C. Kiran[1], M. Gaganakumari[1], S. R. Kavana[1], R. Meghana[1] and R. Varshana[1]

[1]*Department of Information Science and Engineering, Global Academy of Technology, India*

Abstract: The development of the web and advancements in computation and multimedia technologies have led to an increase in the variety of photo databases and the collection of hundreds of images that include medical images, e-libraries, and art galleries. The necessary images from that kind of big collection may demand a lengthy period to retrieve using traditional image extraction techniques like Textual Based Images Retrieval. It is essential to develop an effective image extraction procedure that can handle such vast amounts of data at once. The main objective is to develop a trustworthy tool that effectively generates, uses, and responds to data. It employs a strategy for creating an effective picture retrieval application that enables individuals to ask questions about the software and extract it from a huge database.

Keywords: deep learning, content-based image retrieval, convolution neural network, VGG-16, principal component analysis

1. Introduction

Extracting imagery is a method of dynamically cataloging photographs by extracting basic components, mainly colors and shapes that are utilized for pattern identification. The two very important situations for such realization of the features extraction job are, in fact, the measuring of resemblance and features representation. Researchers from several disciplines have been studying picture retrieval for almost 10 years. Even though many solutions have been proposed, it remains among the biggest challenging continuing studies centered upon an image. The difference between the higher accuracy of sensory systems and the substandard computer-acquired pixels is the main problem with feature extraction. In the last 20 years, multimodal technologies have developed significantly, leading to the formation of enormous quantities of content including electronic photos, songs, or movies. This is still challenging to select appropriate images from a vast database collection. When trying to find photos that are linked to a certain search query that is provided as input, photo resemblance is especially important. In order to calculate picture similarities, lesser visual feature approximations are used. It makes use of a range of visual properties, with coloration, textures, and structure playing a key role. A machine learning (ML) method is appropriate to address such a problem. Recently, improvements in ML methods have been made to address this problem. One important modeling approach is deep learning. It offers a variety of ways to study with the goal of demonstrating high-resolution data-gathering methods using specific methodologies with numerous machine parameters. Deep learning (DL), as opposed to traditional methods, usually uses a predefined framework to enable the development of complex designing architectures that can assess information at many stages of computing and interpretation.

*Corresponding Author: Krupa K S, Department of Information Science and Engineering, Global Academy of Technology, India. Email: krupaks@gmail.com

2. Literature Survey

The extracting effectiveness of content-based image retrieval (CBIR) is significantly impacted by the feature extraction and likeness assessment, as scholars in interactive media have recognized for years [1]. In recent times, emerging ML techniques have exhibited enormous strides. DL is a class of ML method which makes use of various techniques consisting of multiple nonlinear transformations in an effort to generalize knowledge widely. By examining DL techniques for acquiring characteristics at distinct scales via the information voluntarily, a technique can learn complex calculations that immediately transform confidential material intake to the results without relying on manmade qualities that require knowledge of the region. For feature extraction, convolution neural network (CNN) approaches are used in this paper [1]. CBIR plays a significant role in this work because this really is a current research issue in the field of computer perception. Current work focused a lot of emphasis on the fundamental function of this algorithm, which is to find comparable photographs based upon the input query image. We provide a CBIR strategy for detecting objects of absorption in a collection of documents, while CNN's being used to get the portrayal of the recoverable picture in order to go around such problems [2]. The two major components of a CBIR project execution are extracting features and comparable assessment, and researchers are working on these for a very lengthy period. There are many solutions offered, but these remain the most challenging issues in the ongoing CBIR research because of the discrepancy between the high image pixels captured by machines and the lower quality pixels captured by amazing human perceptions. The suggested solution resolves a CBIR characteristic of human-compressed images using DL techniques. This study makes use of many Corel database types [3]. In order to the maximum degree feasible, DL technology does an intelligent extraction of contents again of information in use. Data may be extracted from hundreds of divided photos, thanks to a computational intelligence process

called DL. The recommended strategy relies on DL techniques, which encompass all information and understand the information by separating features to the profound bottom. The system seems to contain its dedicated computing infrastructure, which will only offer a few features [1]. CBIR is a major research challenge in the field of feature extraction. The main driver is the rising need for tools that really can extract images—or specific parts of them—from massive data banks amassed over the past few years in contemporary civilization. First, the writer explored the concept of producing a closed-form expression for speedy picture retrieving by using a pretrained CNN-like features extraction and trying to take into consideration feature map of different dimensions. The other method involved grouping 2 DL models within a pretrained Siam CNN network framework for extraction of characteristics and likeness measurement [4]. DL is gaining popularity because of its outstanding performance in both areas. As DL achieves amazing results in each of the visual and linguistic realms, researchers are starting to look at the conceptual differences between pictures and textual data. Picture-to-text retrieval is a type of continuous image-text cross-modal retrieval that concurrently utilizes a photograph like a query to identify specific information in written texts. The study focuses on DL based cross-modal retrieval strategy for image-text scenario, that has been recently considered as unique DL based methodology with considerably increased efficiency [5]. In response to the picture query, similar images must be found from a sizable set of data. CBIR is the name for this. The basic method involves finding similar photographs by examining just handful image characteristics. The needed information in the photographs should theoretically be represented by such qualities. As an outcome, more complex elements are needed, but several simplified type of data, including input images, are ineffectual. In order to find related photographs, this article used traits taken from a systems method developed long ago utilising a DL CNN model designed for large images categorizing [6]. Textual content binarization that relates to assuming the graphic model for things only with two consecutive frames that can be expanded to numerous data creation techniques that were used sequentially would be the first phase into DIA. And over 90% of most recently DIA techniques published at ICDAR as well as ICPR include classification at a certain point throughout the processing, even if its necessity remains up for debate. To accomplish binarization, numerous heuristics methods have been put forth [7]. The research of documents recognition and classification, with a focus on scanned texts, is known as document segment procedure and [1]. Prior to this, a sizable portion of study has been focused on documents recovery, initially using text information and also with sloppy writing. Power equipment saw widespread popularity in the beginning 1990s, especially in the area of postal technologies. ANN - Artificial Neural Network based methods do have close relationship with the subject of the DIAR investigation [8]. Like the saying states, a comparable image recovery software looks for images one by one, inputs the images that need to get restored, and afterwards creates a number of images that are strikingly similar. Labeling was first applied to the development of picture retrieving methods. The foundation of this system is built on data capture techniques, and it directly identifies the qualities of the picture, including the caption, keywords, writer, and other characteristics. This study shows a novel approach based on specific image generation and neural ML. Local fusion characteristics are used as picture expressiveness to address the problem of stated characteristics being ignored by congested layer characteristics. To constrain attributes and address the element technical problem, proportionate training is introduced to DL. ResNet34, a known CNN architecture, is the classifying technique employed in the paper [9]. A number of computational visual research programs, comprising classification tasks, object tracking, and segment, have successfully employed ResNet as their main design. The goal of this study is to determine whether characteristics produced by a CNN architecture can be used as picture extracting data and whether models that have been trained on high-quality datasets can be applied to low-quality data. According to the research, CNN models develop characteristics that recognise similarities among images belonging to the identical categorization through a series of tests. This information is converted to vectors in the significant professional that support grouping. The proposed methodology is subject to distortion that cannot really be eradicated with simple enhancing methods, tests suggest [10]. The need for an efficient picture retrieval method focused on visual data has been motivated by the vast amount of image datasets and the dearth of viable text-based image extracting approaches. The research from many studies inside the CBIR industry over earlier periods was evaluated for this study. The stages of CBIR architecture and the most recent methods for bridging the semantic gap were also examined in this paper. Finally, this study covers a wide variety of strategies that could facilitate the creation of a distinctive CBIR by emphasizing some of the more crucial factors that influence CBIR effectiveness, creating a method that minimizes processing costs and achieving acceptable retrieval precision [11]. The study has shown the promise of the attention mechanism by demonstrating as the use of standard sequence-based systems can be eliminated by using a systematic attentive approach. Such solutions are high in value, finish the assignment, and take a lot less time to train. They use a sampling strategy that enables quick calculation and retrieval even on huge datasets to demonstrate in this study that the combined distinctive description of photo and word is indeed the fundamental idea behind multimodal understanding. The requirement to maintain the visual significance location is fundamentally another crucial factor of multimodal photo and text learning. In Rao et al. [12], a creative method using groups of distance measures was devised utilizing DConvNet and principal component analysis (PCA). In order to obtain accurate picture interpretations of images, the researchers design a large-scale CNN architecture and use it to produce a CBIR deep convolutional approach. The researchers conducted a thorough sequence of investigations for such complete evaluation of CNNs, utilizing a range of content-based picture retrieving operations in varied scenarios Simran et al. [13], in an effort to understand the characteristics of depictions. The suggested framework has mAP and mAR values of 85.23 and 88.53, accordingly. Because many pertinent data were collected, the simulation outcomes showed how the proposed CBIR strategy was highly effective. The proposed method is centered on deep methods in Simran et al. [13], wherein each piece of information is limited and components are learnt by sorting out the qualities that are most important first. A multiple attribute photo retrieving strategy is provided by combining the traits of histogram equalization, edge, edge directions, edge histogram, and textures based. The content-based picture in such paradigm would have been selected as sample of predetermined image groups. After a few pre-processing steps like choice removal, its attributes are collected and saved as small identification documents. Throughout the similarities test, the range among distinct attributes is measured.

3. Methodology

Due to advancements in DL, various techniques can be utilized to extract the images. The goal is to create a mechanism that will

enable users to request specific images from big datasets. For excellent outcomes, the questioned image is treated to reduce noise as well as improve image quality. The sequence of steps involved in retrieval of images similar to query image from the existing dataset is shown in Figure 1. The input query image is preprocessed and trained using a deep learning model. The deep learning model is trained to retrieve and compare image characteristics such as color, texture, size. As a result, the model outputs the matching images from the dataset.

3.1. VGG-16

A sectioned image is sent to the system as inputs (224, 224, 3). The first two layers include the identical padded and 64 channels having 3*3 filter size. Alongside a max pool layer of stride (2, 2), two layers comprise fully connected levels of 128 sampling frequency and step size (3, 3). A max-pooling stride (2, 2) which is the same as the layer preceding it comes next. Then, 256 filtering with filtration widths of 3 & 3 are distributed over two convolutional layers. After that, there are different pairs of three fully connected layers, and then a max-pooling layer comes next. Every filter is having the similar padding and has 512 filters of size (3, 3). The stacks of two convolutional layers then receives this image. As opposed to AlexNets and ZF-11*11 Net's and 7*7 filters, we use 3*3 filters in the CNN and max pool layers. Additionally, a few of these layers use 1*1 pixel to change the number of input streams. After each convolution layer, 1-pixel pad is applied to prevent the image's spatial feature. After adding a convolutional as well as max pool layer to the stack, a (7, 7, 512) features map is obtained. To construct a characteristic representation with the value 1, 25,088, this result is compressed. There are then three fully connecting layers; 1st layer generates a vector with a size of (1, 4096) using the most recent feature representation as input; the 2nd layer also produces (1, 4096); however, the 3rd layer produces a vector of size (1, 1000), which is used to implement the SoftMax activation function to categorize 1000 classes. ReLU is used by every hidden layer as its activation function. Because ReLU promotes quicker learning and lessens the probability of disappearing gradient issues, it is more computationally efficient. The Architecture map of VGG-16 is shown in Figure 2.

Figure 1
Proposed Methodology

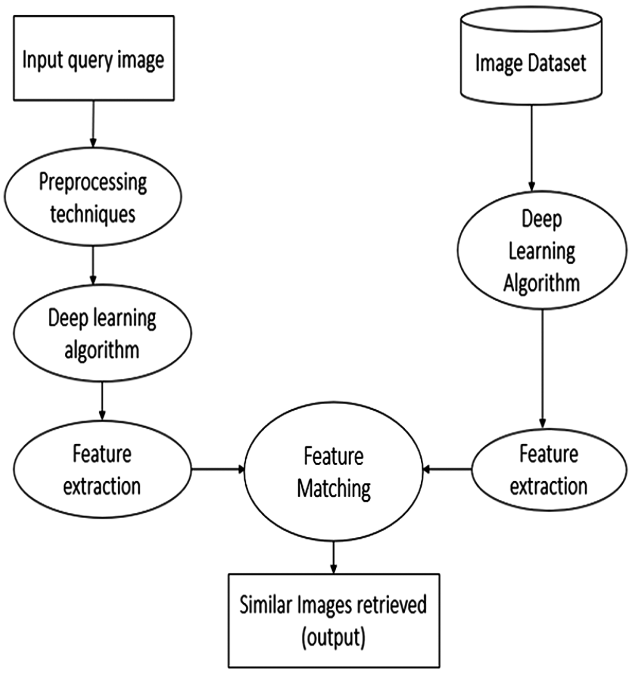
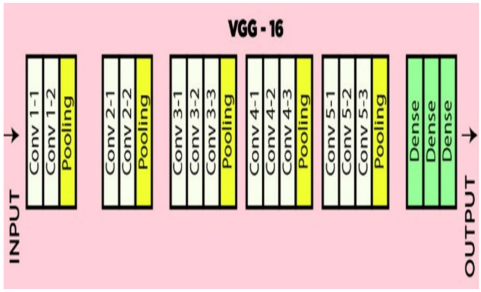


Figure 2
VGG-16 Architecture Map



3.2. Cosine similarity

Cosine similarity a measure that quantifies the similarity between two vectors. Python's cosine Resemblance function can be used to compare two texts for similarity. Cosine similarity treats each data object in a dataset as a vector. The following is the formula to determine the cosine similarity between the two vectors:

Given two vectors $A = [a_1, a_2, \ldots, a_n]$ and $B = [b_1, b_2, \ldots, b_n]$ with n attributes each, the cosine similarity, $\cos(\theta)$ is represented as

$$\cos(\theta) = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \times \sqrt{\sum_{i=1}^{n} b_i^2}}$$

We are given that $a_i, b_i \epsilon \{0, 1\}$ for $i = 1..n$. Since $a_i, b_i \geq 0$,

$$\sum_{i=1}^{n} a_i b_i \geq 0; \sqrt{\sum_{i=1}^{n} a_i^2} \geq 0; \sqrt{\sum_{i=1}^{n} b_i^2} \geq 0$$

3.3. PCA

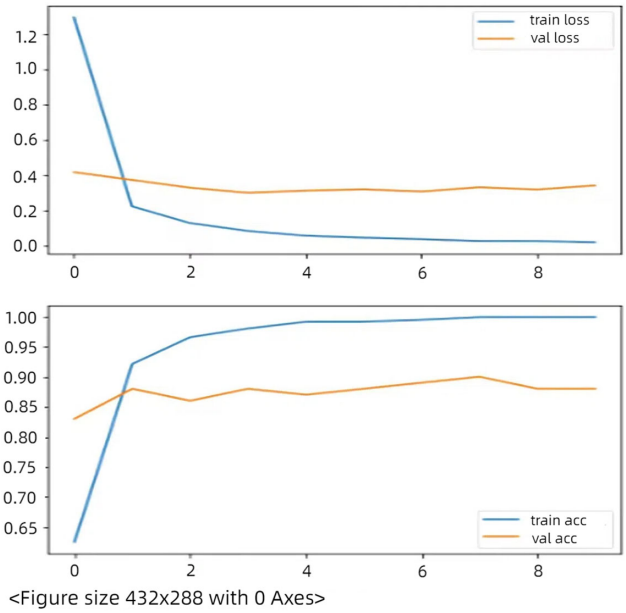
An unsupervised learning approach called PCA is used in ML to decrease dimensions. By using orthogonal transform, a mathematical process is used for transforming the observed correlated attributes together into a set of linear statistically independent data. Principal components are newly modified attributes. One of the widely used tools for exploratory data analysis and prediction modeling is this one. It is a method for identifying significant patterns in the provided dataset by lowering the variances.

Typically, PCA looks for the surface with the lowest dimensionality onto which to reflect the high-dimensional data. Because the high value reveals a good split between the classes and hence reduces dimensionality, PCA works by taking into account the variance of each attribute. Examples of PCA's practical uses include image processing, movie recommendations engines, and channel-specific power optimization. As a feature extraction technique, it includes the crucial variables while excluding the less crucial ones.

Figure 3
Output of Corel Dataset

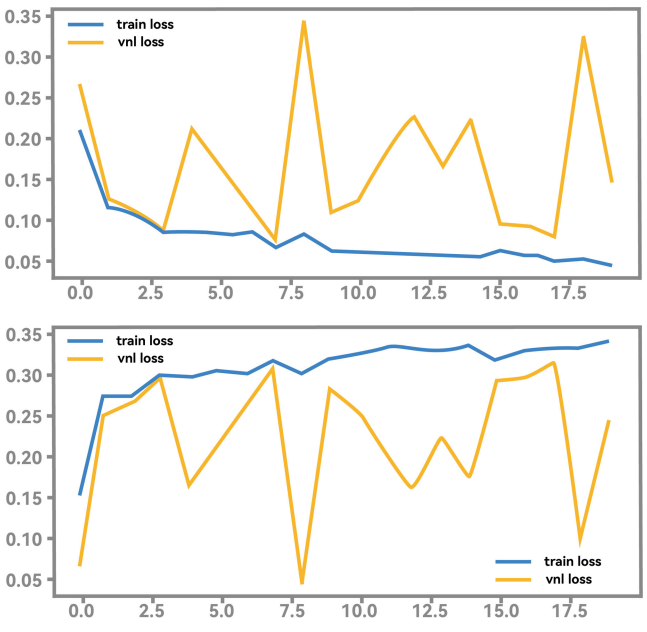
Figure 4
Output of Cifar-10 dataset

Figure 5
Accuracy graph of Corel dataset



<Figure size 432x288 with 0 Axes>

Figure 6
Accuracy graph of Cifar-10 dataset



<Figure size 432x288 with 0 Axes>

4. Results

In this study, we considered two datasets: the Corel dataset and Cifar-10 dataset which was available online (Figures 3, 4, 5, and 6).

The performance analysis of our proposed method using Corel and Cifar-10 dataset is shown in Table 1.

Table 1
Performance analysis

Methodology	VGG-16, PCA, Cosine similarity
Datasets	Corel
	Cifar-10
Result	Corel – 99.8%
	Cifar-10 – 98.26%

5. Conclusion

The main objective is to generate a solution that enables users to request specific images out of massive datasets. For improved outcomes, the queried image is treated to reduce noise as well as improve image clarity. DL algorithms are used to extract picture characteristics out from the query image which is processed for comparison with the characteristics of the images in the dataset. These qualities include color, texturing, and size.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflict of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data available on request from the corresponding author upon reasonable request.

References

[1] Wiggers, K. L., Britto, A. S. Jr., Heutte, L., Koerich, A. L., & oliveria, L. E. S. (2018). Document image retrieval using deep features. In *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE.

[2] Desai, P., Pujari, J., Sujatha, C., Kamble, A., & Kambli, A. (2021). Hybrid approach for content-based image retrieval using VGG16 layered architecture and SVM: An application of deep learning. *SN Computer Science*, *2*(3), 170. https://doi.org/10.1007/s42979-021-00529-4

[3] Saritha, R. R., Paul, V., & Kumar, G. (2018). Content based image retrieval using deep learning process. *Cluster Computing*, *22*(1), 4187–4200. https://doi.org/10.1007/s10586-018-1731-0

[4] Chen, J., Bai, C., Kpalma, K., & Zhang, L. (2020). Review of recent deep learning based methods for image-text retrieval. In *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE.

[5] Maji, S., & Bose, S. (2020). CBIR using features derived by deep learning. *ACM/IMS Transactions on Data Science*, *2*(3), 1–24. https://doi.org/10.1145/3470568

[6] Liwicki, F. S., & Liwicki, M. (2020). Deep learning for historical document analysis. In *Handbook of pattern recognition and computer vision*. World Scientific.

[7] Lombardi, F., & Marinai, S. (2020). Deep learning for historical document analysis and recognition: A survey. *Journal of Imaging*, *6*(10), 110, https://doi.org/10.3390/jimaging6100110

[8] Zhang, Y. (2020). Similarity image retrieval model based on local feature fusion and deep metric learning. In *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*. IEEE.

[9] Li, Y., & Wang, M. (2020). Image retrieval algorithm based on deep learning. In *ICGSP 2020: 2020 The 4th International Conference on Graphics and Signal Processing*. Association for Computing Machinery.

[10] Dang, T. V., Yu, G.-H., Nguyen, H. T., Vo, H. T., Lee, J.-H., & Kim, J.-Y. (2020). Convolutional neural network-based image retrieval with degraded samples. In *SMA 2020: The 9th International Conference on Smart Media and Applications*. Association for Computing Machinery.

[11] Hameed, I. M., Abdulhussain, S. H., Mahmmod, B. M., & Pham, D. T. (2021). Content based image retrieval: A review of recent trends. *Cogent Engineering*, *8*(1), 1927469, https://doi.org/10.1080/23311916.2021.1927469

[12] Rao, S. S., Ikram, S., & Ramesh, P. (2021). Deep learning based image retrieval system with clustering on attention based representations. *SN Computer Science*, *2*(3), 179, https://doi.org/10.1007/s42979-021-00563-2

[13] Simran, A., Shijin Kumar, P. S., & Bachu, S. (2021). Content based image retrieval using deep learning convolutional neural network. *IOP Conference Series: Materials Science and Engineering*, *1084*(1), 012026, https://doi.org/10.1088/1757-899X/1084/1/012026

How to Cite: Krupa, K. S., Kiran, Y. C., Gaganakumari, M., Kavana, S. R., Meghana, R., & Varshana, R. (2022). Deep Learning-Based Image Extraction. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA2202326