Novel CBIR System using CNN Architecture

K Ramanjaneyulu
Research scholar, ECE Department
JNTUCEK, Kakinada,
Andhra Pradesh, INDIA
Associate Professor, ECE Dept.
QISIT, ONGOLE, A.P

K.Veera Swamy
Professor, ECE Department
Vasavi College of Engineering
Telangana, INDIA
kilarivs@yahoo.com

CH Srinivasa Rao
Professor, ECE Department
JNTUKUCEV, Vizianagaram,
A.P, INDIA
chsrao.ece@jntukucev.ac.in

***Abstract*- Development of multi-media technologies large number of images are used in various fields such as video satellite data, medical treatment and digital judicial systems and surveillance systems. An efficient representation of features from an image for retrieval process is a challenging task. In this paper provides the feature extraction of an image using deep learning technique to tackle the differences between low-level features and high-level semantic features of basic CBIR systems. In this technique feature database can be created from each image in the database using VGG 16 model. By using Euclidean distance metrics an image analogous to the image of the query was retrieved by comparing the feature vector of the query image (compute similar to the data base images) and the feature database. The results suggest that the proposed CNN techniques yields better results than the other existed techniques.**

Keywords: CBIR, CNN architecture, Precision, Recall, F-score

I. INTRODUCTION

With rapid growth of electronic devices such as built in cameras & rapidly developing Internet technologies, large numbers of persons are likely to browse for web & photo sharing. The traditional method of an image extraction using text is appears as insufficient for large image database. There are some disadvantages of downloading images based on text, such as marking labels to individual images in large databases, using text is time consuming and depends on language that is valid only for one language at a time. Another disadvantage is the same image that other/different users can set a different label. These disadvantages can be avoided using image content to extract images. This kind of image retrieval is known as content-based image extraction. In CBIR technique image searching is done based on the content of an image such as colour, texture, edges or shape rather than the annotation of an image. In a retrieval system (CBIR), the key problem is to retrieve image features that effectively represent the contents of the image in a database. Such retrieval requires an effective evaluation of the detail of the extraction of the characteristics of the image. For CBIR techniques efficient representation of feature vector and similarity measurement is essential for performance of the extraction. The Semantic gap is a major drawback in CBIR. A semantic gap occurs among the low-level pixels in an image taken by machines & the high-level semantics observed by the human beings [3]. The latest achievements of deep learning methods mainly in CNN dealing with Computer Vision encouraged me to do work in this area to resolve the drawbacks of CBIR using an image data set Basics related to the Convolutional network architecture are introduced in section - II. CBIR using CNN VGG16 model is discussed in section - III. In Section - IV represents the experimentation results. Lastly, Conclusion is in section - V.

II. CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE

The basic foundation for CNN architecture consists a various kinds of layers i.e., convolutional layer, pooling layer & fully-connected layers. Usually, there are several filters (kernels or weights) in every convolution layer that emits the similar number of feature maps by sliding the filters through feature maps of the preceding layer.

A. Convolution Layer:

Convolution layer is one of the main building blocks for convolutional network (CNN). Layer parameters contain a set of selectable filters (kernels) which have a small perception range but covers through the input volume in full depth. Convolution layer executes the convolution operation over the entries of the filter and the input image (width, height & color channels).It creates a two-dimensional (2-D) map to activate this filter.

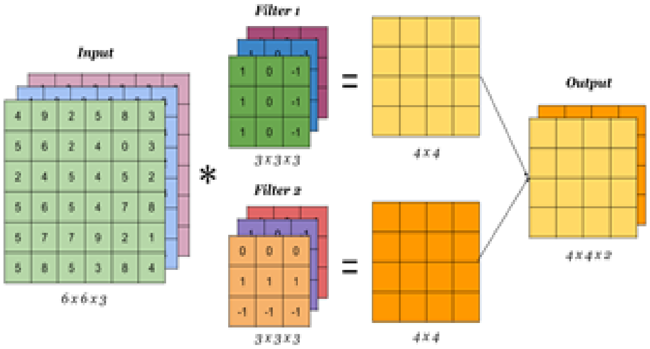


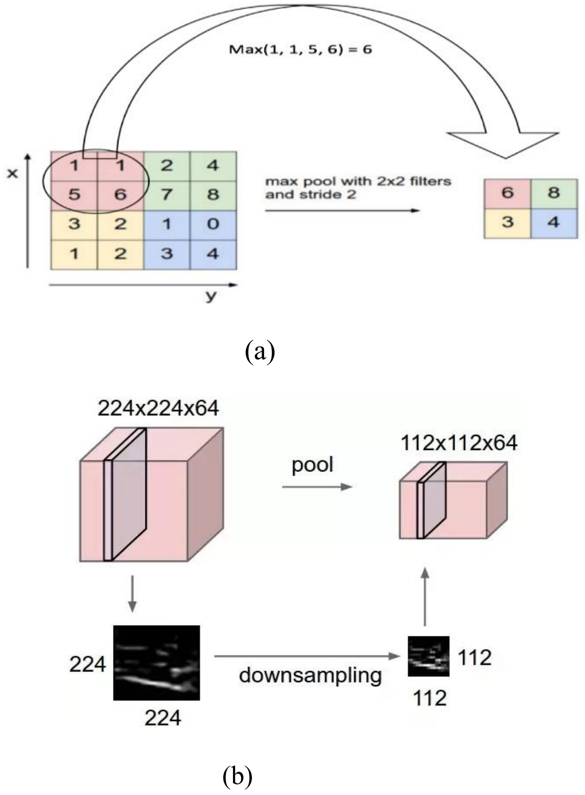
Fig: 2.1 convolution layer process

B. Pooling Layer:

One of the most important concepts of CNNs is Max-pooling which a form of non-linear down-sampling. Here, Input image is divided into groups of rectangles that do not overlap and maximum values are allocated for each sub-region. We use the maximum matching for visibility for the following reasons - The upper layer calculation has been reduced by removing the non-maximal values. In general pooling layer inserts in-between the consecutive Convolutional layers in CNN architecture. Frequently, a pooling layer with 2x2 filters is used with a two-step downward sampling depth. Each MAX

978-1-5386-4985-5/18/$31.00 ©2018 IEEE

operation in this case will be accepted with a maximum of four numbers (small area of 2x2 in any depth). The purpose of pooling layers in CNN is reduces the complexity of the data & resolves the input image size. It makes the strong robustness to the input sample for a neural network in terms of recognition such that CNNs can efficiently retrieve more features from an image [2].



(a)

(b)

Fig:2.2 (a),(b)Max - pooling layer process

C. *Fully Connected Layer:*

Lastly, a numerous convolutional, max pooling layers after the high-level reasoning in the neural network is ended via fully connected layers [8]. In fully connected layer, Neurons have connections to all activations in the preceding layer, as seen in systematic neural networks & their activation is calculated by using multiplication and followed by a bias offset.

III. CBIR USING CNN VGG16 MODEL

In this paper we use VGG16 Architecture for DCNN model training. VGG16 model has16 trained layers .The key features of VGG16 model are network expansion. In this model, firstly the input RGB image is resized in to 224 x 224.latter, resized image is transferred through a five blocks of convolution layers. Every block of layers having an increasing number of filters (3 x 3) size & stride is fixed at one, while the input of the converting layer is shaded so that the spatial resolution is stored after convolution. The blocks are divided by max-pooling layers. Max-pooling is performed with stride 2. The five set of convolution layers are extended by the FC layers (three). The last layer in the architecture is soft max layer which indicates the class probability. The architecture for VGG-16 can be shown below.

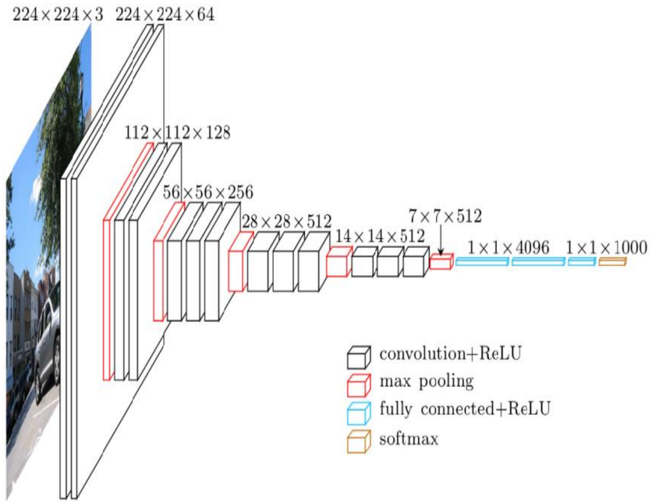


Fig: 3.1 Layered CNN architecture of VGG 16

After the CNN architecture is effectively improved and trained on the set of database images, extract the features vectors of an image from the last 3 fully connected layers using trained architecture. Feature database is constructed from the images of a database. The extraction process has been initiated when a user asks a system using an example image or a drawing of the object. For a query image also feature vector computed using the same process that is used to create an entry database. Similarity measure is used to compute the distance between the feature database images & the query image [1].

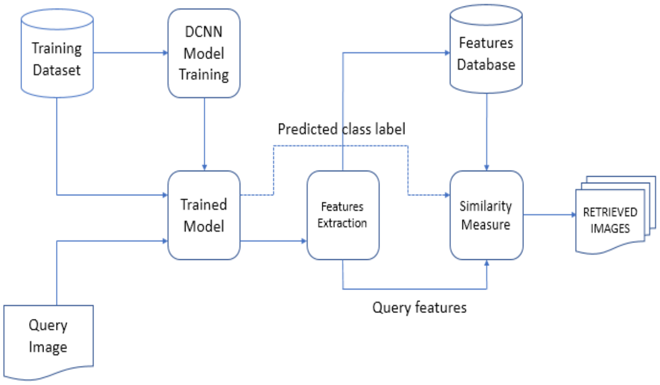


Fig: 3.2 Block diagram for CBIR model using CNN

Similarity Measurement:

The distance between the database feature vector & query image is calculated by using Euclidean distance metrics. Feature vector of an image with lesser distance is most resemblance to the query. Retrieving a suitable set of images and then arranged descending order by estimating their distance. Euclidean distance (ΔD) metric can be computed by following equation

$$\text{EUCLIDEAN DISTANCE } (\Delta D) = \sqrt{\sum_{i=1}^{n}(|Q_i - D_i|)^2}$$

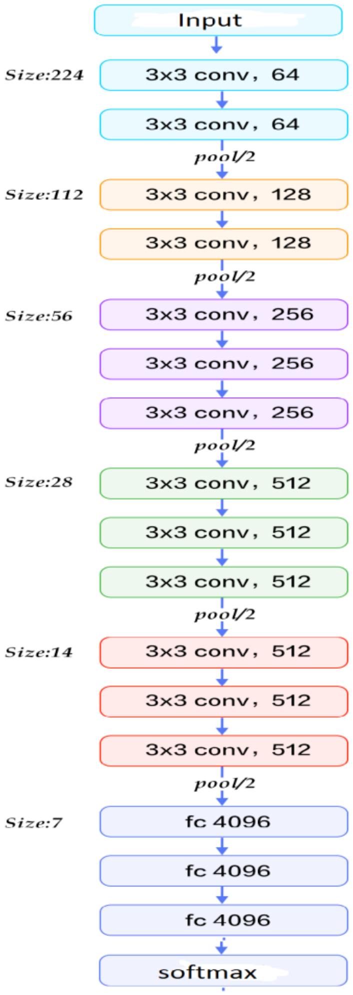


Fig:3.3 Flow chart for VGG 16 algorithm

IV. DATASET & EXPERIMENTATION RESULTS

4.1. Database & performance evaluation

Database we used in our evaluation is a Corel-1k dataset [7] and data includes 10 types of dinosaur, people, buses architectures, and other types. Each type has 100 photos. Experimental results are tested by in python.

4.2 Performance measures:

The system performance can be measured by using Recall, Precision, and F-score. Precision represents the effectiveness of a system where as Recall indicates the exactness of the system.

The following equations are used to calculate Precision & Recall,

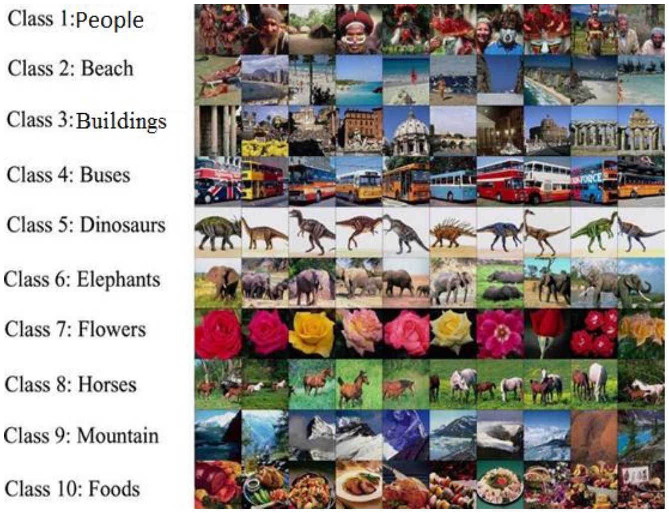


Fig 4.1: Example for different classes of images in Corel database

Precision (P) = X/Y

X=N.O of Relevant images retrieved

Y=Total N.O of images retrieved.

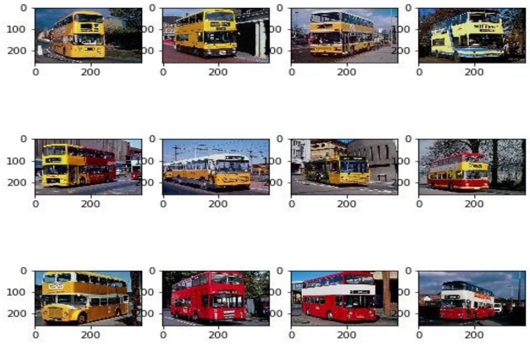
Recall = X/Z

X=N.O of Relevant images Retrieved

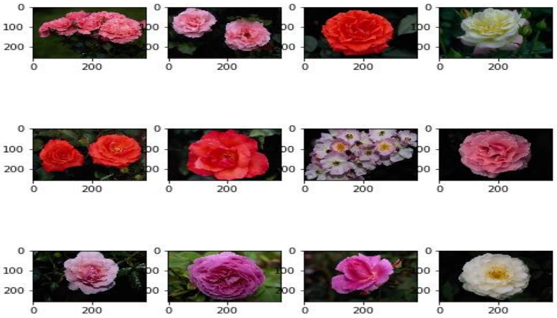
Z=N.O of relevant images in the collection of Dataset.

F-score/F-measure = 2 (P*R) / (P+R)

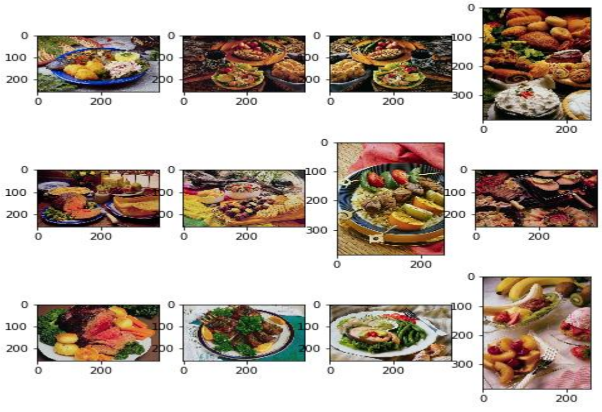
Where P is Precision, R is Recall



(a)

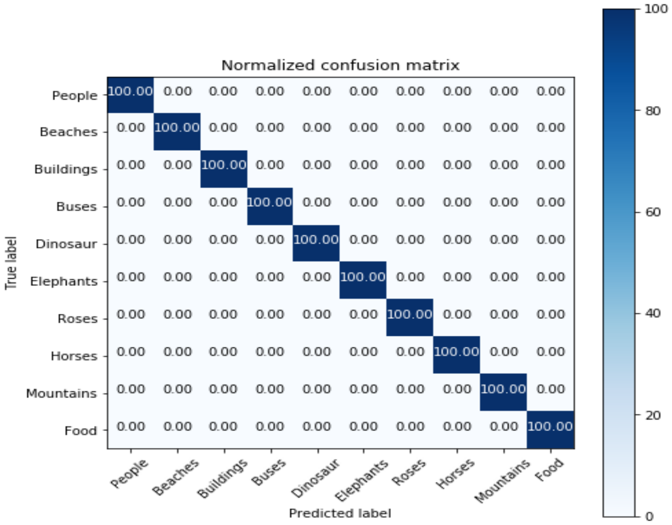


(b)

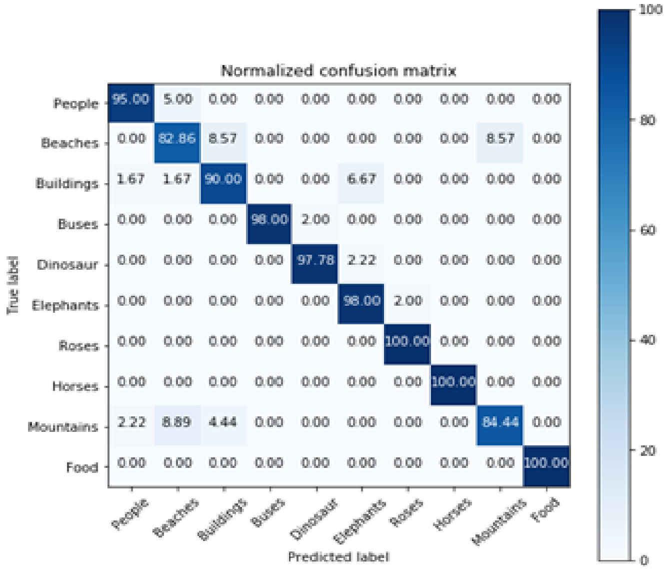


(c)

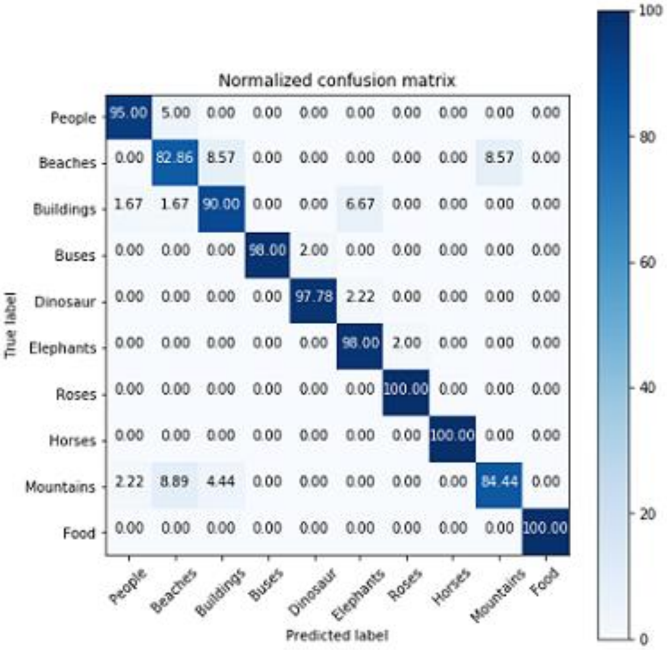
Fig: 4.2 Retrieval results of different categories (a) Dinosaurs (b) Roses (c) foods



(a)



(b)



(c)

Fig: 4.3 (a) Confusion Matrix --- Top 1(b) Confusion Matrix --- Top 5 (c) Confusion Matrix --- Top 10 images

Based on the confusion matrix [6] precision, recall is measured as by using following equations.

$$recall = \frac{true\ positives}{true\ positives\ + false\ negatives}$$

$$precision = \frac{true\ positives}{true\ positives + false\ positives}$$

Table 1: precision (%) for Corel Dataset (VGG16 architecture)

categories	Top '1'	Top '5'	Top '10'
People	100	96.06	93.69
Beach	100	84.19	84.87
Buildings	100	87.3	84.15
Buses	100	100	99.16
Dinosaurs	100	97.99	98.99
Elephants	100	91.6	91.21
Flowers	100	98.03	98.02
Horses	100	100	95.73
Mountains	100	90.7	83.38
Food	100	100	97.53

Table 2 : Recall (%) for Corel Dataset (VGG16 architecture)

categories	Top '1'	Top '5'	Top '10'
People	100	95	85
Beach	100	82.86	77.14
Buildings	100	90	83.33
Buses	100	98	99
Dinosaurs	100	97.78	98.89
Elephants	100	98	98
Flowers	100	100	99.29
Horses	100	100	100
Mountains	100	84.44	83.33
Food	100	100	99

Table 3: F-score (%) for Corel Dataset (VGG16 architecture)

categories	Top '1'	Top '5'	Top '10'
People	100	95.53	89.13
Beach	100	83.52	80.82
Buildings	100	88.63	83.74
Buses	100	98.00	99.08
Dinosaurs	100	97.88	98.94
Elephants	100	94.69	94.48
Flowers	100	99.01	98.65
Horses	100	100.00	97.82
Mountains	100	87.46	83.35
Food	100	100.00	98.26

Comparison of proposed technique with existed technique based on precision for Corel-1k database

	Existed	Proposed	
categories	[1]	TOP ‘1’	TOP ‘5’
People	96	100	96.06
Beach	84	100	84.19
Buildings	86	100	87.3
Buses	100	100	100
Dinosaurs	100	100	97.99
Elephants	94	100	91.6
Flowers	100	100	98.03
Horses	98	100	100
Mountains	86	100	90.7
Food	94	100	100
Over all (%)	93.8	100	96.38

V. CONCLUSION

In this article Novel CBIR system using CNN Architecture is presented here. In this CBIR technique we extract the image features using VGG16 CNN & experiments done on most popular dataset Corel-1k. We compute the performance measures i.e., Recall, Precision & F-score which shows the 100% results for top 1 retrieval image and improves using VGG 16 architecture and the performance of the proposed is raised by 6.6 % for Top-1 similarity of an image & 2.75 % for top-5 similarity of an images compared to the existing technique[1].

REFERENCES

[1] Shah, Amjad & Naseem, Rashid & , Sadia & Iqbal, Shahid & Arif Shah, Muhammad. (2017). Improving CBIR accuracy using convolutional neural network for feature extraction. 1-5. 10.1109/ICET.2017.8281730.

[2] Hailong Liu, Baoan Li, Xueqiang Lv, Yue Huang. "Image Retrieval Using Fused Deep Convolutional Features" , Procedia Computer Science, Volume 107 Issue C, PP.749-754 April 2017, ISSN: 1877-0509 EISSN: 1877-0509

[3] scholarworks.rit.edu

[4] D. G. Lowe, “Distinctive image features from scale invariant keypoints,” Int’l Journal of Computer Vision, vol. 60, pp. 91–11 020 042, 2004.

[5] Alzu’bi, A. Amira, and N. Ramzan, “Semantic content-based image retrieval: A comprehensive study,” Journal of Visual Communication and Image Representation, vol. 32, no. July, pp. 20–54, 2015.

[6] https://en.wikipedia.org/wiki/Sensitivity_and_specificity

[7] Corel database http://wang.ist.psu.edu/~jwang/test1.tar

[8] en.wikipedia.orgA.

[9] X.-Y. Wang, B.-B. Zhang, “Content-based image retrieval by integrating color and texture features,” Multimedia Tools and Applications, vol. 68, no. 3, pp. 545–569, 2012.

[10] J. Z. Zhang, ‘’Content-Based Image Retrieval using color and edge direction features”, 2010 2nd International Conference on Advanced Computer Control, pp. 459–462, 2010.

[11] "Advances in Computer Science and Information Technology. Networks and Communications”, Springer Nature America, Inc, 2012

[12] Nam Vu, Cuong Pham. “Traffic Incident Recognition Using Empirical Deep Convolutional Neural Networks Model" , Springer Nature, 2018

[13] Weixun Zhou, Zhenfeng Shao, Qimin Cheng. "Deep feature representations for high-resolution remote sensing scene classification" , 2016 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA), 2016

[14] Jacob John Foley, Paul Kwan. "Chapter 583- Feature Extraction in Content-Based Image Retrieval" , IGI Global, 2015

[15] Adnan Qayyum, Syed Muhammad Anwar, Muhammad Awais, Muhammad Majid. "Medical image retrieval using deep convolutional neural network" , Neurocomputing, 2017