**ENDSEMESTER REPORT**

**on**

**CONTENT BASED IMAGE RETRIEVAL**

**USING VGG16**

**Submitted by**

DEVYANSHI TIWARI (Enroll. No. R110216060)

SIDDHARTH SHARMA (Enroll. No. R110216154)

SOURADEEP BANERJEE (Enroll. No. R110216159)

**Under the guidance of**

Mr. Ravi Tomar

Assistant Professor (SG), Department of Virtualization



**SCHOOL OF COMPUTER SCIENCE**

UNIVERSITY OF PETROLEUM & ENERGY STUDIES

Bidholi Campus, Energy Acres, Dehradun – 248007.

**May - 2019**

**CONTENTS**

[CANDIDATES DECLARATION 1](#_Toc9417432)

[ACKNOWLEDGEMENT 2](#_Toc9417433)

[CERTIFICATE OF ORIGINALITY 3](#_Toc9417434)

[ABSTRACT 4](#_Toc9417435)

[INTRODUCTION 5](#_Toc9417436)

[PROBLEM STATEMENT 6](#_Toc9417437)

[LITERATURE REVIEW 7](#_Toc9417438)

[METHODOLOGY 8](#_Toc9417439)

[VGG16 8](#_Toc9417440)

[CUDA 10](#_Toc9417441)

[EUCLIDEAN DISTANCE AS SIMILARITY MEASURE 11](#_Toc9417442)

[SCHEDULE 11](#_Toc9417443)

[IMPLEMENTATION 12](#_Toc9417444)

[Index.ipynb 12](#_Toc9417445)

[Query.ipynb 13](#_Toc9417446)

[extract\_cnn\_vgg16\_keras.ipynb-VGG16 model 15](#_Toc9417447)

[RESULTS 16](#_Toc9417448)

[CONCLUSION 18](#_Toc9417449)

[FUTURESCOPE 18](#_Toc9417450)

[REFERENCES 18](#_Toc9417451)



**School of Computer Science**

**University of Petroleum & Energy Studies, Dehradun**

# CANDIDATES DECLARATION

I/We hereby certify that the project work entitled **CONTENT BASED IMAGE RETRIEVAL USING VGG16** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering with specialization in Cloud Computing and Virtualization Technology and submitted to the Department of Virtualization at school of Computer Science, University of Petroleum and Energy Studies, Dehradun, is an authentic record of my/our work carried out during a period from **January, 2019** to **May, 2019** under the supervision of **Mr. Ravi Tomar, Assistant Professor in Department of Virtualization.**

The matter presented in the project has not been submitted by me/ us for the award of any other degree of this or any University.

(**Devyanshi Tiwari, Souradeep Banerjee, Siddharth Sharma**)

**(R110216060, R110216159, R110216154)**

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Date 22 May, 2019) (Name of the Guide)

Mr. Ravi Tomar

**Dr. Deepshika Bhargava**

**(HOD)**

**Department of Virtualization**

# ACKNOWLEDGEMENT

We wish to express our deep gratitude to our guide **Mr. Ravi Tomar**, for all advice, encouragement and constant support he has given us through out our project work. This work would not have been possible without his support and valuable suggestion.

We sincerely thank our Head of the Department, **Dr. Deepshika Bhargava**, for her great support in doing our **CONTENT BASED IMAGE RETRIEVAL USING VGG16** on **SoCS**.

We are also grateful to **Dr. Manish Prateek Professor and Director SoCS and Dr. Kamal Bansal Dean CoES**, UPES for giving us the necessary facilities to carry out our project work successfully.

We would like to thank our friends for their help and constructive criticism during the project work. Finally we have no words to express our sincere gratitude to our parents who have shown us this world and for every support they have given us.

# CERTIFICATE OF ORIGINALITY

This is to certify, that the research project submitted by us is an outcome of our independent and original work. We have duly acknowledged all the sources from which the ideas and extract have been taken. The project is free from any plagiarism and has not been submitted elsewhere for publication.

Name of the Authors: Souradeep Banerjee, Devyanshi Tiwari, Siddharth Sharma

Designation: CSE-CCVT

Affiliated Institution: UPES Dehradun

Title of the Project: CONTENT BASED IMAGE RETRIEVAL USING VGG16

Email: [500052778@stu.upes.ac.in](mailto:500052778@stu.upes.ac.in)

Contact No: 9874525966



**School of Computer Science**

**University of Petroleum & Energy Studies, Dehradun**

**Minor**

****II

**PROJECT TITLE:** CONTENT BASED IMAGE RETRIEVAL IN PYTHON

# ABSTRACT

Using Tensor Flow and Keras we are going to train a VGG16 model, using which we will extract the features of a predefined image data set and store the features in the format HDF5. And then we are going to search our database using the features based on a query image being given by the user. We are using a GPU to run the program in python with the help of CUDA.

# INTRODUCTION

Image Retrieval is an application of Computer Vision wherein the system performs a lookup on the database in order to fetch similar images of the queried image or the input image. Computer Vision is a field of science wherein systems are trained to read or make a vision out of images to extract useful information like human eyes. Basically, there are two ways the system can retrieve images – Text Based Image Retrieval and Content Based Image Retrieval. In text-based retrieval, searching is done on the basis of metadata like keywords, text, etc whereas in content based it will refer the actual content consisting of the fusion of various parameters like color, histogram, shape, texture etc. In this project, the concept of Content Based Image Retrieval is being implemented which includes feature extraction of database and querying image, storing the extracted features of database images in a HDF5 file format and then finding similarity between these features of our query image and images in database and hence displaying the result accordingly.

K. Simonyan and A. Zisserman proposed the VGG16 model in one of their research papers named “Very Deep Convolutional Networks for Large-Scale Image Recognition” in Oxford University. VGG16 is a convolutional neural network model. It was an upgrade over the existing model of AlexNet which has large kernel-sized filters. VGG16 (Fig 1.1) replaced these with multiple 3X3 kernel-sized filters. It has 16 convolutional layers with very small receptive fields of 3X3, 5 max-pooling layers of size 2X2 for spatial pooling followed by 3 connected layers.

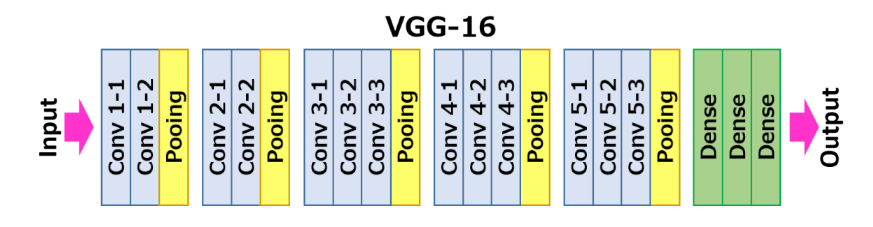


Fig 1.1: VGG16 Model

HDF stands for Hierarchical Data Format. It is a set of file formats for backing up and structuring files containing large amounts of data. It was being originally designed at National Center for Supercomputing Applications by the The HDF Group.

Here we are going to use the HDF5 (Fig 1.2) format which addresses some of the limitations of the HDF4 library.

It includes two major types of object:

* Datasets, in the form of multidimensional arrays (which we are going to use)
* Groups, to hold datasets and other groups.

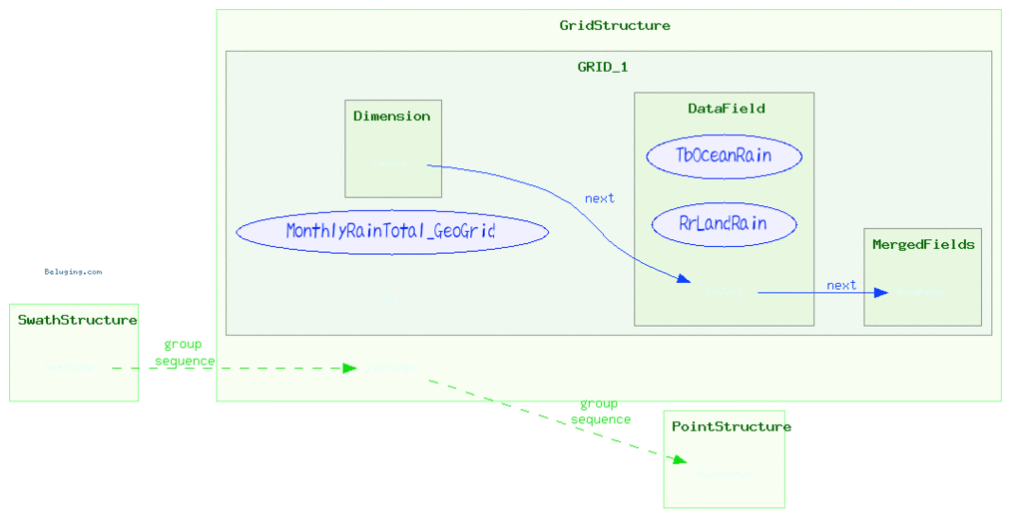


Fig 1.2: HDF5 Format

Moreover, the retrieval system includes similarity techniques includes algorithms to measure how close any two documents are (which can be images too as in our case). Dot product or rather Euclidean distance and cosine similarity are the simplest and common choices of any simple retrieval system.

# PROBLEM STATEMENT

In order to compare an image with other images, we need to study the features of existing image set. For that we need to design a model in order to extract the features from the images, store them in some form of data structure. Then we are going to extract the feature of the query image and compare them with the pre existing features and display the top ranked results.

# LITERATURE REVIEW

In the research paper “Study on Content based Image Retrieval” [1] by Shabnam Kumari, Reema, Yashika Kadian, they have discussed that Content Based Image Retrieval Systems (CBIR) is a topic of great importance in this current era of internet. Filtering images based on content and finding images of best similarity to the given image is a requirement these days. So, we need to find some faster retrieval system that can extract features from a image and can search a database based on the query image.

Guoyong Duan, Jing Yang, Yilong Yang in their paper entitled “Content-Based Image Retrieval Research” have discussed the most important advantages of using CBIR as a key technology [2]. It also shows the disadvantage of the conventional method of CBIR and introduces the use of color, texture and shape for image retrieval. It also focuses on the various algorithms that can be used for feature extraction and representation and thus help in image comparison.

D. Latha1and Dr. Y. Jacob Vetha Raj wrote a paper “A Review on Different Categories of CBIR Methods” [3] in which they decided to discuss about different categories of CBIR available out there rather than discussing about the feature extraction from query image. Classifying the different CBIR categories help us to decide which method is suitable for which environment and application.

V. Ramaya in his research paper “Content Based Image Retrieval System using Clustering with

Combined Patterns” [4], has discussed a special case of CBIR system where the goal is to find a query object in the database of images, that is the images containing the query object. The results given in this research paper gives the efficiency of the system.

“Content based image retrieval using feature extraction with machine learning” is another research paper by Aasia Ali and Sanjay Sharma [5] in which the authors have described the process of the image retrieval using image contents and extraction of the features from the images without human interaction. The system also uses the BFOA (Bacteria Foraging optimizing algorithm) in order to reduce the complexity, cost, etc. of the process.

In the research paper by “Deep Learning for Image Retrieval: What Works and What Doesn't.” by Wang, Huafeng & Cai, Yehe & Zhang, Yanxiang & Pan, Haixia & Lv, Weifeng & Han, Hao [6], the authors have discussed about how Convolutional Neural Network is in a leading position for feature extraction and indexing of images.

“Text-Based and Content-Based Image Retrieval on Flickr” is another research paper being worked on by Juan Manuel Barrios, Diego Diaz-Espinoza, Benjamin Bustos [7], in which the authors have improved the existing CBIR system with processing of both images and text. The system is going to use the different texts present in the title, tag and description of the images.

# METHODOLOGY

Development of a particular system involves a particular life cycle model to be followed in the whole process. The life cycle model that we are following is the Iterative model. Using this model, we design a simple version of the model and then enhance it for further functionalities.

VGG released two CNN models, the first one is 16 layered and the second one is 19 layered model. Here we are going to use the 16 layered model for feature extraction.

This project includes 3 major components: -

1. Feature Extraction
2. Feature Storage
3. Similarity Measures

Fig 4.1- Model of CBIR

## VGG16

VGG16 is a convolutional neural network (CNN is a type of Artificial neural network model being designed especially for images as it works best for pattern recognition and processing the pixel data for object recognition) which is being trained with a known dataset of “Imagenet”.

Imagenet is a project developed in support of researchers in the field of object recognition and includes a large database with over 14 million images that are classified under 20 thousand categories.

Convolutional neural networks are type of feed forward neural networks (much like artificial neural network with no connections between nodes forming up a loop). It works best with image like data. For a 1000 X 1000 image, normal neural network will need 10^6 nodes in the very first input layer. Due to this inefficiency to process each pixel, a new neural network called CNN was proposed to solve it up. They perform a series of operations – Convolution, pooling and normalization to extract high level features of the images and then fed to a fully connected neural network. So, it involves: -

* Convolution layers: It performs convolution on the images (say 7 X 7) with a filter matrix (say 3 x 3) initialized with random weights by multiplying the pixel intensities with corresponding value on filter and is then averaged which is used to replace the value of central pixel that this filter covers on our image. The borders of input image our made 0(since they are not neighbored with enough pixels). Filter is moved from left to right and top to bottom on our image and correspondingly values are updated. This results us with a reduced image size.
* Normalization layers: Since, our filter may contain negative as well as positive values as they are randomly initialized which in a way may somehow reduce some of the pixel intensity values to negative after convolution layer. Therefore, in normalization layer we pass each of the intensity value through ReLU (Rectified Linear Unit). It will change all negative values to 0 ensuring all negative pixel values to be 0.
* Max pooling layer: To remove unnecessary pixel values as all of them will not be needed for predictions, a pooling window of some size is moved across the image left to right without repeating the pixels so as they don’t overlap at all unlike convolutional layer and we take the maximum value from the window, replacing the whole window with a single pixel have that value as its intensity. This reduces the number of pixels by some factor.
* Fully Connected Layer: It flattens the resultant image to a fully connected vector or rather neural network as feature vector of our input image.

These layers are stacked upon each other and are used multiple number of times. And this all works for 3 channels- Red, Blue and Green.



Fig4.1.1: CNN Architecture

In particular, VGG16 has multiple 3X3 kernel-sized filters. It has 16 convolutional layers with very small receptive fields of 3X3, 5 max-pooling layers of size 2X2 for spatial pooling followed by 3 connected layers.

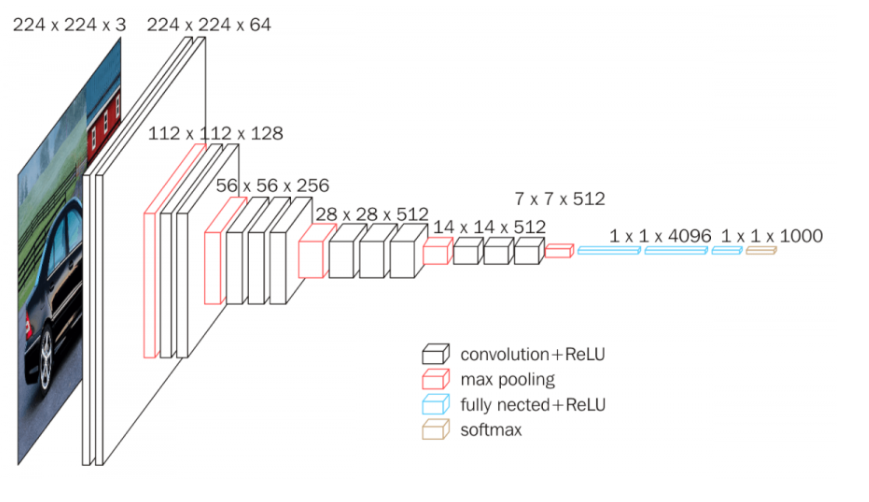


Fig 4.2: The VGG16 model architecture

## CUDA

CUDA is a computer application being developed by the NVIDIA corporation. We are using CUDA in order to run the python program on the CUDA enabled GPU.

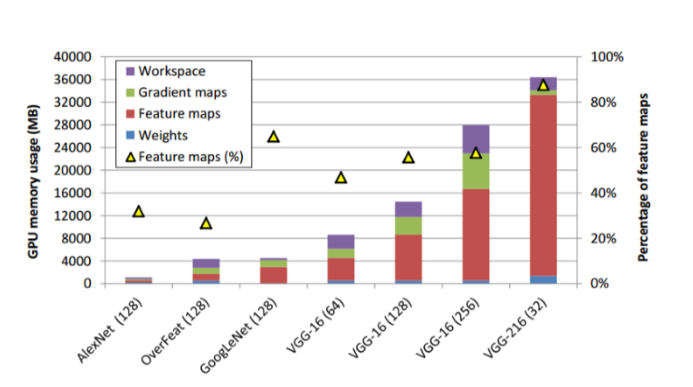


Fig 4.3: GPU Memory Usage of different models

## EUCLIDEAN DISTANCE AS SIMILARITY MEASURE

Similarity is a numerical measure of how alike two data objects are. Euclidean distance is one such metric to calculate it. It is just an ordinary straight line between two points in Euclidean space. The points in the space can be represented by a position vector. So, dot product of any two position vectors results in the distance between them. Lesser the distance, more is the similarity criteria. And in case of image, these position vectors are the feature vector of two different images. So, Euclidean distance can be calculated as:

Distance or Similarity = (Feature vector of image **A**) **.** (Feature Vector of image **B**)

# SCHEDULE

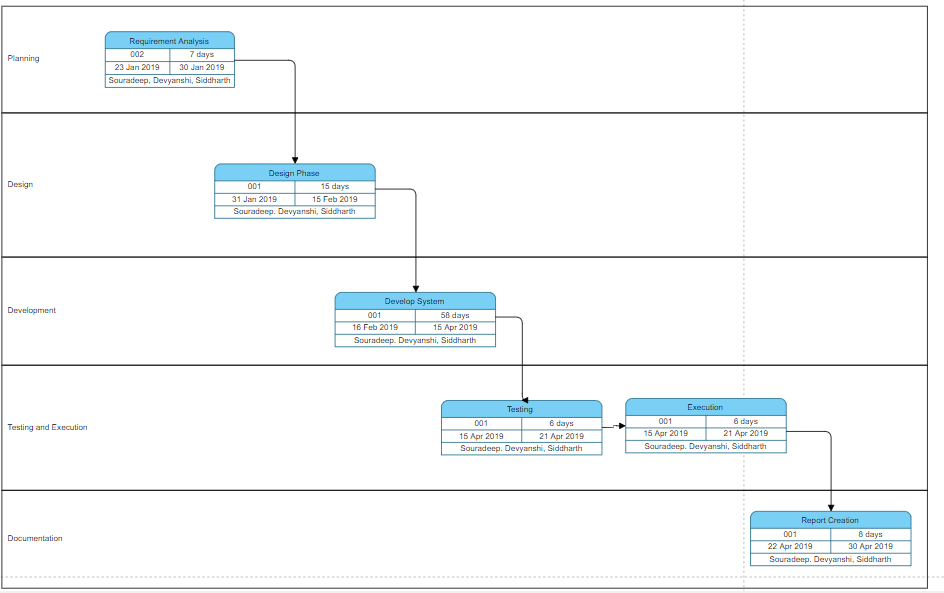


Fig 5.1: Pert Chart

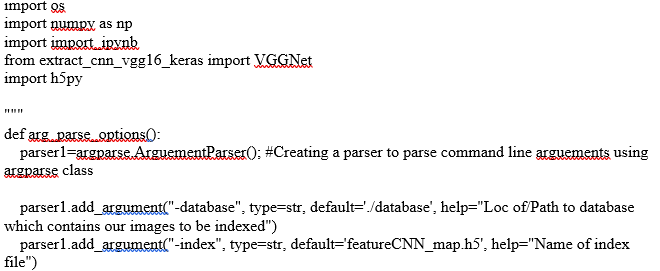
# IMPLEMENTATION

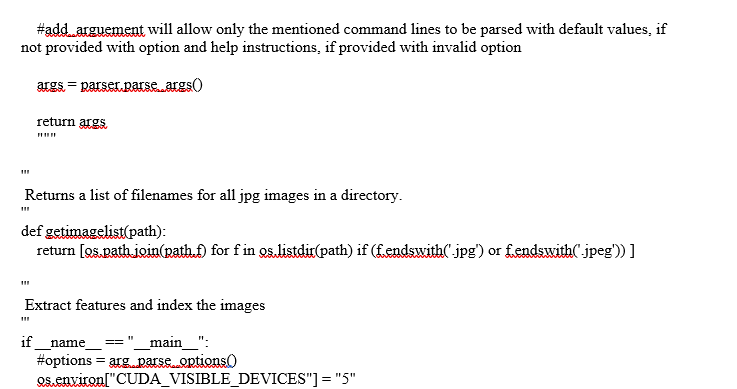
Our whole implementation part is being divided into 3 modules: Index, Query, and extract\_cnn\_vgg16\_keras.

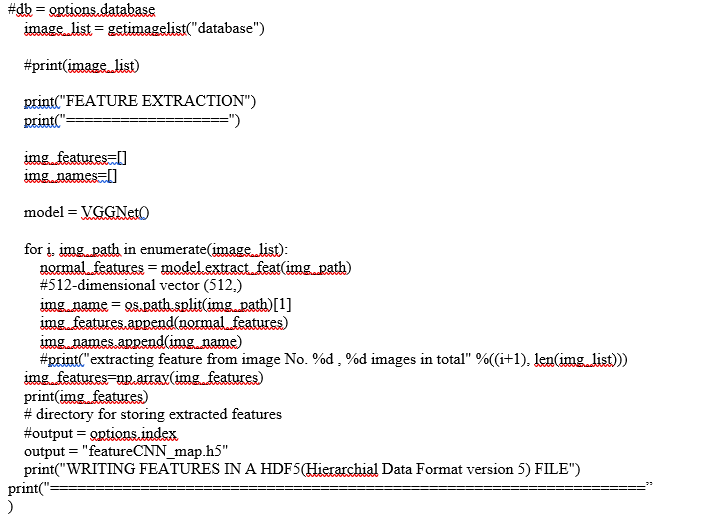
## Index.ipynb

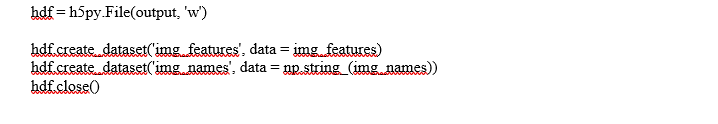
The Index.ipynb notebook is being used to drive the VGG16 model and extract the features from the images in the existing database and store them in a HDF5 format for being used later.

**CODE:**





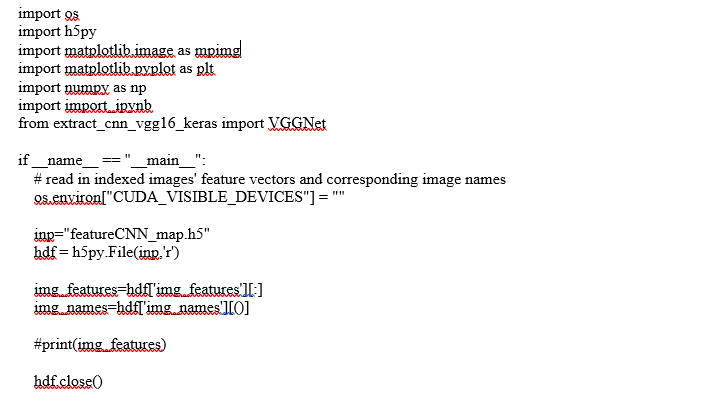


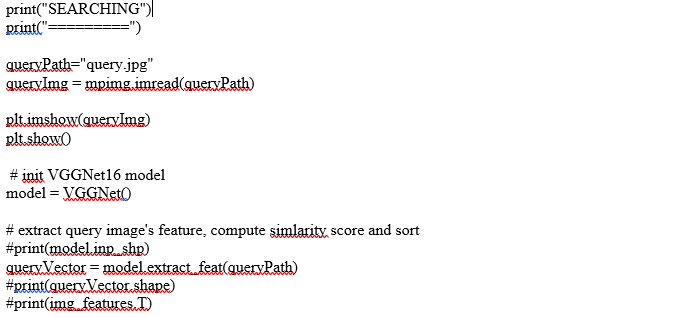


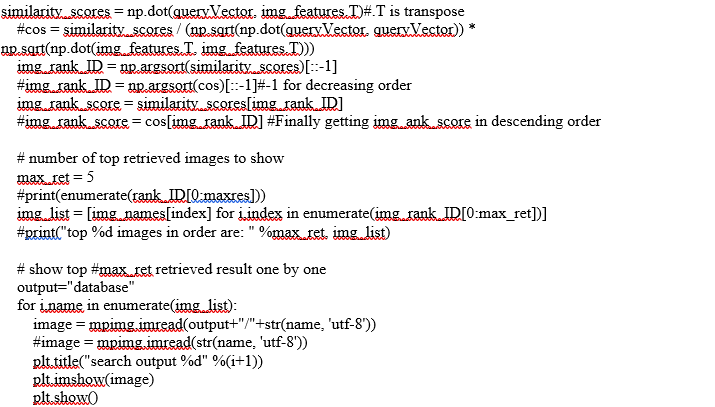
## Query.ipynb

This part of the code is to compare the input image being given by the user based on the features extracted from the input image and then relating it to the pre saved features in the HDF5 format. The top ranked results are being shown afterwards.

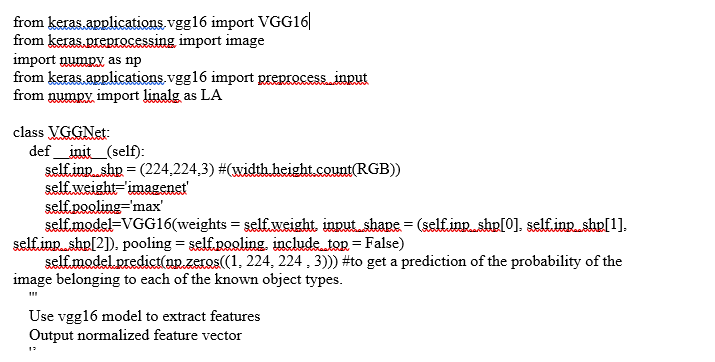
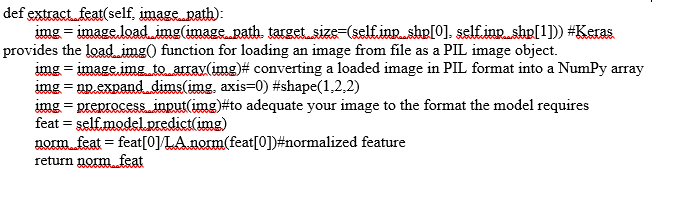
**CODE:**







## extract\_cnn\_vgg16\_keras.ipynb-VGG16 model

# RESULTS

The query image that we are using in our project is the following:



Fig 7.1: Query Image

We are going to find the most similar images in the data set using the pre trained VGG16 model and then display the top ranked results.

The results obtained are as follows:



Fig 7.2: Result 1



Fig 7.3: Result 2



Fig 7.4: Result 3



Fig 7.5: Result 4

# CONCLUSION

Using this project, we are able to compare images with one another on a detailed level using extracted features and can find similar images based on the same rather than doing it manually. We can increase the number of images in our data set in order to increase the accuracy of the system.

There are other models like VGG19, RCNN, etc. which can also be used for the same purpose in order to get higher similarity results and faster processing. In the end since we are using CUDA in this project, a better graphic card would lead to better and faster execution of the program.

# FUTURESCOPE

Using this project, we can make some android application or some web application which will run on the server end and thus relieve the client-side computer of all the heavy graphical processing task, in-order to make the project hardware independent. Again, we can also improve the model for training using some other libraries in python in order to get some better results with some better accuracy.

# REFERENCES

[1] <https://www.academia.edu/36976282/A_Study_on_Content-Based_Image_Retrieval> A Study on Content-Based Image Retrieval- Naveen Balaji

[2] <https://www.sciencedirect.com/science/article/pii/S1875389211007279> - Content-Based Image Retrieval Research by Guoyong Duan, Jing Yang, Yilong Yang

[3] <http://ijsrcseit.com/paper/CSEIT1831256.pdf> - A Review on Different Categories of CBIR Methods by D. Latha1, Dr. Y. Jacob Vetha Raj

[4] <http://ijsrcseit.com/paper/CSEIT1831237.pdf> - Content Based Image Retrieval System using Clustering with Combined Patterns by V. Ramaya

[5] <https://ieeexplore.ieee.org/document/8250625> - A. Ali and S. Sharma, "Content based image retrieval using feature extraction with machine learning," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, 2017, pp. 1048-1053.

[6]<https://www.researchgate.net/publication/300412100_Deep_Learning_for_Image_Retrieval_What_Works_and_What_Doesn't> - Wang, Huafeng & Cai, Yehe & Zhang, Yanxiang & Pan, Haixia & Lv, Weifeng & Han, Hao. (2015). Deep Learning for Image Retrieval: What Works and What Doesn't. 1576-1583. 10.1109/ICDMW.2015.121.

[7] <https://ieeexplore.ieee.org/document/5271936> - . M. Barrios, D. Diaz-Espinoza and B. Bustos, "Text-Based and Content-Based Image Retrieval on Flickr: DEMO," 2009 Second International Workshop on Similarity Search and Applications, Prague, 2009, pp. 156-157