Reverse Image Search Engine

Text and Content Based Image Retrieval

Team Members:

Aswath Rao V.S (16PD05)
Ridhanya M (16PD28)

Abstract:

Recurrent Neural Network for Content Based Image Retrieval Using Image Captioning Model

With the tremendous growth in the collection of digital images by social media, ecommerce applications, medial applications and so on, there is a need for Content based image retrieval **(CBIR)**. Automatic retrieval process is one of the main focus of CBIR, whereas traditional keyword based search approach is time consuming. Semantic based image retrieval is performed using CBIR where the user query is matched based on the perception of the contents of the image rather than the query which is in text format. One of the main research issue in CBIR is the semantic gap which can be reduced by deep learning. In this project we built an image captioning model for image retrieval based on the content using recurrent neural network (LSTM) and convolutional neural network (VGG16) in deep learning.

Applications of this project:

Image Search Engines are used for detecting copyright of images and logos. They are also used by various aerial Industries for detecting similar kind of regions in the entire Globe. Moreover, these applications are very computationally expensive.

Example of Already existing Reverse image search engines:

- **Google reverse image search engine,** which take in an image and returns you the most similar images in a fraction of a second

- **Pinterest** lets us to search the visually similar images of the selected objects.

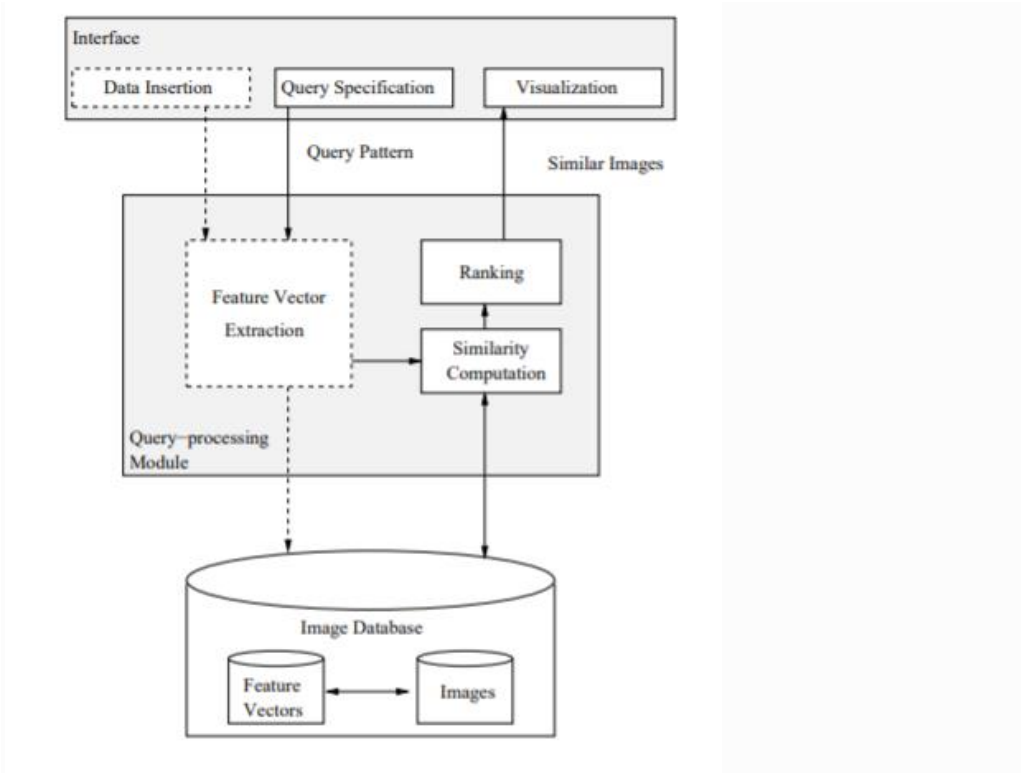
Overview:

Content-Based Image Retrieval (CBIR) aims to locate the specific image in a large collection of images without any meta-data. Current techniques focus on manipulating image pixel spaces (such as shape, color, and texture) but face two challenges:

- Limitation of representation.
- Inaccuracy of image similarity measurement.

We address these two problems by using **image captioning model** to generate rich textual meta-data, which translates the pixel space into text space where many textual similarity measurements and ranking methods can be applied. Our preliminary result shows this approach has great promise.

Architecture of CBIR Systems:



Typical architecture of a content-based image retrieval system

Description:

Two main functionalities are supported:

- ***Data insertion(Crawling and indexing)*** – In this step we use flickr API to crawl and download images from *https://www.flickr.com/* website. Next we used a pre trained CNN model **(VGG16)** for extracting appropriate features from images and used it to predict captions for those images using an **LSTM (Long Short Term Memory)** an Image Captioning Model and stores them into the image database (dashed line in the above architecture). This process is performed off-line.
- ***Query processing*** - In query processing, we have built and interactive user interface which allows user to specify a query by means of text or image and to visualize the retrieved similar images.

  The query-processing module extracts a feature vector from query image using **VGG16** CNN model and predict captions for that image using **LSTM** model and then applies a text similarity measurement **BLUE Score** to evaluate the similarity between the query image and the images in our database. Next, it ranks the database images in a decreasing order of similarity to the query image and forwards the most similar images to the interface module.

Note: The database images are indexed according to their feature vectors speed up retrieval and similarity computation.

Image Captioning Model:

Generating a textual description for an image is called image captioning. It is performed in two steps, Computer Vision and Natural Language processing. Computer vision is used to detect the image and understand the context of the image. Natural Language Processing to turn the image into words called captions with the correct sequence. The automatic generation of captions for images is a challenging problem for researches.

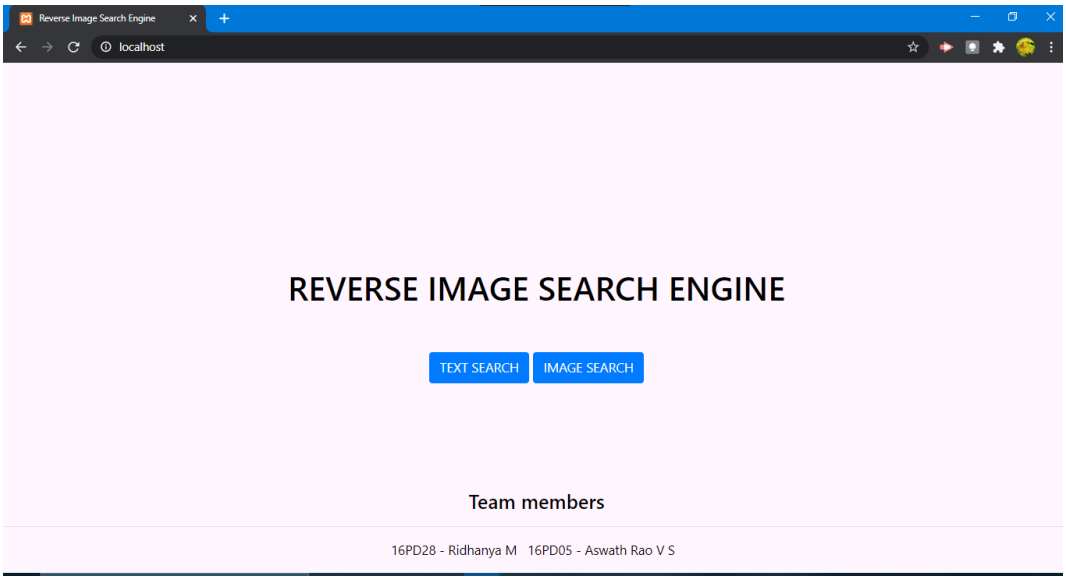
Content Based Image Retrieval (CBIR) which is a **semantic based image retrieval** matches user query based on some perceptual contents rather than text based query. The main issue in CBIR is to reduce the **semantic gap**. The role of image captioning in content based image retrieval is to achieve semantic similarity by capturing the textual data associated with the images.

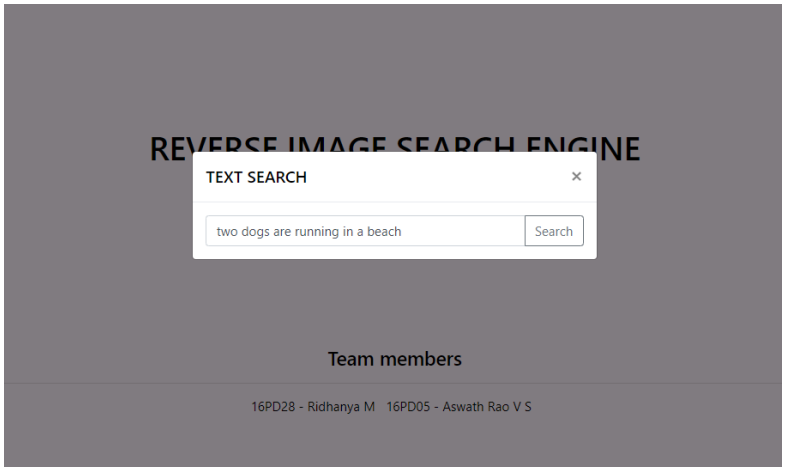
In this project we focus on image caption based processing system using deep learning for CBIR application. Convolutional neural network **VGG16** is used for image classification and Recurrent Neural network **LSTM** is used for caption generation. The resulting caption of the Image is used for retrieving similar images.

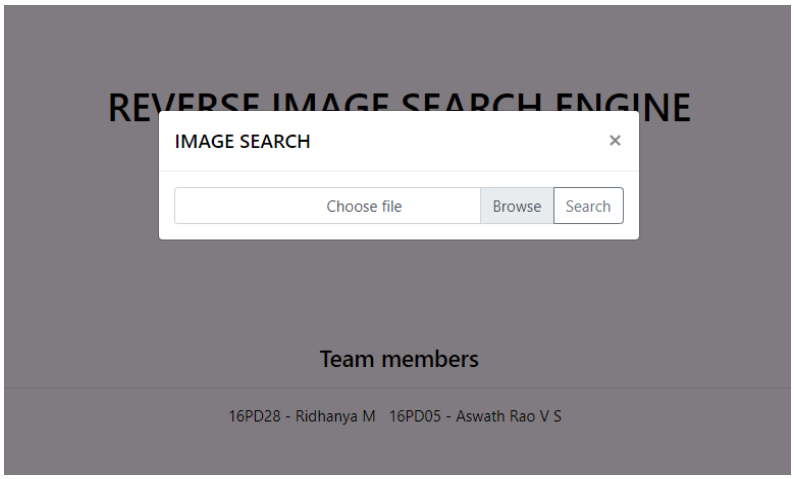
To know more Image Captioning Model, please refer the below link:

[Documentation of Image Caption Generating model]

User Interface:







Result:

Text Search:

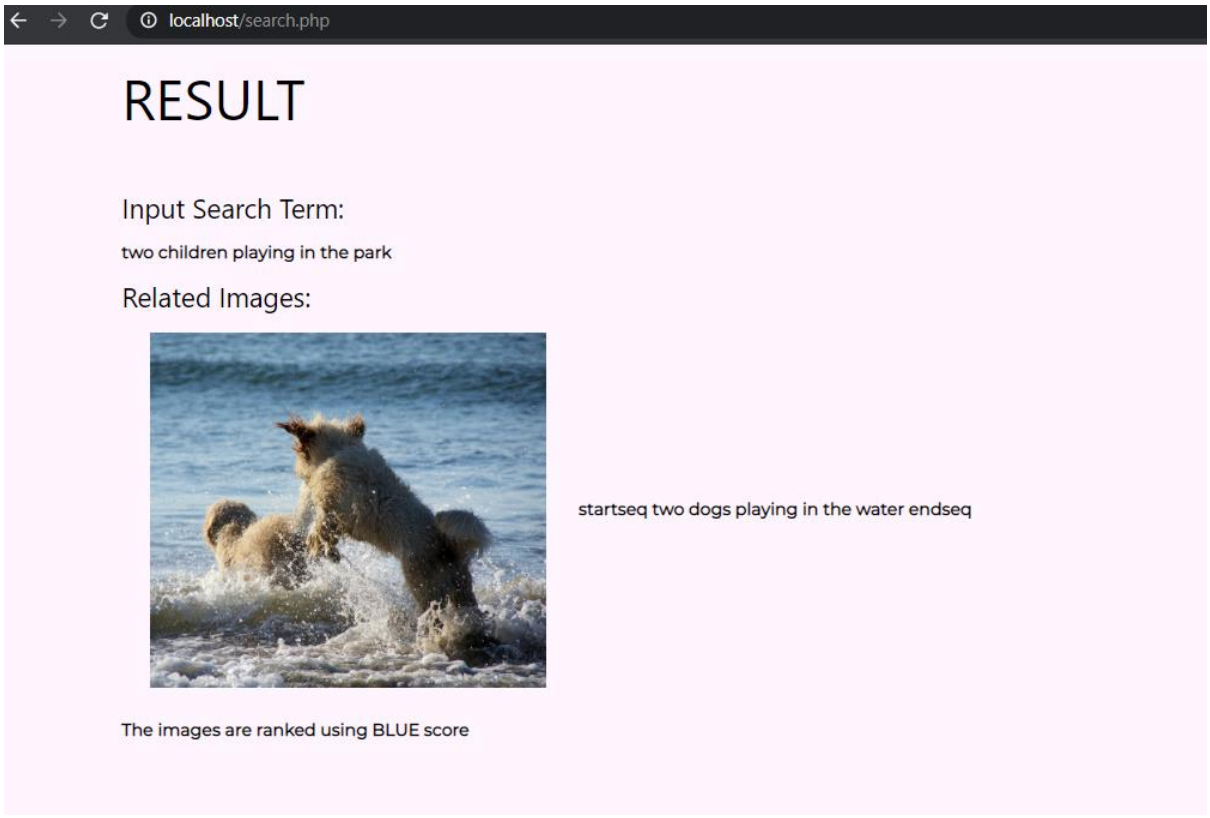
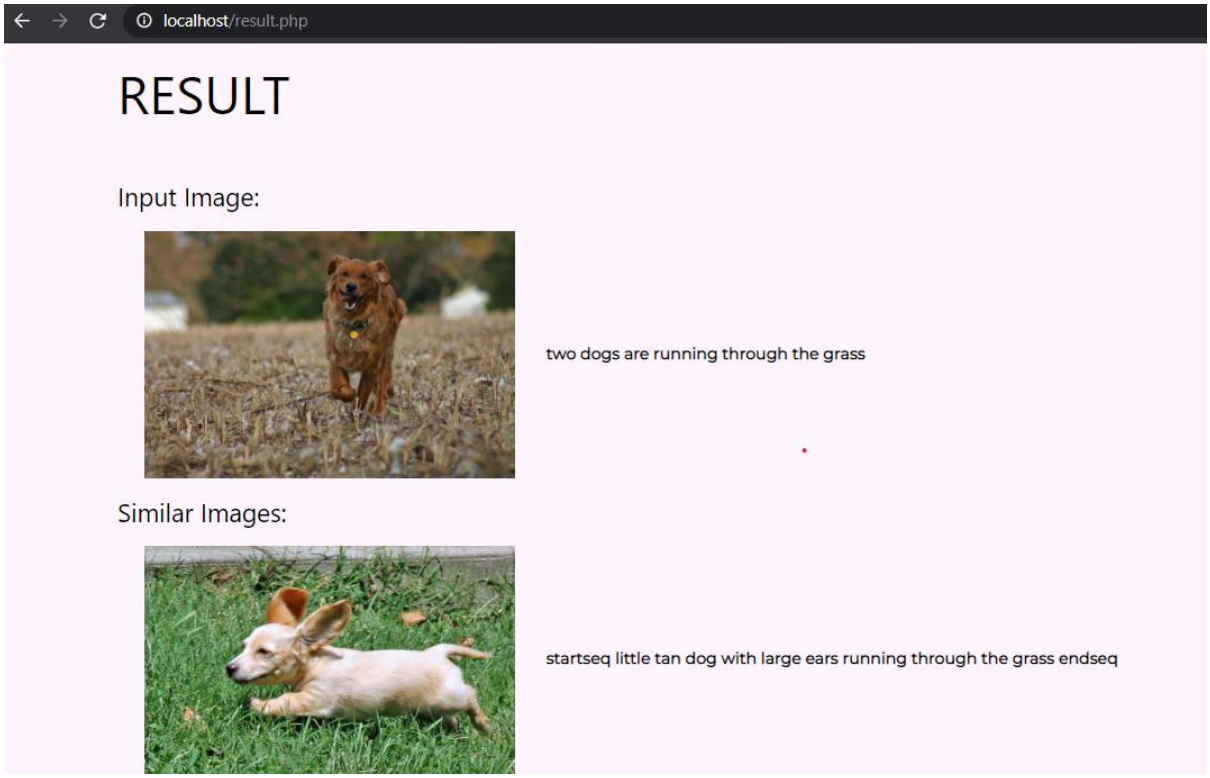
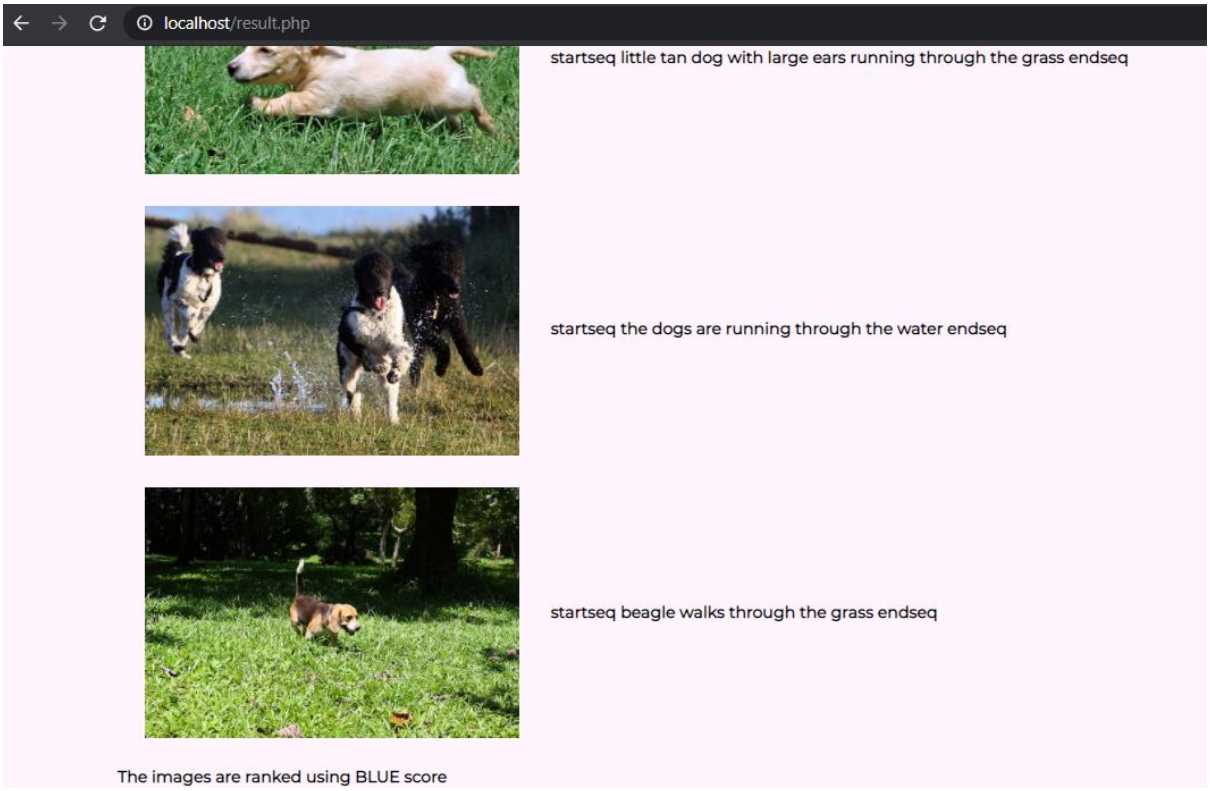


Image Search:





Future Exploration:

- Query by sketch for fashion domain.
- Train better LSTM model by changing model parameter, Using large dataset to improve its predictions.
- Try other approaches for CBIR.
- Build a Reverse image search engine which takes 3 input Query Image, Positive Image and Negative Image. Similar to our model build a CNN which takes those images and produces a feature vectors. Now that we have got three feature vectors, and we have to train our network such that the loss function should maximize the distance between the query image and negative image and minimize the distance between the query image and positive image simultaneously. By defining triplet loss as the model objective function.
- The same technique can be applied to build **facial recognition systems**.

References:

Content-Based Image Retrieval: Theory and Applications

CBIR Research Using Deep-Learning Techniques

Text-Based Image Retrieval Using Deep Learning

A Comprehensive Survey of Deep Learning for Image Captioning

Image Captioning with Deep Bidirectional LSTMs and Multi-Task Learning

A Review of Wrapper Feature Selection in Content Based Image Retrieval Systems