Content-Based Image Retrieval using Convolutional Neural Networks

Multimedia Computing Assignment

By:
Ameesha Mittal (2014A7PS107P)
M Sharat Chandra (2014A7PS108P)

Under the guidance of:
Prof. Sundaresan Raman

Introduction

- CBIR or content-based image retrieval is the application of multimedia and computer vision techniques to retrieve images similar to an image from a large database of images.
- The key challenge has been ascribed to the outstanding "semantic gap" issue that exists between low-level image pixels caught by machines and high-level semantics captured by humans.
- Our approach is to use Convolutional Neural Networks as they appear to give astounding results in the context of content-based Image Classification.

Need of CBIR

- CBIR is desirable because searches that rely purely on metadata are dependent on annotation quality and completeness.
- Having humans manually annotate images by entering keywords or metadata in a large database can be time consuming and may not capture the keywords desired to describe the image.
- Also, different users tend to use different words to describe a same image characteristic.
- Thus, the lack of systematization in the annotation process decreases the performance of the keyword-based image search.

Applications of CBIR

- **Medical Applications**: The number of medical images generated each day at different hospitals and medical centres is far too much to be efficiently used at maintained using metadata.
- **Art collections and Museums**: Another example is the digital museum of butterflies, aimed at building a digital collection of Taiwanese butterflies. This digital library includes a module responsible for content-based image retrieval based on color, texture, and patterns.
- **Commercial Systems**: Many modern commercial systems are based on CBIR. IBM's QBIC, Excalibur's Image RetrievalWare, VisualSEEk and WebSEEk and iSearch - PICT to name a few.
- **World-wide web**: Recent explosive progress of WWW (World-Wide Web),has resulted in a huge image database online.

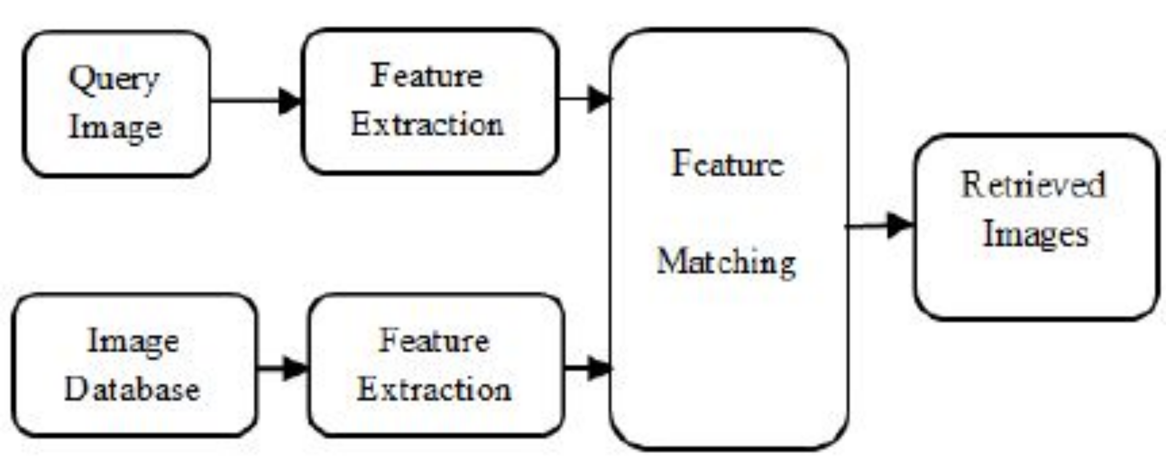
Major challenges in CBIR

- Image Representation (size and resolution)
- Feature extraction
- Image Similarity Evaluation
- Image Annotation/classification
- Image Indexing and Database Organization
- Query Formulation
- Query Result Display and Assessment
- Users’ Feedbacks & Updating

Traditional Approach to CBIR

- Traditional architecture of CBIR systems is quite similar to that of classical retrieval systems.
- It includes two basic modules: indexing module and retrieval module.
    - The former is responsible for data processing and constructing indices, which considerably speed up the search.
    - The latter takes care of the retrieval itself by the user request.
- These approaches try to search independently by different image features.
- Each of image feature is represented by a point in the corresponding feature space.
- These feature spaces are used to then perform the task of retrieval.

Illustration of a typical CBIR Architecture



Developments in the Past Decade

- Approaches to Feature extraction
- Approaches to Retrieval
- Annotation and Concept Detection
- Approach using Neural Networks

Approaches to Feature extraction

- **Color and Texture**: A region based dominant color descriptor indexed in 3-D space along with their percentage coverage within the regions was proposed which is argued to be more efficient than than high dimensional histograms in terms of search and retrieval.
- **Shape**: Shape Context for shape matching is fairly compact yet robust to a number of geometric transformations.
- **Segmentation**: Normalized cuts for image segmentation.
- 2-d multiresolution hidden markov models for characterizing spatial arrangements of color and texture.

Approaches to Retrieval

There has been a large number of fundamentally different frameworks proposed in the last few years like:

- Region based image retrieval
- Vector quantization (VQ) on image blocks
- Windowed search
- Anchoring based image retrieval
- Probabilistic frameworks for image retrieval

Deep Learning for CBIR

- Deep learning refers to a class of machine learning techniques, where many layers of information processing stages in hierarchical architectures are exploited for pattern classification and for feature or representation learning.
- Recently, it has become a hot research topic in computer vision and machine learning. This increased popularity is mainly due to increased chip processing abilities, lowered cost of computing hardware and recent advances in machine learning and signal/information processing research.
- We introduce the a deep learning framework for CBIR, which consists of two stages: (i) training a deep learning model from a large collection of training data and (ii) applying the trained model for learning features of new images. We have adopted the architecture of Convolutional Neural Networks for this project.

Convolutional Neural Networks (CNN)

- CNNs are quite similar to normal Neural Networks. They are made up of neurons having learnable weights and biases.
- Convolutional Neural Network try to model the Visual Cortex part of the brain and hence perform extremely well in the area of Computer Vision.
- However, these networks use a special architecture which is particularly well-adapted to classify images. This, in turns, helps us train deep, many-layer networks, which are very good at classifying images.
- The architecture of a typical CNN is composed of multiple convolutional and subsampling layers. Each layer perform a specific function to transform its input into more useful representation.

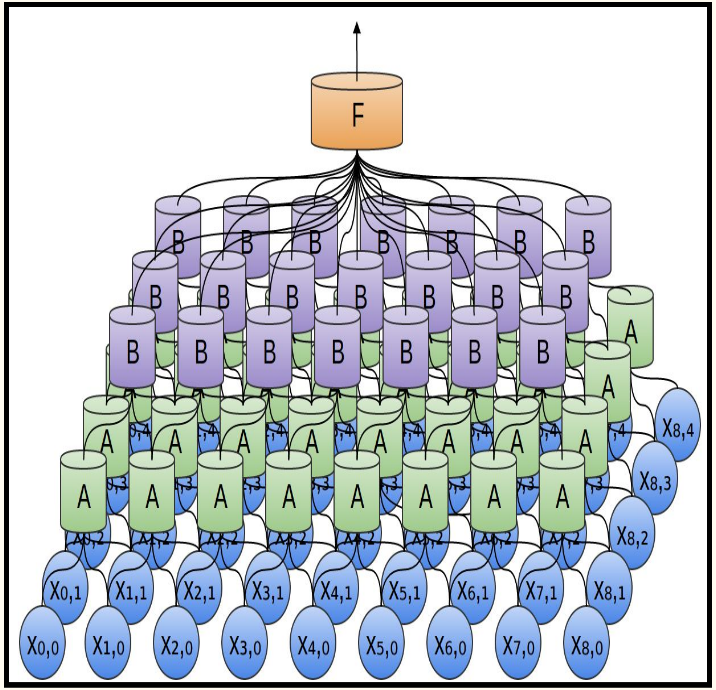
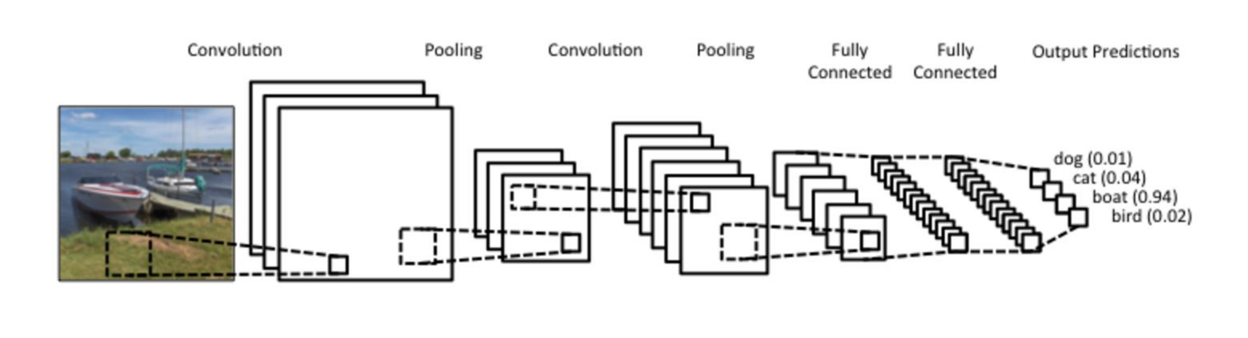


Figure : A 2D Convolutional Neural Network with Convolutional and Subsampling Layers

How CNNs work

There are four main steps in CNN: convolution, subsampling, activation and full connectedness. They are briefly described below.

- **Convolution**: Application of filters over the images.
- **Subsampling**: Done in order to converge the network into smaller portion by extracting the important information from the lower layers.
- **Activation layer**: Controls how the signal flows from one layer to the next, emulating how neurons are fired in our brain. Most common activation function used with CNN is Rectified Linear Unit.
- **Fully connected layer**: It aggregates all the information and is then passed through a softmax function to provide the class probabilities.



Feature Representation for CBIR

Various schemes of proposed feature representation are:

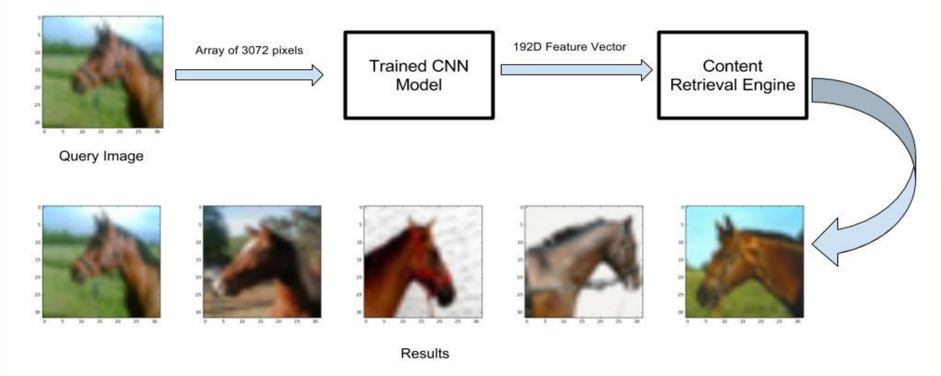
- **Direct Representation**: In this the images are fed directly in new datasets into the input layer of the pre-trained CNN model, and then take the activation values from the last n layers.
- **Refining by Similarity Learning**: In this instead of directly using the features extracted by the pre-trained deep model, we attempt to explore similarity learning (SL) algorithms to refine the features.
- **Refining by Model Retraining**: In this scheme, we will retrain the deep convolutional neural networks on the new image dataset for different CBIR tasks by initializing the CNN model with the parameters of the ImageNet-trained models by the method of Transfer Learning.

Our Approach

- We aim to build a Neural network architecture which will serve as a model to perform the task of Content based Image Retrieval
- We are using a slightly modified version of the Direct Representation Scheme.
- The activation values from the penultimate fully connected layer will be used to generate an image descriptor [A feature vector of 192 dimensions that represents each image].
- This image descriptor will then be used to find out similarity between images.The similarity coefficient will determine the retrieval of the closest matching images for the given query image from our dataset.

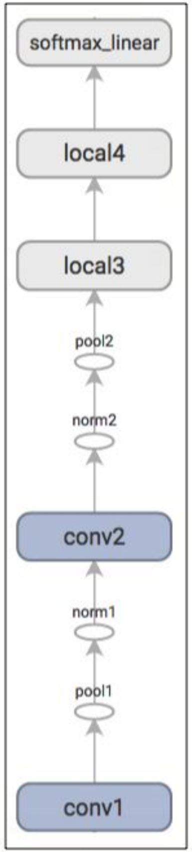
Proposed Architecture of CBIR Model

- We first transform the query image to a 32x32 pixel JPEG image.
- We then feed in the transformed image to the trained CNN model which produces the activation values of the penultimate layer as the feature vector.
- This vector is taken in by the Content Retrieval Engine which looks for the images in the database and retrieves the top matches.



Trained CNN Model

- Our CNN model uses alternating convolutional and normalization layers followed by fully connected layers towards the end. All of these layers have weights and biases which get updated in every iteration.
- While training, images are passed through these layers wherein various filters are applied and the extracted features are carried forward and predictions are made.
- The weights and biases of these layers are modified according to the predicted output class of the image by minimizing cross entropy using stochastic gradient descent algorithm.



Content Retrieval Engine

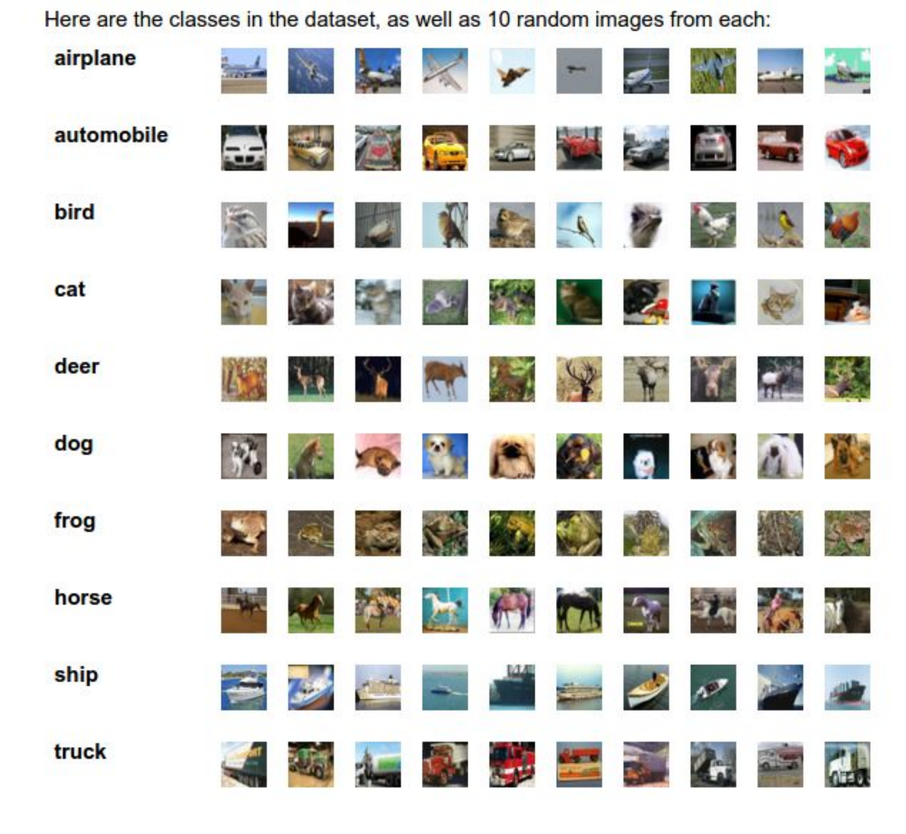
- The Content retrieval engine mainly performs two actions: Database Building and Query Handling.

- Database Building: It builds a database of final feature vectors of images and indexes them properly. KDTree data structure has been used to store the image vectors and their ids which are the indices.

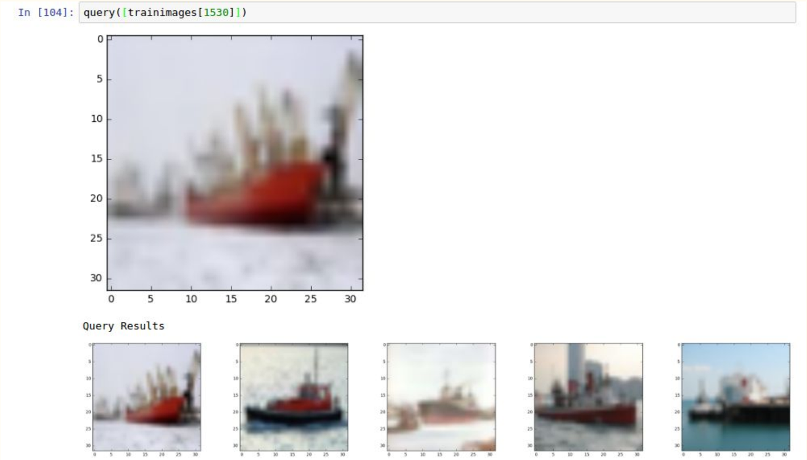
- Query Handling: It provides function calls which take in the feature vector of query image and retrieves similar images based on the features This is done by loading the KDTree saved earlier and supplying the feature vector of the input image to this structure. KDTree returns the ids of similar images which are then retrieved and displayed.

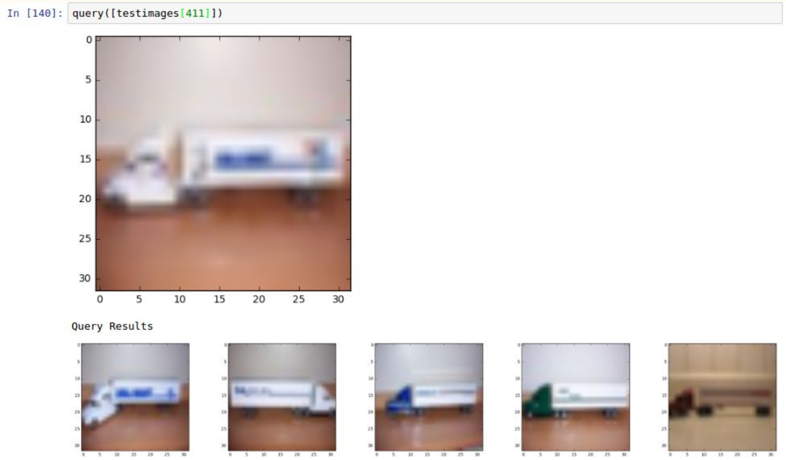
Image Dataset

- The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images.
- Each image is represented by an array of 3072 pixels. These pixels are arranged as 1024 Red, 1024 Green and 1024 Blue pixel values respectively.

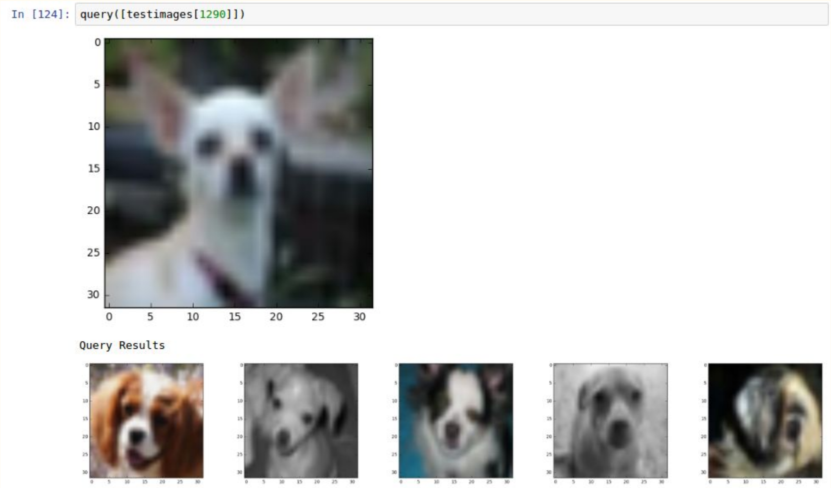


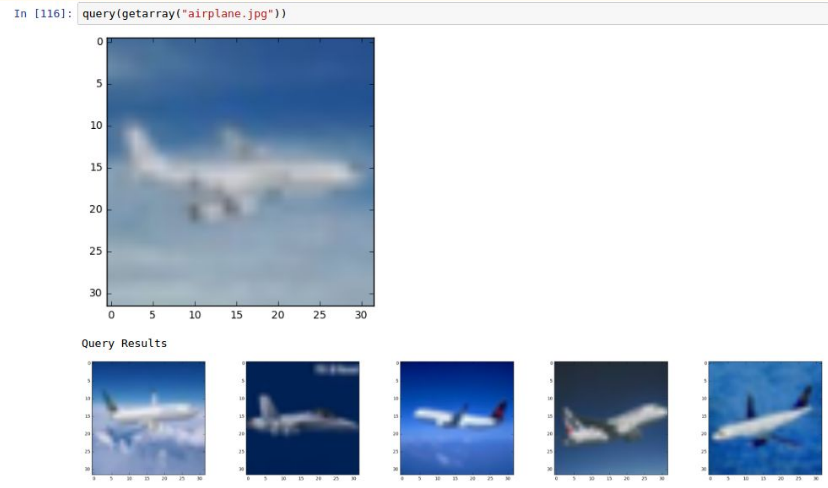
Results


In [104]: query([trainimages[1530]])
Query Results


In [140]: query([testimages[411]])
Query Results

Results


In [124]: query([testimages[1290]])
Query Results


In [116]: query(getarray("airplane.jpg"))
Query Results

Conclusions

- Inspired by recent successes of deep learning techniques, we attempted to address the long-standing fundamental feature representation problem in Content-based Image Retrieval (CBIR).

- In particular, we developed a model of deep learning with application to CBIR tasks by examining a state-of-the-art deep learning method (convolutional neural networks) for CBIR task using the CIFAR 10 dataset.

- The results obtained were quite promising. Though this method was computationally expensive, it achieved really good results with an accuracy of 85%.

Individual Contributions

Ameesha:

Background Study, Understanding the working of CNN architecture, modifying the available template for extracting image vectors, handling new inputs, generating output images from received vectors.

Sharat:

Study regarding deep learning and its application in CBIR, Understanding the CNN architecture, creating properly indexed database for storing image vectors, handling retrieval of similar images.

THANK YOU