Image Based Search Engine Without Metadata*

Bhakti Batra(MT19115) Anmol Jain (MT19005) Shubham Verma(MT19017) Mukul Sharma(MT19042)

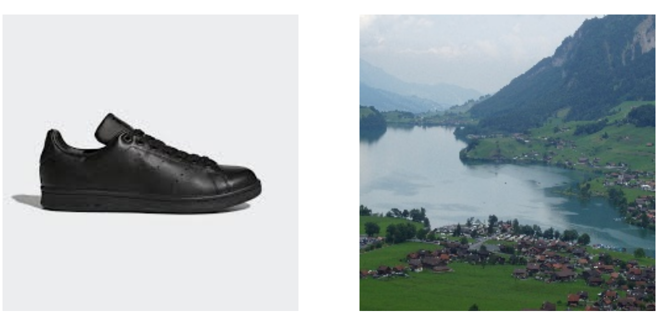
***Abstract*— This project attempts to create an image based search engine which searches the relevant image from the set of images. In this we have used multiple feature extraction techniques along with deep learning model to extract features and applied similarity measures to reach out to most relevant search results.**

I. Problem Statement and Motivation

Content based image retrieval is also known as query by image content. Content based means that the search analyzes the content of the image rather than the metadata such as keyword tags or descriptions associated with the image[1]. To search for a relevant image from an archive is a challenging research problem for computer vision research community. Most of the search engines retrieve images on the basis of traditional text-based approaches that rely on captions and tags or other metadata associated with the image. This approach fails when a tag or a keyword represents multiple entirely different object. For example; apple tag may represent a fruit or logo of Apple Inc. which are entirely different images. Moreover single image may associate with different tags for different users which sometimes leads to inappropriate tagging of images. With increasing number of images in search space, it is not practical to tag all images. Content-Based Image Retrieval extracts all relevant features like colour, colour intensity, object shape and does indexing based on all the relevant features and constructs a multidimensional vector and then all further operations like feature matching, feature transformation and processing are done on this feature vector[2].

II. Method

A. Dataset



(a) Shoe Dataset (b) Holiday Dataset

Fig. 1: Dataset Used

CBIR is implemented on Shoe DataSet that has more than 14000 images belonging to various categories of shoes. The dataset is divided into 4 main categories that are further divided into sub categories. Another dataset that has been used is Holiday dataset. It contained high resolution images of various holiday destinations. A subset of the dataset with more than 800 images was used for image search.

B. Preprocessing

Holiday dataset images were of different dimension and aspect ratio. In order to train models for these images, they were scaled to minimum dimensions present in the dataset (240px X 240px). The images were not square therefore images were cropped from the centre for further implementation.

Images in shoe dataset were clear and did not require much preprocessing. Unlike Holiday Dataset, all images were of same dimensions.

C. Feature Extraction

To minimize search time and get more relevant images , we extracted features of about 1000 dimensions from an image using these feature extraction techniques and got better results on combining some of them. Some techniques and their uses are as follows

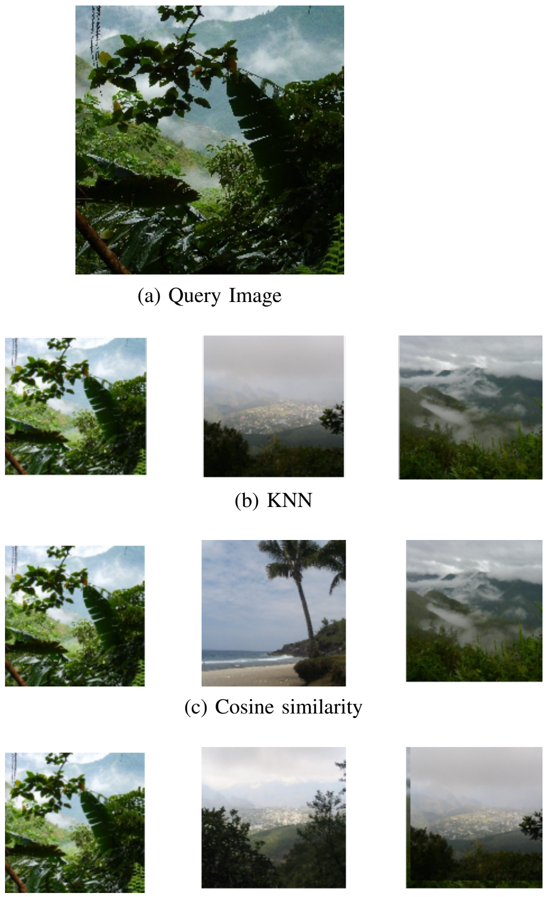
- Histogram matching for images is an important feature for searching similar images as it gives a measure of similarity in color distribution. Similar images have similar color distribution over the image.
- Segmented Histogram feature was used to overcome drawbacks of simple histogram where important features are localized in a segment of an image rather than whole image.
- Hu Moment feature is a global feature extraction technique used to extract shape of an important object from images.
- Hog Feature is an edge detection technique which works on the principal of histogram.
- VGG16 is a convolution neural net (CNN) architecture which is a pretrained deep learning model, which is used to extract overall relevant features from image[3].
- Haralack feature is used to detect texture in an image.

D. Retrieval Techniques

1) K Nearest Neighbours: For a given image, its distance from all images in the dataset is measures and top k nearest images to the query image are retrieved. Analysis was one on multiple distance metrics and it was seen that different query images, show slightly varying results for different metrics.

2) Cosine Similarity: It is a robust and widely used metric to measure similarity between images. Cosine similarity generally produces efficient results compared to basic distance measures like Manhattan Distance.

3) Similarity Voting: Here multiple models like kNN with variations in distance metric or feature set are used and their results are combined to provide more efficient image searching.



(a) Query Image

(b) KNN

(c) Cosine similarity

(d) Voting Similarity

Fig. 2: Top 3 retrievals on Holiday Dataset

III. Experiment and Results

- Various combinations of features were used to retrieve relevant images. Histogram and VGG proved to be most significant features and gave robust results.
- Features such as HOG, Hu Moments and Haralack are not capable of providing efficient results. They can only be used as support to make system more robust.
- Degree of importance of a feature or a feature set varies for each query image or the object in a particular image.
- Lack of prominence of an object (lack of present of any important feature) may lead to inefficient query results of images.
- Increasing size of training dataset, improves the efficiency of the result. In other words, efficiency increases with increase in size of the training dataset.

(a) Query Image

(b) KNN

(c) Cosine similarity

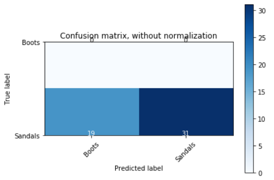
(d) Voting Similarity

Fig. 3: Top 3 retrievals on Shoe Dataset

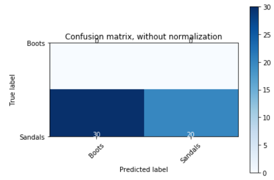
- Shoe dataset is divided into 4 major categories and each category is assumed to be ground truth class for image belonging to that particular class. This is used as a measure of evaluation of relevance of results.
- Model predicts 62% accuracy for k-NN, 60% for cosine and reports 74% accuracy for Voting Image Retrieval.
- Precision for k-NN is 0.7, Cosine is 0.68 and for Voting is 0.74.
- Recall for k-NN is 0.6, Cosine is 0.54 and for Voting is 0.62.
- Voting Image Retrieval is state of the art relevant image retrieval for CBIR

IV. Observation

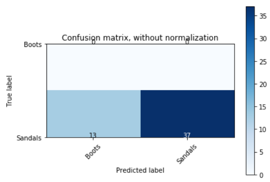
- Histogram is an important feature for retrieving relevant images.
- Segmented histogram has an advantage over normal histogram as it incorporates localization of color distribution in images.
- VGG is a powerful feature extraction technique that give promising results.
- When used in combination with VGG or histogram, other features such as hog and hu moments help minimize noise. It other words, they make ranking of images more efficient.
- Cosine similarity measure is capable of producing more



(a) KNN



(b) Cosine similarity



(c) Voting Similarity

Fig. 4: Confusion matrix of Shoe Dataset

relevant images on top. However, as list size increases, its efficiency decreases.

- kNN in itself may be very naive but when multiple kNN are used in combination, results produced are more reliable. Two types of distance measures are used for clustering - manhattan and euclidean distance and performance is evaluated.

References

[1] Swapnalini Pattanaik, D.G.Bhalke, “Beginners to Content Based Image Retrieval”,International Journal of Scientific Research Engineering Technology (IJSRET),Volume 1 Issue2, pp 040-044, May 2012.

[2] Z. Rian, V. Christanti and J. Hendryli, "Content-Based Image Retrieval using Convolutional Neural Networks," 2019 IEEE International Conference on Signals and Systems (ICSigSys), Bandung, Indonesia, 2019, pp. 1-7, doi: 10.1109/ICSIGSYS.2019.8811089.

[3] S. Jain and J. Dhar, "Image based search engine using deep learning," 2017 Tenth International Conference on Contemporary Computing (IC3), Noida, 2017, pp. 1-7, doi: 10.1109/IC3.2017.8284301.