Search Engine Architecture - Project Proposal

Objective:

Our goal is to provide the most relevant set of images when we query using another image. This is helpful in finding other websites potentially using similar images, finding similar images of different resolutions, and understanding the content the image represents.

Below are the main components for this task.

1. **Dataset**: We plan start by restricting our scope to Flickr dataset and MS Coco dataset and making this reverse-searchable. This contains a set of images along with their meta data like captions, tags that could also be used in feature generation. The Flickr dataset is available under the Creative Commons license.
2. **Image representation**: Images can be arbitrarily large, hence, making them searchable requires us to reduce the dimensionality of these images, so that the similarity searches can be performed efficiently, at a high speed. Some of the traditional techniques that are used to reduce dimensionality are PCA-SIFT, SIFT, and SURF. We will experiment with some of the newer nonlinear dimensionality reduction techniques like stacked autoencoders. To obtain a more semantic understanding of the objects/actions contained in an image, we would also explore Neural networks like Fast R-CNN, to obtain feature vectors to be used in similarity searches.
3. **Indexing:** Once an image representation is obtained, clustering algorithms can be used to determine the shards the images should belong to, and similar techniques are used in image segmentation. As we add new images to our dataset, we can re-index by re-running the clustering algorithm. Simple K-NN clustering algorithms are expected to work reasonably well. We will also explore CNN based hashing techniques, k-d trees and best-bin-first algorithms for K nearest neighbor problems.
4. **Retrieval:** When this dataset is queried the query image is represented in the same format in which the images are stored, and top k similar images will be displayed along with a snippet of their metadata.
5. **Potential extensions:** Once our baseline system is working we will explore extending the amount of indexed data and potentially investigate caching to speed up query time. Given time, we would like to include multi-modal search - hybrid queries that include images and text.
    a. Querying with an image and text : Eg. image of a cat and “cats on a tree in New York”

        i. The approach here would be to generate query representations which would most accurately capture both objects in the image along with ideas that the text implies.
    b. Querying with an image and style : Eg. image of a cat with a particular style.
        i. The approach here would be to detect the style of an image using deep learning techniques like Neural Style transfers.

6. **External tools:** PyTorch (nonlinear dimensionality reduction), scikit learn (clustering) , openCV(SIFT, SURF)

References:

1. Ranking : http://users.eecs.northwestern.edu/~jwa368/pdfs/deep_ranking.pdf
2. CNN based hashing: https://arxiv.org/pdf/1509.01354.pdf
3. K-d trees: https://en.wikipedia.org/wiki/K-d_tree
4. Applying kNN on image classification: http://cs231n.github.io/classification/#summaryapply