Constrained Clustering of Images with Textual (or Visual) Explanation

Advanced Topics in Machine Learning- Semester Project, 2022

Motivation

The goal is to make constrained clusters on a PASCAL Visual Object Classes Challenge (VOC 2007) data set. We must extract features of the images given in the dataset so that we have numerical features of images and can be compared. The user provides a number of images to be clustered upon and a query of image. kNN is used to filter out the number of images on which clustering mut take place.

The problem: The main problem is that our data is image, and it is not something which the machine understands and must be converted to something on which computations can take place. We must rely on feature extraction algorithms to help describe an image and must assume that the algorithms are providing information which will help us in our clustering process. The questions which we ask is that which features best help describe the image and how the constraints and clustering change with change in featuring method.

Data, Constraints Creation & Constraints Clustering

Dataset: The PASCAL Visual Object Classes Challenge (VOC 2007) data set consists of about 9,000 images of 20 classes. The twenty object classes that have been selected are:

- person
- bird, cat, cow, dog, horse, sheep
- aeroplane, bicycle, boat, bus, car, motorbike, train
- chair, dining table, potted plant, sofa
- tv monitor

Constraints Creation: Here we will talk about how the constraints are being generated for the clustering process. We first create a distance matrix between the candidate images (i.e candidate images are the images that have been selected using kNN for a given query image) and then set the number of must link neighbourhoods (i.e a must link neighbourhood is a set of images between which must links are to be generated) to be created. We are setting the number of must link neighbourhoods to the number of object classes found in the candidate images. We are creating these neighbourhoods by taking any image at random from our candidate images and then using 2NN on the distance matrix. We put the random image and its two nearest neighbours as one must link neighbourhood.

Constraints Clustering: Used algorithms are COP-K-means and PCK-means. We are choosing COP-K-Means for demonstrating clusters with hard constraints and PCK-means for demonstrating clusters with soft constraints. We chose PCK-means over MPCK-means for hard constraints because of the explanation factor.

Feature Extraction

1. Bag of Visual Words (BOVW)

The general idea of bag of visual words (BOVW) is to present an image as set of features. Features consists of keypoints and descriptors. Keypoints are “stand out” points in an image and descriptor is the description of the keypoints.

The steps involved are as follows-
Extracting features locally: defining the size of the patch and the number of patches per image and using LBP features as a base descriptor. Obtaining features of LBP for each patch.

Cluster the local features: Using the clustering algorithm, KMeans, defining the number of visual words

Histograms from Bag of features: check the frequency of each visual word in each training image. First, computing features for each image and predicting the number of patches of images. Finally, computing

2. Moving Picture Experts Group (MPEG7)

MPEG-7 is a standard for describing features of multimedia content. The goal of the MPEG-7 standard is to provide a rich set of standardized tools to describe multimedia content.

The steps involved are as follows-
Image partition - Divide the image into 64 blocks, a grid of 8x8. The 8x8 grid ensures invariance to resolution or scale.

Representative color selection - A single color is selected as the representative of that block from each block.

Get a tiny image of 8x8 in YCbCr color space by converting color space from RGB to YCbCr.

DCT transformation is applied to each matrix of Y, Cb and Cr components for each block method of Opencv.

Scan each matrix in a zigzag fashion to group the low-frequency coefficients of the matrices.

3. Scale-invariant feature transform (SIFT)

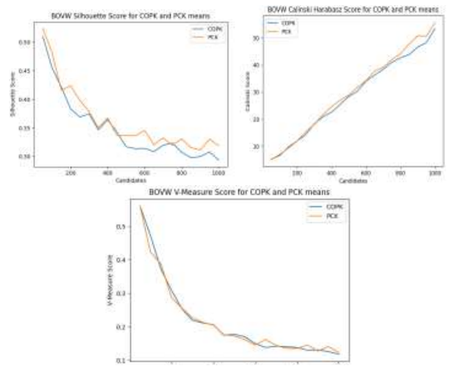
SIFT helps locate the local features in an image, commonly known as the 'keypoints' of the image. These keypoints are scale & rotation invariant that can be used for various computer vision applications, like image matching, object detection, scene detection, etc.

The steps involved are as follows-
Formation of Scale space
Detection of Keypoint using Difference of Gaussians (DoG)
Estimation of keypoint orientation using image gradients and direction histograms
Extraction of description of keypoints

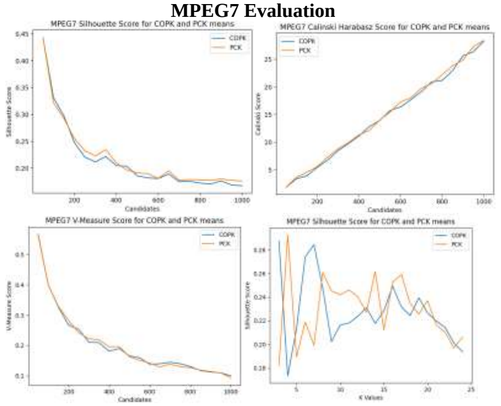
Evaluation

To evaluate performance of clusters on COPK and PCK means, we are using Silhouette method, V-Measure and Calinski-Harabasz Index.
Silhouette Score: In silhouette score, data set is used for measuring the mean of the Silhouette Coefficient for each sample belonging to different clusters.
V-Measure: It is defined as the harmonic mean of homogeneity and completeness of the clustering.
Calinski-Harabasz Index: It is defined as the ratio between the within-cluster dispersion and the between-cluster dispersion.
Query Image: “000036”

BOVW Evaluation



MPEG7 Evaluation



SIFT Feature Evaluation