Indian Institute of Technology Kanpur
CS783: Visual Recognition, 19–20

Student Name: Sarthak Mittal, Gurpreet Singh
Roll Number: 150640, 150259
Date: February 5, 2019

ASSIGNMENT 1

1. Problem Statement

We are given a dataset of images of several objects based on which we have to train a model that can recognize instances of different objects in any given test image. Basically, for any query image we have to give a similarity ranking over all images in training set. For query images with multiple objects, we have to retrieve images from the training data with any object from the query image present.

2. Preliminaries

2.1. SIFT/SURF

Most problems in the field of Computer Vision deal with local features in images. However, it is difficult to extract such local features, especially features which are invariant to certain transformations. The scale-invariant feature transform (SIFT) is one popular feature detection algorithm used to detect and describe such local features in images.

The working of the SIFT algorithm is primarily divided into two parts. The points of interests of objects are first extracted from a set of reference images, and, secondly, an object is recognized in a new image by individually comparing each feature from the new image to this stored set of points and finding candidate matching features based on a defined distance metric (most commonly Euclidean distance) of their feature vectors.

The set of matches is refined by thresholding on the distance from the points of interest. These features are representative of the location, scale, and orientation of the objects in the images. In order to compute the features, the SIFT algorithm uses cascaded filters to detect scale-invariant characteristic points, where the difference of Gaussians (DoG) is calculated on rescaled images progressively.

Speeded up robust feautes (SURF) algorithm is another popular object detection algorithm used for extracting local features from images. SURF, in contrast to SIFT, uses square shaped filters which are faster to use as compared to DoG. Other than this, the SURF algorithm is more robust to different transformations than SIFT.

2.2. Brute Force Matching

It is a simple matching algorithm that takes the descriptor of one feature in the first set and is matched with all the other features in the second set using some distance metric. For SIFT and SURF, L_2 distance serves to be a good distance metric.

2.3. FLANN Matching

FLANN stands for Fast Library for Approximate Nearest Neighbors. It contains a collection of algorithms optimized for fast nearest neighbor search in large datasets and for high dimensional features. It works faster than the Brute Force Matching algorithm for large datasets.

2.4. ResNet

ResNets are deep neural networks with skip-connections in between layers to learn residuals at each step. There are many standard ResNets (e.g. Res-101, Res-152, etc.) used in practice for classification and detection tasks. They are also able to transfer very well to other datasets by using pre-trained ResNet models and switching its last fully connected layer with a custom one and just training the last layer on the custom dataset.

3. Approaches

3.1. SIFT/SURF with Feature Matching

To get a ranking over training images, we used an algorithm that is based on feature extraction and then feature matching. In essence, we perform the following steps to get a ranking score between a test and a train image:

- Since all the training images have objects right at the centre, we crop them to remove other non-important objects from the image.
- We get the features from both the cropped training image and the query test image using one of the two feature extraction algorithms, SIFT and SURF.
- We them perform feature matching between the two images using either Brute Force Matching algorithm or FLANN algorithm.
- After using K-Nearest Neighbour matching between features, we perform the Lowe's Ratio Test to extract out the good matches.
- The number of good matches between a train image and the query image is the raw match score for the train image.
- The ranking is generated based on the images with the highest match score as well as the class with the highest total score (details in section 4).

3.2. Classification with ResNet

This approach is based on learning a classification model and then using the probabilities of classification into different classes as a ranking score over the classes, which implicitly defines a ranking over training images. In essence, we do the following:

- We take a pre-trained ResNet (Res-101) that is trained over the ImageNet data.
- We swap the last layer of the pre-trained ResNet with our own layer and then train the last layer using the assignment dataset.
- Having obtained the model, given a query image we can feed it to the model and get a probability ranking over the classes, that defines a ranking over the training images in the sense that all training images of the class with higher probability have higher ranking score.

4. Results

We observed best results on the sample test data using the SIFT feature extraction algorithm with the Brute Force matcher. Although we include code for our other experiments along with the submission, we report the results on this combination only.

Implementation Details The hyperparameters for the Brute Force matcher include the number of neighbours (k) to consider while computing the score for each match. We chose $k = 2$ as this is reported to work well with the Lowe's Ratio test. The ratio for the Lowe's test is set to 0.4, which seemed to work best considering the scores of different matches for the sample test data provided.

As mentioned earlier, we crop the train images before computing their SIFT features as the objects of interest are centered in each image with a lot of unimportant space. We crop the images to have a height of 175 pixels and a width of 75 pixels (centered). In order to save time, the SIFT features of the train data are computed and stored (pickled) in a file.

For each test image, the SIFT featuers are computed and are matched with the features of each train file for each class using the Brute Force matcher, as mentioned earlier. The match score of each training image, along with the class of the train image is stored. Any image with a higher match score is given a higher rank. However, since many images result in the same match score, we introduce another parameter to this ranking. In order to decide the order of rankings for images with the same match score, priority is given to the classes with a higher sum of the scores of the top N images, *i.e.* , the sum of the N highest match scores are taken for each class and are used to order images.

The top classes for the sample test data are reported in table 1.

Test File	**Actual Objects**	**Top 3 Predicted Objects**
Test 1	palmolitive	palmolitive; adhesive spray; listering
Test 2	coca cola	coca cola; anti-breakage shampoo; palmolitive
Test 3	listerine	listering; adhesive spray; coca cola
Easy single 1	almonds	almonds; cheddar; apple cinnamon
Easy single 2	volumizing shampoo	volumizing shampoo; anti-breakage shampoo; detergent
Easy single 3	chocolate chips	chocolate chips; apple cinnamon; red marker
Easy Multi 1	hot sauce; almonds; noodle soup	hot sauce; almonds; noodle soup
Easy Multi 2	red marker; noodle soup	red marker; noodle soup; volumizing shampoo
Easy Multi 3	red marker; volumizing shampoo	red marker; volumizing shampoo; syrup
Hard Single 1	adhesive spray	adhesive spray; chocolate chips; cheddar
Hard Single 2	red marker	detergent; pringles; coca cola
Hard Single 3	noodle soup	noodle soup; detergent; pringles
Hard Multi 1	hot sauce; chocolate chips; red marker	hot sauce; chocolate chips; detergent
Hard Multi 2	pringles; volumizing shampoo	pringles; volumizing shampoo; anti-breakage shampoo
Hard Multi 3	pringles; noodle soup	pringles; noodle soup; chocolate chips

Table 1: Top classes obtained using feature matching on sample test data

Note. We do not have a checksum/hash value of the model as there is no trained model. WE simply compute the SIFT features of each test image and match them with those of train images