Image Object Classification

Akshay Tayal (201350878)
Ambika Kaul (201450855)
Senthilkumar Ramalingam (201250854)

1. Introduction

This project is to find a method to classify objects in a raw image and categorize them in specific groups for identification. Image objects are a group of pixels that are similar to one another based on some measure of spectral properties such as size, shape, texture, etc., in context with the neighboring or surrounding pixels. The traditional ways of identifying an object in an image is time consuming, cumbersome and expensive. We would like to apply the statistical methods that are available to automate this process and provide more accuracy to the object classification problem.

2. Methodology

In our project, the initial step is to select a dataset. Inspired by the reference paper of Stanford, we evaluated the datasets Caltech-101 and Caltech-256. For our purpose, we have taken the Caltech-101 dataset for this purpose. The Caltech-101 has about 101 images that are chosen carefully with low clutter or noise for the classification problem where the images vary in size within and across categories, each category having a minimum of 31 to a maximum of 800 images. Some of the images are purposefully rotated or kept in specific direction for easy interpretation. For our study, we have chosen 50 categories each containing 40 images in the training dataset and the rest as testing dataset.

The images are in raw form and the size varies between images within categories and also across categories. The rough estimate of the size of the images has been between 300X200. Hence in the pre-processing step, we tried to rescale all the images to a standard size. For this purpose, we calculated the average length and width of all images across categories and resized them to 279X249 as the standard size.

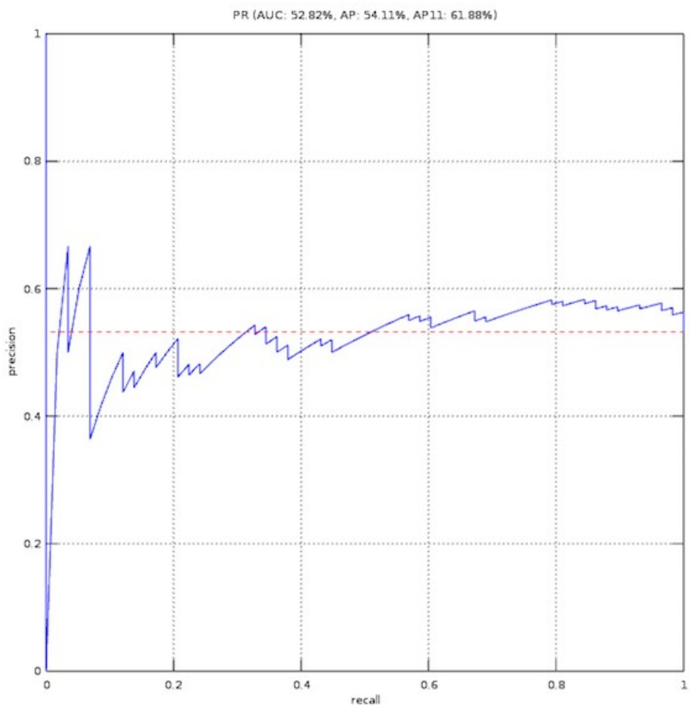
These images are then used to compute the dense SIFT key points and descriptors. We used the standard computer vision algorithms provided by VLFeat for most of our project work. Using the keypoints and descriptors, we quantize them into visual words and create a vocabulary for the images. This vocabulary data is then used to create the histogram of the image using the visual words of the image. Then we apply them to the list of images to get the histogram data for all the images in the training data. This data is then sent as an input to the linear SVM classifier to learn the images in the training dataset.

Once the SVM learning is done on the training dataset, the same preprocesses and all the above steps are applied on the separate testing dataset and these images are fed to the SVM to see how they are getting classified. The below ranked images and the precision recall chart will tell us more about the efficiency of the classifier in image object classification.

3. Results

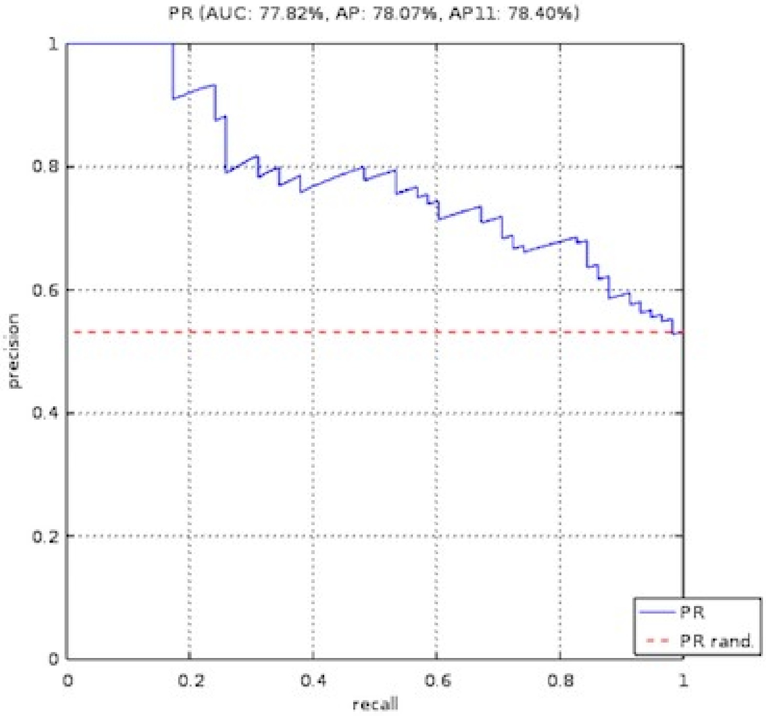
For brain (pstv) – butterfly (ngtv) :

fraction of training samples = 0.5



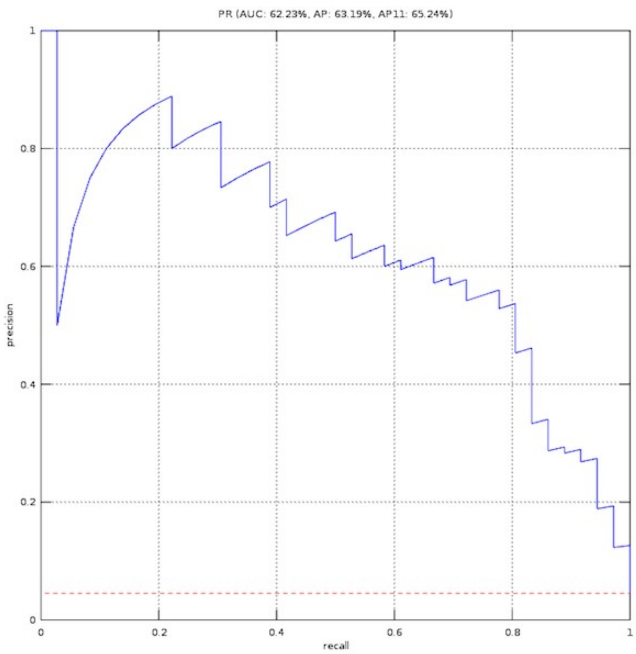
Accuracy – 54.11 %

fraction of training samples = +inf



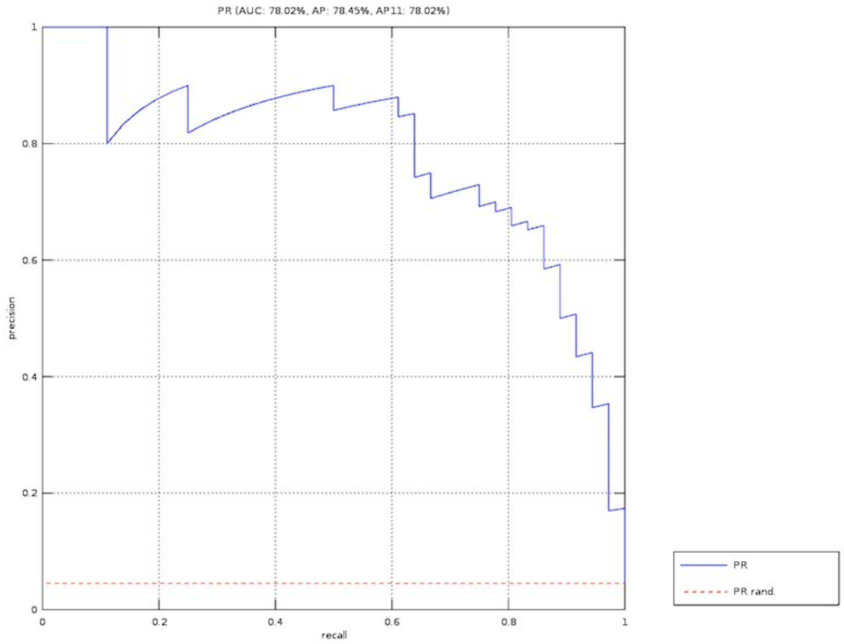
For minaret (pstv) – airplanes (ngtv) :

fraction of training samples – 0.5



Accuracy – 63 %

fraction of training samples – +inf



Accuracy – 78 %