Method Summary: Visual Bag Of Words (BOW) can be applied to image classification by treating image features as words. In document classification, a bag of words is a sparse vector of occurrence counts of words; that is, a sparse histogram over the vocabulary. In computation, a bag of visual words is a vector of occurrence counts of a vocabulary of local image features.

1) Software Used: - python with scikit for machine learning

2) Feature Extraction Process (*):- Visual Bag of Words

3) Similarity/Distance Measures (*):- Mean of all incorrect labels

3) Classifier:- SVM with ‘rbf’ kernel

Algorithms details:

1) Feature Extraction step (*):- Extracting local features of all the data set images and using the dense sift features as the features and then generating a codebook of visual words of each of the training images and feeding the image to the classifier to train a model.

2) Training Algorithm:

a. Extracting local features of the image to be classified.

b. Aggregating the histograms of the visual words for the image using the prior codebook.

c. Feeding the histogram to the classifier to predict a class for the image.

Input Format:- Matrix of features and labels to training system.
Tunable Parameters:- C and gamma are the tunable parameters in the SVM
Output Format:- Prediction labels

3) Validation and Parameter Tuning (*) : We tried to tune our C and gamma parameters on the validation data so that our model is not susceptible to over fitting.

4) Prediction Algorithm: Suppport Vector Machines are supervised learning models with associated learning algorithms that analyze and recognize patterns used for classification. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

Interpretation of results on validation data:

Why do you think your algorithm got the accuracy that it did on the validation data? Are there scope for improvements?

Our algorithm was stable on the validation set but there were some images with some shadows that our algorithm were unable to classify and rotational invariance should be considered for achieving better results.