Computer Vision, IIIT Sricity (Spring 2018)

Dr. Shiv Ram Dubey

Programming Assignment – 2

Release Date: 27-Feb-2018, Deadline: 15-March-2018 (5.00 pm)

Scene Categorization

The goal of this assignment is to introduce you to image categorization. We will focus on the task of scene categorization. In the first part, your task is to implement image features, train a classifier using the training samples, and then evaluate the classifier on the test set. In the second part, you have to use the pre-trained AlexNet and other models for transfer learning.

Dataset: In the supplemental material, an outdoor scene database is supplied with images from the 8 categories: coast, mountain, forest, open country, street, inside city, tall buildings and highways. The dataset has been split into a train set (1888 images) and test set (800 images), placed in train and test folders separately. The associated labels are stored in "train_labels.csv" and "test_labels.csv", for example, label id of 42.jpg in the training folder corresponds to 42nd entry in "train_labels.csv". The SIFT word descriptors are also included in "train_sift_features" and "test_sift_features" directories. For each image there is a separate .csv file. For example, the file "1_train_sift.csv" has the SIFT descriptors for 1st image of training set. The row in this file corresponds to the different SIFT keypoints (i.e. no. of row = no. of detected regions using SIFT detector). The last 128 values in ith row is the SIFT descriptor for ith region. The first four values in a row correspond to the center of region, scale of region and orientation of region.

A. Bag of visual words model and nearest neighbor classifier

- Implement K-means cluster algorithm to compute visual word dictionary. The feature dimension of SIFT features is 128.
- Use the included SIFT word descriptors included in "train_sift_features" and "test_sift_features" to build bag of visual words as your image representation.
- Use nearest neighbor classifier (kNN) to categorize the test images.
    - ❖ Work with different number of visual words.
    - ❖ Display the confusion matrix and categorization accuracy.

B. Bag of visual words model and a discriminative classifier

- Use the bag of visual word representation.
- Replace the nearest neighbor classifier with SVM classifier. Use 1 vs. all SVM for training the multi-class classifier.
    - ❖ Report the training time and testing time for SVM
    - ❖ Display the confusion matrix and categorization accuracy.

C. Transfer Learning and Fine-tuning

- Apply transfer learning with pre-trained AlexNet model over ImageNet database. Replace only class score layer with a new fully connected layer having 8 nodes for 8 categories.
- Freeze the weights of all layers except last layer, i.e. replaced one. Fine tune only last layer (i.e. retrain only weights of last layer).
    - ❖ Report the accuracy
    - ❖ Compare the results with previous two approaches

D. (Extra credit) Transfer Learning and Deeper Fine-tuning

- Experiment with different number of layer (from last) for fine tuning, i.e. 2, 3, 4, 5, etc. In part C, only one layer i.e. last layer was fine tuned.
- Fine tune last few layers of other models also such as VGG16 and GoogleNet and observe the performance.

Acknowledgments: Some parts are taken from the projects developed by Derek Hoiem (UIUC).