Report - Evaluation Assignment 1

Classification and spatial pyramid matching

Digified Company

By: Zynab Smaan

Period: 3 days

From 5/14/2019 12:06 AM to 5/17/2019 12:06Am

1.Introduction

This work shows the performance of using Spatial pyramid matching (SPM) on the dataset as a pre-processing step for images before feeding to a classifier in just recognition task. It's obvious that pre-preparing data before training has large impact on the performance. SPM is more advanced than Bag Of words, unlike using unorder features, SPM uses features and preserves their localized information. SPM also is considered a deeper technique as; it contains information about the image from different levels. So it's more useful in detection tasks more than in recognition tasks.

2. Workflow

The figure (figure 2) shows the workflow in brief. Let's describe each step in details.

2.1 Get the data

There is a lot of free datasets on the internet. One of them is CIFAR-10. It consists of 60000 32x32 color images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images. The CIFA dataset has a training set and a testing set with split ration equal to '.2' for the testing set. But in this task, 500 images just are used in both training and testing phase to fasten the training.

2.2 Preprocessing

First, features are extracted from the training images using Scale-Invariant Feature Transform

(SIFT). SIFT is a technique used to quantize the key points that are stable and invariant so, it's called descriptors. Then, the descriptors (features) from all images are gathered and divided into groups

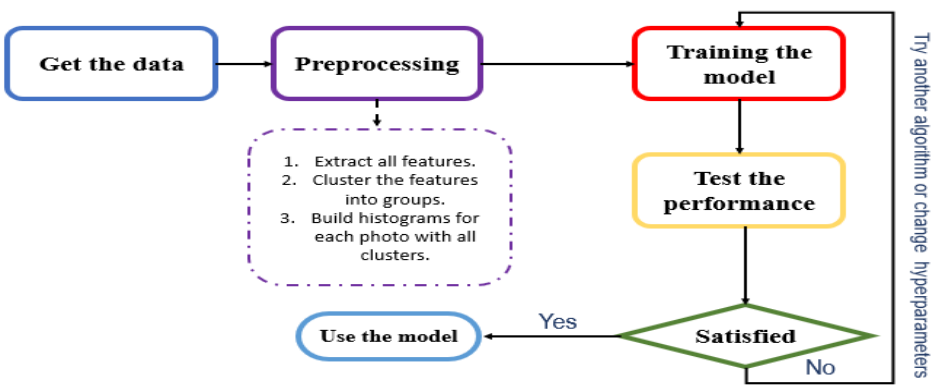

Figure 1 - workflow

according to their similarities by using K-means algorithm with k=60 which, means that the number of features now becomes 60. After that, building the histogram of each image (for both training and testing) with applying "spatial pyramid matching". SPM works as follows:
a) Dividing the image into subregions, the division step is related to the current level and follows the formula: 2^l. For example: if the level l = 0, the number of subregions is equal 1for each dimension. The first level does exactly the same as Bag Of Words. But if $l = 1$ the results will be 2 each dimension so, the number of subregions will be 4 and so on (figure2).

b) Building the histograms of each subregion for each level and stack them together to represent one image. SPM is a pretty good choice in detection, matching pictures and retrieval similar images that have the same composition. For example, beach images, are more likely to have the sky at the top, the sea at the middle and sand at the bottom.

Figure 2 - How SPM divides the image according to each level.

We should care about increasing the levels as it increases the computational cost and time. Choosing reasonable L is a must. Finally, the data needs to be split in two sets training and testing to be feed to the model.

2.3 training the model

This step takes the most amount of time because it is kind of try and error process. Various techniques should be tried but due to short time. Support vector machine (SVM) classifier is used in this work. The batch size is 500. The accuracy improved on dealing with gray scale images unlike RGB and also the training time is fewer.

3.Results

K = 60, L=2, batch size = 500

	SVM - RGB	SVM-GRAY
Training acc	13.75%	72.5%
Testing acc	10%	18%

4.References

1 – Lazebnik1, Schmid2 Ponce1 et al. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories [2006].

2 – This repo helps me in understanding how implementing SPMhttps://github.com/TrungTVo/spatial-pyramid-matching-scene-recognition.

3 – The official website for CIFAR-10 dataset https://www.cs.toronto.edu/~kriz/cifar.html.