Subject: CSC445: Neural Networks
Final Project
Lecturer: Prof. Dr. Mostafa Gadal-Haqq
TAs: Mohamed Gawish & Marwa Salah

Ain Shams University
Faculty of Computer and Information Sciences
Computer Science Department
2nd Semester 2015/16

Final Project Title

Object Detection and Recognition

1. Introduction

An essential part of the behavior of humans is their ability to recognize objects. Humans are able to recognize large numbers of other humans, letters, digits, and so on.
The object recognition problem can be defined as a labeling problem based on models of known objects. Formally, given an image containing one or more objects of interest (and background) and a set of labels corresponding to a set of models known to the system, the system should assign correct labels to regions, or a set of regions, in the image.

2. Objective

The goal of this project is to build an object recognition system that can pick out and identify objects from an inputted camera image, as shown in Figure 1, based on the registered objects.

Figure 1. Desired performance of the project.

3. System Architecture

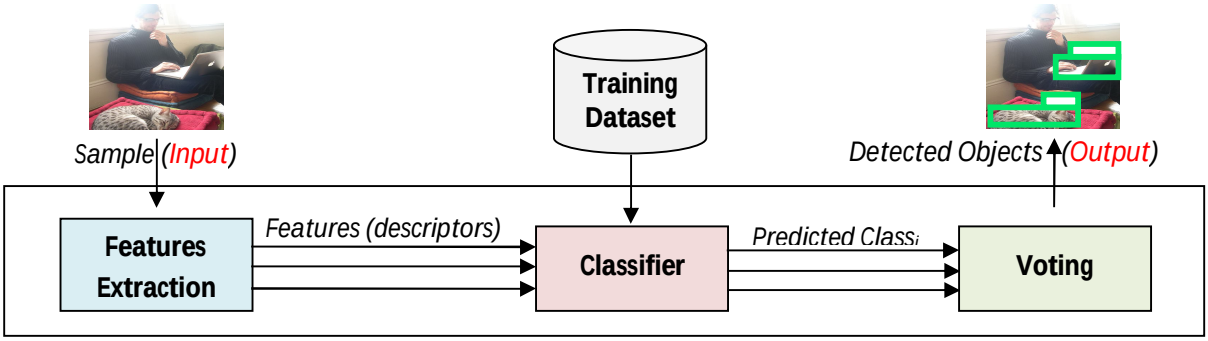

Figure 2. System Architecture

4. Features Extraction

- Use Scale Invariant Feature Transform (SIFT) algorithm [1] to extract features of an image.
- SIFT describes image features that have many properties that make them suitable for matching differing images of an object or scene.
- The features are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint.
- An important aspect of this approach is that it generates large numbers of features that densely cover the image over the full range of scales and locations.
- As shown in Figure 3, given an image, SIFT generates a set of keypoints, each keypoint consists of its location, scale, orientation, and a set of 128 descriptors.

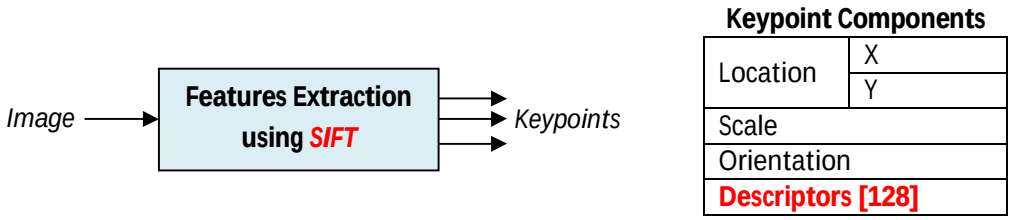

Figure 3. Features Extraction Using SIFT

- Keypoints are the samples, and their features are the 128 element feature vector (descriptors) for each keypoint.
- You can use the implementation of SIFT in VLFeat library [2] or in OpenCV [3].

5. Dataset

- The dataset is real-world data, gathered from ImageNet [4]. **ImageNet** is an image database organized according to the WordNet hierarchy.
- The training dataset has 5 objects (classes) as shown in Figure 4, 5 different models each.
- The testing dataset has 14 different images, each image may contain one or more objects as shown in Figure 5.
- Each training model/testing image is given as colored image in .jpg format.
- You can perform any useful preprocessing routines on images, such as converting them to grayscale.

Figure 4. Objects (Cat, Laptop, Apple, Car, and Helicopter) [4]

Figure 5. A test image contains Helicopter, Cat, and Apple.

6. Classification Algorithms

Implement the following three learning algorithms for recognizing objects in **ONE** package:

No.	*NN Architecture*	*Learning Algorithm*
1	Multilayer Perceptron	Back-propagation
2	Radial-Basis Function	Least Mean Square
3	-	Support Vector Machine

- For each ***NN Architecture***, the number of neurons in input layer is 128 neurons, and in output layer is 5 neurons as shown in Figure 6. For hidden layer, try using different numbers of layers and neurons, and different network parameters to achieve a maximum performance.

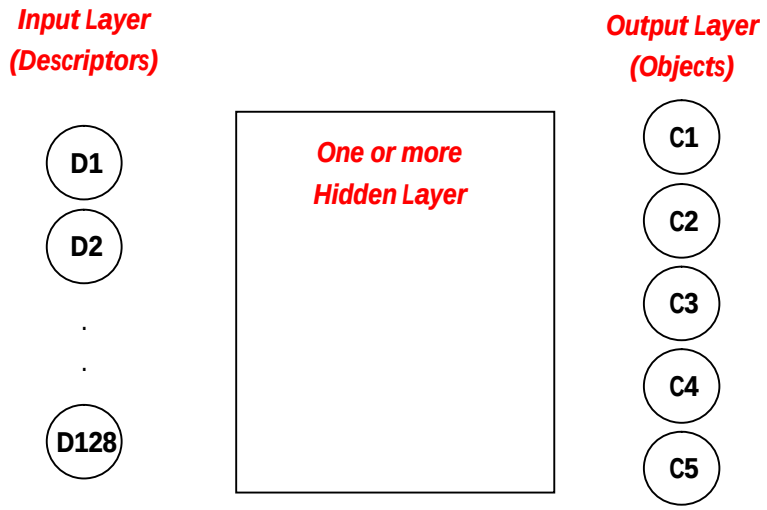

Figure 6. NN Architecture.

7. Voting (*in testing phase only*)

After classifying each keypoint of the inputted test image, count the number of keypoints that belong to each class. If the number of keypoints that belong to class(i) is greater than 3, then the object of class(i) exists in the image.

8. Ouput

- There are a two options of identifying the detected objects on the image,
  - ***Option(1)*** (Bonus): drawing a rectangle around each one on the image, as shown in Figure 7(a).
    - For drawing the rectangle, use the location of keypoints. So, you need to find the minimum X and Y, and the maximium X and Y. Minimum X and Y represent the lower-left point, (maximum X – minimum X) is the rectangle width and (maximum Y – minimum Y) is the rectangle height.
  - ***Option(2)***: drawing only the keypoints' descriptors as shown in Figure 7(b).
    - For drawing the keypoints' descriptors under Matlab, use "*vl_plotsiftdescriptor*" function from VLFeat.

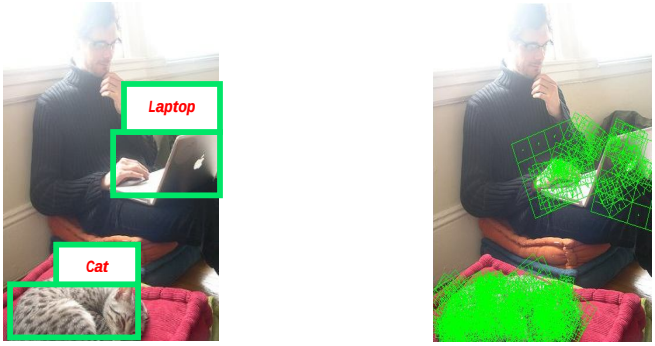

Figure 7. (a) Option(1), (b) Option(2).

9. System Performance Evaluation

Separately, evaluate the performance of the three classifiers by the following techniques:

1) Confusion Matrix

use it to describe the performance of a classifier on *test images*, by representing actual and predicted classifications done by that classifier.

2) Overall Accuracy

use it to obtain the accuracy of a classifier from confusion matrix, by counting the number of *correct classifications*, and then dividing this by the total number of *classifications*.

3) Time Complexity for training phase

4) Time Complexity for testing (recognizing) a sample

5) Memory Complexity for training phase

6) Memory Complexity for testing (recognizing) a sample

10. Requirements

- The user must be able to insert an input (image) to the application, and the application has to identify objects on the inputted image.
- Using the test images you will test your classifier. And find the performance of your classifier using the Overall Accuracy (OA) and Confusion Matrix.
- A comparative study showing the difference in applying the three classification algorithms based on the six evaluation measures mentioned above. Thus, a report template will be provided to you for filling it.
- Also, the report must be provided showing the different NN architectures and different parameters you used, and their effect on the training and testing results.

11. Bonus

Conduct a comparative study of different feature extraction algorithms such as SIFT, PCA-SIFT [5], and SURF [6] to show up their effects in improving classification performance of the project's objective based on the six evaluation measures mentioned above. (A report template will be provided for that)

12. Rules

- The same groups of lab. tasks (any change in group members is not allowed).
- **All** the group members must be aware of the project (anyone can be asked in any part of the project).
- **IT'S NOT ALLOWED TO USE ANY AVAILABLE SOURCE CODE or LIBRARIES IN IMPLEMENTING CLASSIFICATION ALGORITHMS (1&2).**
- You can use any programming language such as **(C++, C#, Java, or Matlab)**

13. Deadlines

Delivering the whole project during the Neural Networks practical exam time.

14. Support

There will be a support labs for answering your inquiries.

15. References

[1] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 60, 2, pp. 91-110, 2009.

[2] A. Vedaldi and B. Fulkerson, VLFeat: An Open and Portable Library of Computer Vision Algorithms, 2008, http://www.vlfeat.org/

[3] OpenCV, http://opencv.org/

[4] ImageNet, http://image-net.org/index

[5] Ke, Yan, and Rahul Sukthankar. "PCA-SIFT: A more distinctive representation for local image descriptors." *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on.* Vol. 2. IEEE, 2004.

[6] Bay, Herbert, et al. "Speeded-up robust features (SURF)." *Computer vision and image understanding* 110.3, pp.346-359, 2008.

Good luck! ☺