You may do this homework in groups of one or of two. The answers are due by email, to daf@cs.uiuc.edu, with CS 498 in the title, by **November 16**.

In this homework, you will build a simple classifier to distinguish between adverts and business graphics on the one hand, and images of real scenes on the other hand. Very typical of adverts and graphics is the use of strong, bright colors; these tend not to appear in images of real scenes.

1) Collect a training set of at least 100 ads/business graphics and at least 100 real images. Using google's image search is traditional for this sort of thing, but by no means mandatory. You are welcome to use more images, and to share training sets (but if k groups share training sets, there should be k x 2 x 100 training images in the pool).

2) Collect a test set of at least 10 ads/business graphics, and at least 10 real images. Do not share these.

3) Build code that represents an image as a 3D color histogram. Use 8 bins in each direction, for a total of 512 bins. Your code should turn this histogram into a 512 dimensional vector.

4) Now train a classifier, using logistic regression, using your training data. Use cross-validation to select a regularization weight.

5) Now evaluate your classifier on your test set.

6) Hand in, by email, (a) your evaluation on your test set (b) your test set and (c) an alias that your team wants to be known by. I will then construct a reference test set out of all the submitted test sets, and send that out. You will then run your code on the reference test set, and I'll publish a league table, by aliases, of performance.