The University of Texas
at Austin

CS381V Visual Recognition

Coding Assignment 02

Edited by LaTeX

Department of Computer Science

STUDENT
Jimmy Lin
xl5224

INSTRUCTOR
Prof. Kristen Grauman

TASSISTANT
Kai-Yang Chiang

RELEASE DATE
Sep. 16 2016

DUE DATE
Sep. 30 2016

TIME SPENT
15 hours

October 2, 2016

Overall

In this assignment, we will implement an image recognition system that trains models for the 25 provided categories, and can predict the class label for a novel test image.

Datasets We will be working on a subset of the SUN Database, whose natural images containing 25 scene categories. Unfortunately, some iamge data cannot get loaded. Missed training images are as follows:

```
/work/01932/dineshj/CS381V/SUN397/h/hospital_room/sun_ahokhhxjiclpxqqa.jpg
/work/01932/dineshj/CS381V/SUN397/b/bow_window/indoor/sun_awnrlipyxpgxxgxz.jpg
/work/01932/dineshj/CS381V/SUN397/b/bus_interior/sun_adhktvidwzmodeou.jpg
/work/01932/dineshj/CS381V/SUN397/c/chicken_coop/outdoor/sun_abcegmmdbizqkpgh.jpg
 /work/01932/dineshj/CS381V/SUN397/c/chicken_coop/outdoor/sun_amaonsnnkskxwmrj.jpg
/work/01932/dineshj/CS381V/SUN397/b/bow_window/indoor/sun_asmvdfnjlulewkpr.jpg
/work/01932/dineshj/CS381V/SUN397/f/formal_garden/sun_bjvrlaeatjufekft.jpg
/work/01932/dineshj/CS381V/SUN397/b/bow_window/indoor/sun_abeugxecxrwzmffp.jpg
```

Missed testing images are as follows:

```
 /work/01932/dineshj/CS381V/SUN397/l/landing_deck/sun_aizahnjfkuurjibw.jpg
/work/01932/dineshj/CS381V/SUN397/l/landing_deck/sun_acodgoamhgnnbmvr.jpg
/work/01932/dineshj/CS381V/SUN397/g/garage/indoor/sun_akbocuwclkxqlofx.jpg
```

Testing We refer to the open-sourced online material Convnet_test.py (https://gist.github.com/axel-angel/b2af7d980eb217a0af07) to compute predictions of all testing images.

Challenge A: Model Without Pre-training

We use the non-pre-trained 16-layer network, sourced from Caffe Model Zoo (https://github.com/BVLC/caffe/wiki/Model-Zoo).

The result yielded from the model without pre-training is disappointing. Overall, the training/testing accuracy remains floating around 0.04, meaning that the network fails to distinguish various categories. The network is not sufficiently evolved because

- The allowed training time at TACC computing nodes is limited.
- The given training data is not enough (given the network is not pre-trained).

The confusion matrix is as follows:

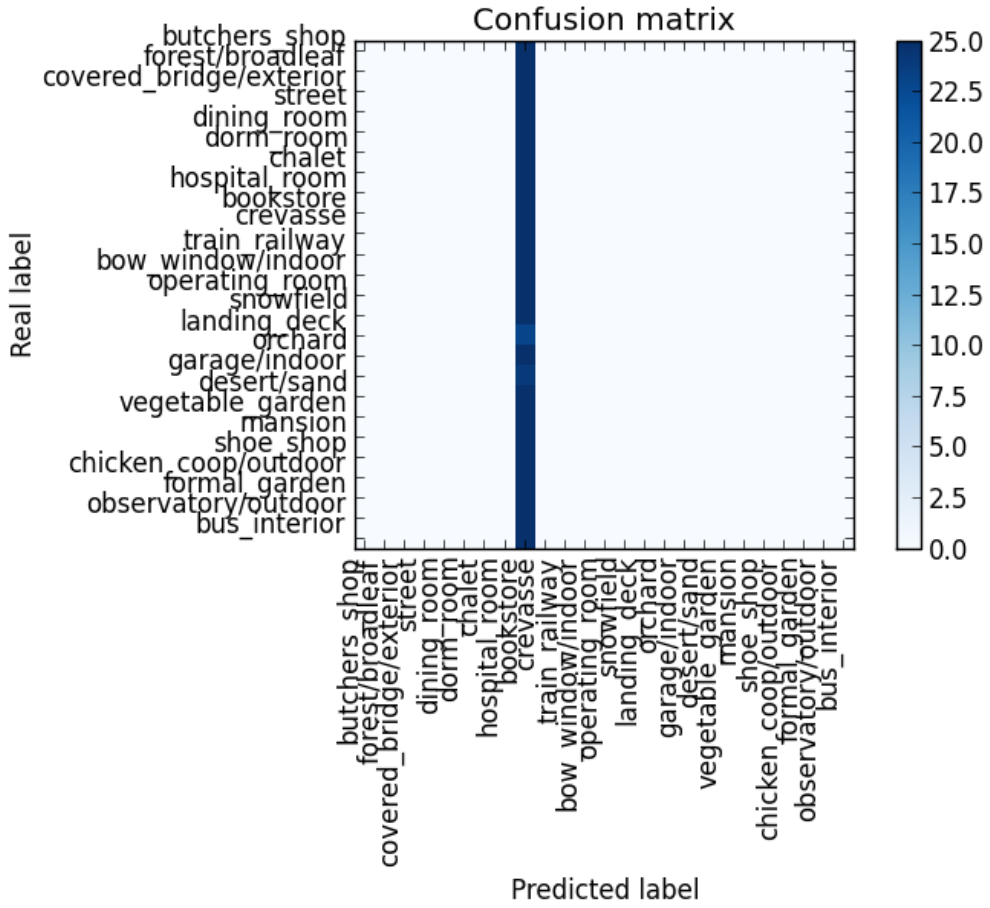

Figure 1: Confusion Matrix of Non-Pre-trained Model (800 iterations of training, accuracy: 4%)

Remark Obviously, all testing images are predicted as one particular class. This means that the entire model needs more learning. For this task, it is suggested that more computational resources can be allocated and pre-train the network with a dataset of larger size.

Challenge B: Model With Pre-training

We choose the same 16-layer network with its pre-trained performance at 7.5% top-5 error on ILSVRC-2012-val, and 7.4% top-5 error on ILSVRC-2012-test, sourced from Caffe Model Zoo (https://github.com/BVLC/caffe/wiki/Model-Zoo).

We fine-tuned the pre-trained network with 200 iterations and it came out with *accuracy*@1 = 0.89 and *accuracy*@5 = 0.99 at the test data set. And more specifically, the training set loss is 0.110429 and the training set accuracy 0.96875 (obviously overfitting).

The confusion matrix is as follows:

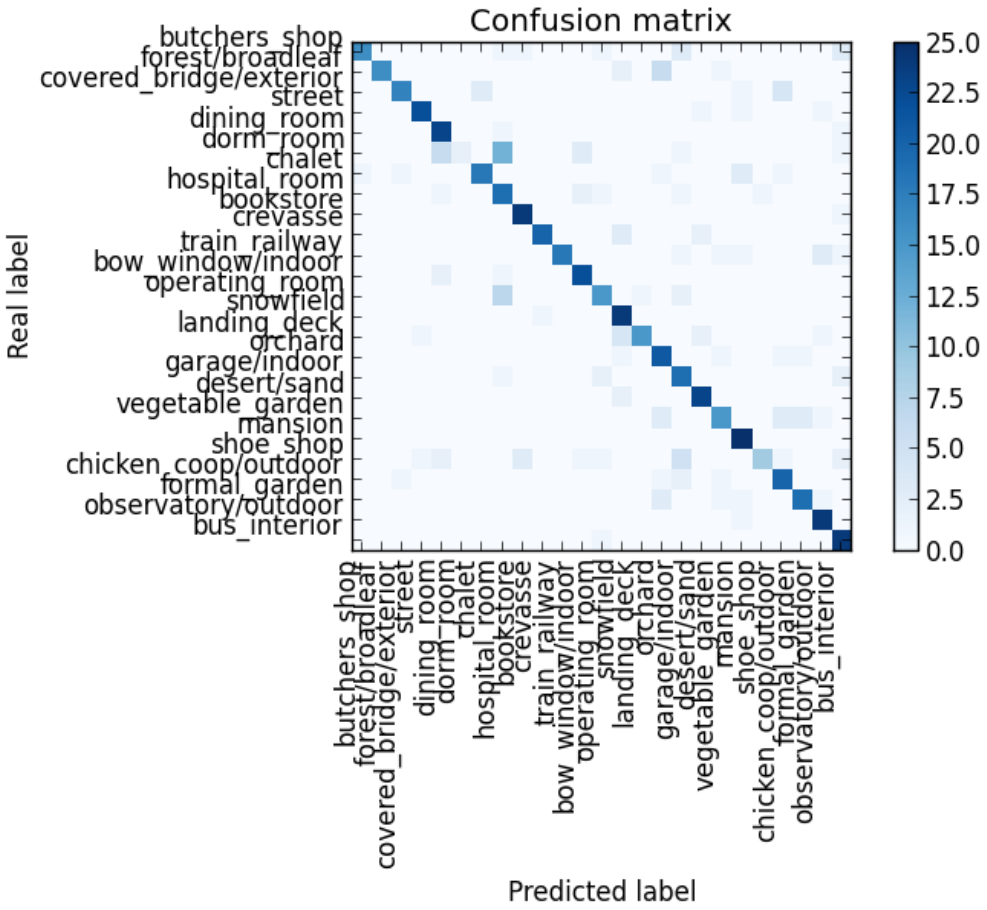

Figure 2: Confusion Matrix of Pre-trained Model (100 iterations of fine-tuning, accuracy: 75.6%)

Remark The pretrained model correctly predicts most of testing images. Also, we found that as the learning progresses after 100 iterations, the testing accuracy decreases. So it is likely that the entire network overfits.