Scene Category Classification for TV Series
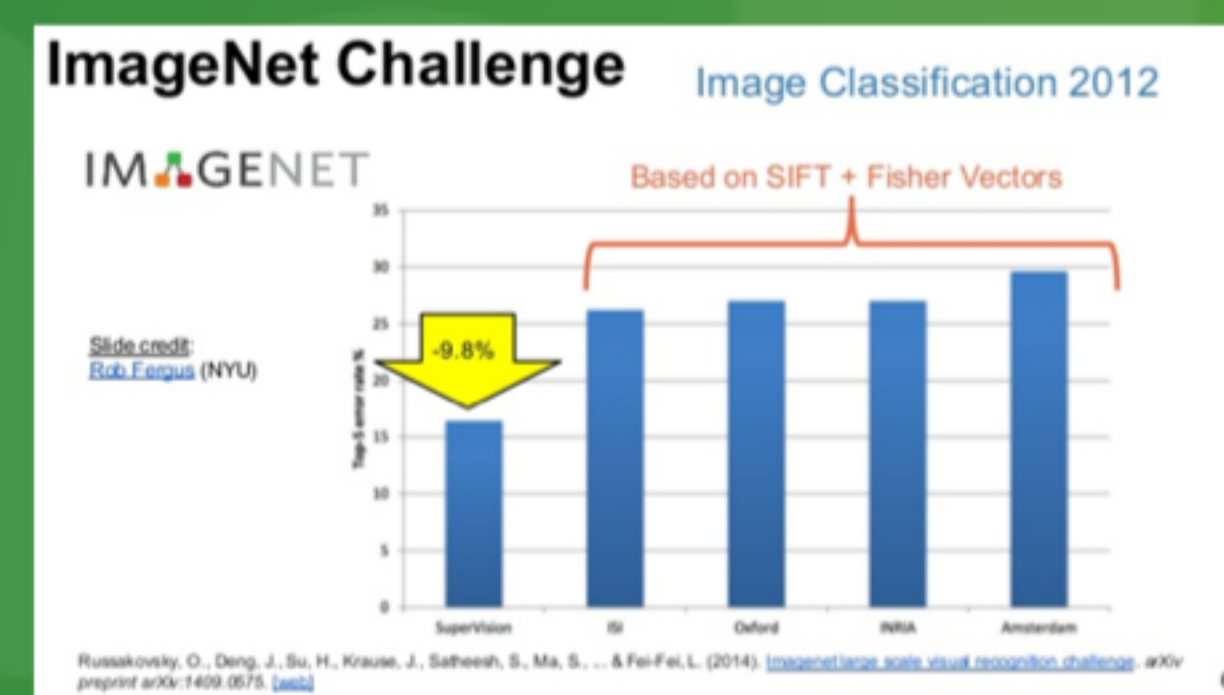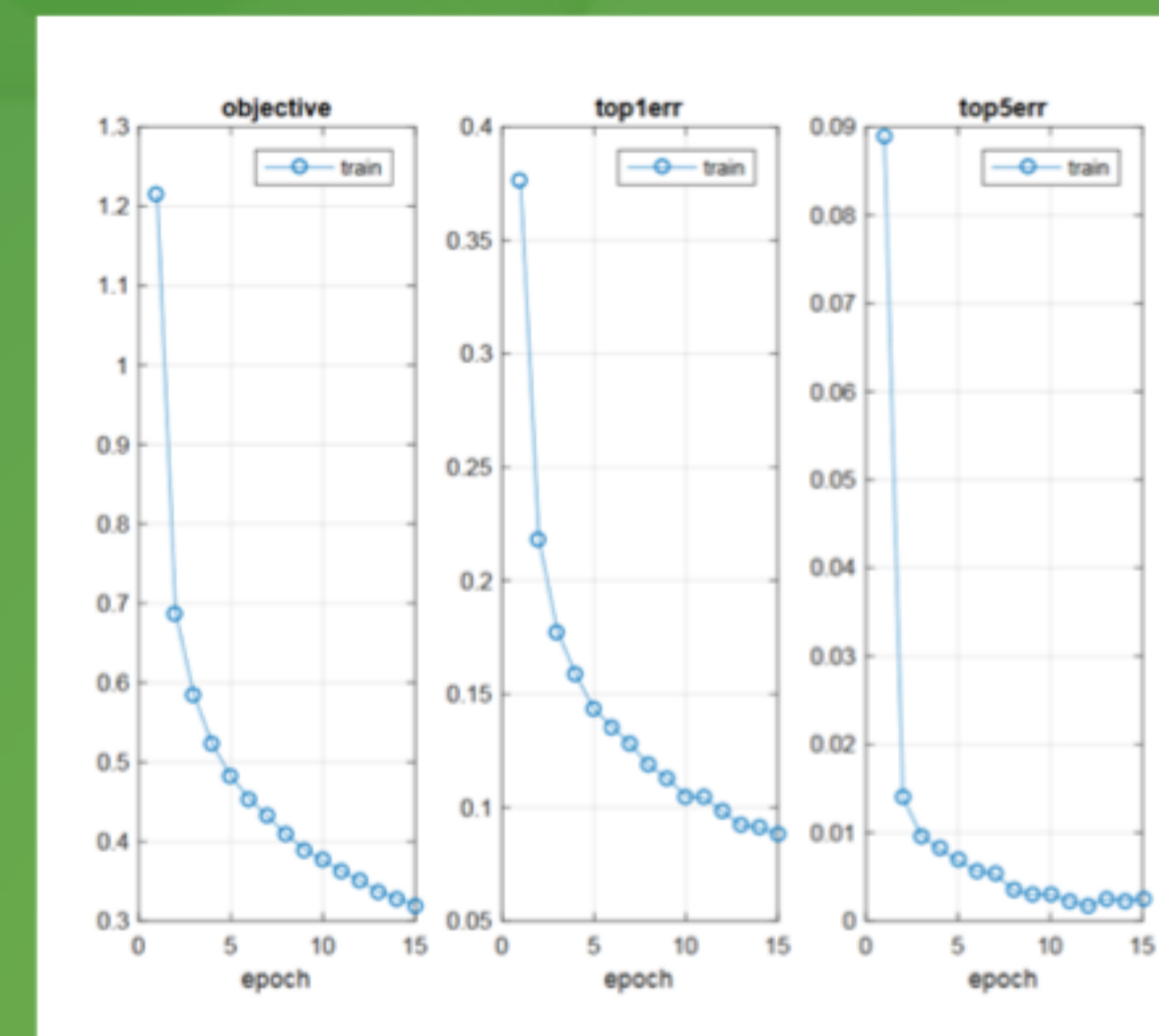
Abhinandan Dubey and Russell Walker

Background Information

▪ Image recognition research and techniques have made huge strides in just a few short years.

▪ Much of this improvement is thanks to a return to the neural network model, Convolutional Neural Networks in particular.

▪ Starting from 2012 CNNs have largely replaced engineered high-level features with many layers of learned features in a deep learning format

Goal & Method

▪ Our objective was to use state-of-the-art techniques to classify sets of images from a TV show based on the background scene of each

▪ Two major stages: testing and training

Training Process

Load training set images and sample 10 patches for each → Using VGG's Very Deep '16 ConvNet, convert each image to a set of image vectors → Using VLFeat's MatConvNet, train parameters that will predict classifications

Testing Process

Load test set images and sample 10 patches for each → Using VGG's Very Deep '16 ConvNet, convert each image to a set of image vectors → Predict each image's class using the mean of all patches' predicted label through MatConvNet
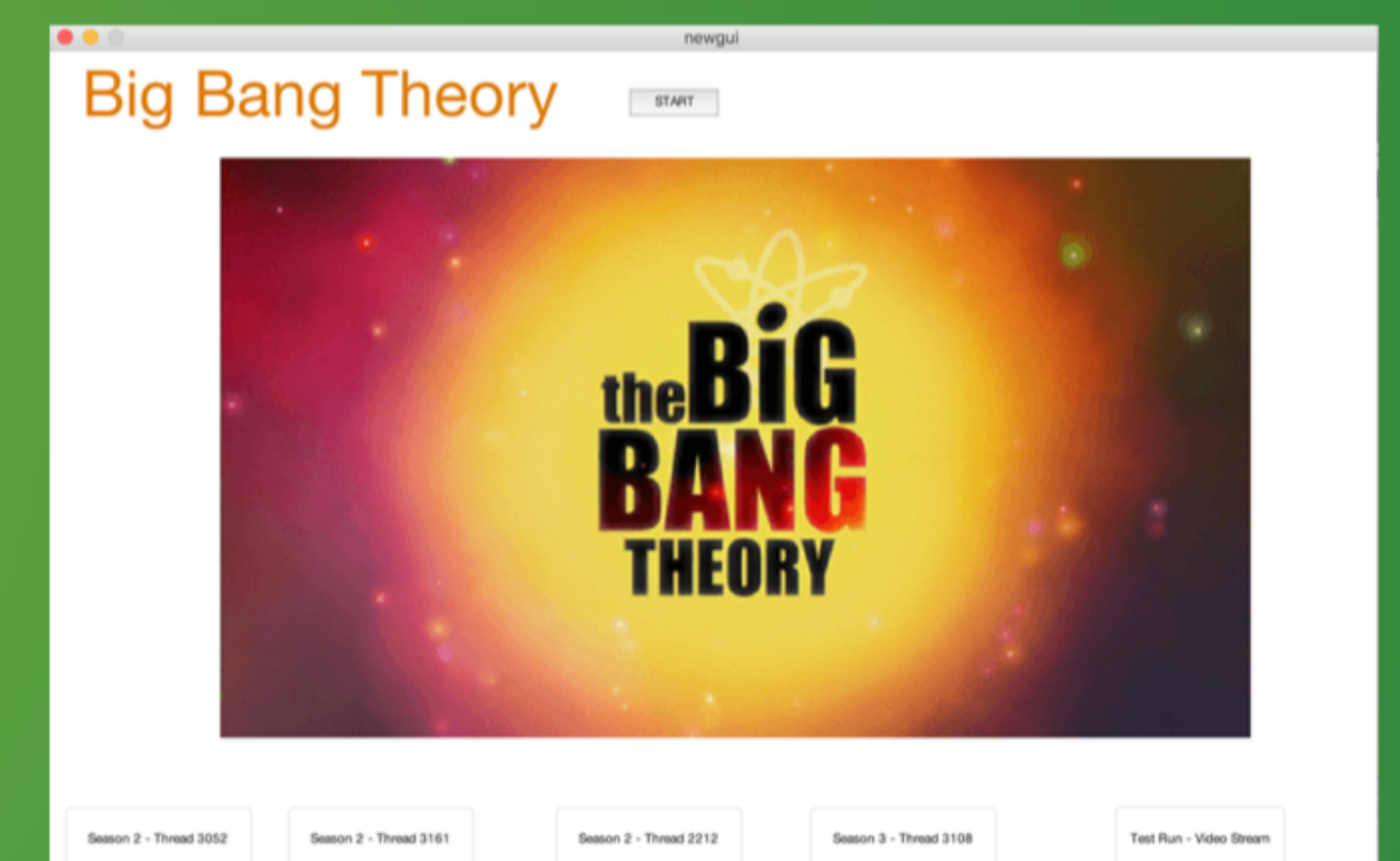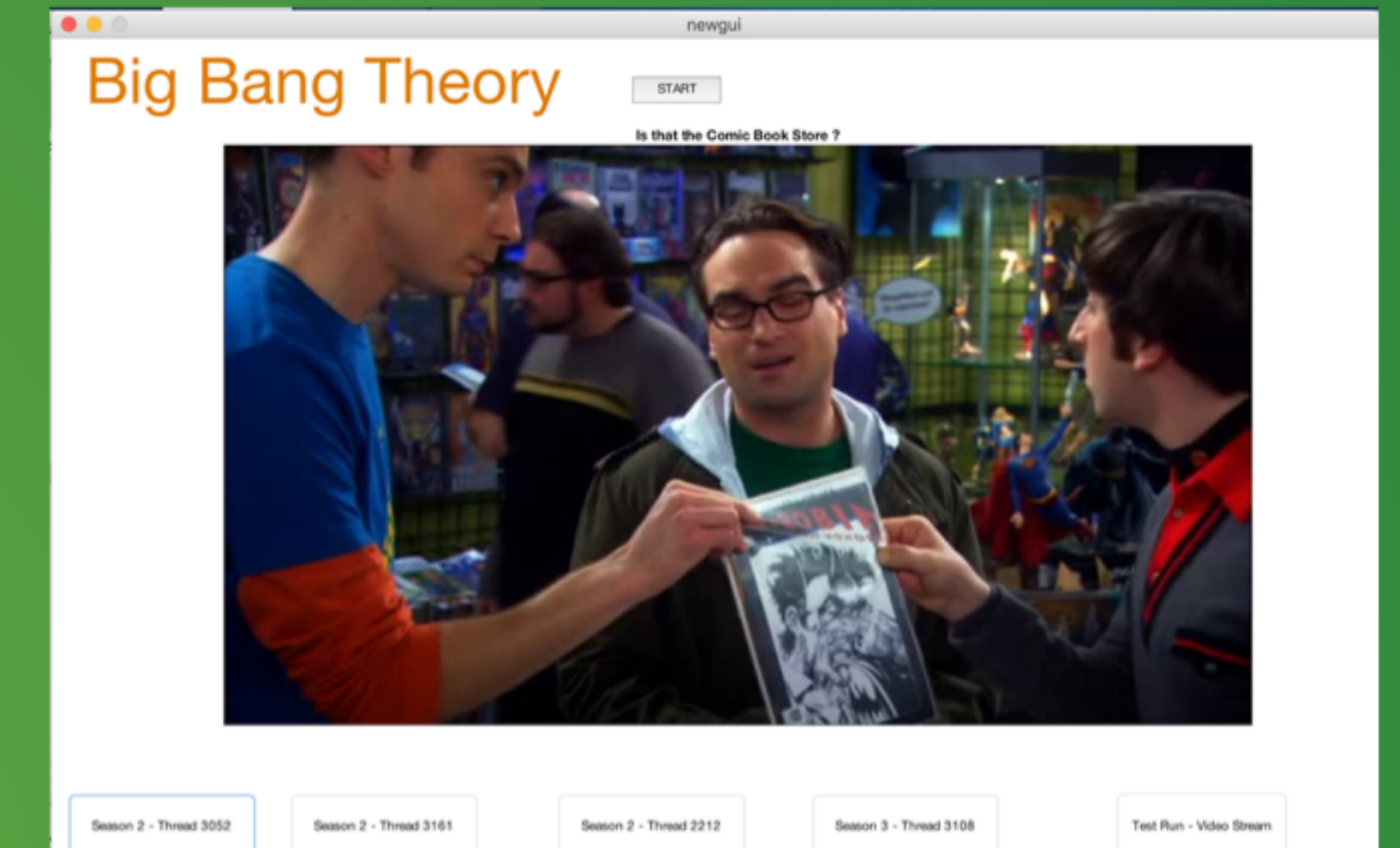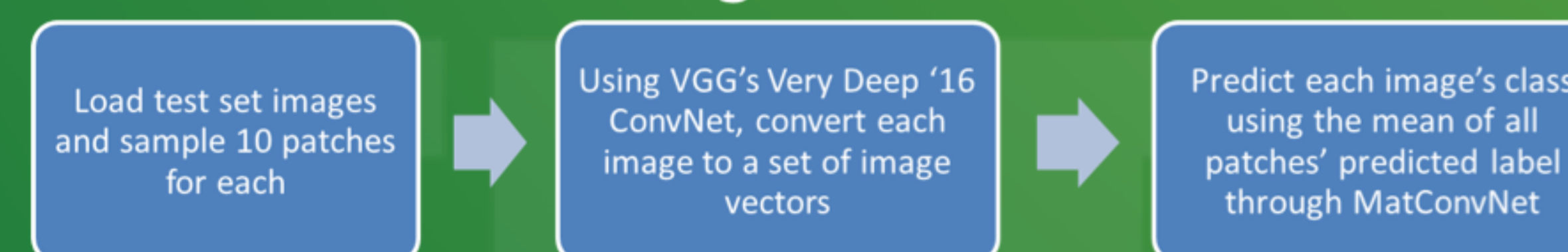
Learning on Training Set

▪ Learning the image classification parameters is an iterative process, each epoch reduces the error rate in the training set

▪ Learning is accomplished via stochastic gradient descent with momentum, which evaluates the direction in which error decreases the fastest for each iteration

▪ After a set number of epochs, the resulting parameters are used to predict labels for the test set and accuracy is evaluated

Data & Results

▪ Data used: frame data from The Big Bang Theory TV show Seasons 2 & 3

▪ Data is broken up into shots and threads, representing break points in the frame data, although the classification model does not change its behavior based on this supervision.

▪ 3000 threads for each season were used as training data, with a batch size of 100 and 15 epochs during learning.

▪ The remaining 564 S02 and 468 S03 threads were used as test data

▪ Our current accuracy on the test set is 80.81%

▪ Current training is still ongoing and may improve these results after more computation time

Sample Output

Libraries & References

▪ VGG16 Very Deep trained ConvNet model: credit K. Simonyan, A. Zisserman

▪ VLFeat MatConvNet Matlab toolbox for CNN implementation: credit A. Vedaldi, K. Lenc, A. Gupta

▪ Karen Simonyan and Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition"

▪ Demir Gokalp and Selim Aksoy, "Scene Classification Using Bag-of-Regions Representations"