

16-824 Visual Learning and Recognition - Assignment 1

Animesh Ramesh - animesh1

The reason why we cannot use a simple “Reshape/View” is because the data to be extracted is not sequential. That is, the input is of size (batchSize, 40, 16, 16) and there is an explicit need to extract the 16x16 images every batch and then rearrange it to (batchSize x 16 x 16, 40).

Part 1 : Finetuning FCN using ImageNet-pretrained AlexNet

Results after 10 epochs:

Mean: 29.488; **Median:** 24.991; **11.25:** 22.228; **22.50:** 45.532; **30:** 58.232; **45:** 77.412

The qualitative results of the pretrained model are shown in the next page.

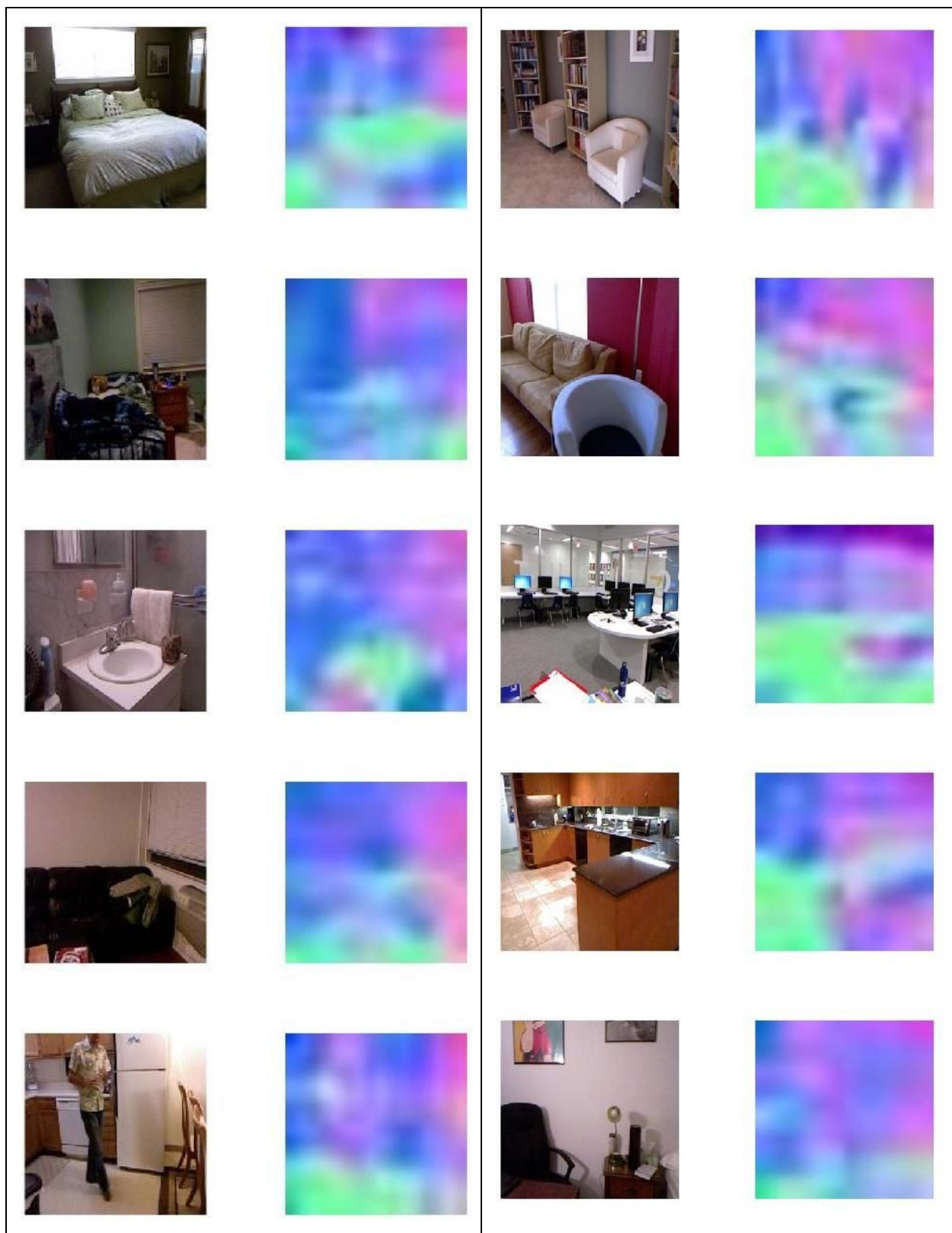


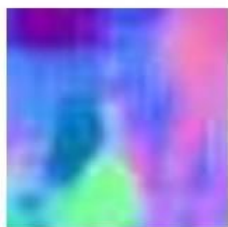
Figure 1 : Qualitative results using the pretrained model

Part 2 : Training FCN from scratch

Results after 30 epochs

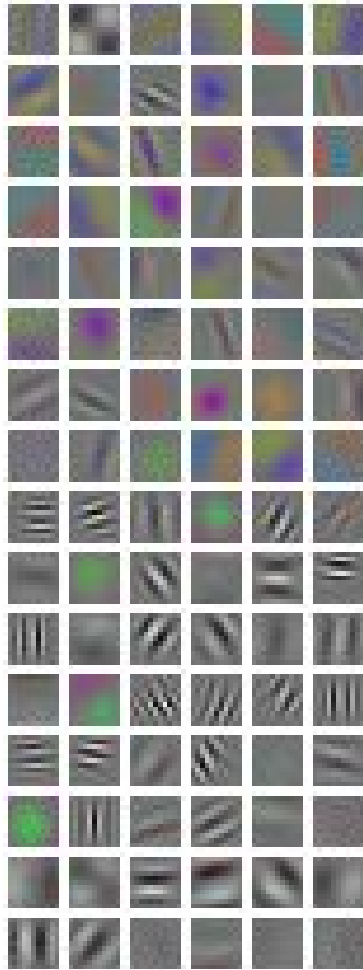
Mean: 30.225; **Median:** 24.868; **11.25:** 22.675; **22.50:** 45.915; **30:** 58.00; **45:** 75.893

The qualitative results of the model trained from scratch is shown in the next page.

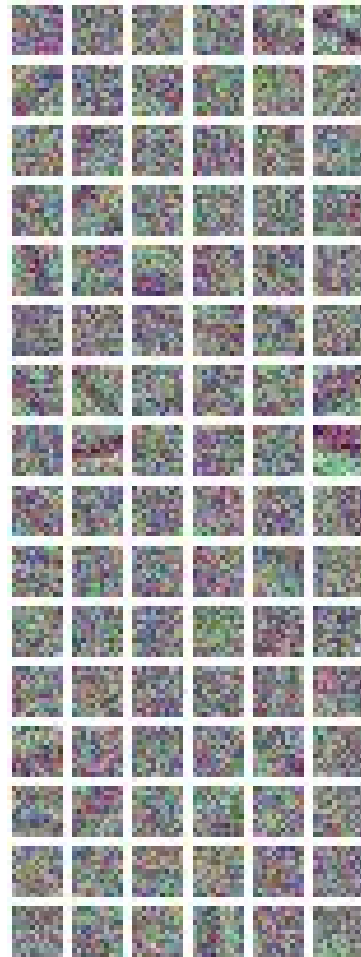


Visualization of Conv1 features

As we can see below, the features of the pretrained AlexNet model are much sharper and cleaner compared to those of the model trained from scratch. (since it has been pretrained on ImageNet)



Visualization of the 96 features in the pretrained model



Visualization of the 96 features in the scratch model