HBM 516E – Term Project

Mustafa Can Uslu

702161008

In this project, I visualized the outputs of artificial neural network layers to investigate which neuron is responsible for which feature (edge detection, face recognition, color detection etc.).

Convolutional Neural Networks are the state-of-the-art technology for visual problems. The connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. Similar to human eye, some layers and neurons are specialized on different tasks. If those layers or neurons detected, we can use that specific layers or neurons for other networks due to it's modular structure. Visualization of outputs of network will be the key factor to detect them.

Firstly, I generated a dataset from ResNet-50 neural network of Microsoft Research Team, which won the ImageNet competition in 2015. 5 of ImageNet classes are taken with 5 example images for each class. Those images supplied to the network and each layer's output recorded. Since, layers produce many outputs; the ones with the most activation selected. Then, I concatenated all the layers outputs. Structure of the dataset as follows:

- X, value in x (image is 2D)
- Y, value in y (image is 2D)
- level, network depth

Then, netCDF files generated for each of the 25 examples by using Python. Also, I normalized the outputs of each layer from 0 to 1. The reason of that is, preventing auto rescaling during the visualization state.

Finally, netCDF files dumped to Paraview. I took slices to investigate outputs. Visualization completed as expected and desired results obtained. Due to structure of CNN, in the first layers input images can be seen with some modification. However, when we are diving into the middle of the network; abstraction phase begins and there is no human detectable visualization. Just before the classification, last layers of network, we regain the ability to understanding of outputs. Moreover, last layer indicates why the network classified the image as a class with highlights.

I was planning to do more investigation on this subject. One of the my goals was investigating how image properties affect the classification. Unfortunately, creating dataset and visualizing it with Paraview did not come up with satisfying results.