Jason Luu

CS 478 – Intro to Deep Learning

Professor Gheni Abla

Assignment 5

1. ML Web Server Description:

In this assignment, I have an opportunity for creating an web server and client for ML inference system. The mechanism of this web server is that the client will run the client.py file with the argument of the name of the test image. Then, the server side will read that image and load the CNN model from multiclass\_image\_classification.ipynb in order to predict whether the image is rock, paper or scissor. After the prediction, the server will response the result back to the client side.

1. Testing the Web Server and Client:

In order to test the web server, you first need to pull it from my github repository. Then, we need to run the server by “python3 server.py”. By excuting this command, the server is running on the localhost with the port at 5000 like 127.0.0.1:5000. Next, in order to test the client side, you first need to run the client by “python3 client.py”. By excuting this command, the client is sending the request to server side to display the information to the console including the welcome statement, my name, and current date. Similarly, you can run the command like “python3 client.py <imagefile>”. By excuting this command, the client will request the server with the particular image file, then the server will do the prediction and response back to the client with additional information of the prediction.

1. Screen Capture of Test Run:

Here is a screen capture of the command “python3 server.py”:

Text

Description automatically generated

Here is a screen shot of the command “python3 client.py” (right)with the response from the server.py (left):

Graphical user interface, text

Description automatically generated

Here is a screen capture of the command “python3 client.py mysecret.png”(right) with the response from server.py(left): Text

Description automatically generated