DSC 481 Final Project

Due: Wednesday, December 11, 2019 11:59 PM

Part 1 – Team Members:

Yuan Lu (ylu50), He Huang (hhuang41), Yuexi Wang (ywang307)

Part 2 – Files:

There are totally 13 files and folders included in the zip file of `FinalProject_ylu50_hhuang41_ywang307.tar.gz`:

1. `Code_1_Data Preparation.ipynb` is the code to prepare the datasets (i.e. training, validation, and testing datasets). The preparation process is clearly stated in the report.
2. `Code_2_KNN.ipynb` is the code for the KNN model.
3. `Code_3_XGBoost.ipynb` is the code for the XGBoost model.
4. `Code_4_CNN_resnet18.ipynb` is the code for the CNN (Resnet-18) model.
5. `JPEGImages` is the folder containing all the original images prior to processing. It is used by `Code_1_Data Preparation.ipynb` to prepare the datasets.
6. `labels.csv` is the labels for images in `JPEGImages` folder. It is used by `Code_1_Data Preparation.ipynb` to prepare the datasets.
7. `train` is the folder containing the prepared training datasets. There are 4 subfolders, each of which contains the images of that cell type. It is used by `Code_2_KNN.ipynb`, `Code_3_XGBoost.ipynb`, and `Code_4_CNN_resnet18.ipynb` to train the models.
8. `validation` is the folder containing the prepared validation datasets. There are 4 subfolders, each of which contains the images of that cell type. It is used by `Code_2_KNN.ipynb`, `Code_3_XGBoost.ipynb`, and `Code_4_CNN_resnet18.ipynb` to validate the models.
9. `test` is the folder containing the prepared test datasets. There are 4 subfolders, each of which contains the images of that cell type. It is used by `Code_2_KNN.ipynb`, `Code_3_XGBoost.ipynb`, and `Code_4_CNN_resnet18.ipynb` to test the models.
10. `hpt_Adam_29` is the folder containing the best weights after tuning. It is first generated and later used by `Code_4_CNN_resnet18.ipynb` to test the CNN model.
11. `DSC481_Project_Readme_ylu50_hhuang41_ywang307.pdf` is the readme file.
12. `DSC481_Project_Report_ylu50_hhuang41_ywang307.pdf` is the detailed report file.
13. `DSC481_Project_Slide_ylu50_hhuang41_ywang307.pptx` is the well-noted presentation file.

Part 3 – Run Time Instruction:

1. All files and folders included in the zip file shall be extracted to the same folder for the code to properly run.

2. `Code_1_Data Preparation.ipynb` uses `JPEGImages` and `labels.csv` to prepare the datasets. The prepared datasets will be stored in a folder called `Train_Aug_only`. However, since the code uses random augmentation, these datasets generated by running `Code_1_Data Preparation.ipynb` will not be the same as the ones we generated earlier to train and test our models. The code and folders are included for demonstration purpose.

3. The datasets included in folders of `train`, `validation`, and `test` are the ones we generated earlier using `Code_1_Data Preparation.ipynb`. These datasets were used for `Code_2_KNN.ipynb`, `Code_3_XGBoost.ipynb`, and `Code_4_CNN_resnet18.ipynb` to make the models comparable.

4. The code blocks within each code file shall be run in their natural order.