General Instructions

1. All code for this project can also be found in my github repo(Have a look at the README for this repo as well to understand my entire code more easily)
2. The report contains a lot of hyperlinks to specific files in my github repo. In the results section I have also added hyperlinks to some of the plots due to the space limitation of 6 pages in the report.
3. Even though my github repo is very modular, each notebook is self-contained (i.e all of the classes are explicitly redefined as opposed to importing them). The main reason for this is because I was not able to import other python files in a colab instance which I was using for training.
4. It is recommended that all jupyter notebooks mentioned in this project are run on google colab. Reason being, each notebook has logic which can mount the google drive with your google colab instance.
5. All model weights and processed datasets can be found in this drive link. (It is recommended to download this entire folder as is and then upload this folder to the root directory of your google drive so as to run all colab notebooks without any further edits)

How to test the models ?

As long as step #5 from the “General instructions” is followed the colab notebooks should work as is without any further modifications.

In all the WBC notebooks (all 8 of them) the very last section involves a piece of code which loads my best model weights and then performs the testing on the test dataset yet again.

Note that this section will have the title “**Loading the final model weights and testing the code (Run only this after running all the function definitions for direct evaluation)**”

It is also recommended that all cells are run except for the “**Training code**” section as that section can be skipped to directly run the final section mentioned before

How to train the models ?

As long as step #5 from “General instructions” is followed from the previous section , there should be no need to change any of the code cells in the ipynb notebooks.

However, if you still wish to process the original dataset in the same manner I did kindly follow the following steps (which include applying masks etc)

a. Camelyon Model:

* Take the zip file "Camelyon.zip" mentioned in the drive link and place it in your google drive root path. (This zip contains the mask information as well)

* Run the colab notebook as is
(or)
* Take the dataset path and run the ./data/maskification/apply_mask.py script locally which will apply the masks present and add this image as a new datapoint and save it in a specified output path
* Upload this new dataset with the mask information to drive (either as a zip or the whole dataset)
* Once the drive gets mounted to colab and change the Config.datasets_path appropriately (based on whether a zip was uploaded or the whole dataset was uploaded in the previous step)

b. pRCC Model

* Take the zip file "pRCC.zip" mentioned in the drive link and place it in your google drive root path
* Run the colab notebook as is
(or)
* Upload the dataset to drive (as is or zip) and mount it to colab and change the Config.datasets_path

c. WBC Model & WBC Pretrained model

* There are 4 different splits of this dataset (1,10,50,100) which is referred to as "n" below.
* Take the zip file "WBC_{n}.zip" mentioned in the drive link and place it in your google drive root path.
* (OR) Take the zip file "WBC_{n}_balanced.zip" mentioned in the drive link at the bottom of the README.md and place it in your google drive root path. This is the balanced dataset obtained by running the balancing script (./data/balancing/wbc.py) to augment under-represented classes.
* Run the colab notebook as is
(or)
* Take the dataset path and run the ./data/maskification/apply_mask.py script locally which will apply the masks present and add this image as a new datapoint and save it in a specified output path
* Take the new dataset with masks and run the ./data/balancing/wbc.py script locally in case you want to balance all the classes of the dataset as the original dataset is imbalanced.
* Upload the dataset to drive (as is or zip) and mount it to colab and change the Config.datasets_path. (You can choose to use the original imbalanced dataset or the balanced dataset)
* The dataloader code varies for the base wbc model and the pretrained wbc model, but the same WBC dataset can be used in both colab notebooks