Hands on Activity IV: Data Augmentation effects on CNN classifier ("Big Data")

Description of Activity

This activity aims to provide students with an idea of data augmentation to create a larger dataset for classifier training. Students will run scripts to rotate and flip the original images and use these new data as input to the CNN classifier.

After this activity, students should be able to:

1. Run script to process augmentation of image data.
2. Run script to train a CNN classifier.
3. Appreciate the pros and cons of training a large dataset.

Task 1: Image dataset augmentation

1.1. Open "impro_aug.m" in MATLAB.
1.2. Change the "imgPath" and "outPath" in lines 6, 7, 39 and 40 to match your folder directory. (Hands on Activity III, Task 1, Step 1.5)
1.3. The script uses the function "imrotate" and "fliplr" to rotate and flip the images respectively. After processing, "imresize" is called to ensure the size of the resultant image is 120 x 140.
1.4. Click "Run" to run the script and you can view the results in the output folder "Processed Data Aug"

Task 2: Train a CNN classifier with augmented dataset

2.1. Open "CNN_CT_aug.m" in MATLAB.
2.2. Change the path directories in line 14 and 18 to match your folder directory. (Step 1.2 in Task 1).
2.3. Click "Run" to run the script.
2.4. Note how much slower the script runs compared to the previous activity with a smaller dataset.