

# 95-891 Homework 3 – Computer Vision

Due **March 17, 2022**

## Task

You will be provided images of multiple categories of flowers. For instances, the two rows below show two categories of followers.



The goal is to classify images into one of the 102 categories above.

## Dataset

You are provided three files – a training file, a test file, and a zipped folder containing the images referenced in the train/test files. The images files are already organized into train/test-class structure for your convenience

File descriptions in flowers.tar.gz

- An archive containing all images, train, validation and test.

## Software Instructions

You can use PyTorch to complete this assignment.

If you want to install PyTorch, visit : <https://pytorch.org/get-started/locally/>

First thing if you are using any of the platforms is to make sure you can iterate through the data. For PyTorch, it can be efficiently done by creating your own data loader. For TensorFlow, there are a lot of online guides to read images.

Since this is a Transfer Learning task, PyTorch will need to download weights of a pre-trained network and then modify the classifiers before you start training.

### **PyTorch:**

```
from torchvision import models  
model = models.alexnet(pretrained=True)
```

A good document about transfer learning in

**PyTorch:** [https://pytorch.org/tutorials/beginner/transfer\\_learning\\_tutorial.html](https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html) .

You should not be needing heavy computational resources for this homework. A good result can be achieved by training a model for 30 mins- 1 hour. A baseline result of 70% was achieved by training a model using AlexNet for 30 mins. Depending on your hardware, the training time can vary. So make sure you do not choose a very large network to train unless you have the time to train it!

In this homework, you will be using "Transfer Learning" but you are free to train a neural network from scratch if you believe you can get better performance out of it.

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. It is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in performance that they provide on related problems.

## Submission Instructions

**Submit your documented code in the Jupyter notebook with a .ipynb extension. Please answer the questions directly in the notebook—no separate writeup is needed.** The documentation doesn't need to be extensive, but it should be easy to understand what the various parts of the code are doing by reading the code comments. Without the accompanying code, you will not obtain credit for answers to the below questions which depend on your code.

In addition to the code submission, please answer the following questions:

1. What is the effect of random flip and rotation in image data transformation (see code for more information)
  - a. How does it impact the training?
  - b. What do not we apply them randomly at the validation and the test time?
2. Explain the meaning of batch dimensions in the dataloader (see code for more information).
3. Why are we using the Dropout layer in neural networks?
  - a. Explain its usage briefly
  - b. How does the dropout layer work during the prediction (evaluation) stage?
4. Report the best model's validation accuracy.
5. Briefly describe the plots of the training process.
6. Report the best model's test accuracy.
7. Provide a confusion matrix between the ground truth and predicted labels in the test dataset. The rows should correspond to the ground truth categories and the columns to predicted categories. The confusion matrix here should have raw datapoint counts and should not be normalized as percentages.
8. Now normalize the raw confusion matrix reported in the previous question by dividing each row by the sum of all its elements. Include this normalized confusion matrix in your report.
9. What does the diagonal denote in the normalized confusion matrix? What is the meaning of the number in the row 1 column 2?
10. Transfer learning basics:
  - a. What is the purpose of a pretrained model and why should it work?
  - b. Why should we freeze the features' gradient in the pretrained model but only update the classifier?
  - c. What may happen if we do not freeze the features' gradient?