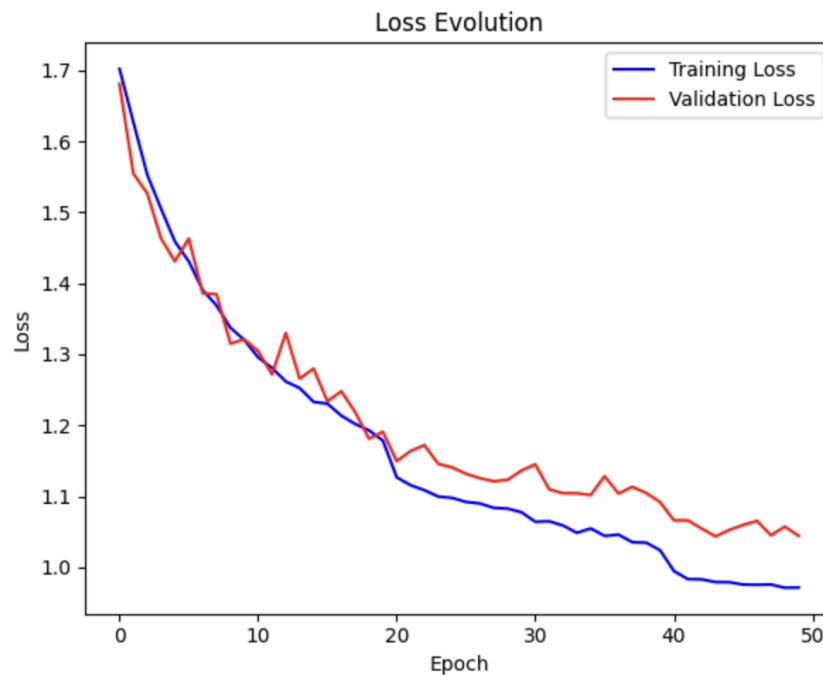


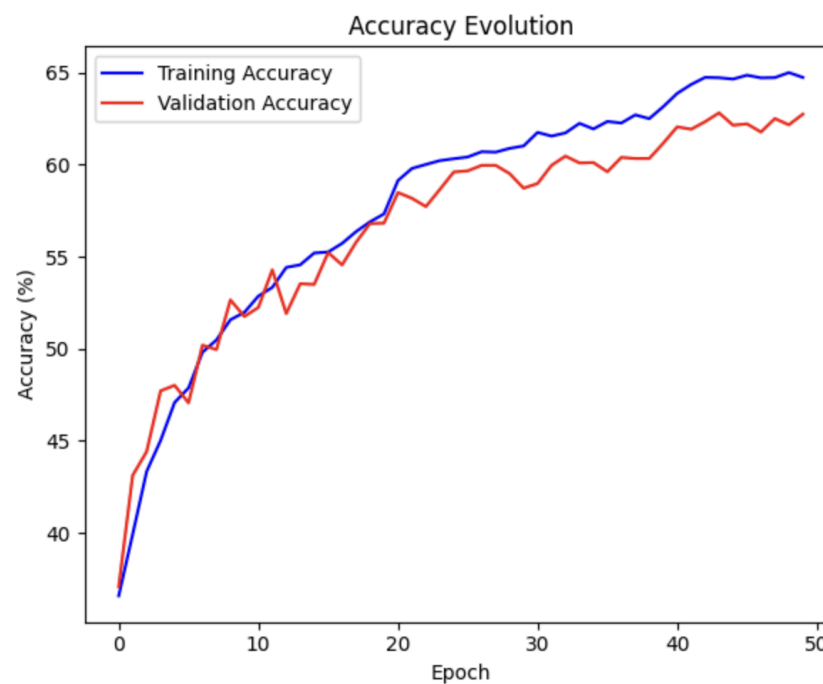
1. Read the dataset and create data loaders in 'load_data_cifar_10()' definition.
2. The model follows the architecture consisting of Stem, Backbone (Ns of blocks) and Classifier based on Convolutional neural Networks.
3. used CrossEntropyLoss for the loss function, and SGD optimizer with learning rate sets to 0.01

4.

- a. the curves for the evolution of loss in 50 epochs



- b. the curves for the evolution of training and validation accuracies in 50 epochs



c. all training details including hyper-parameters used.

❖ batch_size: 128

- if too small (eg. 32), there are more noisy gradients, it will provide a better generalisation but less stable updates.
- if too large (eg. 256), this is a more accurate gradient, it will provide a faster and more stable training, but may have a poor generalisation.

❖ lr (learning rate): 0.01

- if too high (eg. 0.1), the training may be unstable, the loss may fluctuate, or even diverge, because it takes too large steps and skipping over the optimal point
the observation of using lr=0.1 is that, the accuracy didn't improve beyond 20-30%

Epoch 1/50

Training accuracy: 26.756. Testing accuracy: 27.53. Duration: 78.819s.

Epoch 2/50

Training accuracy: 26.806. Testing accuracy: 27.66. Duration: 77.257s.

Epoch 3/50

Training accuracy: 26.794. Testing accuracy: 27.52. Duration: 78.573s.

Epoch 4/50

Training accuracy: 26.788. Testing accuracy: 27.48. Duration: 78.494s.

- if too low (eg. 0.0001), the training will become very low, and the model might be stuck in a local minimum.

Epoch 1/50

Train Acc: 42.49 | Test Acc: 42.29 | Time: 5518.02 secs

Epoch 2/50

Train Acc: 47.97 | Test Acc: 47.60 | Time: 5188.79 secs

Epoch 3/50

❖ momentum: 0.9

- It helps to accelerate gradients in the appropriate direction while smoothing out updates.

❖ num_epochs: 200

- if too few, it will underfitting
- if too many, it may overfitting

❖ K (the number of experts per block): 4

- if the K value is too low (eg. K=1), it will provide a simple and fast training but lower performance
- if too high(eg. K=8), it will provide a slower but more powerful

❖ N(the number of blocks in backbone): 3

- if N too low (eg. N=1), it is a shallow network, and there is less powerful and my underfit
- if N too high(eg. N=5), it is a deep network, and having a higher capacity but slower training

❖ optimizer: SGD

❖ scheduler: StepLR

❖ loss function: CrossEntropy

5. Final model accuracy on CIFAR-10 Validation Set: 70.23% in epochs 200.

Final validation accuracy: 70.23%