# ZZEN9444 -Neural Networks, Deep Learning Assignment 3

## Image transformations:

There are several transformations we can utilise, such as Horizontal Flips, Random Cropping, and Colour jittering. Unfortunately, our dataset is quite small and as such utilising too many of these may dilute the pool too much to be useful. As such I will be using horizontal flip and normalization in my model. Horizontal flipping is used only on the training set while normalisation is applied to both the training and validation/test sets. We also apply conversions to an 80\*80 tensor to ensure that data is in the expected format.

### Initial model approach:

There are two approaches that I considered for developing the model. The first is to develop a model totally from scratch, the second is to use a pretrained model as a starting point and then building on top of that. I have chosen to proceed with the second approach.

I have chosen resnet-18 as the initial model to try building off. It offers a good level of performance without being overly complex or too large for our 50mb size limit. If necessary, it will be easy enough to move to a more complex model such as resnet-34.

We use a simple initial model with 3 layers. The first is a linear layer that connects the pretrained output to 256 new features with a relu activation followed by a 0.5. Finally, we take another fully connected layer that reduces from 256 nodes down to the 8 classes.

As we are still in the initial phase of development we will use the Adam optimiser. Adam is often an improvement of standard SGD. If necessary, we can try alternatives once the model is further developed. For the loss function we will use cross entropy loss, once again this is a relatively standard choice for classification, and we can consider alternatives as necessary. We will use the default weight initialisation and scheduler.

We will use a smaller batch size of 64 and 20 epochs. I have chosen to allow more epochs as we are transferring knowledge from a pre-existing model.

The initial basic model as described above reaches a training accuracy of approximately 67% and a test accuracy of approximately 60%. I think that while this a good starting point for a model there is a lot of room for improvement.

### Model Improvement:

The first thing I will try is expanding the number of layers and the number of nodes in each layer. Keeping the general structure we will introduce a second linear layer with ReLU optimisation. We will also expand the number of nodes from 256 to 512. I will also add a batch normalisation layer to correspond with each linear layer. We also remove the dropout layers. While they do make overfitting less likely I think at this model development stage they make it harder to tell what is creating meaningful improvement. If required we can add them again towards the end of model development. With these changes the model reaches a training accuracy of 72% and a test accuracy of 57%.

I now am going to add further image transformations. I will add random cropping, rotation of 15%, and colour jitter to the model. With these three added the model reaches a training accuracy of 55% and a test accuracy of 55%. It its possible that with more data including all of these would produce a more accurate model. But I believe that with the training sample being so small it is undermining the quality of the training data. I will remove colour jitter and rotation and assess the accuracy. With now only cropping added the model reaches 66% train accuracy and 60% test accuracy. As adding this does not seem to have adversely affected the model I will keep cropping as it is likely to provide better performance on unseen data.

I will now implement learning rate scheduling. Initially I will try stepLr with a gamma decay of 0.1 every 5 epochs. In tandem with this I will increase the number of epochs to 30 to allow more room for this decay to aid model development. After 20 epochs this gives a training accuracy of 63.34% and test accuracy of 60%.

I am now going to test the same model with GELU activation function instead of the RELU activation function to see if that yields any improvement.