Lab 5 Exercise - A little Linear Regression

Amishapriya Singh
Student id: 33333904
as6u21@soton.ac.uk

1. An initial attempt

For image-to-vector regression task, the loss function will be **mean squared error (MSE)** as our model is doing regression to predict the straight line parameters.
From Table 1, we can see that this model does well on training set however has a huge loss on both validation and test set. This is caused by the model overfitting to the training set. This can be verified by looking at the number of learnable parameters of the model. We find that the first linear layer has a total of 9.8M learnable parameters leading to model overfitting and unable to generalise in validation and test set.

2. A second attempt

Model #2 uses Global Max Pooling to reduce the number of parameters in the fully connected layers. The total number of learnable parameters of this model is 27.8K. As a result, this model gives better generalisation compared to Model #1 (see Table 1). We can see that the train loss for Model #2 is higher indicating that it has not overfitted the training data.

3. Something that actually works?

Of the three models, Model #3 gives the best performance (see Table 1). It gives training loss comparable to our overfitted model but unlike Model #1 maintains the same performance on both test and validation set indicating good generalisation.
The rationale behind Model #3 is the addition of two interleaved channels to our input data for better feature learning. The reason for that include:

- With pooling shrinking the data size, we can maintain the same amount of information but make it more relevant to the problem by keeping more channels.
- More channels allow model to capture more complex features that would otherwise not be learnt due to large receptive fields of deeper layers.

Model	Train loss	Validation loss	Test loss
#1	0.819	16.60	17.31
#2	12.40	11.40	14.58
#3	1.24	1.75	1.98

Table 1. Loss for each model

Model	Learnable parmeters
#1	9.8M
#2	27.8K
#3	28.7K

Table 2. Number of learnable parameters in each model