Question 1
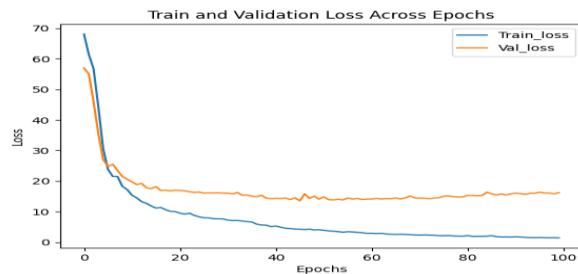
The loss function used was mean squared error loss since it is a regression problem. The performance is as below:

The model appears to be overfitting to the training data as the training loss and validation loss start to diverge after around 50 epochs. This may be because the model is too complex for the small dataset and there are too many parameters in the final hidden layer and the model is not able to generalize well to validation data.

Question 2

The performance of the model is as below:
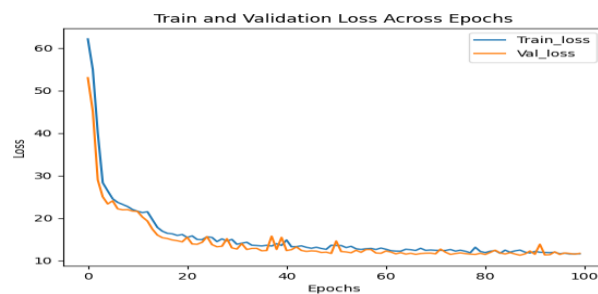
The model appears to produce lower validation loss compared to model in question 1. However, the model still takes a long time to converge; at around 40 epochs where the mean squared loss is around 14. The model does not appear to overfit the training data and the difference between the training loss and validation loss appears to be smaller for each epoch.

Question 3
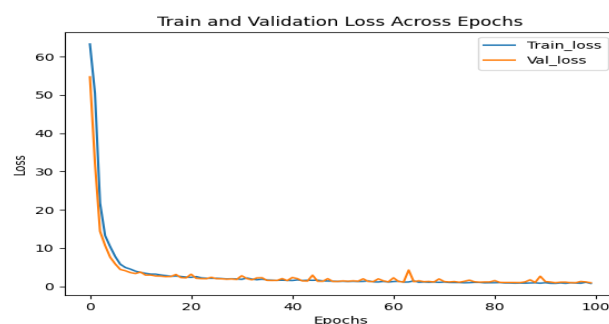
Model 3 concatenates 2 channels with values from -5.0 to 5.0 with existing input channel before passing as inputs into the network. This model produces the highest performance compared to the other 2 models and it converges the fastest, at around 20 epochs. The difference between training loss and validation loss is the smallest for each epoch. The addition of 2 extra channels may provide extra information for the network to better its learning of features present in input images. This may cause faster convergence of the model as the model is able to generalize well to the sparse original training data with additional information in form of 2 extra channels.