The reconstruction loss $L(x, x^)

The former term encourages our model to be sensitive to inputs, and the latter discourages overfitting. A scaling parameter is typically included in front of the

\verb|torch.nn.Linear| has by default He/Kaiming uniform initialisation, as confirmed here.

As images are normalised between 0 and 1 by \verb|transforms.ToTensor|, a sigmoid activation has been added to the output layer to get values that match this input value range.

Binary cross-entropy was found to offer marginal improvements over mean-squared, though subjectively.

Spectral Normalisation was also trialled.