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. It was found to improve the convergence rate slightly, from around $250$ epochs to $200$.

# Might want to graph those two. As proof, y’know.

The larger the compressed encoding of the image, the closer a combination of two encodings is to simply naively combining two images.

A successful approach therefore *must* find the latent space.

Using \verb|LeakyReLU| activation functions with a negative slope of 0.2 appears to be current best practice, with use in both ACAI and PIONEER.

Minor data augmentation also proved useful. Specifically, a random crop, and a random horizontal flip. The introduction of noise was deemed a complicating factor; it would have necessitated a re-design of the AE into a *denoising* AE.