SODA: Bottleneck Diffusion Models for Representation Learning - Optimization

Shivam Sharma (210983, sshivam21@iitk.ac.in)

1 Optimization

The original paper uses a Resnet architecture as the encoder, which was used in the previous submission. Now, it has been replaced with a more SOTA Resnet variation with attention, as described in https://arxiv.org/abs/1704.06904. This new encoder consists of multiple attention modules stacked between the standard residual layers, each having 2 branches, a trunk mask and a soft mask branch.

2 Results

The new model gave better reconstruction results and linear probe results,

Encoder	Linear Probe Top 1	Linear Probe Top 3
Resnet	56%	85%
Attention Resnet	63%	90%

Table 1: Linear Probe Results

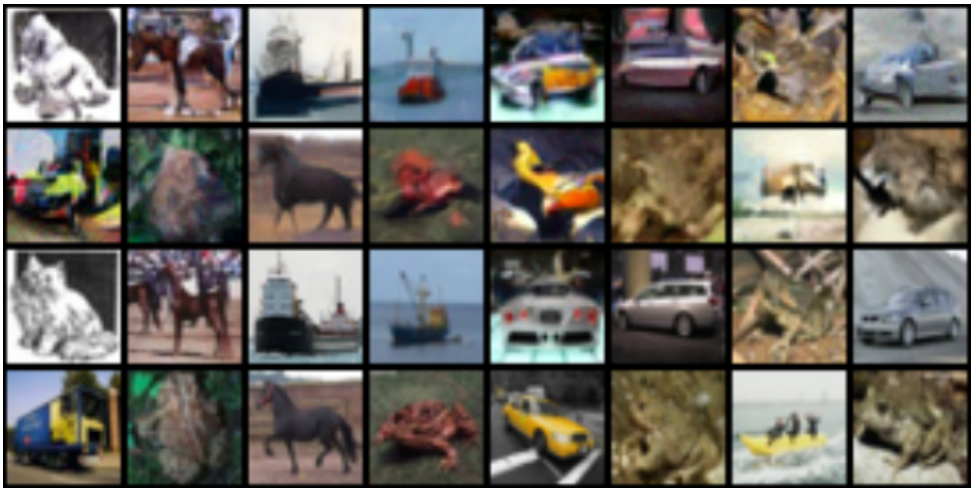

Figure 1: The upper two rows are the reconstructions for the botton two by the denoiser

Also, the model was able to achieve this accuracy with just 35 epochs of training, as compared to about 100 epochs for the Resnet version.