Optimization for SODA - Bottleneck Diffusion For Representation Learning

Indian Institute of Technology Kanpur

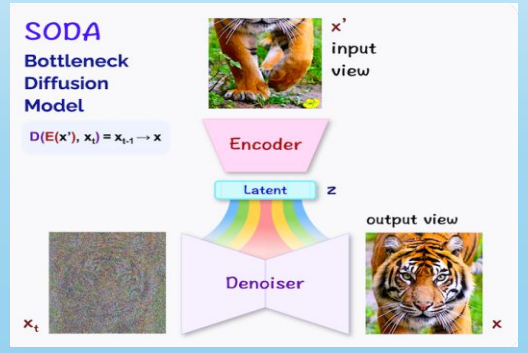
Model Objective

- To get high quality, low dimensional embeddings of images which can be used for downstream tasks like classification, without knowledge of class labels.
- To train a diffusion denoiser, conditioned on these embeddings for controlled image generation

Method

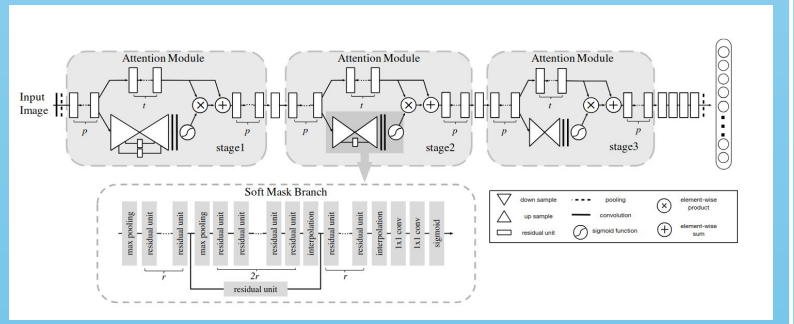
- An encoder network like Resnet is used to map images into a latent representation which is used to guide the denoiser in reconstructing views.
- A tight bottleneck between the encoder and decoder restricts the latent space, forcing the encoder to focus on high-level semantics and thus creating compact and informative representations.
- A UNet based denoiser, with Adaptive Group Normalization layers which have the embeddings and timestep as a parameter.
- Train the denoiser and the encoder together, with the standard diffusion Markov Chain process.

Architecture



Optimization

- The original paper uses a Resnet for the encoder architecture, which can be replaced by more SOTA attention based architectures like Residual Attention Network, which can be more effective in capturing representations.

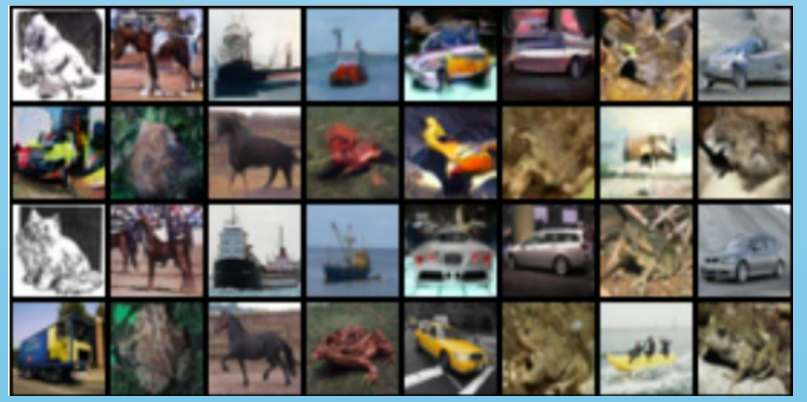


Linear Probe Results

Results for linear probe test performed on CIFAR10 for both variants

Encoder	Top 1	Top 3
ResNet	56%	85%
Attention ResNet	63%	90%

Denoiser Regeneration



Denoiser Regeneration(top 2 rows) from the original images(bottom 2) guided by their latent representation.