DC-AE 1.5: Accelerating Diffusion Model Convergence with Structured Latent Space

Supplementary Material

A. Improving Diffusion Model Scaling Curve

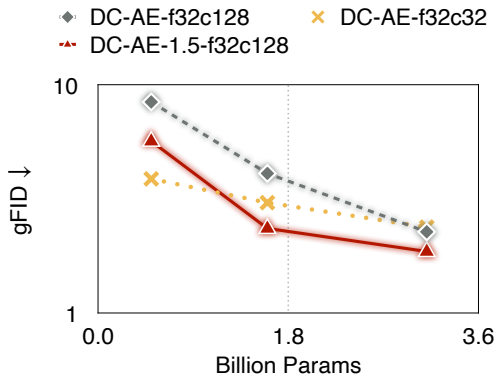


Figure 8. **Diffusion Model Scaling Results on ImageNet 256×256 with USiT.** DC-AE-1.5-f32c128 delivers a better scaling curve than DC-AE-f32c32 and DC-AE-f32c128.

Besides Figure 6, we show the FID metric for the diffusion model scaling experiments in Figure 8. The observation is similar, the generation quality of diffusion models using DC-AE-f32c128 remains inferior to DC-AE-f32c32 until the model is scaled to USiT-3B, while DC-AE-1.5-f32c128 can achieve superior Inception Score than DC-AE-f32c32 on USiT-2B and USiT-3B.

B. Evaluation Details

We follow the common practice to evaluate our autoencoders and latent diffusion models on the ImageNet [11] dataset.

For image reconstruction experiments, we compute the metrics using the 50,000 validation images and their reconstruction results.

For image generation experiments, we generate 50,000 images and compute the metrics with the training split.

The evaluation metrics include FID [23], Inception Score [26], Precision [17], and Recall [17]. The specific implementation of these metrics are provided in our codebase.

C. Complete Latent Space Visualization

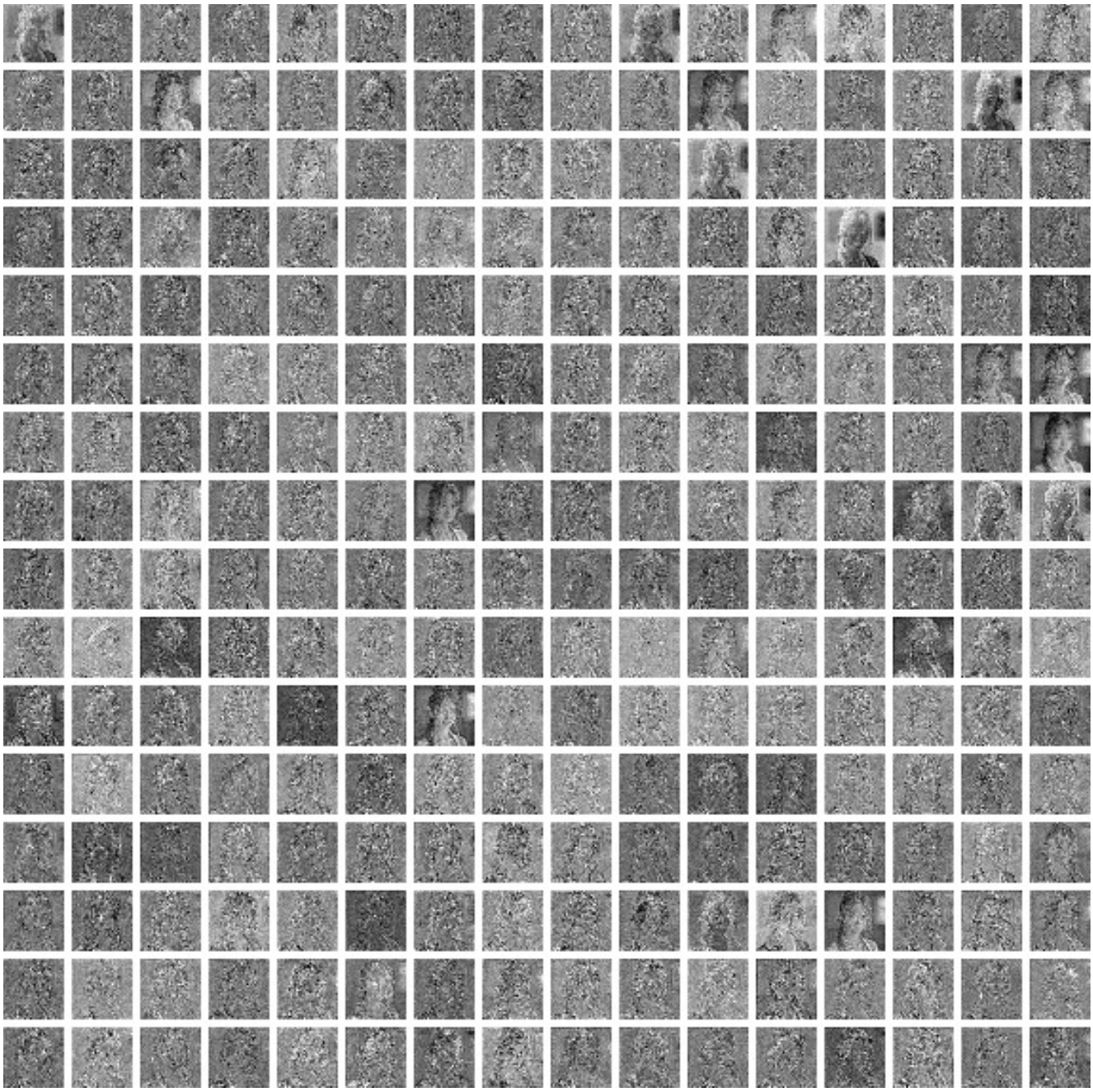
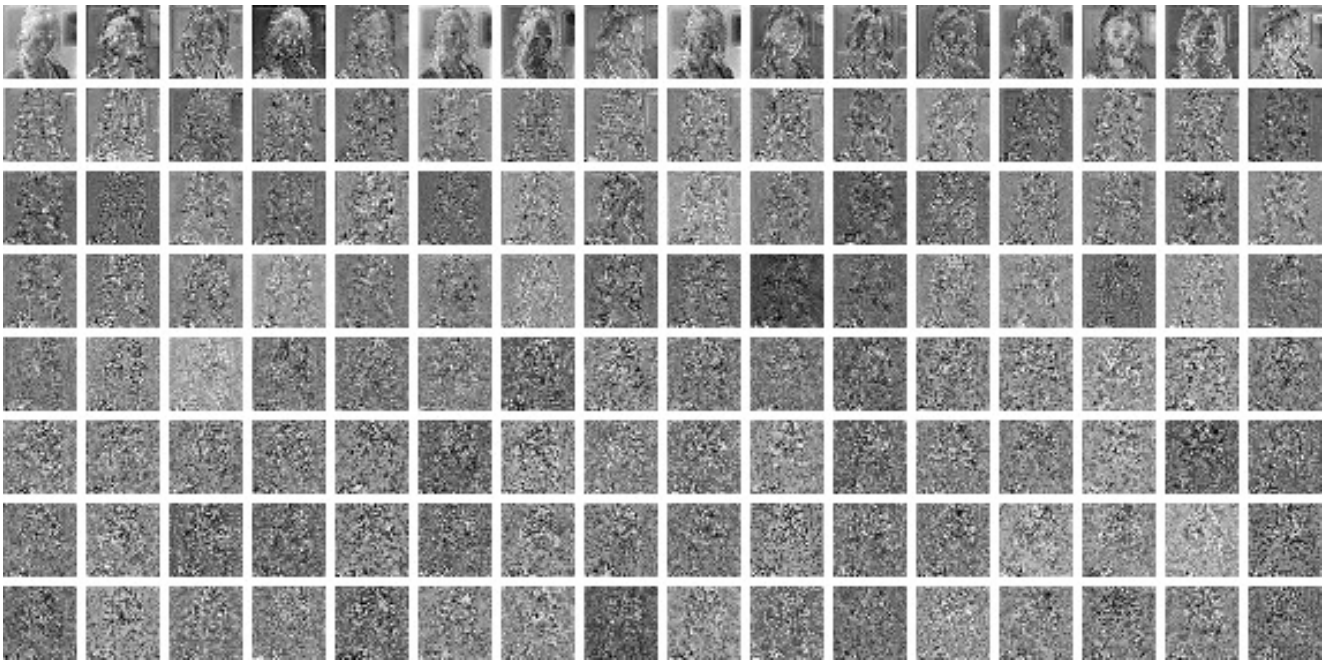


Figure 9. **Complete Latent Space Visualization of DC-AE-f32c256.**

(a) DC-AE-f32c32

(b) DC-AE-f32c128



(c) DC-AE-1.5-f32c128 with structured latent space

Figure 10. **Complete Latent Space Visualization of DC-AE-f32c32, DC-AE-f32c128, and DC-AE-1.5-f32c128.**