Original Image
1024 × 1024

nearest downsample

Original Image
256 × 256

ViT-VQGAN

frozen

Reconstructed Image
256 × 256

Super-Resolution Upsampler

trainable

Reconstructed Image
1024 × 1024