NoiseDiffusion: Correcting Noise for Image Interpolation with Diffusion Models beyond Spherical Linear Interpolation

Pengfei Zheng[1], Yonggang Zhang[2], Zhen Fang[3], Tongliang Liu[4], Defu Lian[1*], Bo Han[2]

[1]University of Science and Technology of China [2]TMLR Group, Hong Kong Baptist University
[3]University of Technology Sydney [4]Sydney AI Centre, The University of Sydney

Background

Spherical linear interpolation method The diffusion models can be used for interpolating images, which holds great potential in generating videos and advertising creativity. The most commonly used image interpolation method in diffusion models is spherical linear interpolation.

$$\mathbf{x}_T^{(\lambda)} = \frac{\sin((1-\lambda)\theta)}{\sin(\theta)}\mathbf{x}_T^{(0)} + \frac{\sin(\lambda\theta)}{\sin(\theta)}\mathbf{x}_T^{(1)} \quad (1)$$

However, when we apply spherical linear interpolation to natural images, the interpolation performance significantly deteriorates.

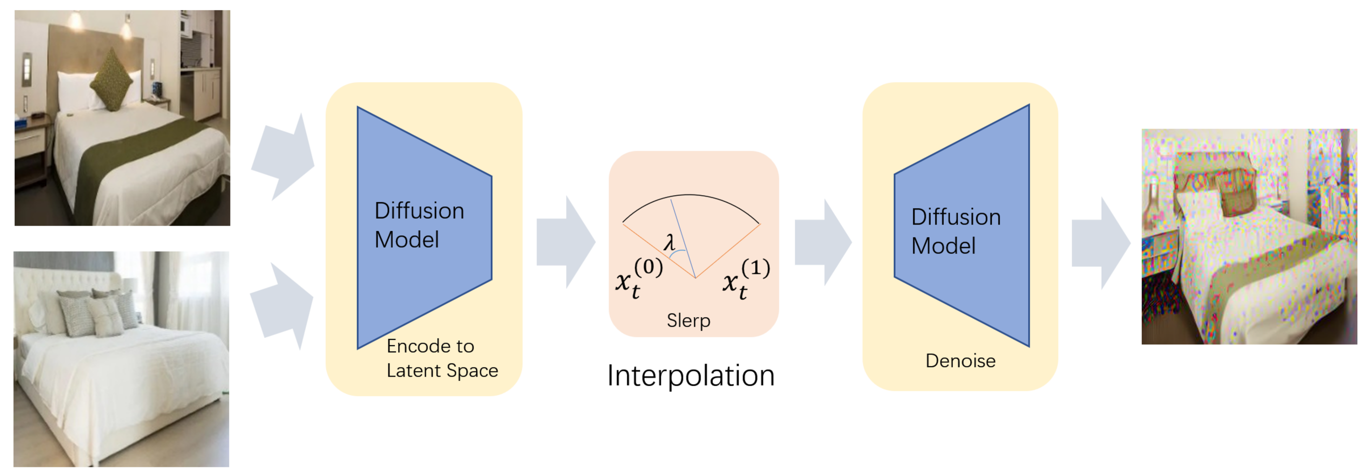

Figure 1: The interpolation of natural images.

Analysis

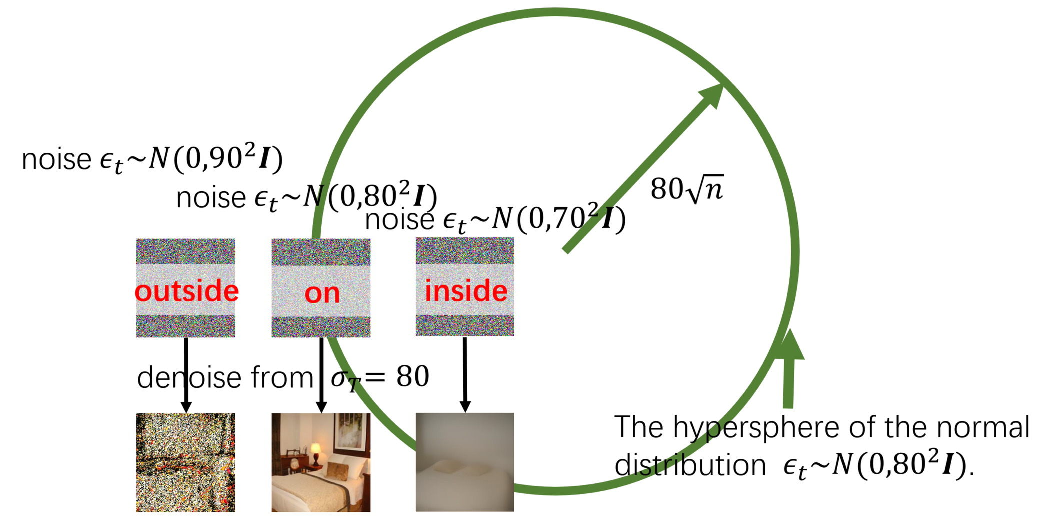

Figure 2: The effect of noise levels.

When the level of Gaussian noise matches the level of denoising (the middle), can we obtain higher-quality images. We use Theorem 1 to explain this phenomenon:

Theorem 1. *The standard normal distribution $\mathcal{N}(0, I_n)$ in high dimensions is close to the uniform distribution on the sphere of radius $\sqrt{n}$.*

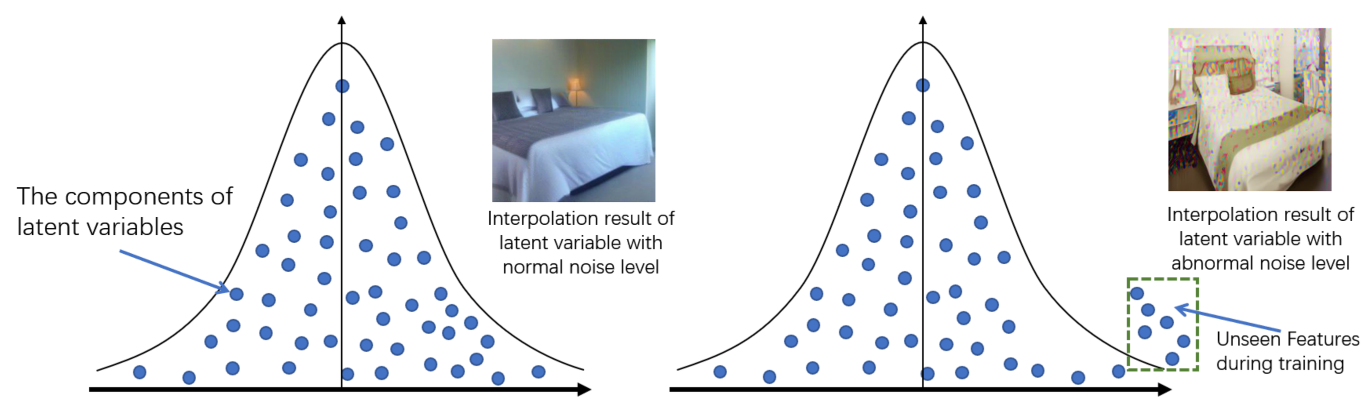

Figure 3: The reason for the failure of natural image interpolation.

Method

Here, we elaborate on our method called NoiseDiffusion.

Design 1

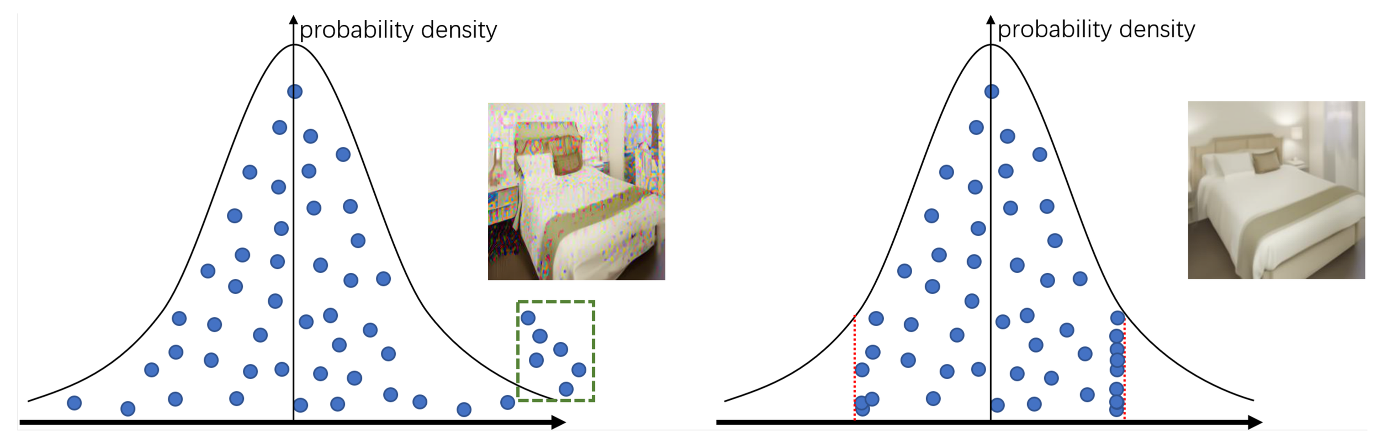

Figure 4: Constraining the extremes of latent variables.

The Gaussian noise at levels higher than the denoising threshold significantly impacts the interpolation results. According to the $68-95-99.7$ rule, components beyond a certain range can be considered outliers. Based on these analyses, we constrain the extremes of latent variables to control their influence.

Design 2

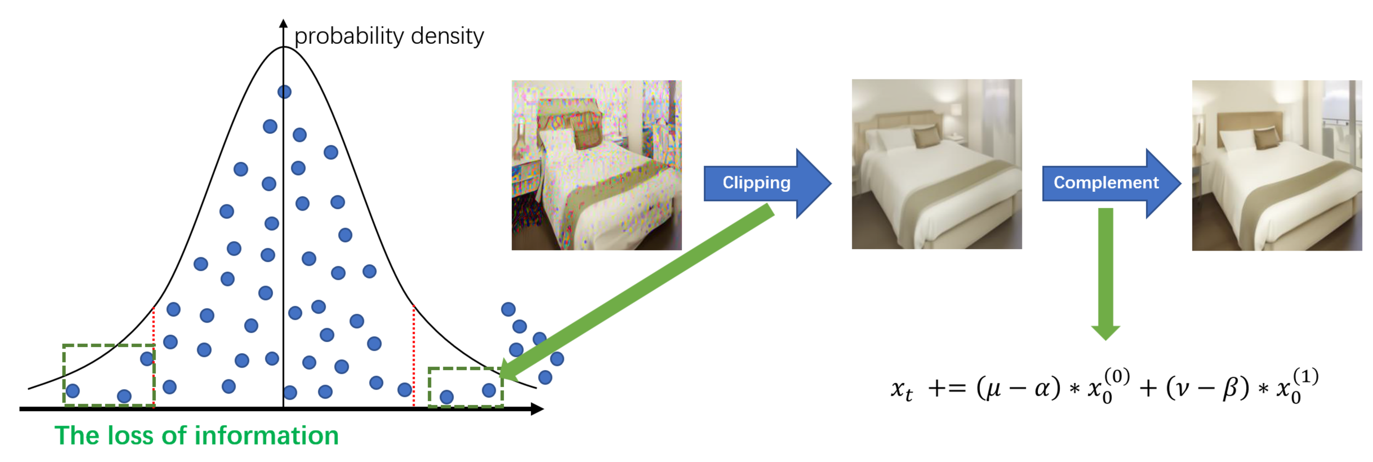

Figure 5: Introducing original image information.

When applying constraints to latent variables, we may inadvertently affect some normal components, leading to information loss. To compensate for this potential loss of information, we introduce original image information as a supplement.

Design 3

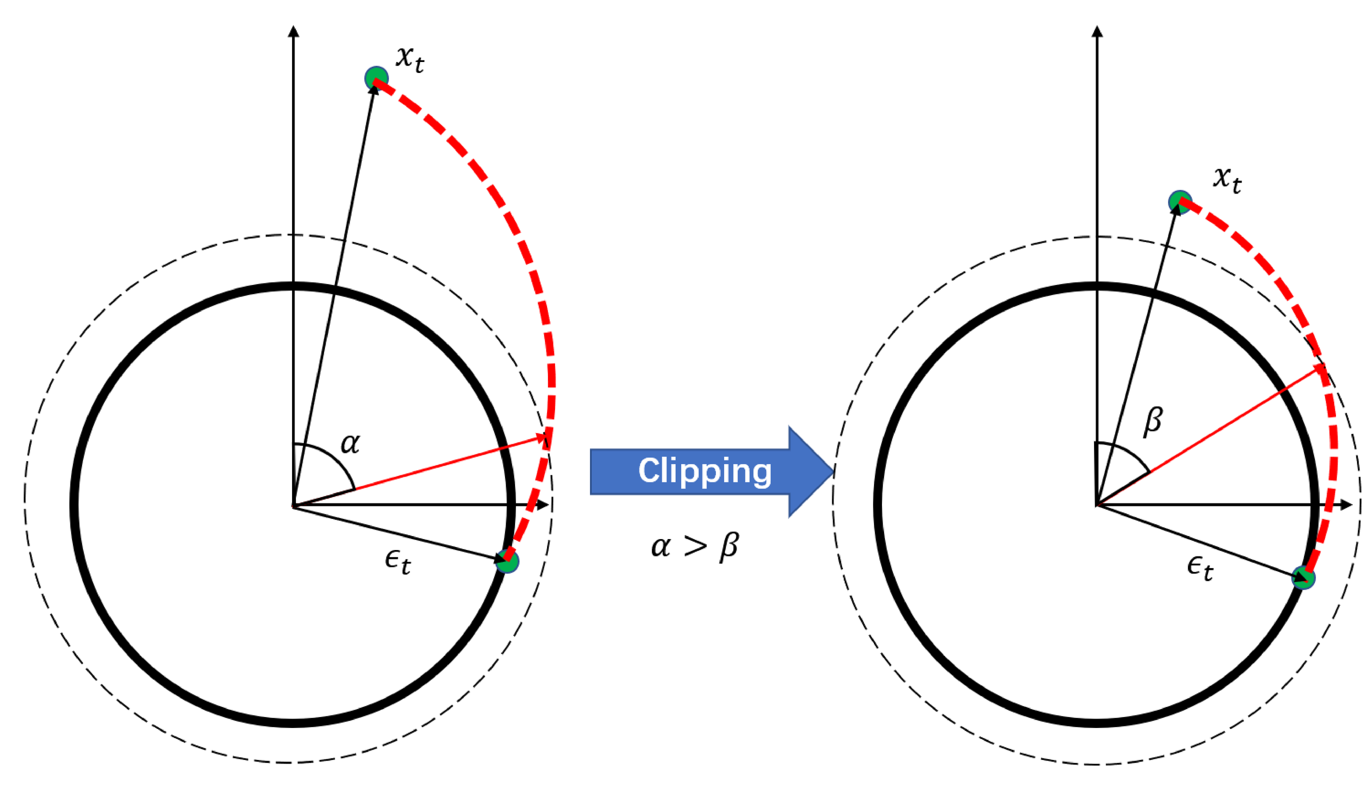

Figure 6: Introducing Gaussian noise.

Method

In practica, we observe that the latent variables are nearly orthogonal. To elucidate this phenomenon, we introduce Theorem 2 as theoretical support.

Theorem 2. *In high-dimensional spaces, independent and isotropic random vectors tend to be almost orthogonal.*

Combining Theorem 1, we can flexibly combine latent variables and Gaussian noise to improve the interpolation results. Specifically, we can use the following formula for combination: $\mathbf{x}_t = \alpha\mathbf{x}_t^{(0)} + \beta\mathbf{x}_t^{(1)} + \gamma\epsilon_t$, where α, β, and γ are combination coefficients, and satisfy $\sqrt{\alpha^2+\beta^2+\gamma^2} = 1$.

By constraining the extremes, we can reduce the amount of introduced noise, enabling us to enhance image quality with minimal introduction of additional information.

Experiments

Figure 7: Comparison with spherical linear interpolation method.

Figure 8: Comparison with spherical linear interpolation method with Stable Diffusion.

Due to the highly unstructured latent space of Stable Diffusion, it is difficult to obtain smooth interpolations at $t = T$ (the pictures in the top row). Therefore, we consider interpolating at smaller time steps (the pictures in the middle row), which could better preserve the features of the original images but may result in lower-quality interpolation. To address this, we apply our method to correct the latent variables (the pictures in the bottom row).