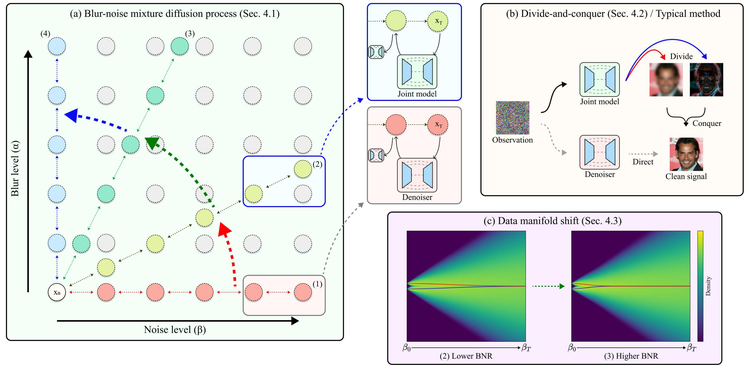
Warm Diffusion: Recipe for Blur-Noise Mixture Diffusion Models

Hao-Chien Hsueh, Wen-Hsiao Peng, Ching-Chun Huang

國立陽明交通大學 NATIONAL YANG MING CHIAO TUNG UNIVERSITY

ICLR

Introduction



Hot Diffusion

- Noise-driven generation
- Performs well across multiple data modalities
- ✘ Struggles to link high-frequency detail with low-frequency structures

Cold Diffusion

- Based on deterministic blurring
- ✘ Prone to out-of-manifold issues
- ✘ Leads to reduced sample diversity & image quality

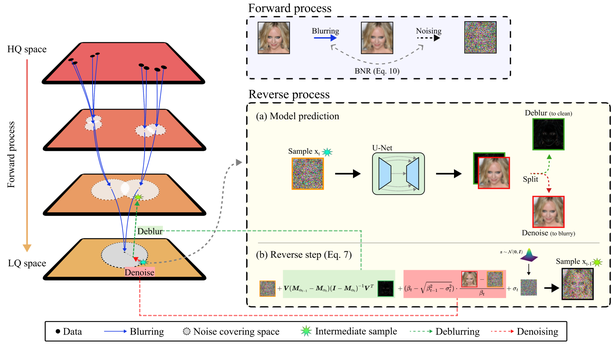
Warm Diffusion (Ours)

A hybrid approach combining the best of both worlds:

- ☑ Spectral Analysis
- ☑ BNR (blur–to–noise ratio) Control to:
  - – Balance spectral dependency
  - – Avoid out-of-manifold pitfalls
- Superior image quality
- Robust across resolutions

Method

Workflow



(a) Divide-and-Conquer Strategy:

- Joint learning of deblurring and denoising

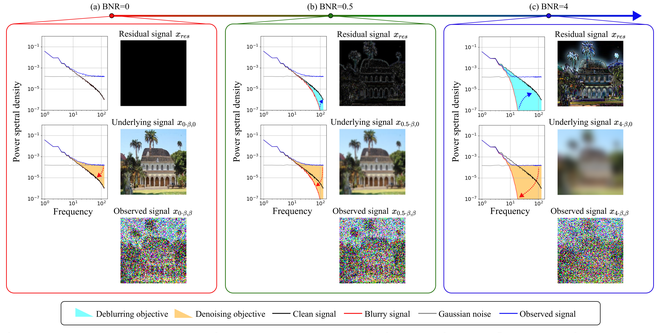
(b) Step-by-Step Refinement:

- Gradual denoising and sharpening the image, leads toward a cleaner, high-quality image space

Blur-to-Noise Ratio (BNR)

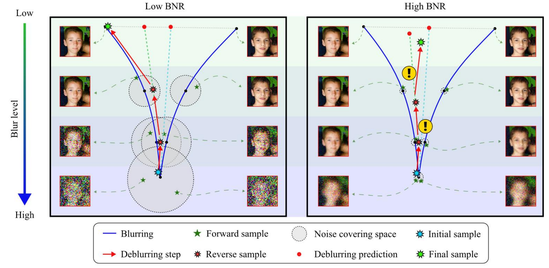
- $BNR = \frac{Blur\ Level}{Noise\ Level} = \frac{\alpha}{\beta}$ → Key lever for controlling the diffusion process

Impact of BNRs on Model Behavior



- Higher BNR → Stronger reliance on spectral dependency for signal reconstruction, but causes a greater shift of the data manifold

Out-of-Manifold Issue



- Due to the ill-posed nature of the deblurring task, samples with higher BNR values are more likely to deviate from the data manifold during the transition, leading to a decline in generation quality

Selecting BNR

- The discrepancy between natural image signals and Gaussian white noise in the frequency spectrum offers an important clue for preserving the data manifold

Experiment Results

Comparison of Generation Quality

Methods	NFE ↓	FID ↓	IS ↑
Unconditional CIFAR-10			
Cold Diffusion (Blur) (Bansal et al., 2022)	50	80.08	-
IHDM (Rissanen et al., 2023)	200	18.96	-
Blurring Diffusion (Hoogeboom & Salimans, 2024)	1000	3.17	9.51
EDM (Karras et al., 2022)	35	1.97	9.78
EDM-ES (Ning et al., 2024)	35	1.95	-
STF (Xu et al., 2023b)	35	1.92	9.79
PFGM++ (Xu et al., 2023a)	35	1.91	-
Ours	35	**1.85**	**10.02**
Class-conditional CIFAR-10			
EDM (Karras et al., 2022)	35	1.79	-
EDM-ES (Ning et al., 2024)	35	1.80	-
PFGM++ (Xu et al., 2023a)	35	1.74	-
Ours	35	**1.68**	**10.19**

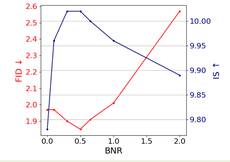
Methods	NFE ↓	FID ↓	IS ↑
FFHQ 64 × 64			
EDM (Karras et al., 2022)	79	2.53	-
PFGM++ (Xu et al., 2023a)	79	2.43	-
Ours	79	**2.29**	**3.41**

Methods	NFE ↓	FID ↓
Unconditional LSUN-church 128 × 128		
Number of samples = 10k		
Denoising Diffusion (Hoogeboom & Salimans, 2024)	1000	4.68
Blurring Diffusion (Hoogeboom & Salimans, 2024)	1000	3.88
Ours	511	3.47
Number of samples = 50k		
IHDM (Rissanen et al., 2023)	400	45.06
Ours	511	2.56

Ablation Studies

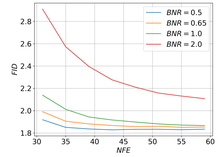
Dataset: Unconditional CIFAR-10

Varied BNRs



Data manifold shifts apparently after BNR > 0.5, leading to a decline in sample quality

BNRs and NFEs



When data manifold shifts, the higher BNR settings require more sampling steps to prevent the out-of-manifold issue