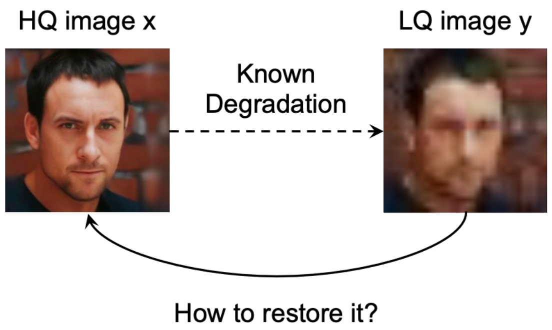
Accelerating Diffusion Models for Inverse Problems through Shortcut Sampling

Gongye Liu, Haoze Sun, Jiayi Li, Fei Yin, Yujiu Yang
Tsinghua University

IJCAI JEJU 2024

Introduction

[Task Definition]



Given a known degradation operator H and the low-quality image y, inverse problems aims to restore the high-quality image x from $y = Hx + n$. Corresponding image restoration tasks including SR, colorization, deblurring, inpainting, …

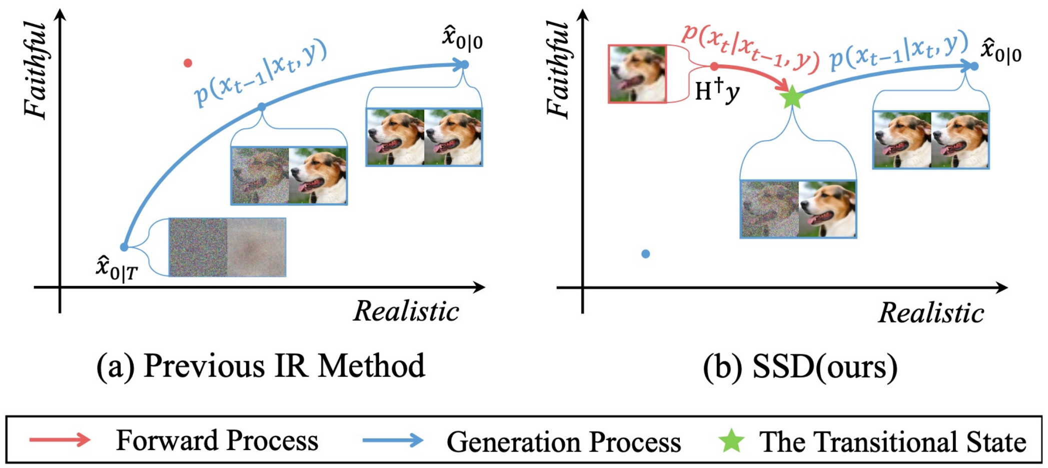
This work mainly focus on Zero-Shot IR with pretrained DM.

What Previous Works Do: Modify the posterior sampling process, from $p(x|z)$ to $p(x|z, y)$, treat it as a conditional generation task(DDRM, DPS, DDNM), which requires more steps to restore.

What We Do: Not all the sampling steps are required. The early stage of denoising can be replaced by a specific "inversion process"

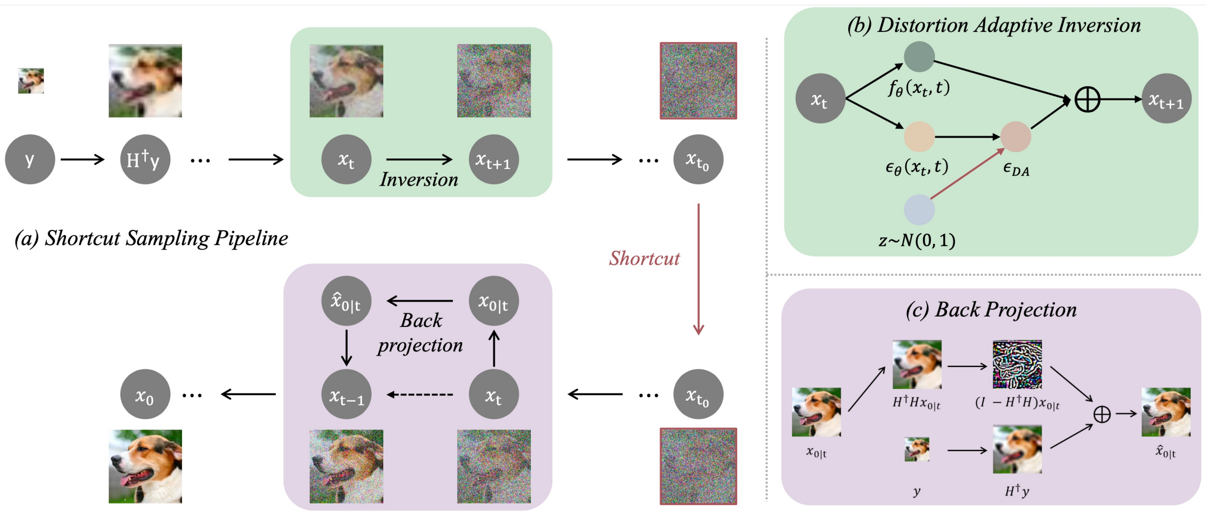
What We Achieve:

- An inversion process that preserve the layout and structure of LQ images;
- A novel method that achieve SOTA performance with fewer steps;



(a) Previous IR Method (b) SSD(ours)

Method



a). Shortcut Sampling Pipeline

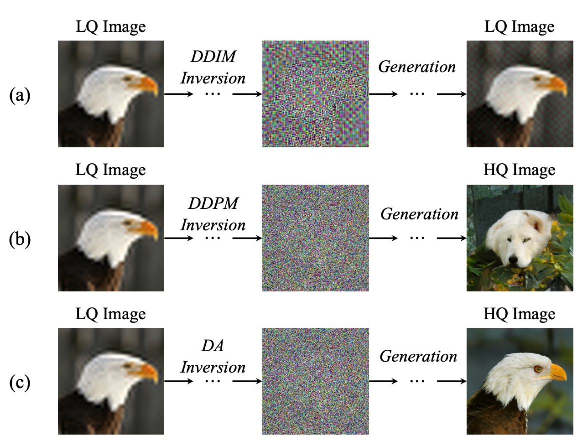
We propose Shortcut Sampling for Diffusion(SSD). Different from previous methods that initiate from noise, We introduce Distortion Adaptive Inversion to replace the early stage of denoising, along with back projection during denoising to force consistency.

b). Distortion Adaptive Inversion

Criteria (i): Contain information from the input image;

Criteria (ii): Retain the capacity for generating HQ images;

DDIM Inversion cannot satisfy Criteria (ii), due to the predict noise deviates from $\mathcal{N}(0, I)$



We propose distortion adaptive inversion to satisfy both:

$$x_{t+1} = \sqrt{\alpha_{t+1}} f_\theta(x_t, t) + \sqrt{1 - \alpha_{t+1} - \eta\beta_{t+1}}\,\epsilon_\theta(x_t, t) + \sqrt{\eta\beta_{t+1}}\,z, \qquad z \sim \mathcal{N}(0, I)$$

c). Back Projection

Denoising Step: $$x_{0|t} = \frac{x_t - \sqrt{1-\alpha_t}\,\epsilon_\theta(x_t, t)}{\sqrt{\alpha_t}}$$

Back Projection Step: $$x_{t-1} = (I - H^{\dagger}H)x_{0|t} + H^{\dagger}y$$

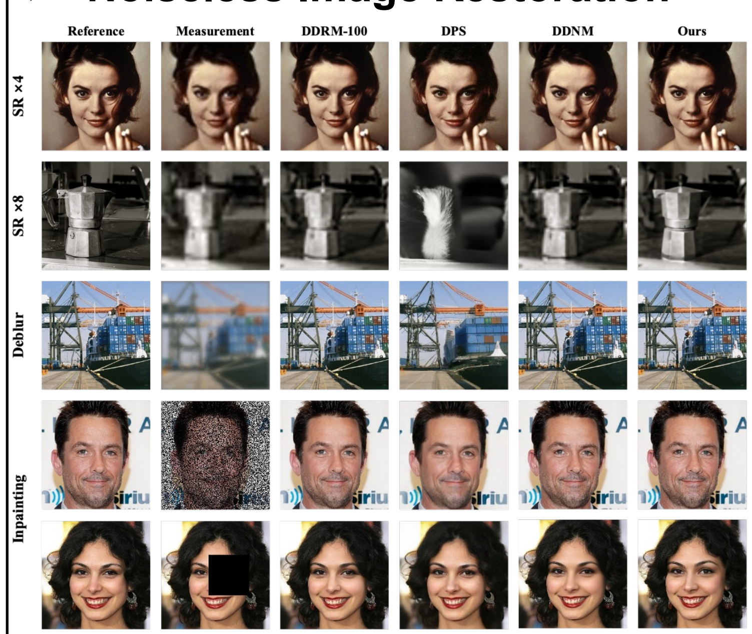
Results

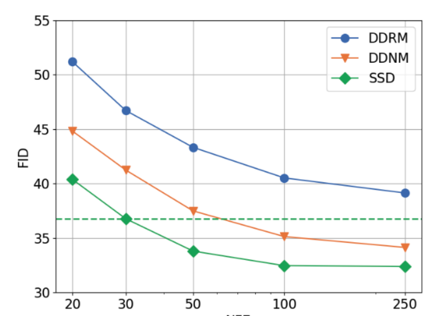
Quantitative Evaluation

CelebA Method	SR × 4 PSNR↑ / FID↓ / LPIPS↓	SR × 8 PSNR↑ / FID↓ / LPIPS↓	Colorization FID↓ / LPIPS↓	Deblur (gauss) PSNR↑ / FID↓ / LPIPS↓	NFEs↓
$H^{\dagger}y$	28.02 / 128.22 / 0.301	24.77 / 153.86 / 0.460	43.99 / 0.197	19.96 / 116.28 / 0.564	0
DDRM-100	28.84 / 40.52 / 0.214	26.47 / 45.22 / 0.273	25.88 / 0.156	36.17 / 15.32 / 0.119	100
DPS	24.71 / 34.69 / 0.304	22.38 / 41.01 / 0.348	N/A	24.89 / 32.64 / 0.288	250
DDNM-100	28.85 / 35.13 / 0.206	26.53 / 44.22 / 0.272	23.65 / 0.138	38.70 / 4.48 / 0.062	100
SSD-100 (ours)	28.84 / 32.41 / 0.202	26.44 / 42.42 / 0.267	23.62 / 0.138	38.62 / 4.36 / 0.060	100
DDRM-30	28.62 / 46.72 / 0.221	26.28 / 49.32 / 0.281	27.69 / 0.214	36.05 / 15.71 / 0.122	30
DDNM-30	28.76 / 41.36 / 0.213	26.41 / 48.25 / 0.277	25.25 / 0.184	37.40 / 6.65 / 0.084	30
SSD-30 (ours)	28.71 / 36.77 / 0.208	26.32 / 44.97 / 0.271	24.11 / 0.159	38.34 / 4.98 / 0.065	30

ImageNet Method	SR × 4 PSNR↑ / FID↓ / LPIPS↓	SR × 8 PSNR↑ / FID↓ / LPIPS↓	Colorization FID↓ / LPIPS↓	Deblur (gauss) PSNR↑ / FID↓ / LPIPS↓	NFEs↓
$H^{\dagger}y$	26.26 / 106.01 / 0.322	22.86 / 124.89 / 0.4690	27.40 / 0.231	19.33 / 102.33 / 0.553	0
DDRM-100	27.40 / 43.27 / 0.260	23.74 / 83.08 / 0.420	36.44 / 0.224	36.48 / 11.81 / 0.121	100
DPS	20.34 / 72.33 / 0.485	18.38 / 76.89 / 0.538	N/A	24.89 / 32.64 / 0.288	250
DDNM-100	27.44 / 39.42 / 0.251	23.80 / 80.09 / 0.412	36.46 / 0.219	40.48 / 3.33 / 0.041	100
SSD-100 (ours)	27.45 / 37.69 / 0.248	23.76 / 82.11 / 0.409	35.40 / 0.215	40.32 / 3.07 / 0.039	100
DDRM-30	27.17 / 46.14 / 0.269	23.50 / 84.53 / 0.426	36.48 / 0.237	35.90 / 13.35 / 0.130	30
DDNM-30	27.22 / 40.12 / 0.256	23.53 / 74.60 / 0.414	36.46 / 0.229	37.67 / 6.91 / 0.081	30
SSD-30 (ours)	27.13 / 38.24 / 0.251	23.44 / 76.35 / 0.411	36.22 / 0.223	39.23 / 4.64 / 0.053	30

Noiseless Image Restoration





Noisy Image Restoration

We restrict the utilization of back projection to the middle stage to adapt to noisy tasks