STRUCT Group Paper Reading

Reconciling Stochastic and Deterministic Strategies for Zero-shot Image Restoration using Diffusion Model in Dual

CVPR 2025

Chong Wang[1], Lanqing Guo[2], Zixuan Fu[1], Siyuan Yang[1], Hao Cheng[3], Alex C. Kot[1], Bihan Wen[1*]
[1]Nanyang Technological University, Singapore
[2]The University of Texas at Austin, USA [3]Hebei University of Technology, China

Presented by Zejia Fan
2026.6.7

Super resolution

- Downsample formulation

$$y = (x * k) \downarrow_s + n$$

y : LR

x : HR

- Consider probability

$$p(x|y)$$

- Bayesian

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

$$p(x|y) \propto p(y|x)p(x)$$

Likelihood

Prior

Super resolution

- Generate HR from LR

Super resolution

- Downsample formulation $$y = (x * k) \downarrow_s + n$$

y : LR

x : HR

- Consider probability $$p(x|y)$$

- Bayesian $$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

$$p(x|y) \propto \underbrace{p(y|x)}_{\text{Likelihood}}\underbrace{p(x)}_{\text{Prior}}$$

Super resolution

- Maximum A Posteriori (MAP)

$$x^* = \arg\max_x p(x|y)$$

$$x^* = \arg\min_x \left[-\log p(y|x) - \log p(x)\right]$$

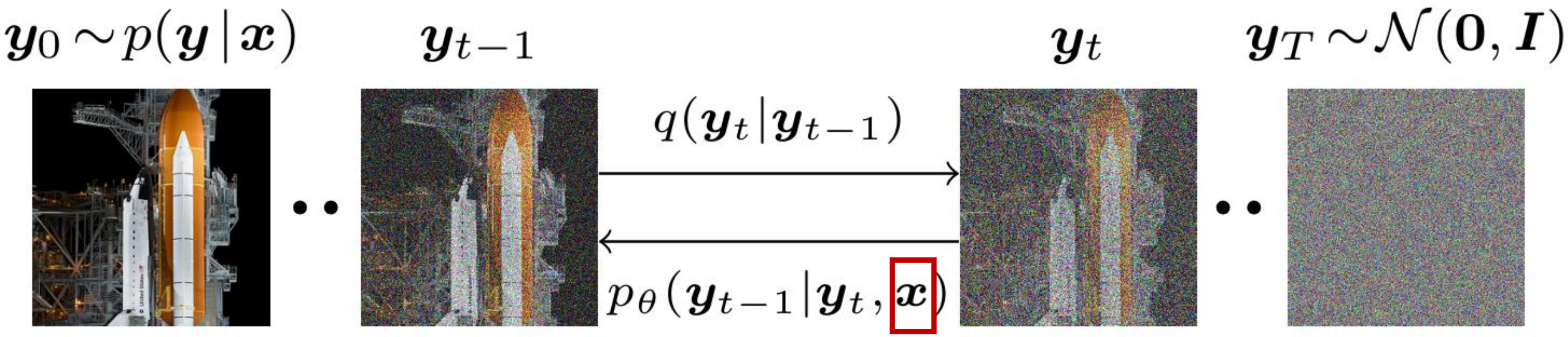
Diffusion

- **Forward / noising process**
  - Sample data $p_0(\mathbf{x}_0) \rightarrow$ turn to noise

- **Reverse / denoising process**
  - Sample noise $\mathbf{x}_T \rightarrow$ turn into data

Diffusion for Image Restoration

- Image Super-Resolution via Repeated Refinement (SR3)



- Conditional denoising diffusion model
- Forward diffusion process $\boldsymbol{q}$, reverse inference process $\boldsymbol{p}$
- Source image $\boldsymbol{x}$, target image $\boldsymbol{y}_0$

Chitwan Saharia, et al. Image Super-Resolution via Iterative Refinement, IEEE TPAMI, 2022

Diffusion for Image Restoration

- Image Super-Resolution via Repeated Refinement (SR3)

Conditioned reverse inference process

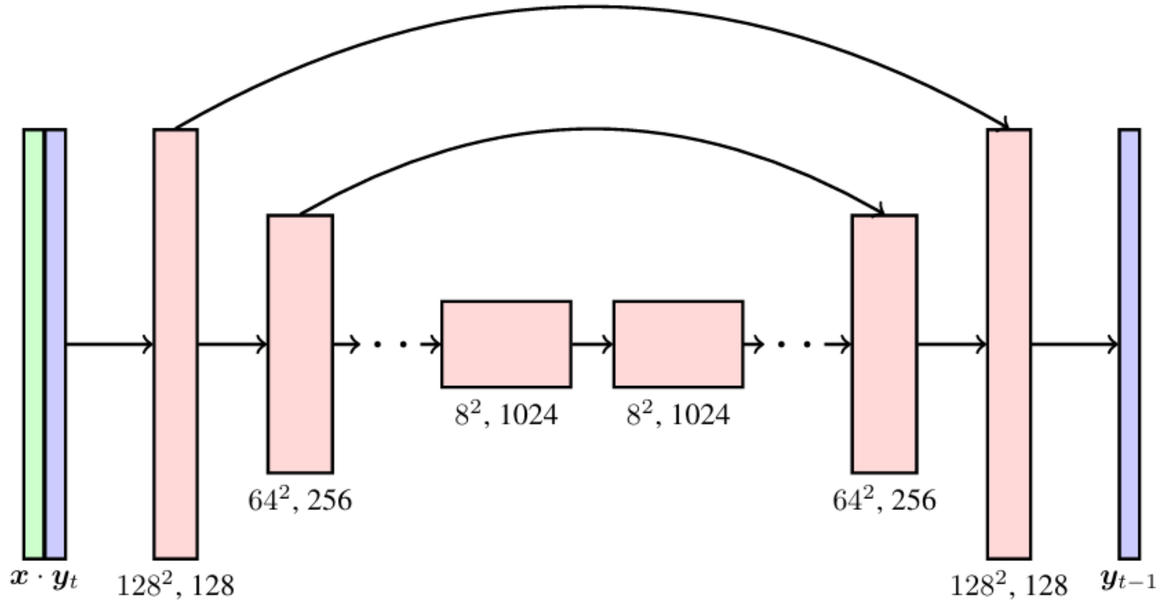
$$
\begin{aligned}
p_\theta(\boldsymbol{y}_{0:T}|\boxed{\boldsymbol{x}}) &= p(\boldsymbol{y}_T)\prod\nolimits_{t=1}^{T} p_\theta(\boldsymbol{y}_{t-1}|\boldsymbol{y}_t,\boldsymbol{x}) \\
p(\boldsymbol{y}_T) &= \mathcal{N}(\boldsymbol{y}_T \mid \mathbf{0}, \boldsymbol{I}) \\
p_\theta(\boldsymbol{y}_{t-1}|\boldsymbol{y}_t,\boxed{\boldsymbol{x}}) &= \mathcal{N}(\boldsymbol{y}_{t-1} \mid \mu_\theta(\boldsymbol{x},\boldsymbol{y}_t,\gamma_t), \sigma_t^2\boldsymbol{I})
\end{aligned}
$$

  - The reverse inference process conditioned on source image $\boldsymbol{x}$
  - $p_\theta(\boldsymbol{y}_{t-1}|\boldsymbol{y}_t,\boldsymbol{x})$ learned by the neural network

Chitwan Saharia, et al. Image Super-Resolution via Iterative Refinement, IEEE TPAMI, 2022

Diffusion for Image Restoration

- Image Super-Resolution via Repeated Refinement (SR3)
  - U-Net architecture
  - x interpolated to target high resolution
  - x concatenate with noisy image y_t

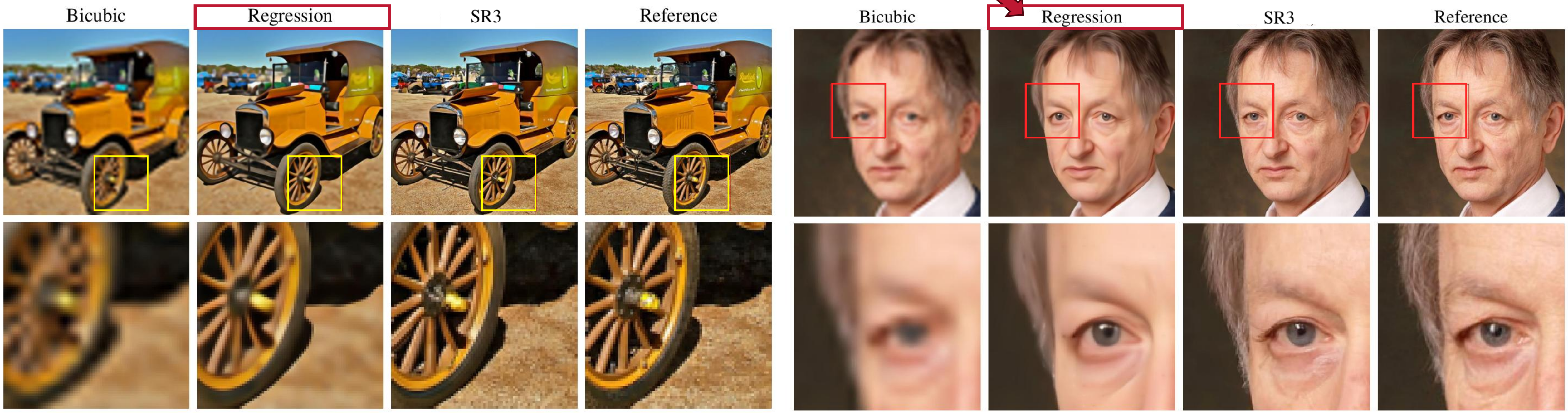


Chitwan Saharia, et al. Image Super-Resolution via Iterative Refinement, IEEE TPAMI, 2022

Diffusion for Image Restoration

- Image Super-Resolution via Repeated Refinement (SR3)

$$x^* = \arg\min_x \left[\boxed{-\log p(y|x)} - \log p(x)\right]$$



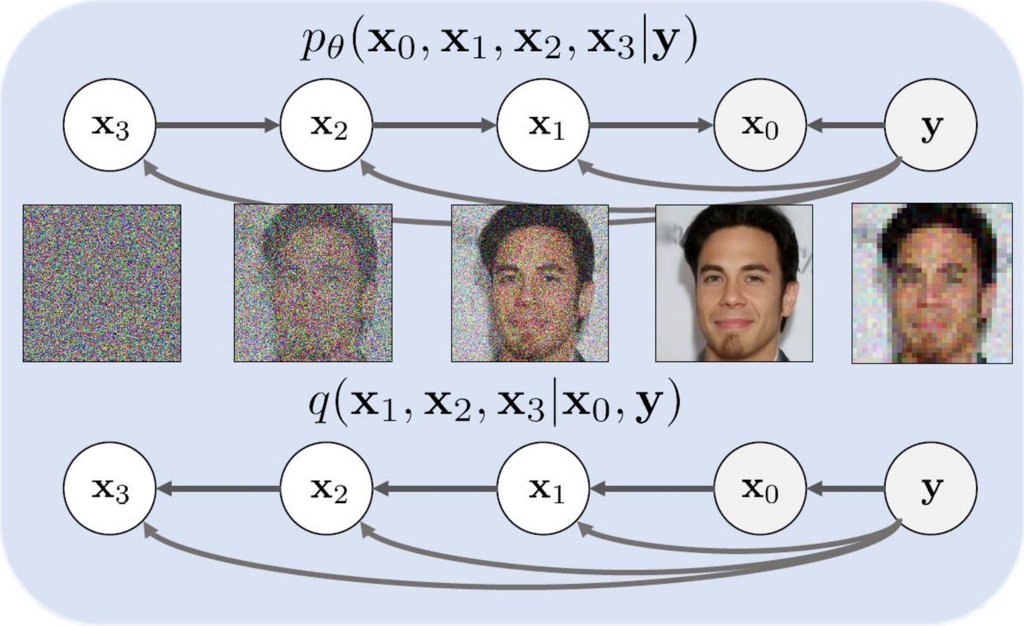
Sampling-based Image Restoration

- Sampling restored images directly from pre-trained models
  - Take the degraded image as guidance

- Example:

Denoising Diffusion Restoration Models

(DDRM)



Denoising Diffusion Restoration Models

Bahjat Kawar, et al. Denoising Diffusion Restoration Models, NIPS, 2022

Sampling-based Image Restoration

- Denoising Diffusion Restoration Models
  - A simplified degradation model in general:

$$\mathbf{y} = \boldsymbol{H}\mathbf{x} + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \sigma_{\mathbf{y}}^2 \boldsymbol{I})$$

  - The target is to find:

$$p_\theta(\mathbf{x}_{0:T} | \mathbf{y}) = p_\theta^{(T)}(\mathbf{x}_T | \mathbf{y}) \prod_{t=0}^{T-1} p_\theta^{(t)}(\mathbf{x}_t | \mathbf{x}_{t+1}, \mathbf{y})$$

  - where θ is the parameter set of a pre-trained diffusion model.

Sampling-based Image Restoration

- Denoising Diffusion Restoration Models
  - The conditional distributions are demonstrated as:

Perform SVD $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$

Correct Sampling Distribution

Distribution of the initial noise

$$\boxed{p_\theta^{(T)}(\bar{\mathbf{x}}_T^{(i)}|\mathbf{y})} = \begin{cases} \mathcal{N}(\bar{\mathbf{y}}^{(i)}, \sigma_T^2 - \frac{\sigma_{\mathbf{y}}^2}{s_i^2}) & \text{if } s_i > 0 \\ \mathcal{N}(0, \sigma_T^2) & \text{if } s_i = 0 \end{cases}$$

$$p_\theta^{(t)}(\bar{\mathbf{x}}_t^{(i)}|\mathbf{x}_{t+1}, \mathbf{y}) = \begin{cases} \mathcal{N}(\bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1-\eta^2}\sigma_t \frac{\bar{\mathbf{x}}_{t+1}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{t+1}}, \eta^2\sigma_t^2) & \text{if } s_i = 0 \\ \mathcal{N}(\bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1-\eta^2}\sigma_t \frac{\bar{\mathbf{y}}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{\mathbf{y}}/s_i}, \eta^2\sigma_t^2) & \text{if } \sigma_t < \frac{\sigma_{\mathbf{y}}}{s_i} \\ \mathcal{N}((1-\eta_b)\bar{\mathbf{x}}_{\theta,t}^{(i)} + \eta_b\bar{\mathbf{y}}^{(i)}, \sigma_t^2 - \frac{\sigma_{\mathbf{y}}^2}{s_i^2}\eta_b^2) & \text{if } \sigma_t \geq \frac{\sigma_{\mathbf{y}}}{s_i} \end{cases}$$

where $\bar{\mathbf{x}}_t$ is the set of vectors in spectral space: $\bar{\mathbf{x}}_t = \boldsymbol{V}^\top \mathbf{x}_t$. Requires known degradation.

Sampling-based Image Restoration

- Denoising Diffusion Restoration Models
  - The conditional distributions are demonstrated as:

Perform SVD $\boldsymbol{H} = \boldsymbol{U\Sigma V}^\top$

Correct Sampling Distribution

$$p_\theta^{(T)}(\bar{\mathbf{x}}_T^{(i)}|\mathbf{y}) = \begin{cases} \mathcal{N}(\bar{\mathbf{y}}^{(i)}, \sigma_T^2 - \frac{\sigma_{\mathbf{y}}^2}{s_i^2}) & \text{if } s_i > 0 \\ \mathcal{N}(0, \sigma_T^2) & \text{if } s_i = 0 \end{cases}$$

Distribution of one single denoise step

$$\boxed{p_\theta^{(t)}(\bar{\mathbf{x}}_t^{(i)}|\mathbf{x}_{t+1}, \mathbf{y})} = \begin{cases} \mathcal{N}(\bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1-\eta^2}\sigma_t \frac{\bar{\mathbf{x}}_{t+1}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{t+1}}, \eta^2\sigma_t^2) & \text{if } s_i = 0 \\ \mathcal{N}(\bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1-\eta^2}\sigma_t \frac{\bar{\mathbf{y}}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{\mathbf{y}}/s_i}, \eta^2\sigma_t^2) & \text{if } \sigma_t < \frac{\sigma_{\mathbf{y}}}{s_i} \\ \mathcal{N}((1-\eta_b)\bar{\mathbf{x}}_{\theta,t}^{(i)} + \eta_b\bar{\mathbf{y}}^{(i)}, \sigma_t^2 - \frac{\sigma_{\mathbf{y}}^2}{s_i^2}\eta_b^2) & \text{if } \sigma_t \geq \frac{\sigma_{\mathbf{y}}}{s_i} \end{cases}$$

where $\bar{\mathbf{x}}_t$ is the set of vectors in spectral space: $\bar{\mathbf{x}}_t = V^\top \mathbf{x}_t$. Requires known degradation.

Sampling-based Image Restoration

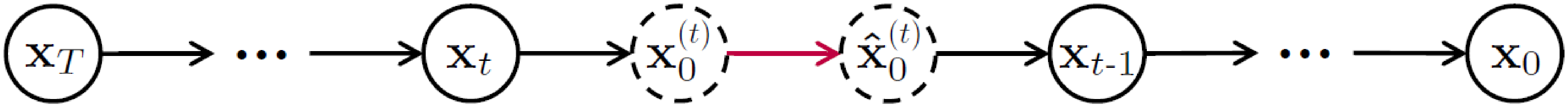
- Denoising Diffusion Restoration Models

(a) Inpainting results on cat images.

(b) Deblurring results ($\sigma_{\mathbf{y}} = 0.05$) on bedroom images.

Sampling-based Image Restoration

- Denoising Diffusion Models for Plug-and-Play Image Restoration (DiffPIR)



Employ a refinement during sampling process:

$$\hat{\mathbf{x}}_0^{(t)} = \arg\min_{\mathbf{x}} \boxed{\|\mathbf{y} - \mathcal{H}(\mathbf{x})\|^2} + \rho_t \|\mathbf{x} - \mathbf{x}_0^{(t)}\|^2$$

Fidelity to degraded image

where $\mathbf{x}_0^{(t)}$ is the predicted noise-free image at timestep t. Requires known degradation.

Yuanzhi Zhu, et al. Denoising Diffusion Models for Plug-and-Play Image Restoration, CVPR, 2023

Sampling-based Image Restoration

- Denoising Diffusion Models for Plug-and-Play Image Restoration

$$\mathbf{x}_T \longrightarrow \cdots \longrightarrow \mathbf{x}_t \longrightarrow \mathbf{x}_0^{(t)} \longrightarrow \hat{\mathbf{x}}_0^{(t)} \longrightarrow \mathbf{x}_{t\text{-}1} \longrightarrow \cdots \longrightarrow \mathbf{x}_0$$

Employ a refinement during sampling process:

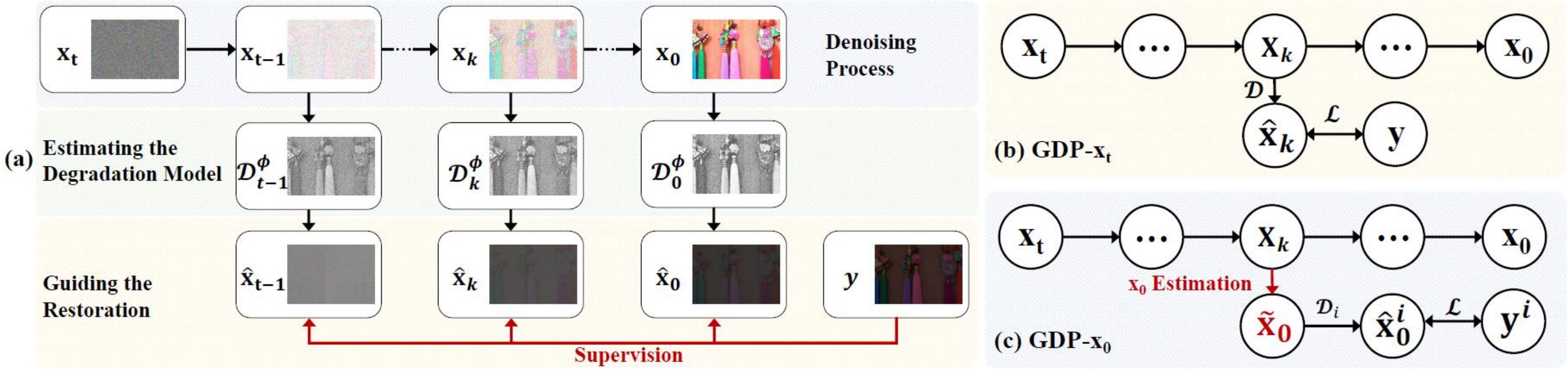
$$\hat{\mathbf{x}}_0^{(t)} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathcal{H}(\mathbf{x})\|^2 + \boxed{\rho_t \|\mathbf{x} - \mathbf{x}_0^{(t)}\|^2}$$

Regularization of generated image

where $\mathbf{x}_0^{(t)}$ is the predicted noise-free image at timestep t. Requires known degradation.

Sampling-based Image Restoration

- Generative Diffusion Prior (GDP)



Learn the degradation model during the sampling process.

Ben Fei, et al. Generative Diffusion Prior for Unified Image Restoration and Enhancement, CVPR, 2023

Sampling-based Image Restoration

- Generative Diffusion Prior (GDP)

At the t-th step, the degradation model will be updated via:

$$\mathcal{L}_{\phi,\tilde{\boldsymbol{x}}_0}^{total} = \mathcal{L}(\boldsymbol{y}, \mathcal{D}_\phi(\tilde{\boldsymbol{x}}_0)) + \mathcal{Q}(\tilde{\boldsymbol{x}}_0)$$
$$\phi \leftarrow \phi - l\nabla_\phi \mathcal{L}_{\phi,\tilde{\boldsymbol{x}}_0}^{total}$$

The noisy image at $(t-1)$-th step will be sampled with guidance provided by the loss:

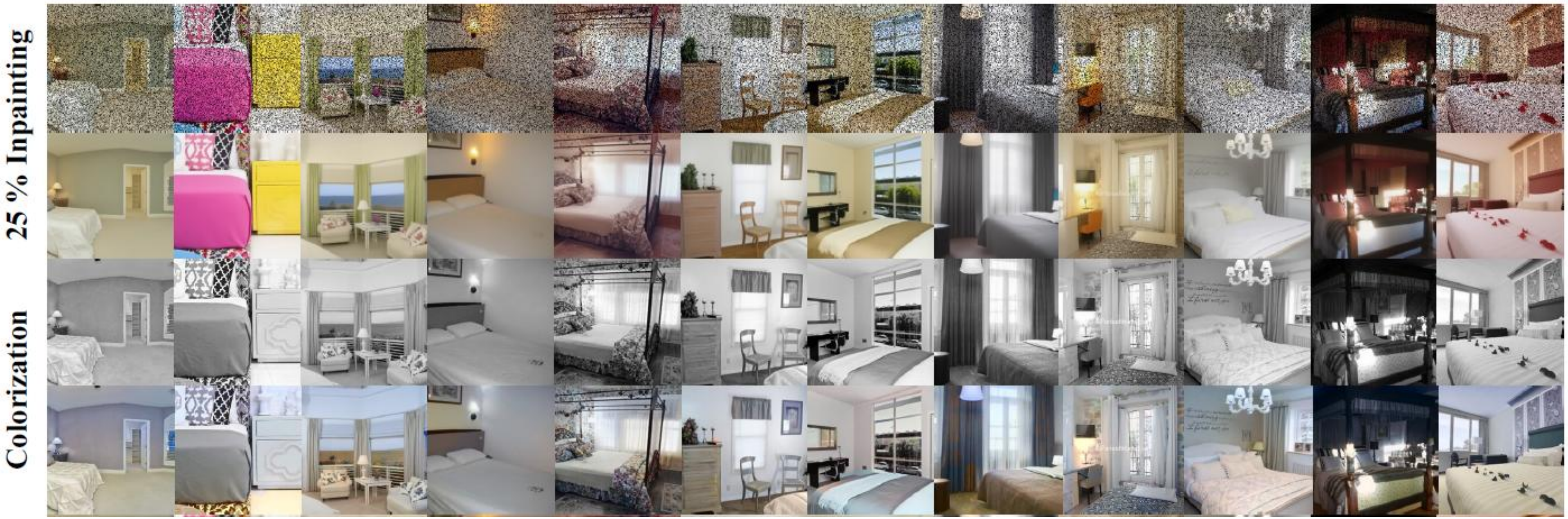
$$\text{Sample } \boldsymbol{x}_{t-1} \text{ by } \mathcal{N}\left(\mu + s\nabla_{\tilde{\boldsymbol{x}}_0}\mathcal{L}_{\phi,\tilde{\boldsymbol{x}}_0}^{total}, \Sigma\right)$$

Ben Fei, et al. Generative Diffusion Prior for Unified Image Restoration and Enhancement, CVPR, 2023

Sampling-based Image Restoration

Ben Fei, et al. Generative Diffusion Prior for Unified Image Restoration and Enhancement, CVPR, 2023

Sampling-based Image Restoration



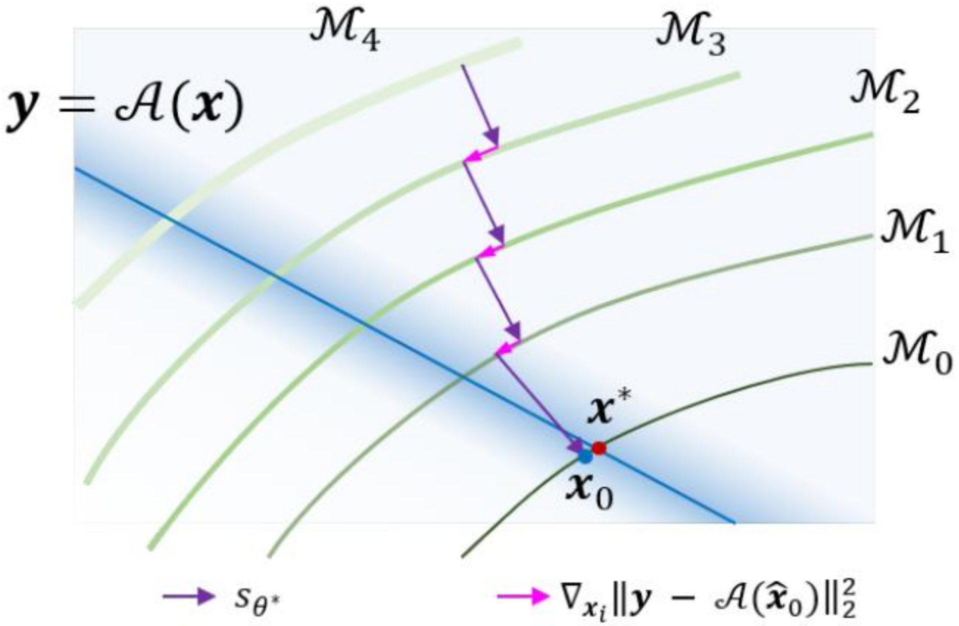
Ben Fei, et al. Generative Diffusion Prior for Unified Image Restoration and Enhancement, CVPR, 2023

Sampling-based Image Restoration

- Diffusion Posterior Sampling (DPS)

refine the noisy state at sampling process via:

$$\boldsymbol{x}_{i-1} \leftarrow \boldsymbol{x}'_{i-1} - \zeta_i \nabla_{\boldsymbol{x}_i} \|\boldsymbol{y} - \mathcal{A}(\hat{\boldsymbol{x}}_0)\|_2^2$$



Hyungjin Chung, et al. Diffusion Posterior Sampling for General Noisy Inverse Problems, ICLR, 2023

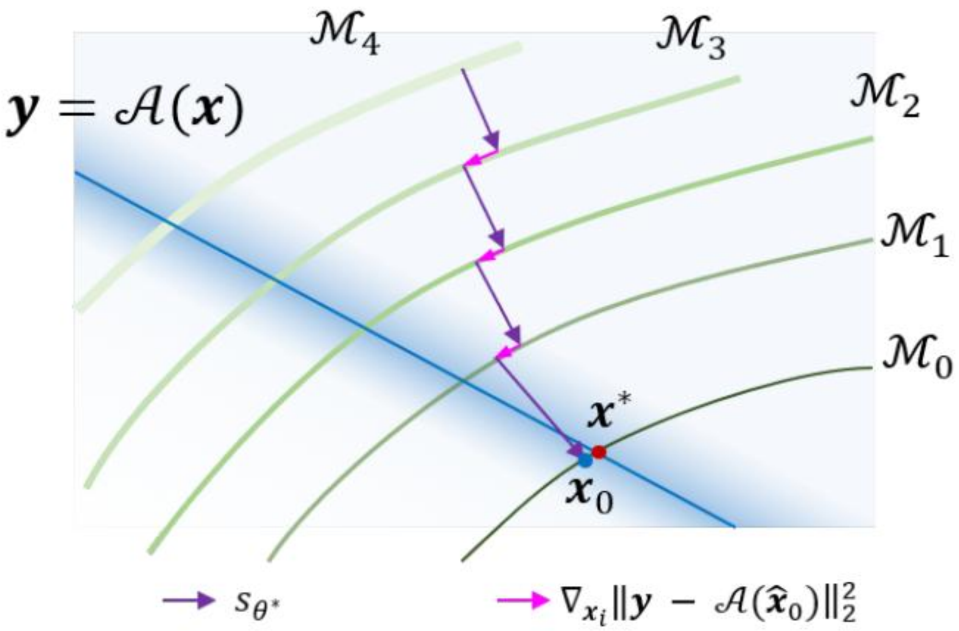
Sampling-based Image Restoration

- Diffusion Posterior Sampling (DPS)

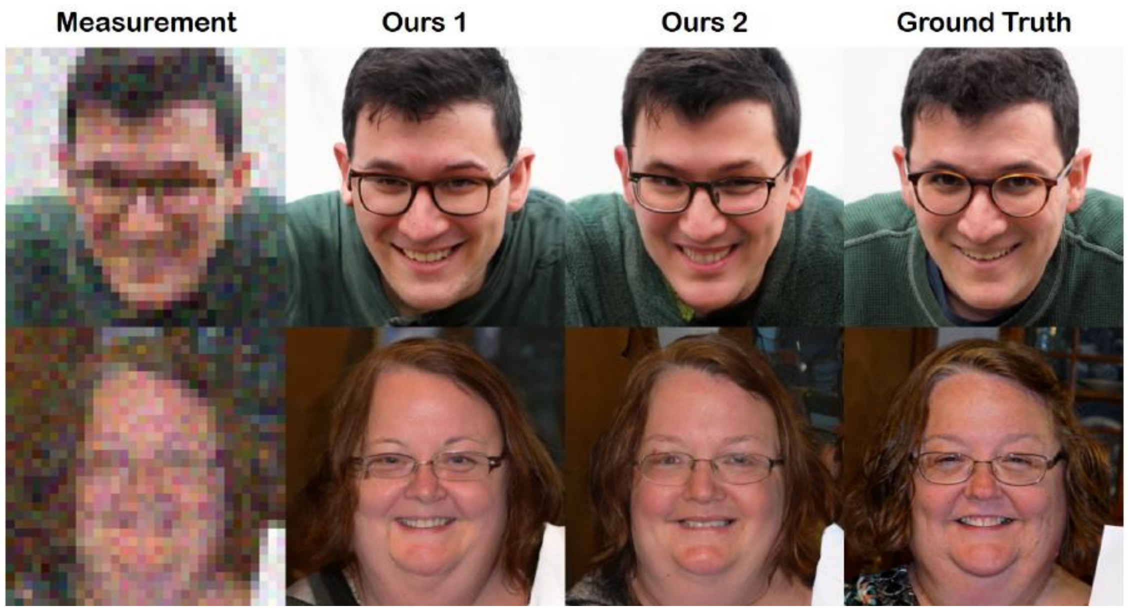
refine the noisy state at sampling process via:

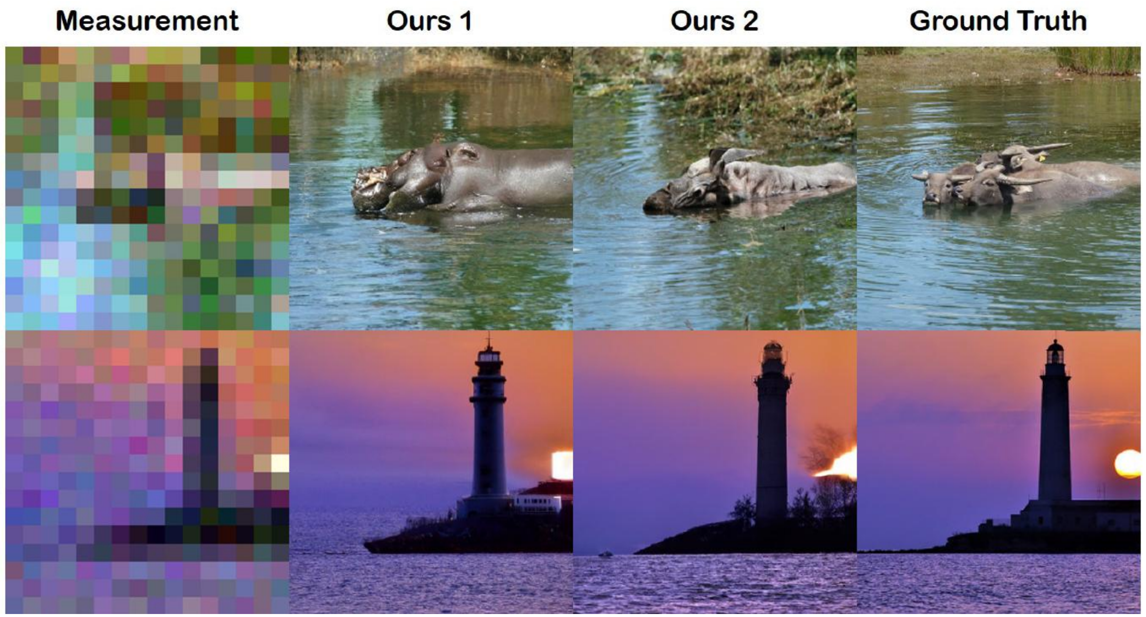
$$x_{i-1} \leftarrow x'_{i-1} - \zeta_i \boxed{\nabla_{x_i} \|y - \mathcal{A}(\hat{x}_0)\|_2^2}$$

Correction Item

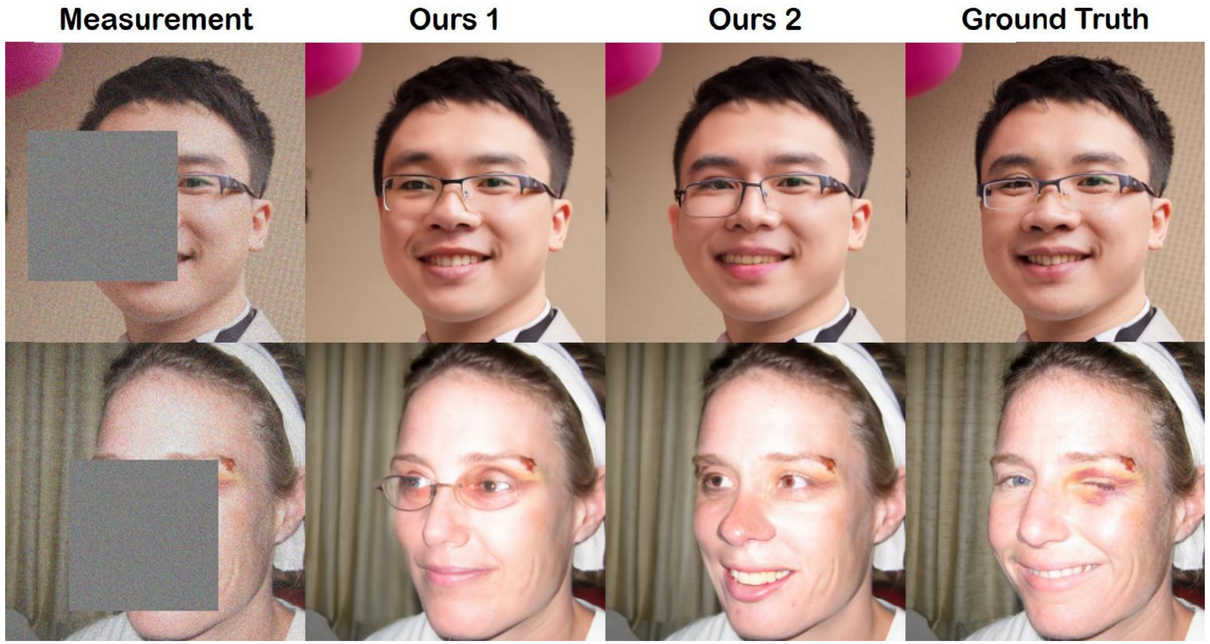


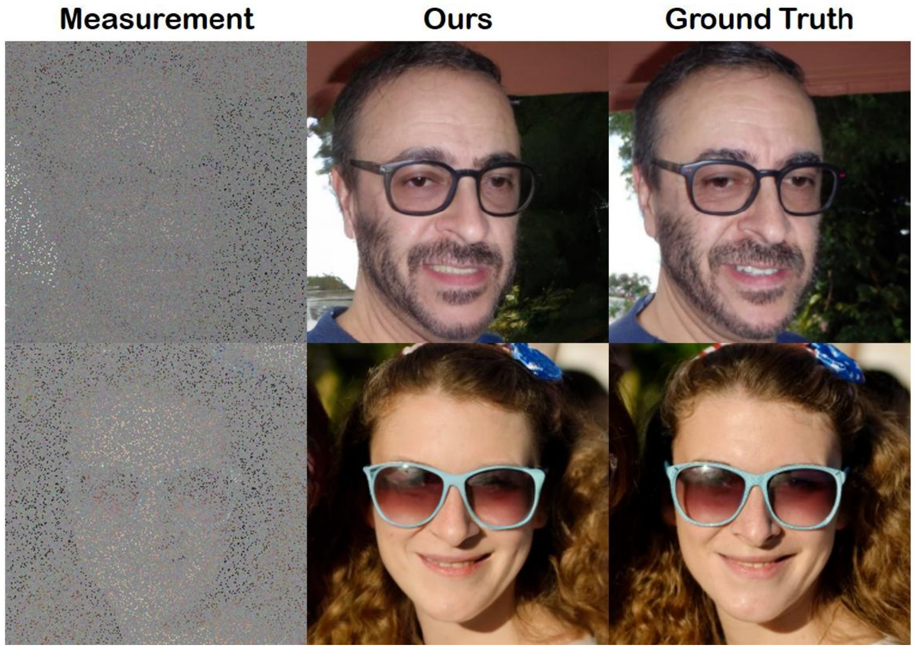
Sampling-based Image Restoration





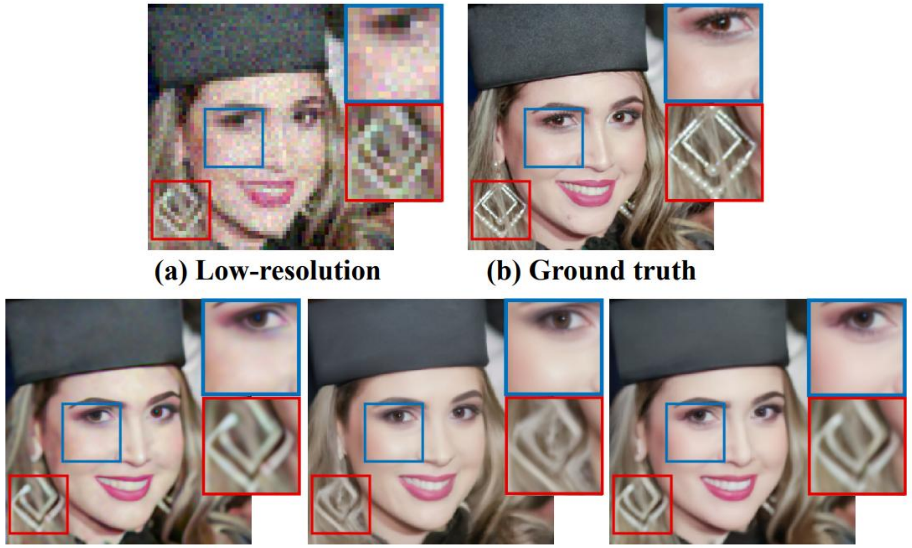
Sampling-based Image Restoration





Reconciling Diffusion Model in Dual

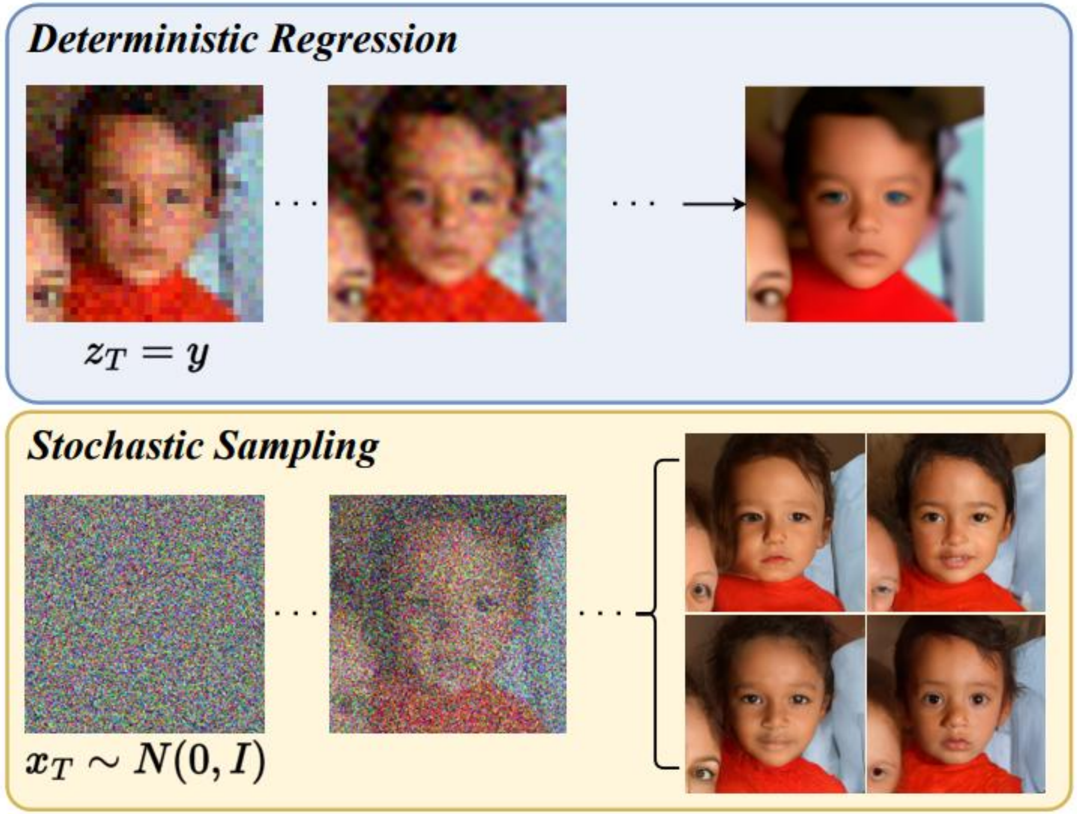
- Based on pre-trained diffusion model
- Half quadratic splitting: two complementary regularizers, stochastic sampling and deterministic regression



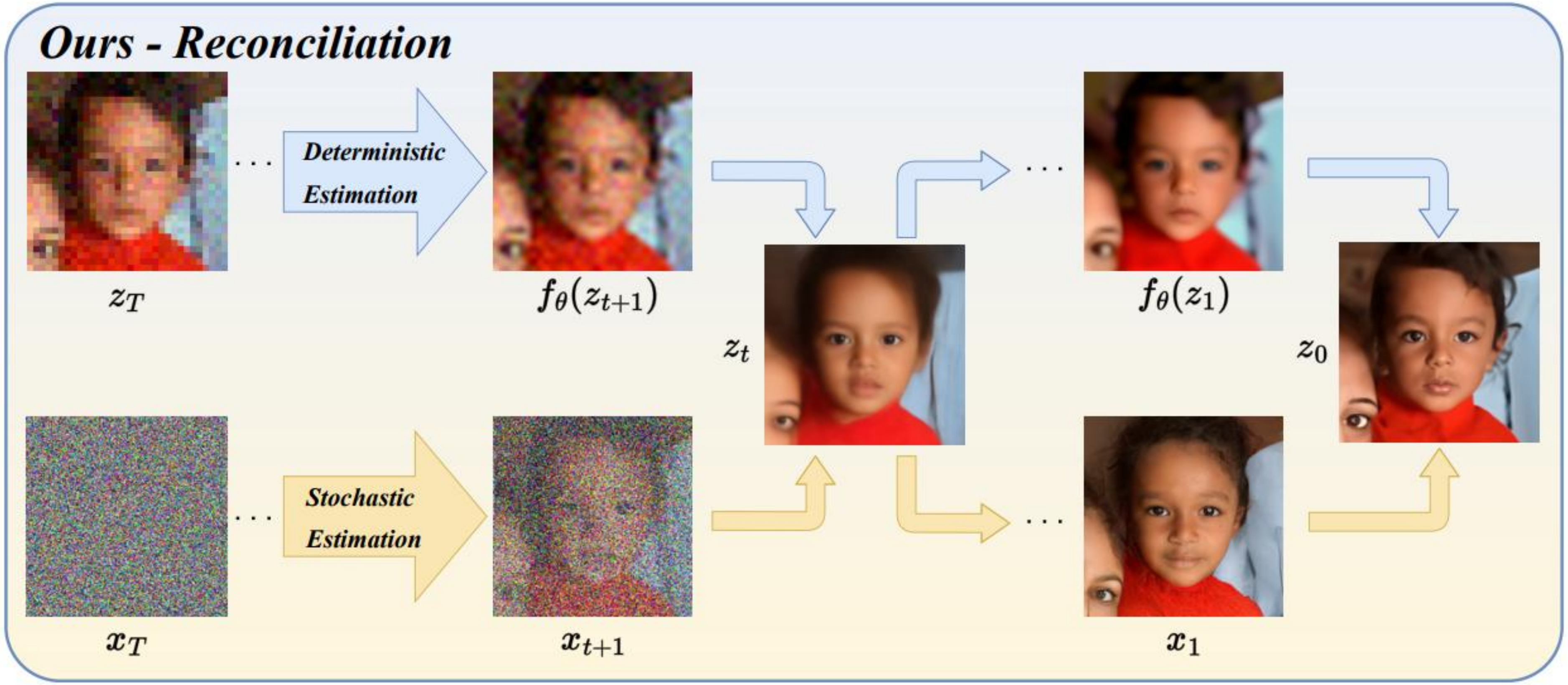
(a) Low-resolution (b) Ground truth

(c) Deterministic (d) Stochastic (e) Reconciled

Reconciling Diffusion Model in Dual



Reconciling Diffusion Model in Dual



Reconciling Diffusion Model in Dual

- HQS algorithm

$$\boldsymbol{x}_t = \underset{\boldsymbol{x}}{\operatorname{argmin}} \frac{1}{2(\sqrt{\tau\lambda/\mu})^2}\|\boldsymbol{x}-\boldsymbol{z}_t\|_2^2 + \mathcal{R}_\theta^S(\boldsymbol{x}),$$

$$\boldsymbol{z}_{t-1} = \underset{\boldsymbol{z}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{y}-A\boldsymbol{z}\|_2^2 + \frac{\mu\sigma_n^2}{2}\|\boldsymbol{z}-\boldsymbol{x}_t\|_2^2 + (1-\tau)\lambda\sigma_n^2\mathcal{R}_\theta^D(\boldsymbol{z}).$$

- Overall

$$\min_{\boldsymbol{z},\boldsymbol{x}} \frac{1}{2\sigma_n^2}\|\boldsymbol{y}-A\boldsymbol{z}\|_2^2 + \frac{\mu}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2 + \boxed{\tau}\lambda\mathcal{R}_\theta(\boldsymbol{x}) + \boxed{(1-\tau)}\lambda\mathcal{R}_\theta(\boldsymbol{z})$$

Fidelity & perceptual

Reconciling Diffusion Model in Dual

Algorithm 1 RDMD

Require: $\boldsymbol{y}, \mathcal{N}_\theta, \lambda, \eta, \tau, \zeta, \{\sigma_t'\}_{t=1}^T$ defined in Section 3.1.

Initialization: $\boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \boldsymbol{z}_T = A^\top \boldsymbol{y}$

for $t = T, \ldots, 1$ **do**

 # pre-calculate t' for deterministic resgression

$t' = \mathrm{argmin}_{i \in [1,T]} |\sigma_t' - \frac{\sqrt{1-\bar{\alpha}_i}}{\sqrt{\bar{\alpha}_i}}|$

 # stochastic estimation

$\boldsymbol{x}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(\boldsymbol{x}_t - \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t; t) \right)$

 # deterministic estimation

$f_\theta(\boldsymbol{z}_t; t') = \frac{1}{\sqrt{\bar{\alpha}_{t'}}} \left(\boldsymbol{z}_t - \sqrt{1 - \bar{\alpha}_{t'}} \boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t; t') \right)$

 # main update process

$\boldsymbol{z}_{t-1} = \boldsymbol{z}_t - \eta \big(\nabla_{\boldsymbol{z}_t} \|\boldsymbol{y} - A\boldsymbol{z}_t\|_2^2 + \tau\lambda \frac{\sigma_n^2}{\sigma_t^2} (\boldsymbol{z}_t - \boldsymbol{x}_{0|t}) + (1-\tau)\lambda\sigma_n^2 (\boldsymbol{z}_t - f_\theta(\boldsymbol{z}_t; t')) \big)$

 # renoising

$\hat{\boldsymbol{\epsilon}}_t = \frac{1}{\sqrt{1-\bar{\alpha}_t}} (\boldsymbol{x}_t - \sqrt{\bar{\alpha}_t} \boldsymbol{z}_{t-1})$

$\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$

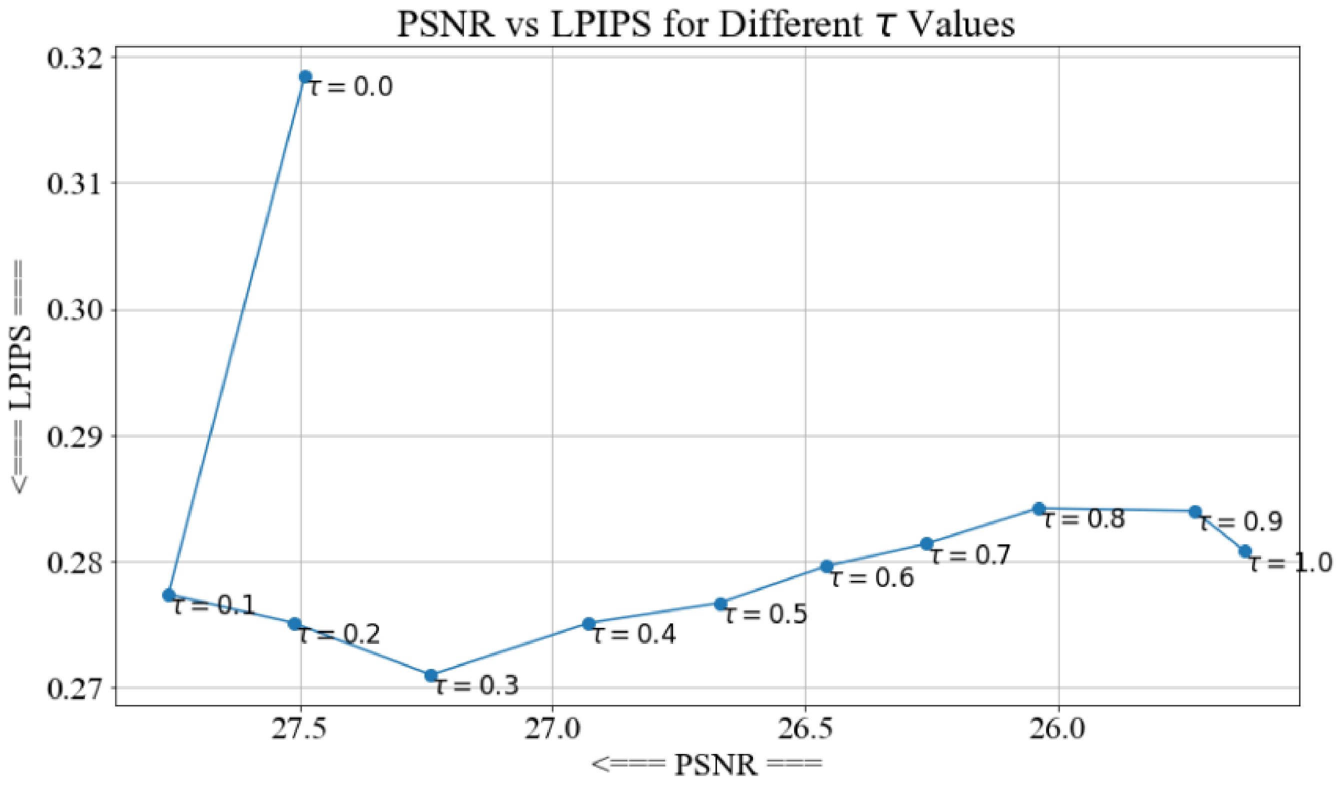
$\boldsymbol{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \boldsymbol{z}_{t-1} + \sqrt{1 - \bar{\alpha}_{t-1}} (\sqrt{1-\zeta} \hat{\boldsymbol{\epsilon}}_t + \sqrt{\zeta} \boldsymbol{\epsilon}_t)$

end for

return $\boldsymbol{x}_0$

Experiment

- Fidelity & perceptual



Experiment

- Fidelity & perceptual

Experiment

- Quantitative Comparison - FFHQ

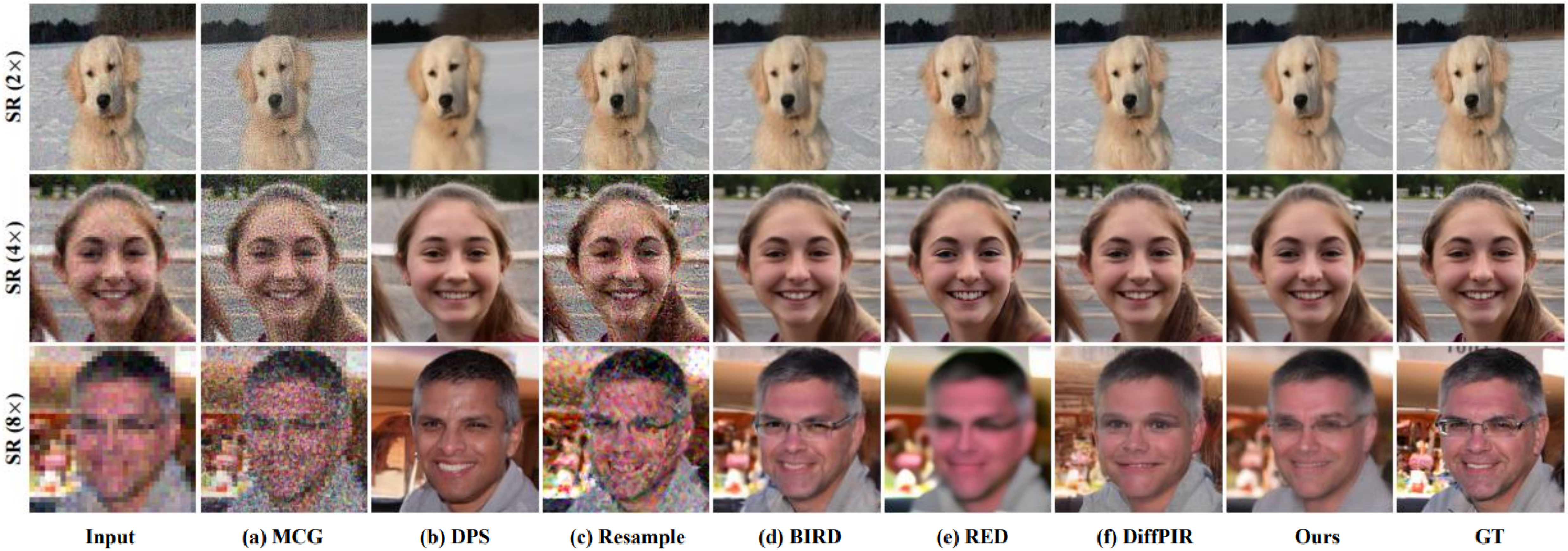
FFHQ	Deblur (*Gaussian*)			Deblur (*motion*)			Super-resolution (2×)			Super-resolution (4×)			Super-resolution (8×)		
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
MCG[8]	11.19	0.0968	0.8042	11.33	0.0831	0.8088	19.57	0.3482	0.5648	17.34	0.2153	0.7190	16.81	0.1829	0.7676
DPS[9]	23.78	0.6776	0.2829	20.64	0.5732	0.3461	25.06	0.7418	0.2738	22.85	0.6503	0.3215	20.54	0.5597	0.3726
Resample[29]	23.50	0.4962	0.4612	18.60	0.2566	0.6455	23.07	0.4754	0.4828	18.28	0.2563	0.6691	18.74	0.2712	0.6807
BIRD[7]	26.14	0.7009	0.2928	24.21	0.6419	0.3412	27.04	0.6150	0.3310	22.51	0.3870	0.5190	19.67	0.3070	0.6340
RED[25](Ours - deterministic)	27.84	**0.8096**	0.3170	26.09	0.7372	0.3613	27.87	0.7919	0.2562	27.05	0.7960	0.3277	22.22	0.6431	0.4883
DiffPIR[43](Ours - stochastic)	27.30	0.7672	**0.2372**	26.63	0.7399	**0.2529**	29.39	0.8060	0.2172	26.27	0.7332	**0.2674**	21.08	0.5634	0.3666
Ours	**28.29**	0.8057	0.2770	**27.05**	**0.7776**	0.2767	**31.03**	**0.8715**	**0.1807**	**27.84**	**0.8012**	0.2768	**23.31**	**0.6524**	**0.3614**

Experiment

- Quantitative Comparison - ImageNet

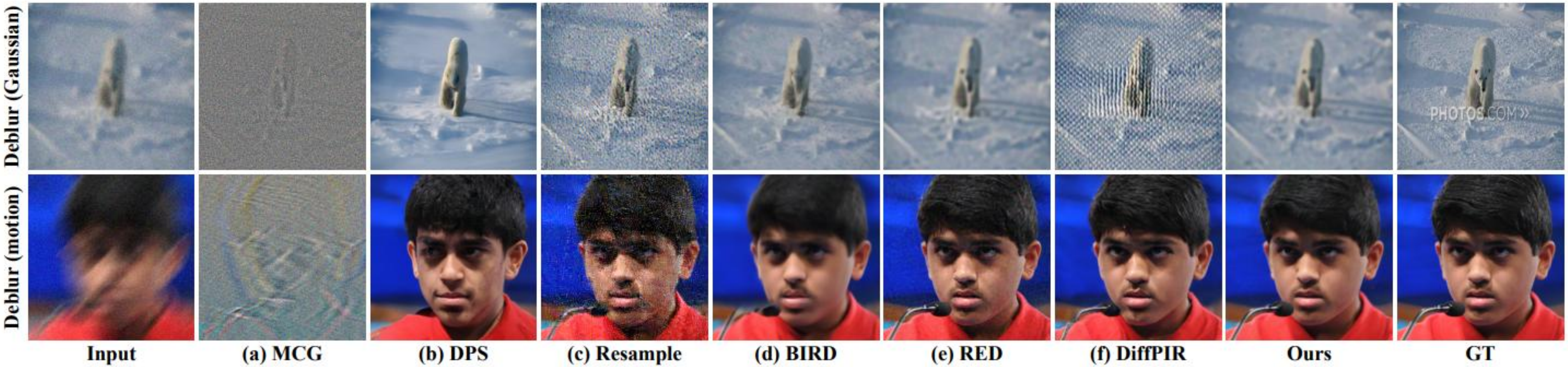
ImageNet	**Deblur (*Gaussian*)**			**Deblur (*motion*)**			**Super-resolution (2×)**			**Super-resolution (4×)**			**Super-resolution (8×)**		
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
MCG[8]	11.11	0.0982	0.7338	11.05	0.0362	0.7535	15.12	0.2217	0.6259	16.31	0.2150	0.6727	15.87	0.1507	0.7271
DPS[9]	20.09	0.4774	0.4272	20.14	0.4871	0.4336	21.31	0.5514	0.4392	19.47	0.4510	0.4961	17.65	0.3522	<u>0.5505</u>
Resample[29]	18.88	0.2957	0.5579	15.49	0.1857	0.6571	21.01	0.4674	0.4334	16.46	0.2235	0.6163	17.55	0.2207	0.6968
BIRD[7]	22.04	0.5360	0.4107	20.79	0.4976	0.4499	24.94	0.6841	0.3047	21.37	0.4521	0.5001	19.67	0.3073	0.6334
RED[25](Ours - deterministic)	<u>23.24</u>	**0.6090**	0.4282	22.52	0.5674	0.4963	<u>25.87</u>	<u>0.7199</u>	0.3023	<u>23.16</u>	<u>0.6027</u>	0.4324	<u>20.42</u>	<u>0.4634</u>	0.6024
DiffPIR[43](Ours - stochastic)	21.97	0.5363	**0.3827**	**23.34**	**0.6276**	**0.3736**	25.33	0.7092	<u>0.2900</u>	22.80	0.5967	**0.3871**	19.44	0.4292	**0.5411**
Ours	**23.30**	<u>0.6004</u>	<u>0.4079</u>	<u>22.77</u>	<u>0.6037</u>	<u>0.4247</u>	**25.94**	**0.7563**	**0.2648**	**23.28**	**0.6203**	<u>0.4000</u>	**20.83**	**0.4736**	0.5690

Experiment



SR (2×)
SR (4×)
SR (8×)
Input
(a) MCG
(b) DPS
(c) Resample
(d) BIRD
(e) RED
(f) DiffPIR
Ours
GT

Experiment



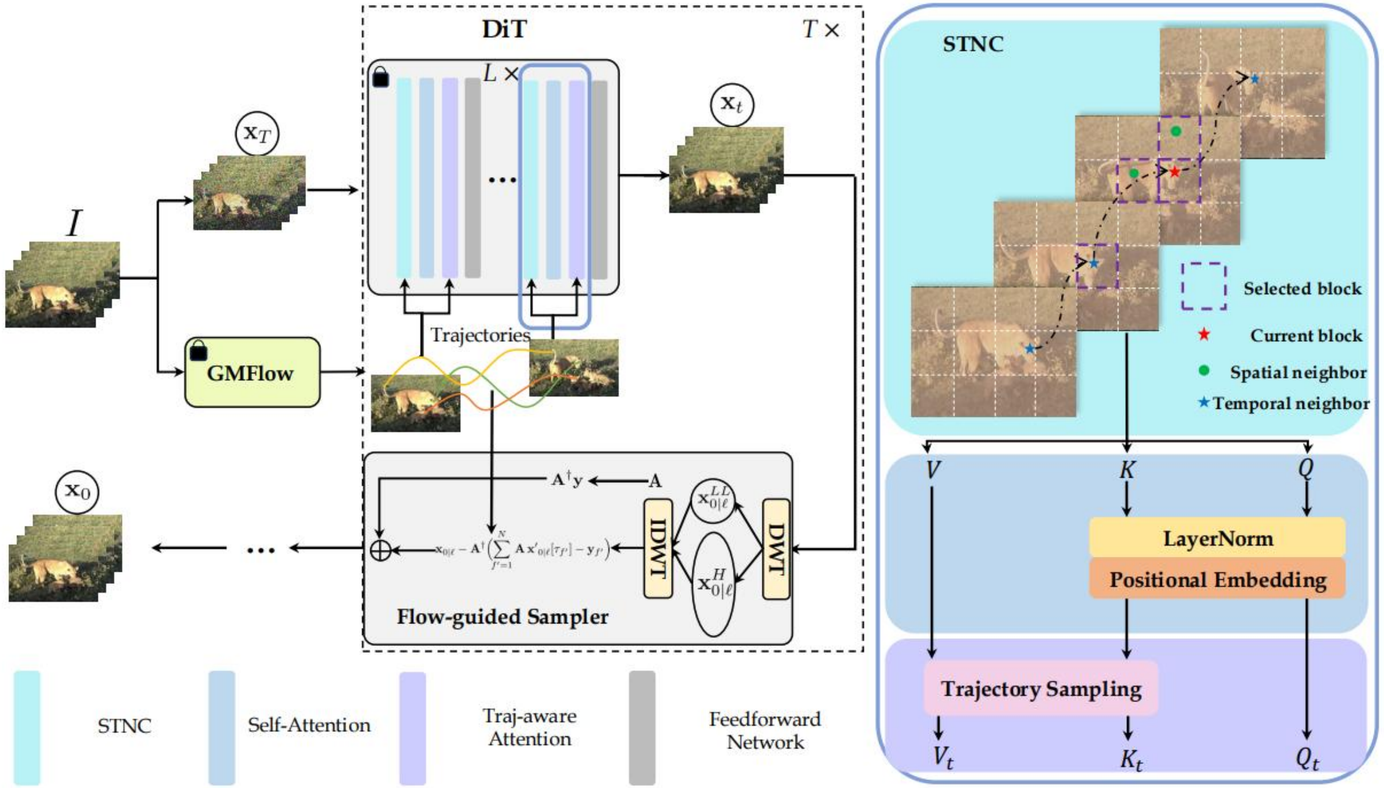
Deblur (Gaussian)
Deblur (motion)
PHOTOS.COM
Input
(a) MCG
(b) DPS
(c) Resample
(d) BIRD
(e) RED
(f) DiffPIR
Ours
GT

Sampling-based Restoration

DiTVR: Zero-Shot Diffusion Transformer for Video Restoration

- CVPR 26

Sicheng Gao Nancy Mehta Zongwei Wu Radu Timofte
Computer Vision Lab, CAIDAS & IFI, University of Würzburg
sicheng.gao@uni-wuerzburg.de



Conclusion

- Restoration problem can be formulated as 2 main parts
  - Likelihood - fidelity
  - Prior – perceptual quality
- Diffusion prior can act as prior regularization
- Balance between fidelity and perceptual quality is still tricky

Thanks for your listening!