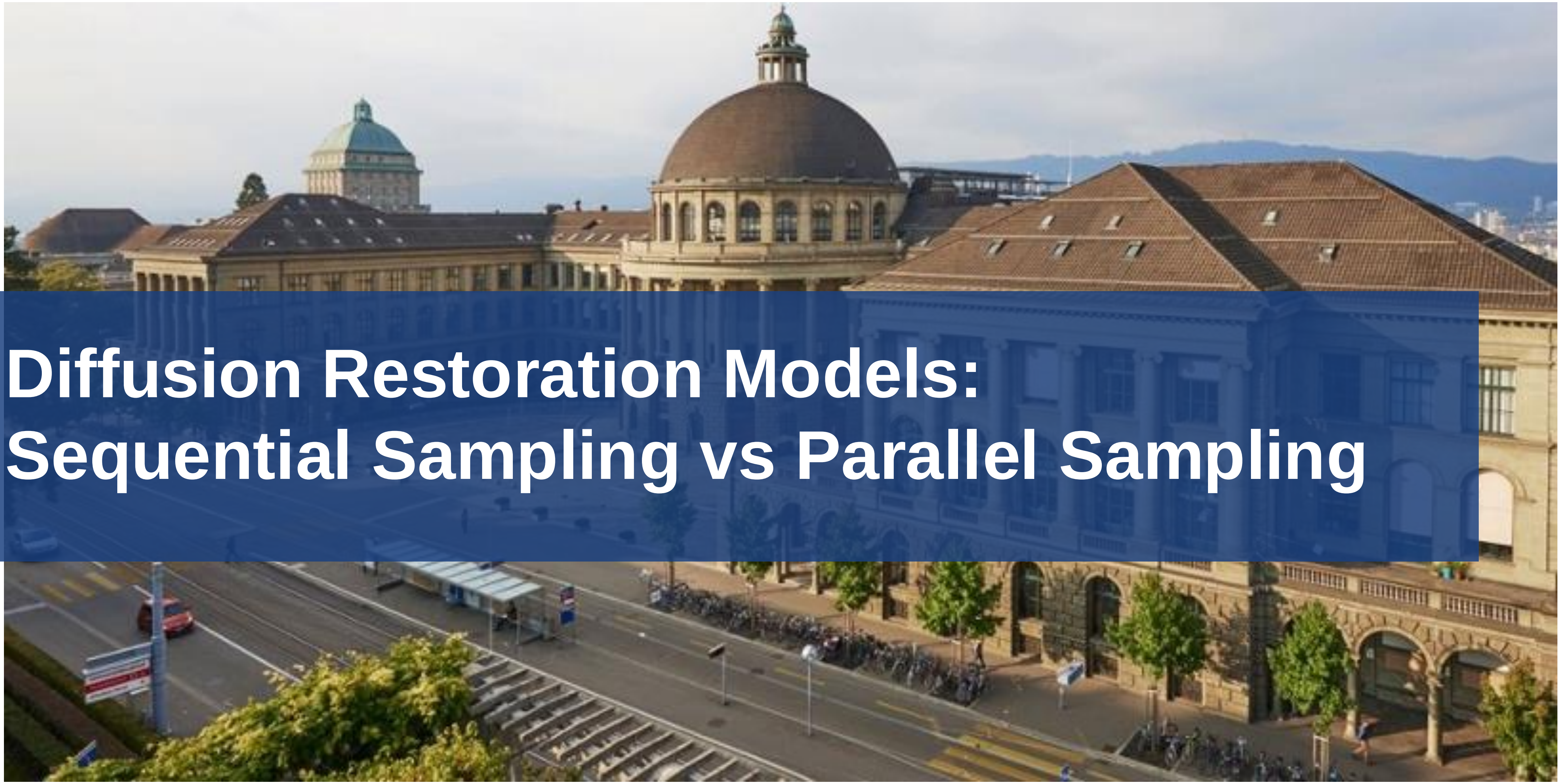
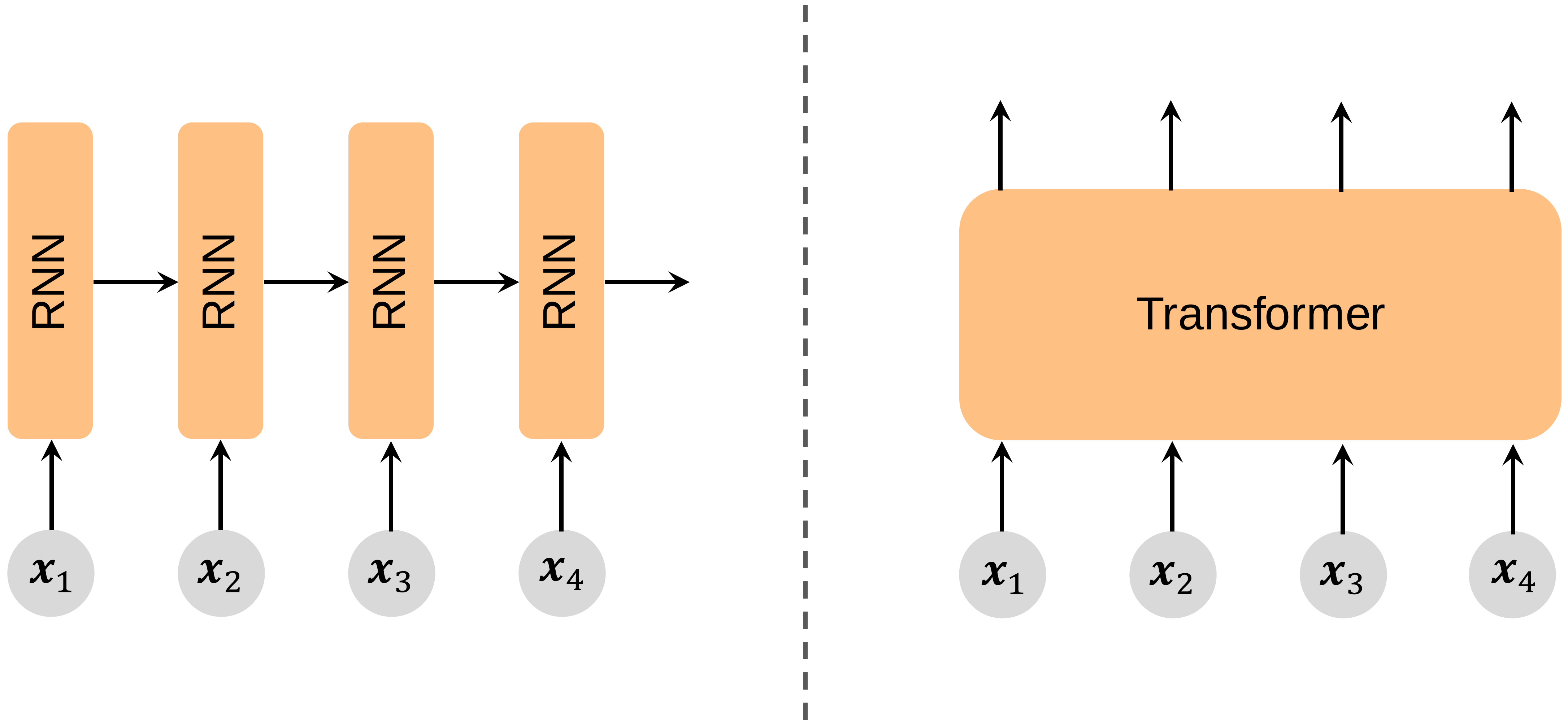
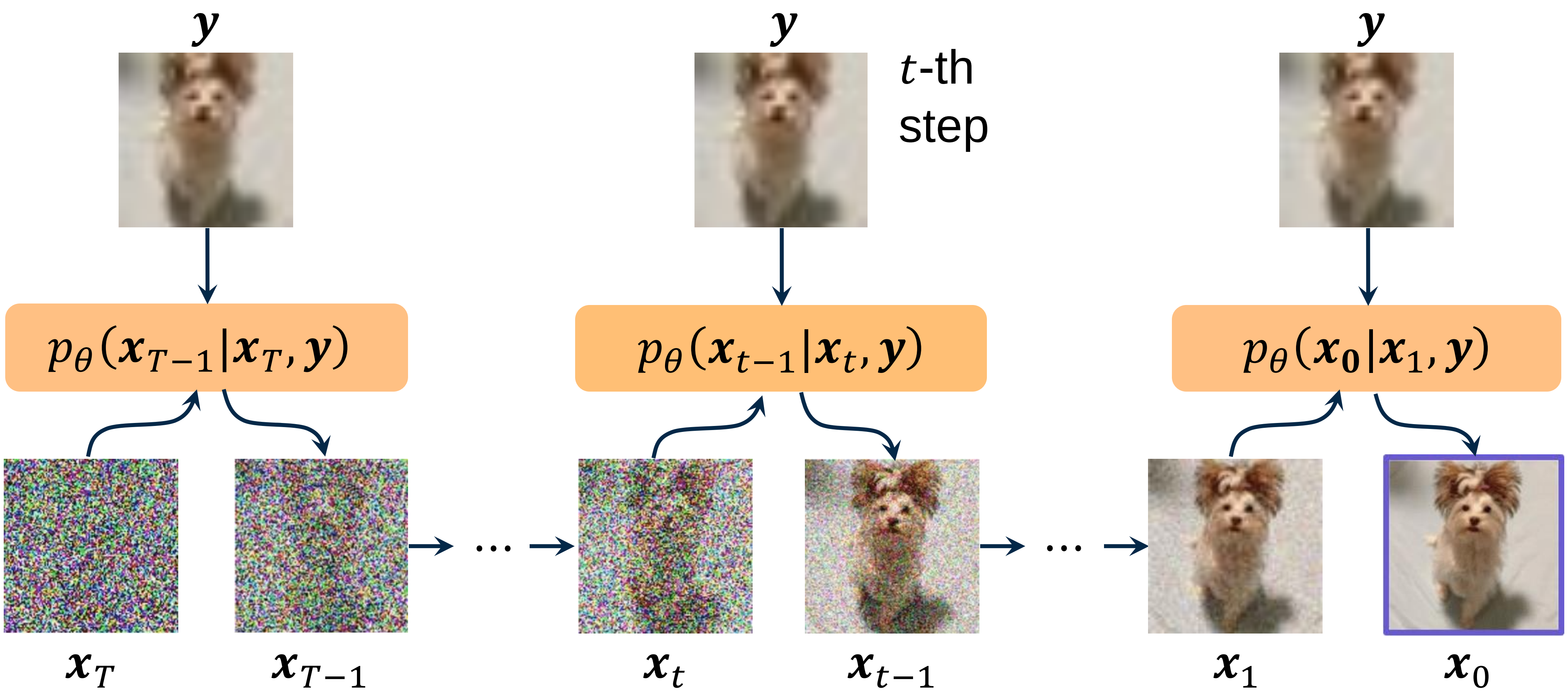
Diffusion Restoration Models:
Sequential Sampling vs Parallel Sampling



RNN vs Transformer

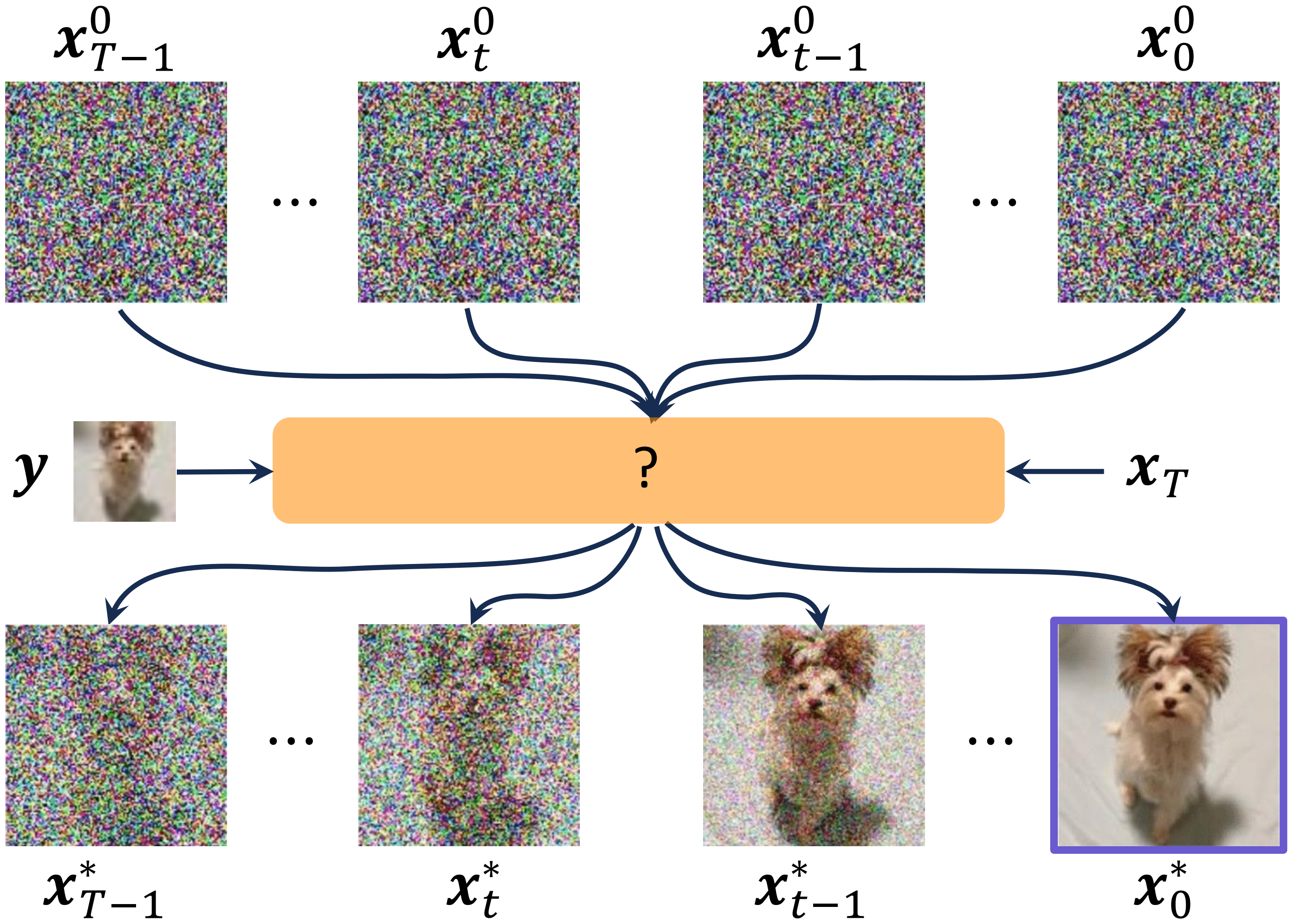


Most Existing Diffusion Model-based IR



- Long sequential sampling has expensive sampling time and high computation costs
- Hinder understanding the relationship between the restoration results and the inputs

Diffusion with Parallel Sampling



Background: Diffusion Restoration

Forward process:

$$q(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}\left(\boldsymbol{x}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, (1-\bar{\alpha}_t)\boldsymbol{I}\right),$$

Reverse process:

$$q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) = \mathcal{N}\left(\boldsymbol{x}_{t-1}; \tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0), \tilde{\sigma}_t^2 \boldsymbol{I}\right)$$

$$\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\boldsymbol{x}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{x}_t$$

$$\hat{\boldsymbol{x}}_{0|t} = \boldsymbol{A}^{\dagger}\boldsymbol{y} + (\boldsymbol{I} - \boldsymbol{A}^{\dagger}\boldsymbol{A})\boldsymbol{x}_{0|t}$$

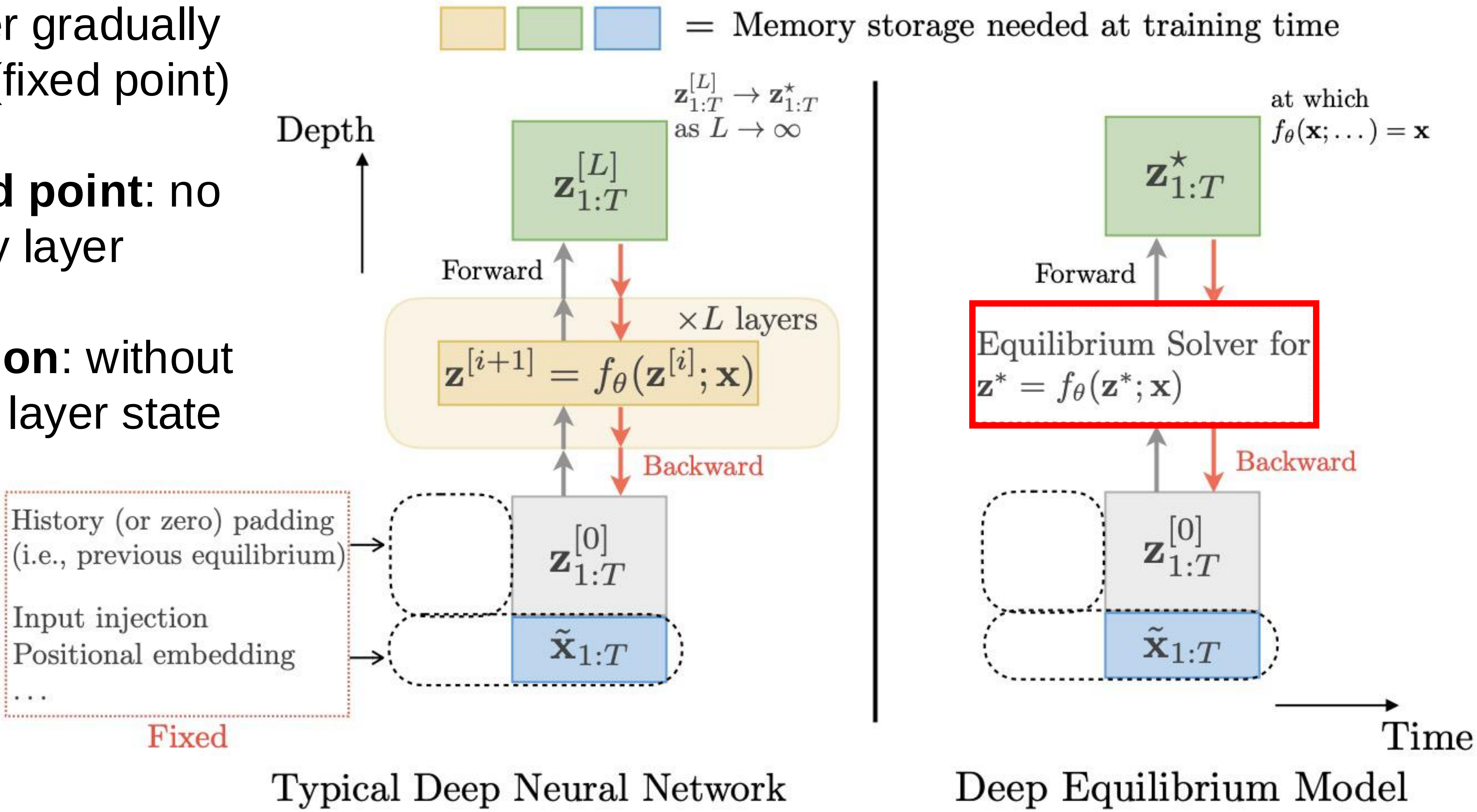
$$\boldsymbol{x}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\right)$$

Background: Deep Equilibrium Models

Assumption: each layer gradually tends to a stable value (fixed point)

Directly solve the fixed point: no longer calculate layer by layer

Implicit backpropagation: without storing the intermediate layer state



Bai, Shaojie, J. Zico Kolter, and Vladlen Koltun. "Deep equilibrium models." *NeurIPS* 2019

Deep Equilibrium Modeling

$$\boldsymbol{x}_{0:T-1} = F\left(\boldsymbol{x}_{0:T-1}; (\boldsymbol{x}_T, \boldsymbol{y})\right)$$

$$\Downarrow \qquad \Downarrow$$

$$\begin{bmatrix} \boldsymbol{x}_{T-1} \\ \boldsymbol{x}_{T-2} \\ \vdots \\ \boldsymbol{x}_{T-k} \\ \vdots \\ \boldsymbol{x}_0 \end{bmatrix} = \begin{bmatrix} f(\boldsymbol{x}_T; \boldsymbol{y}) \\ f(\boldsymbol{x}_{T-1:T}; \boldsymbol{y}) \\ \vdots \\ f(\boldsymbol{x}_{T-k+1:T}; \boldsymbol{y}) \\ \vdots \\ f(\boldsymbol{x}_{1:T}; \boldsymbol{y}) \end{bmatrix} \Rightarrow$$

Closed-form solution:

$$\boldsymbol{x}_{T-k} = \frac{\sqrt{\bar{\alpha}_{T-k}}}{\sqrt{\bar{\alpha}_T}}\left(\boldsymbol{I} - \boldsymbol{A}^\dagger\boldsymbol{A}\right)\boldsymbol{x}_T + \boldsymbol{A}^\dagger\boldsymbol{A}\boldsymbol{z}_{T-k+1} + \sum_{s=T-k}^{T-1} \frac{\sqrt{\bar{\alpha}_{T-k}}}{\sqrt{\bar{\alpha}_s}}\left(\boldsymbol{I} - \boldsymbol{A}^\dagger\boldsymbol{A}\right)\boldsymbol{z}_{s+1}$$

Anderson acceleration solver:

$$\boldsymbol{x}^*_{0:T-1} = \mathrm{RootSolve}\left(F\left(\boldsymbol{x}_{0:T-1}; (\boldsymbol{x}_T, \boldsymbol{y})\right) - \boldsymbol{x}_{0:T-1}\right)$$

Initialization Optimization via DEQ Inversion

DEQ Inversion:

Loss function: $\mathcal{L} = \ell\big(\phi(\boldsymbol{x}_0^*); \varphi(\boldsymbol{y})\big)$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}_T} = -\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}_{0:T}^*} \underbrace{\left(J_g^{-1} \Big|_{\boldsymbol{x}_{0:T}^*} \right)}_{\text{inverse Jacobian}} \frac{\partial F\big(\boldsymbol{x}_{0:T-1}^*; (\boldsymbol{x}_T, \boldsymbol{y})\big)}{\partial \boldsymbol{x}_T}$$

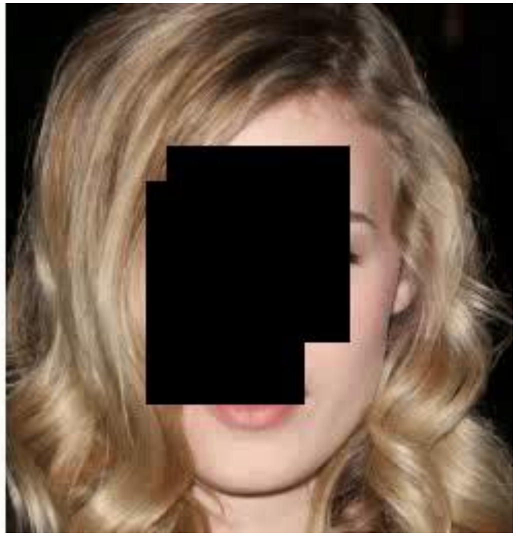
Gradient descent: $\boldsymbol{x}_T \leftarrow \boldsymbol{x}_T + \lambda \cdot \partial \mathcal{L} / \partial \boldsymbol{x}_T$

A New Zero-shot Diffusion Restoration

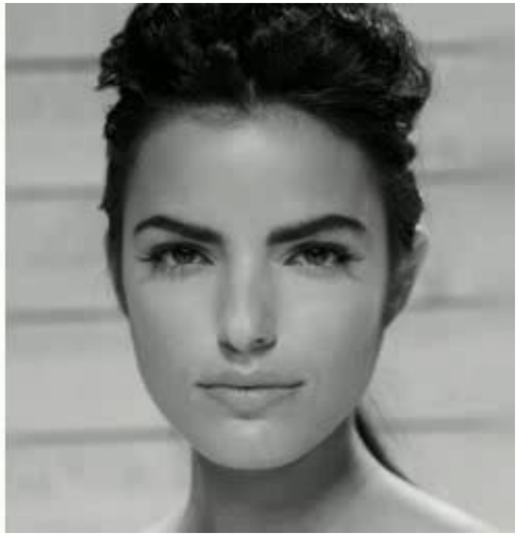
Super-Rresolution

Deblurring

Inpainting



Colorization



Quantitative Comparisons

Super-resolution

Datasets	Methods	2×SR PSNR↑	2×SR SSIM↑	2×SR LPIPS↓	4×SR PSNR↑	4×SR SSIM↑	4×SR LPIPS↓
ImageNet	Baseline	29.63	0.875	0.165	25.15	0.699	0.351
	DGP [55]	22.32	0.583	0.426	18.35	0.398	0.529
	DPS [20]	22.40	0.597	0.405	20.34	0.488	0.464
	ILVR [19]	23.36	0.613	0.334	22.76	0.583	0.383
	DiffPIR [87]	27.16	0.790	0.214	24.31	0.649	0.350
	DDRM [41]	31.43	0.906	0.117	26.21	0.745	0.288
	DDNM [68]	31.81	0.908	0.097	26.49	0.753	0.266
	DeqIR (Ours)	**32.35**	**0.913**	**0.082**	**27.47**	**0.781**	**0.230**
CelebA-HQ	Baseline	35.87	0.953	0.099	30.12	0.857	0.240
	DGP [55]	28.61	0.809	0.279	25.25	0.690	0.405
	DPS [20]	28.71	0.818	0.219	25.01	0.710	0.282
	ILVR [19]	27.31	0.783	0.234	27.09	0.775	0.245
	DiffPIR [87]	32.51	0.882	0.156	28.60	0.795	0.228
	DDRM [41]	**36.76**	0.953	0.074	31.91	0.880	0.149
	DDNM [68]	36.37	0.950	0.065	31.86	0.876	**0.136**
	DeqIR (Ours)	36.63	**0.954**	**0.062**	**32.22**	**0.889**	0.155

Deblurring

Deblur (Gaussian) PSNR↑	Deblur (Gaussian) SSIM↑	Deblur (Gaussian) LPIPS↓	Deblur (anisotropic) PSNR↑	Deblur (anisotropic) SSIM↑	Deblur (anisotropic) LPIPS↓	NFEs /Iters
18.22	0.529	0.433	20.86	0.544	0.480	-
21.81	0.522	0.472	20.77	0.459	0.504	1500
22.04	0.569	0.394	21.82	0.561	0.381	1000
-	-	-	-	-	-	100
25.32	0.673	0.296	23.37	0.535	0.439	100
40.70	0.978	0.040	37.69	0.964	0.057	20
43.83	**0.989**	**0.018**	38.40	0.970	0.038	100
43.42	0.987	0.021	**39.47**	**0.973**	**0.036**	15
18.94	0.704	0.337	23.16	0.727	0.354	-
27.02	0.738	0.372	25.73	0.663	0.426	1500
27.56	0.775	0.229	26.91	0.754	0.234	1000
-	-	-	-	-	-	100
30.63	0.835	0.197	29.32	0.802	0.232	100
43.06	0.983	0.036	41.27	0.976	0.053	20
46.99	0.991	0.021	43.43	0.983	0.037	100
47.18	**0.992**	**0.019**	**43.57**	**0.984**	**0.036**	15

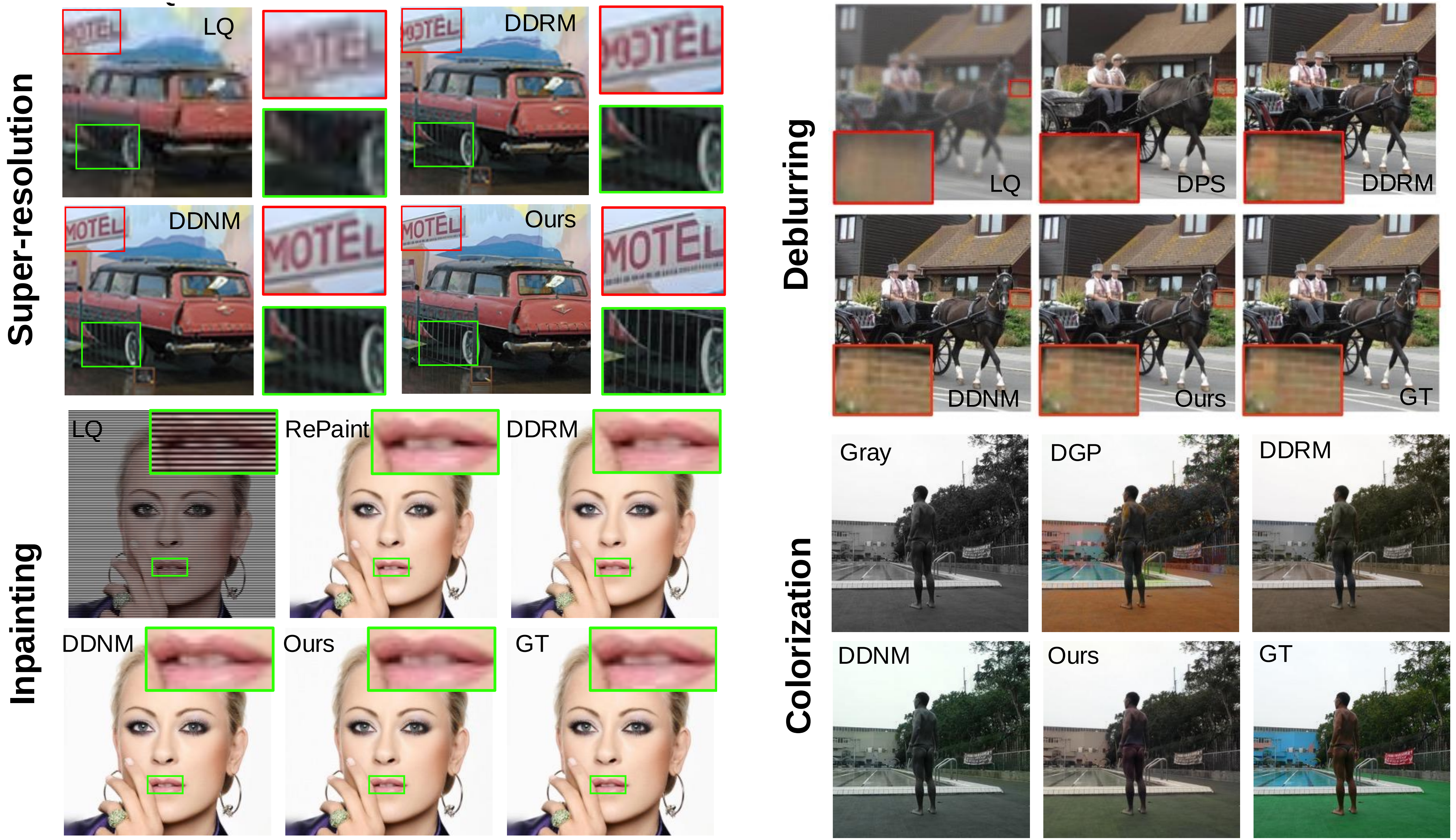
Inpainting

Methods	Text mask PSNR↑	Text mask SSIM↑	Text mask LPIPS↓	Stripe mask PSNR↑	Stripe mask SSIM↑	Stripe mask LPIPS↓
Baseline	14.55	0.642	0.515	9.02	0.131	0.730
Palette [59]	38.09	0.978	0.027	25.91	0.733	0.343
DDRM [41]	37.25	0.969	0.223	34.34	0.933	0.223
RePaint [53]	38.54	0.974	0.039	36.25	0.951	0.086
DDNM [68]	39.45	0.980	**0.023**	36.75	**0.957**	**0.076**
DeqIR (Ours)	**39.72**	**0.981**	0.026	**36.99**	0.948	0.091

Colorization

Methods	ImageNet Cons↓	ImageNet LPIPS↓	ImageNet FID↓	CelebA-HQ Cons↓	CelebA-HQ LPIPS↓	CelebA-HQ FID↓
Baseline	0	0.196	90.93	0	0.210	70.69
DGP [55]	-	0.256	99.86	-	0.218	73.24
DDRM [41]	265.08	0.223	79.42	472.25	0.245	57.29
DDNM [68]	45.07	0.186	77.21	51.43	0.139	45.73
DeqIR (Ours)	**43.15**	**0.171**	**70.94**	**50.16**	**0.092**	**43.98**

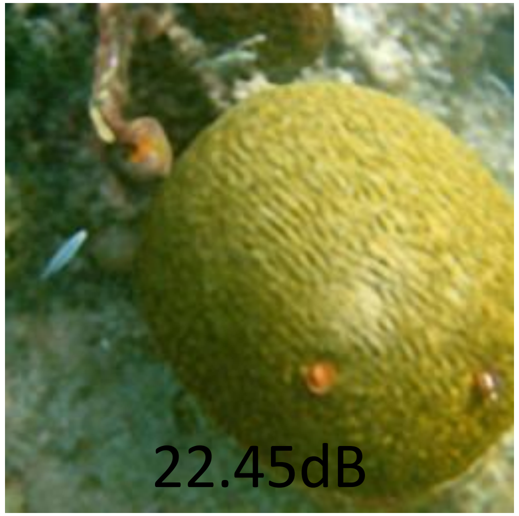
Qualitative Comparisons



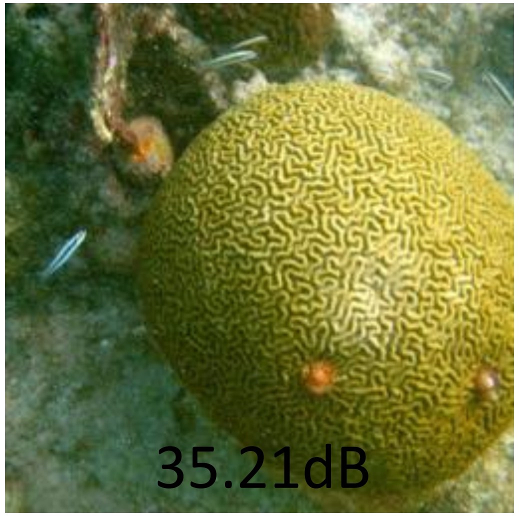
Interesting Applications of DEQ Inversion

LR input

w/o inversion



w/ inversion



GT

gray input

reference

w/o inversion

w/ inversion

PSNR Improvement with DEQ Inversion

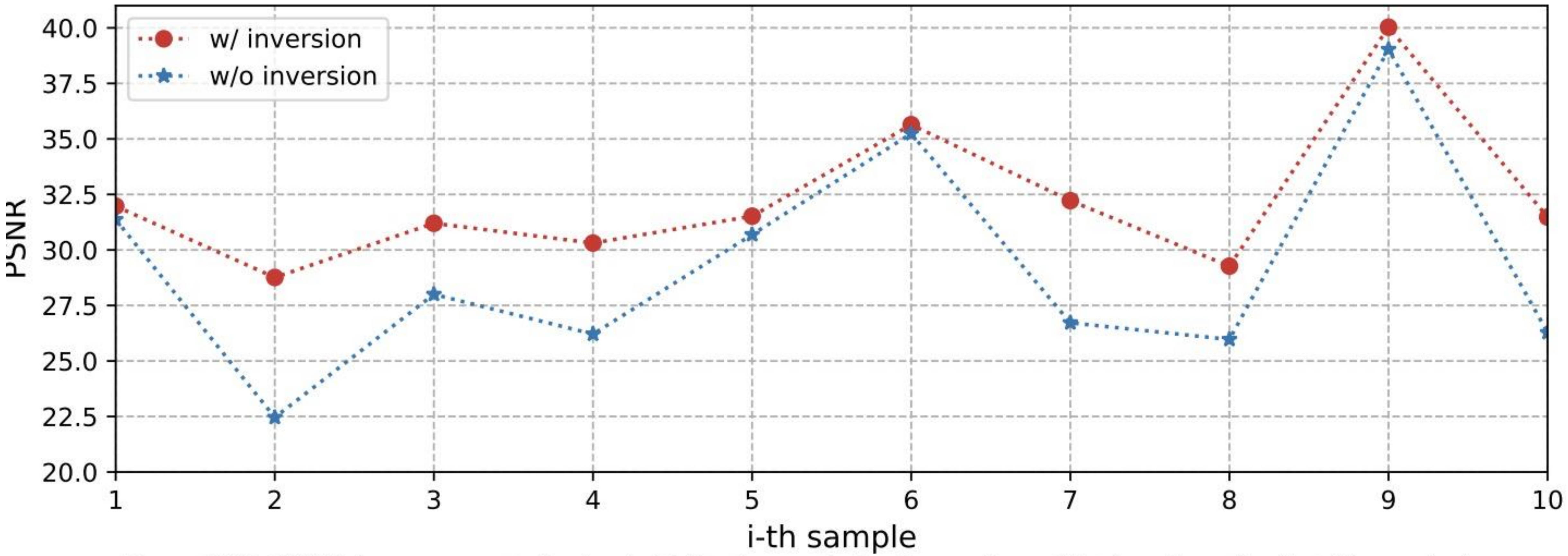
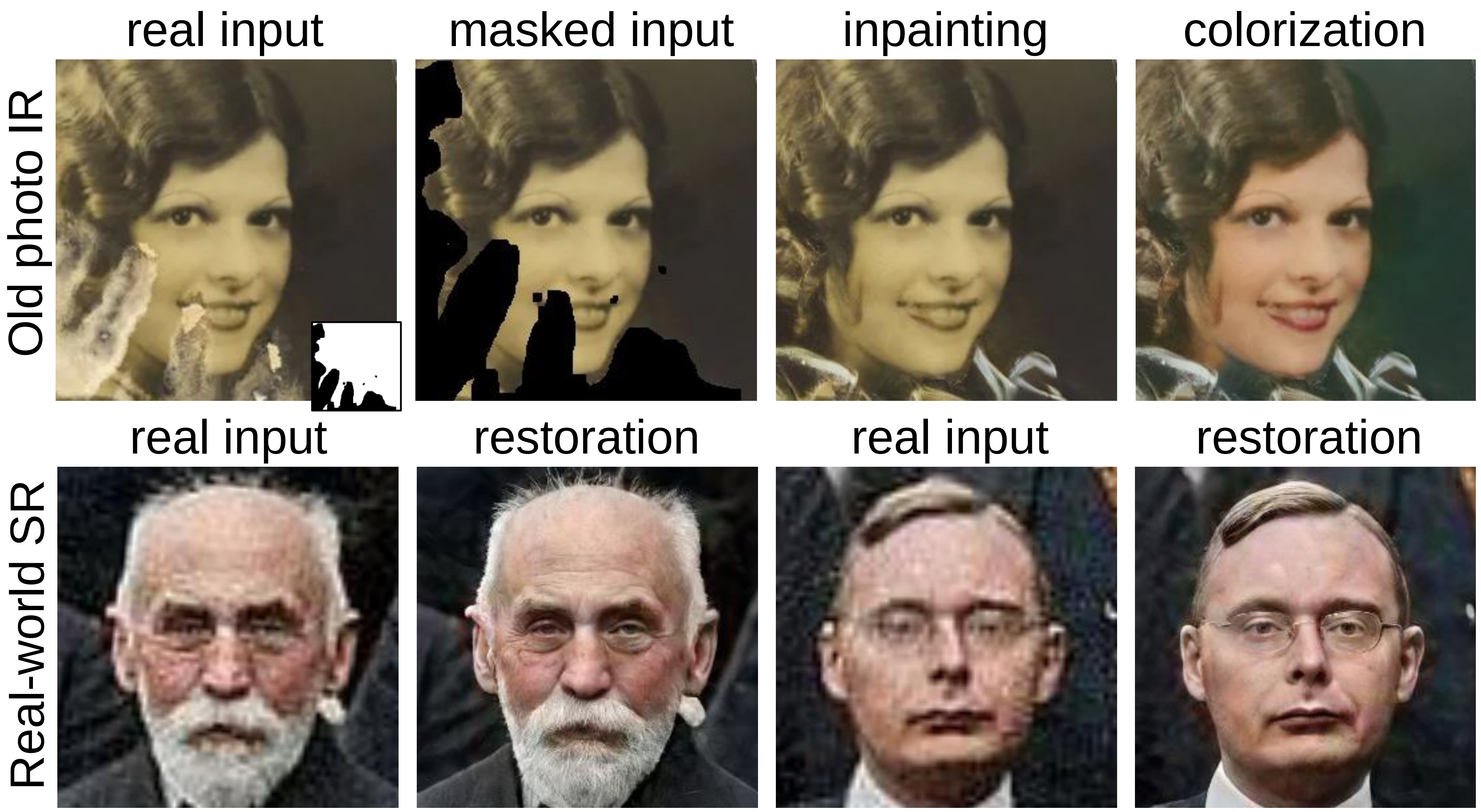


Figure D20. PSNR improvement of using initialization optimization on ImageNet (we show the first 10 samples).

Real-World Applications



Ablation Study

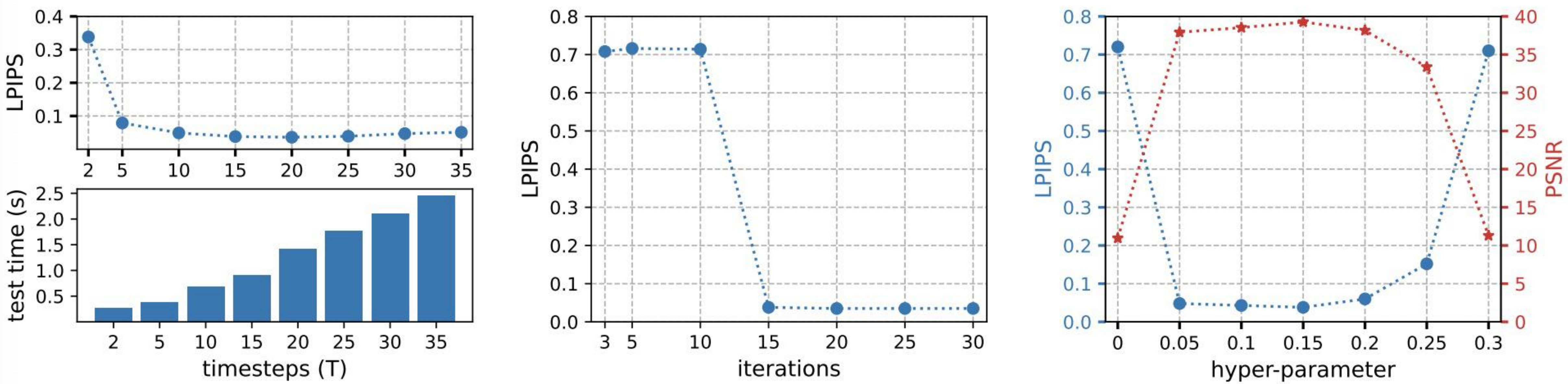
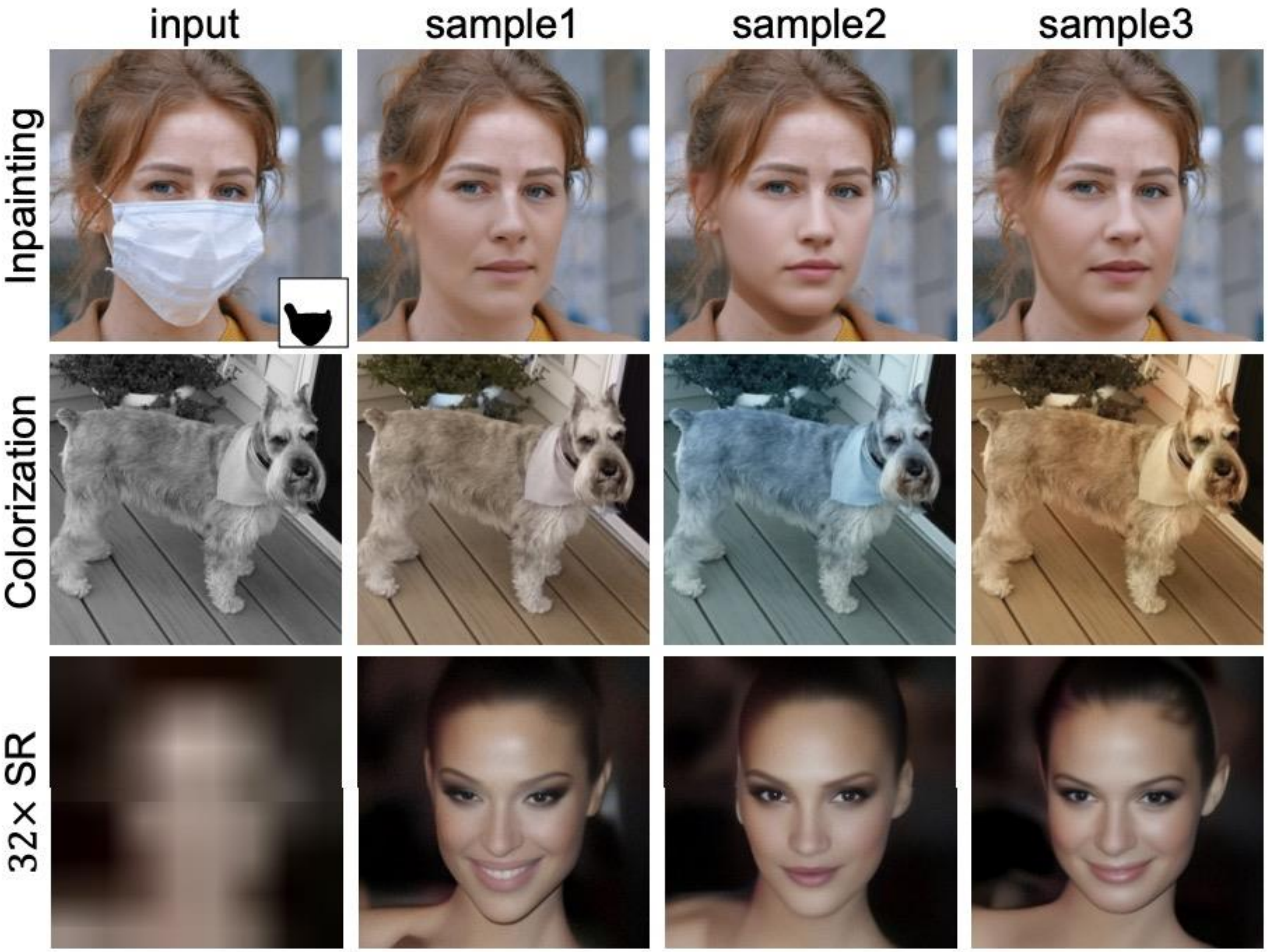


Figure 8. Ablation study of timesteps (left), iteration (middle) and hyper-parameters (right) for anisotropic deblurring on ImageNet.

Diversity of Restoration



Thanks for your attention!