NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

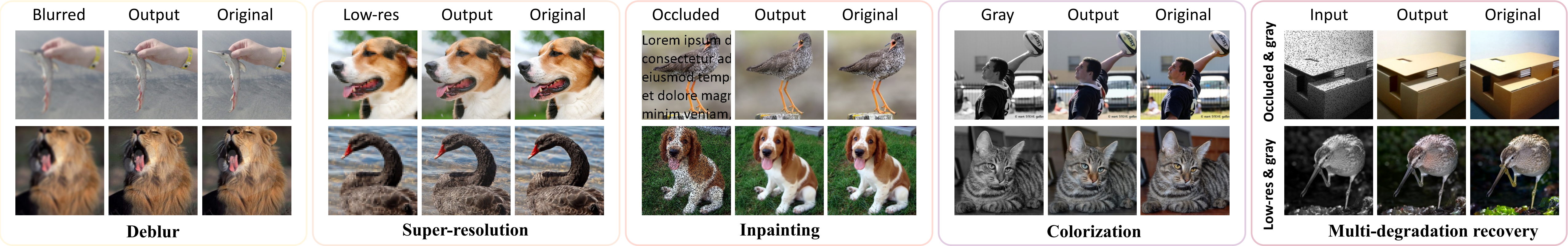
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

WED-AM-163

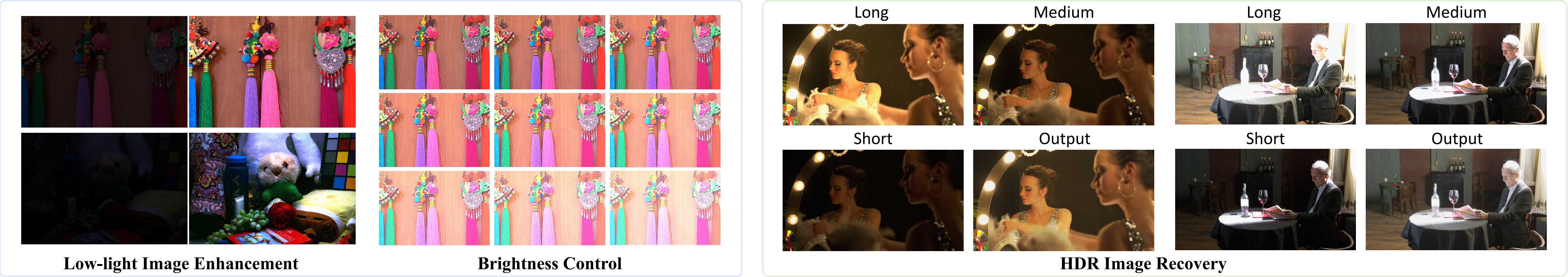
Generative Diffusion Prior for Unified Image Restoration and Enhancement

Ben Fei[1,2,*], Zhaoyang Lyu[2,*], Liang Pan[3], Junzhe Zhang[3],
Weidong Yang[1,†], Tianyue Luo[1], Bo Zhang[2], Bo Dai[2,†]
[1] Fudan University, [2]Shanghai AI Laboratory, [3]S-Lab, Nanyang Technological University

Overview

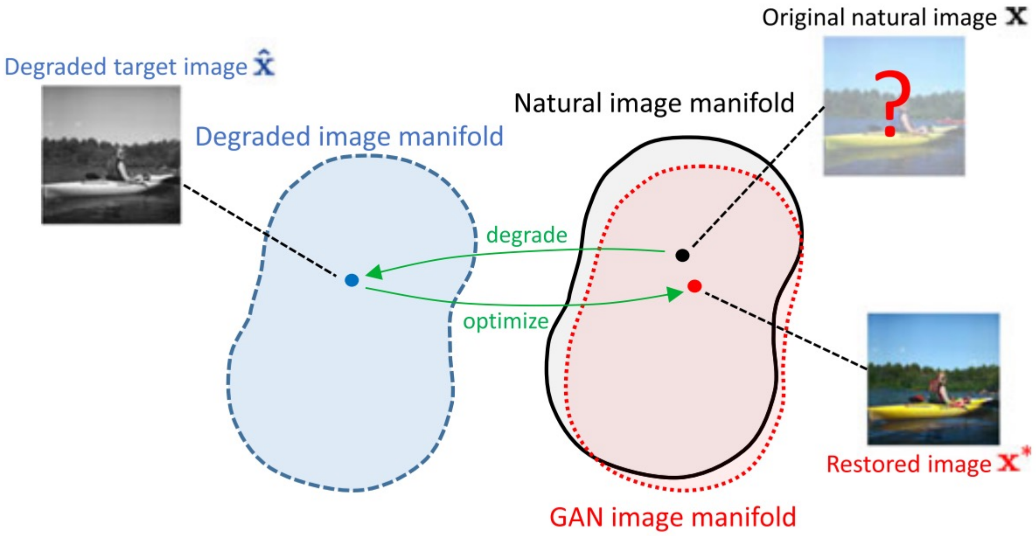


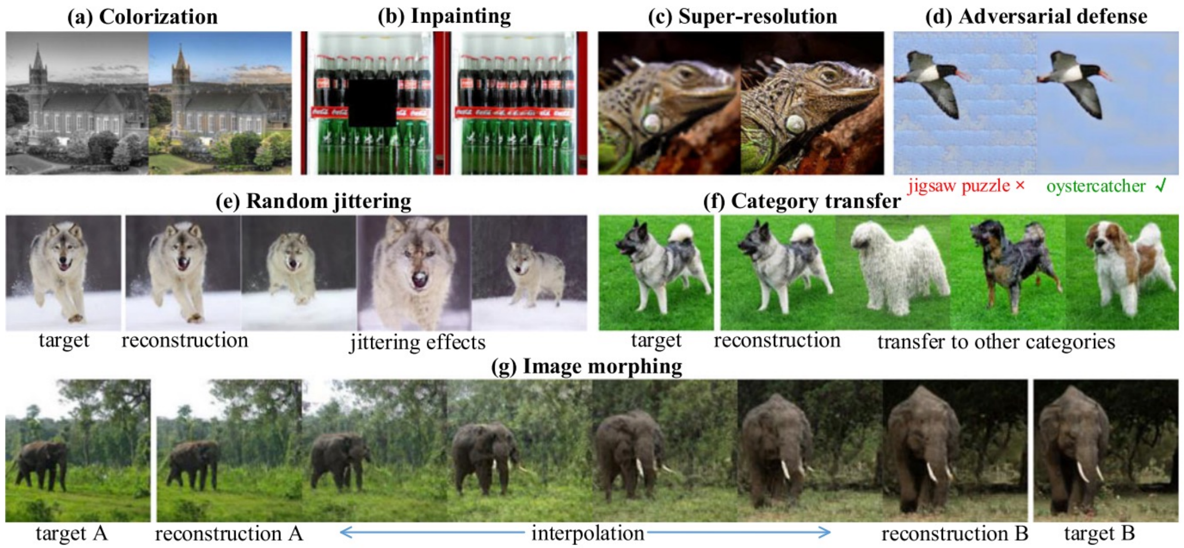
(a) Linear and Multi-Linear Image Restoration



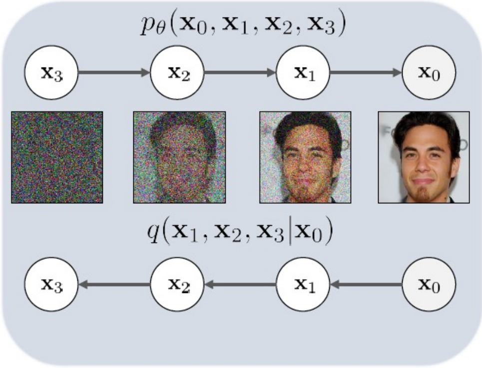
(b) Blind, Non-linear, Multiple-guidance or Any-size Image Restoration

Related work



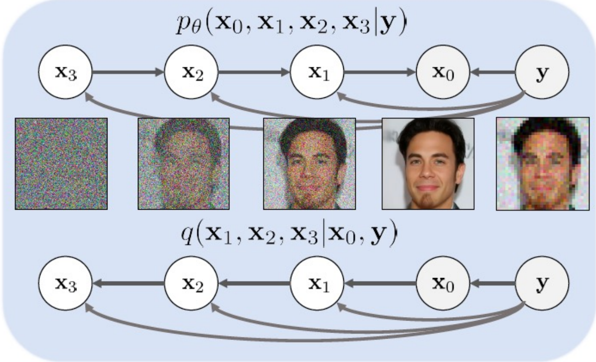


Deep Generative Prior (ECCV20, TPAMI21)

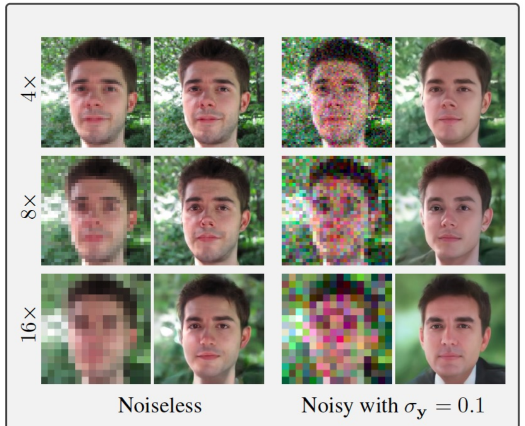


Denoising Diffusion Probabilistic Models
(Independent of inverse problem)

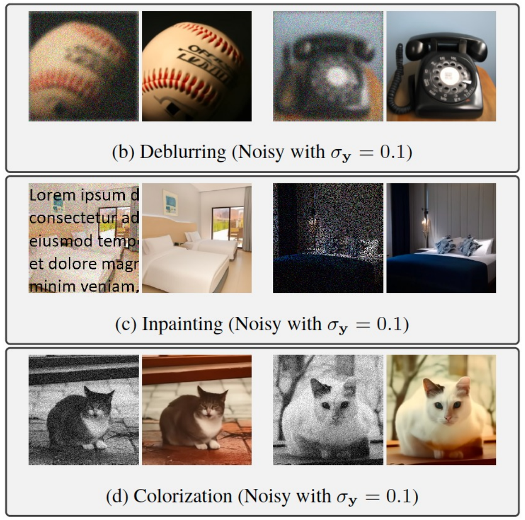
Use pre-trained models for linear inverse problems



Denoising Diffusion Restoration Models
(Dependent on inverse problem)



(a) Super-resolution



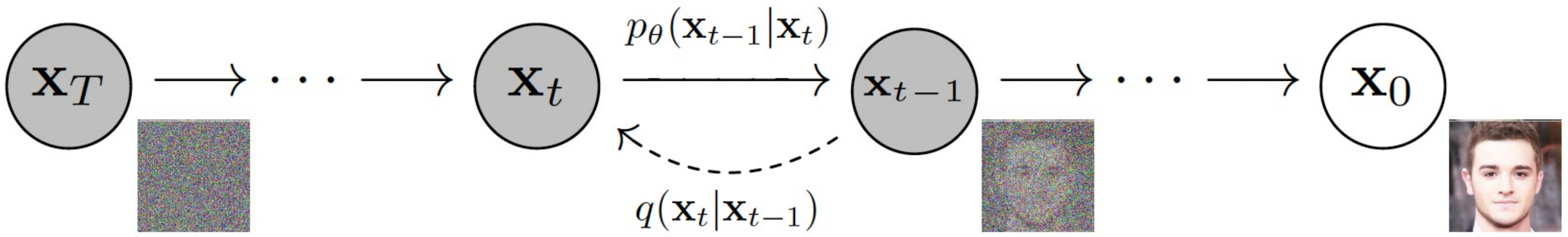
(b) Deblurring (Noisy with $\sigma_y = 0.1$)

(c) Inpainting (Noisy with $\sigma_y = 0.1$)

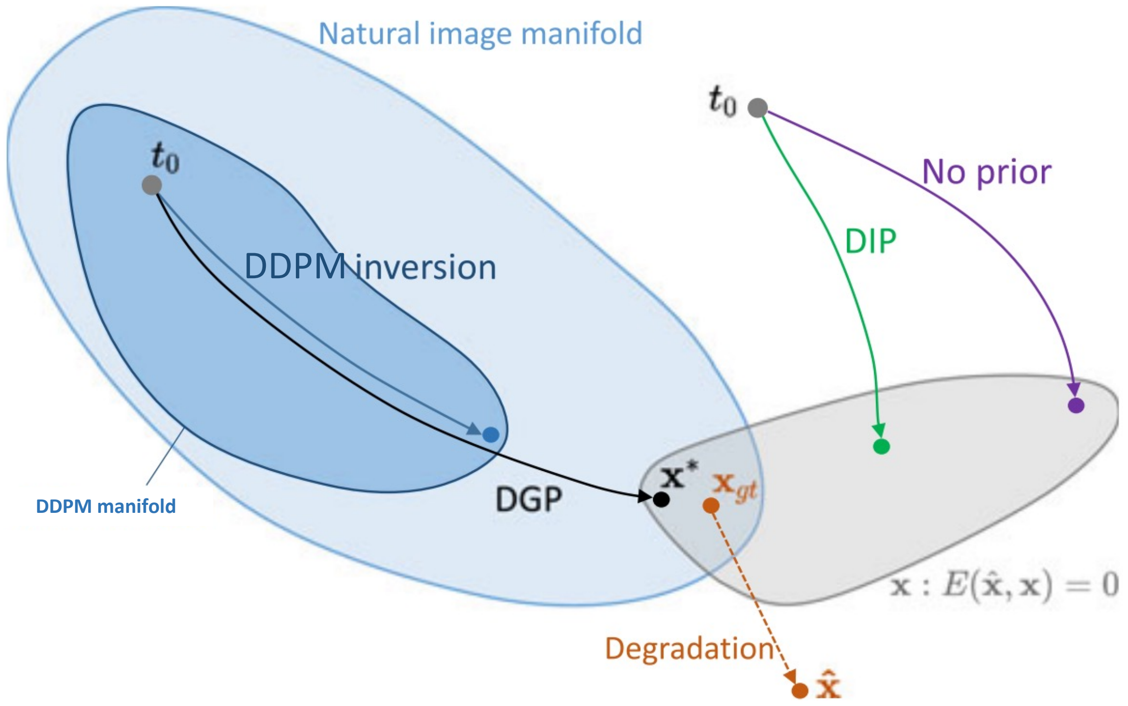
(d) Colorization (Noisy with $\sigma_y = 0.1$)

Denoising Diffusion Restoration Models (NIPS22)

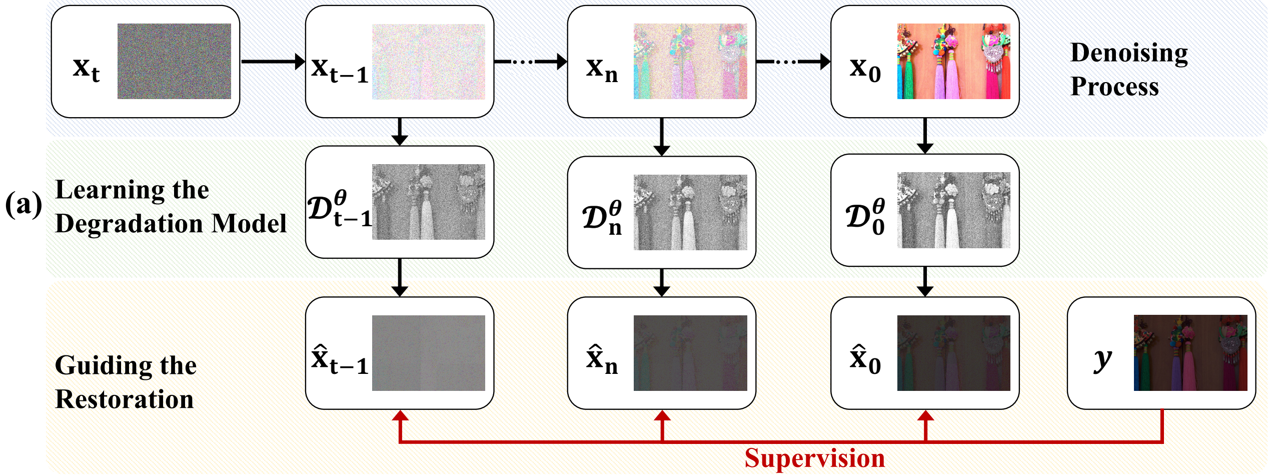
Motivation

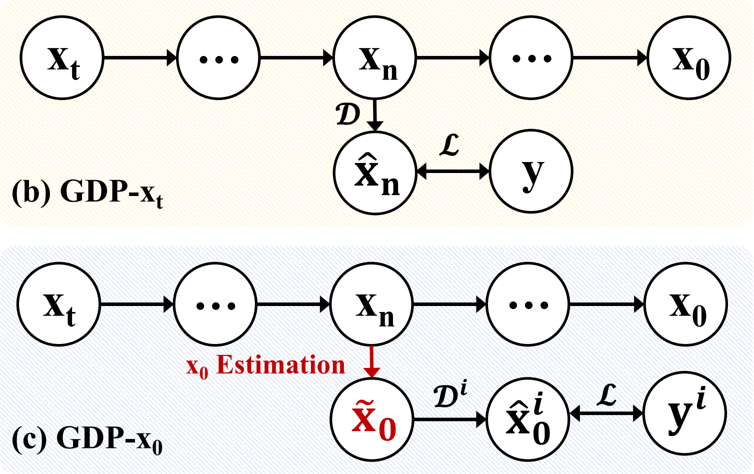


- GDP exploit pre-trained DDPMs with variational inference, and achieve satisfactory results on multiple restoration tasks
- The reconstructed image is consistent with the degraded images
- Better generalization ability
- Can tackle the multi-degradation problem and blind problem
- Achieve arbitrary size image generation



Generative Diffusion Prior





Denoising Process

$\boldsymbol{y}$: Guidance image

$\boldsymbol{y}^{\boldsymbol{i}}$: i^{th} Guidance image

$\hat{\boldsymbol{x}}_{\boldsymbol{n}}$: n^{th} noisy image + degradation

$\hat{\boldsymbol{x}}^{\boldsymbol{i}}_{\boldsymbol{0}}$: $\tilde{\mathrm{x}}_0$ + i^{th} degradation

$\mathcal{D}$: Degradation model

$\mathcal{D}^{\boldsymbol{i}}$: i^{th} Degradation model

$\mathcal{L}$: Loss function

$$-\left(s\Sigma\nabla_{\boldsymbol{x}_t}\mathcal{L}\left(\boldsymbol{x}_t,\boldsymbol{y}\right) + \lambda\Sigma\nabla_{\boldsymbol{x}_t}\mathcal{Q}\left(\boldsymbol{x}_t\right)\right)$$

Algorithm 1: GDP-$\boldsymbol{x}_t$ with fixed degradation model: Conditioner guided diffusion sampling on $\boldsymbol{x}_t$, given a diffusion model $\left(\mu_\theta\left(\boldsymbol{x}_t\right),\Sigma_\theta\left(\boldsymbol{x}_t\right)\right)$, corrupted image conditioner $\boldsymbol{y}$.

Input: Corrupted image $\boldsymbol{y}$, gradient scale s, degradation model $\mathcal{D}$, distance measure $\mathcal{L}$, optional quality enhancement loss $\mathcal{Q}$, quality enhancement scale λ.
Output: Output image $\boldsymbol{x}_0$ conditioned on $\boldsymbol{y}$
Sample $\boldsymbol{x}_T$ from $\mathcal{N}(0,\mathbf{I})$
for *t from T to 1* **do**
- $\mu,\Sigma=\mu_\theta\left(\boldsymbol{x}_t\right),\Sigma_\theta\left(\boldsymbol{x}_t\right)$
- $\mathcal{L}^{total}_{\boldsymbol{x}_t}=\mathcal{L}(\boldsymbol{y},\mathcal{D}\left(\boldsymbol{x}_t\right))+\mathcal{Q}\left(\boldsymbol{x}_t\right)$
- Sample $\boldsymbol{x}_{t-1}$ by $\mathcal{N}\left(\mu+s\nabla_{\boldsymbol{x}_t}\mathcal{L}^{total}_{\boldsymbol{x}_t},\Sigma\right)$

end
return $\boldsymbol{x}_0$

Algorithm 2: GDP-$\boldsymbol{x}_0$: Conditioner guided diffusion sampling on $\tilde{\boldsymbol{x}}_0$, given a diffusion model $\left(\mu_\theta\left(\boldsymbol{x}_t\right),\Sigma_\theta\left(\boldsymbol{x}_t\right)\right)$, corrupted image conditioner $\boldsymbol{y}$.

Input: Corrupted image $\boldsymbol{y}$, gradient scale s, degradation model $\mathcal{D}_\phi$ with randomly initiated parameters ϕ, learning rate l for optimizable degradation model, distance measure $\mathcal{L}$, optional quality enhancement loss $\mathcal{Q}$, quality enhancement scale λ.
Output: Output image $\boldsymbol{x}_0$ conditioned on $\boldsymbol{y}$
Sample $\boldsymbol{x}_T$ from $\mathcal{N}(0,\mathbf{I})$
for *t from T to 1* **do**
- $\mu,\Sigma=\mu_\theta\left(\boldsymbol{x}_t\right),\Sigma_\theta\left(\boldsymbol{x}_t\right)$
- $\tilde{\boldsymbol{x}}_0=\frac{\boldsymbol{x}_t}{\sqrt{\bar{\alpha}_t}}-\frac{\sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\boldsymbol{x}_t,t)}{\sqrt{\bar{\alpha}_t}}$
- $\mathcal{L}^{total}_{\phi,\tilde{\boldsymbol{x}}_0}=\mathcal{L}(\boldsymbol{y},\mathcal{D}_\phi\left(\tilde{\boldsymbol{x}}_0\right))+\mathcal{Q}\left(\tilde{\boldsymbol{x}}_0\right)$
- $\phi\leftarrow\phi-l\nabla_\phi\mathcal{L}^{total}_{\phi,\tilde{\boldsymbol{x}}_0}$
- Sample $\boldsymbol{x}_{t-1}$ by $\mathcal{N}\left(\mu+s\nabla_{\tilde{\boldsymbol{x}}_0}\mathcal{L}^{total}_{\phi,\tilde{\boldsymbol{x}}_0},\Sigma\right)$

end
return $\boldsymbol{x}_0$

4x Super-resolution

Low-res GDP-x_t GDP-x_0 Original

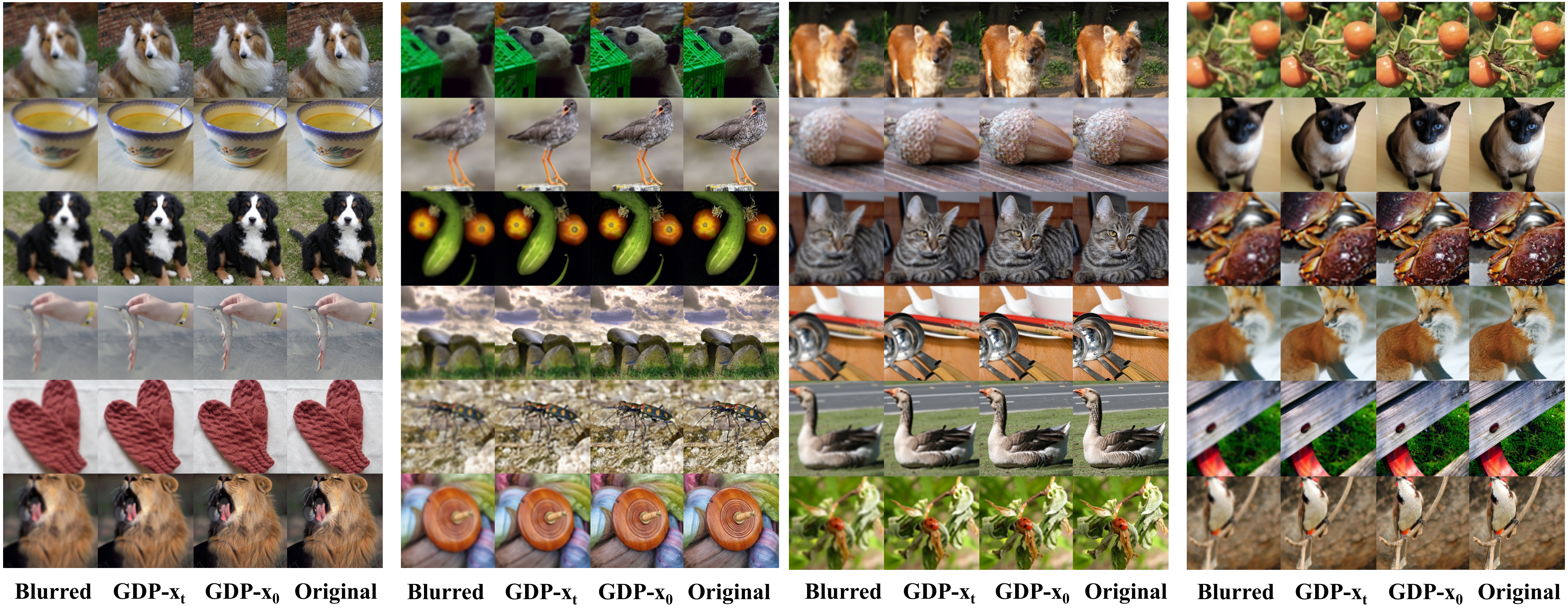
Low-res GDP-x_t GDP-x_0 Original

Low-res GDP-x_t GDP-x_0 Original

Low-res GDP-x_t GDP-x_0 Original

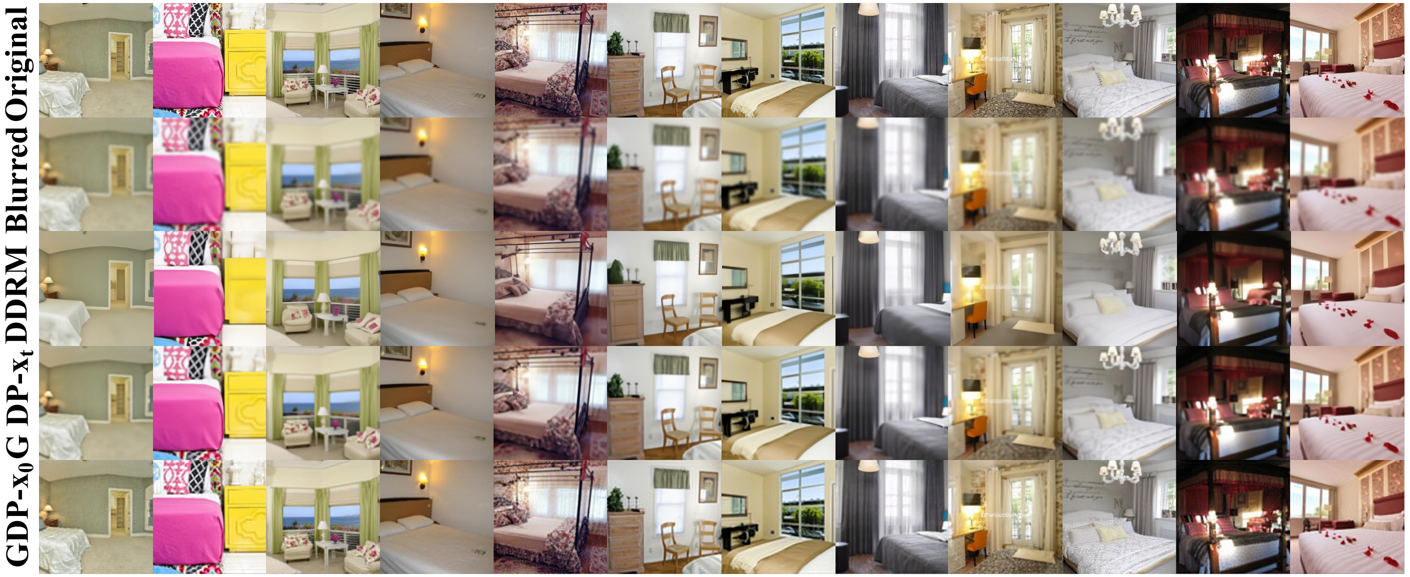
4x Super-resolution

Deblurring


Blurred GDP-x_t GDP-x_0 Original
Blurred GDP-x_t GDP-x_0 Original
Blurred GDP-x_t GDP-x_0 Original
Blurred GDP-x_t GDP-x_0 Original

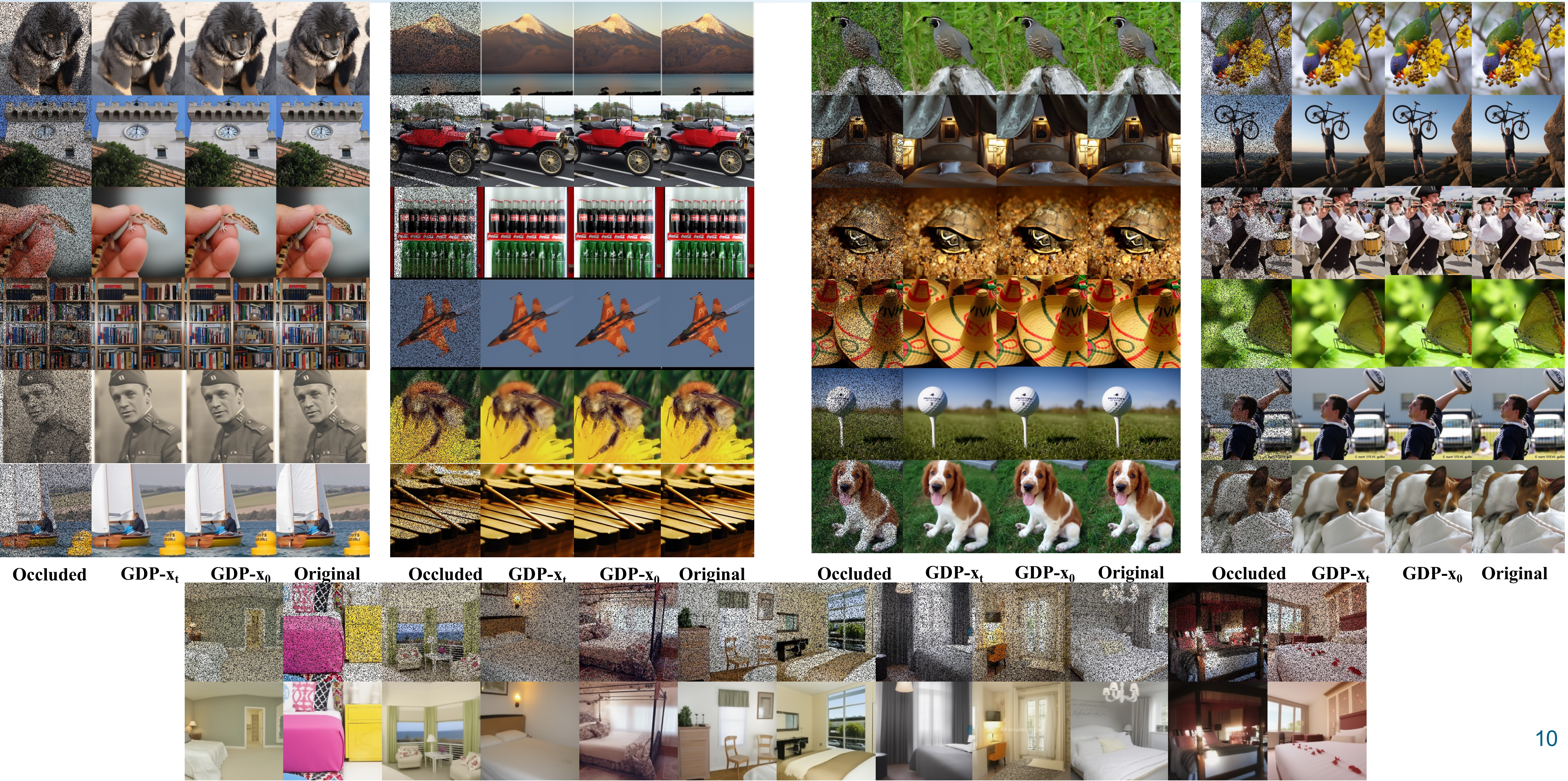
Deblurring


Original
Blurred
DDRM
GDP-x_t
GDP-x_0

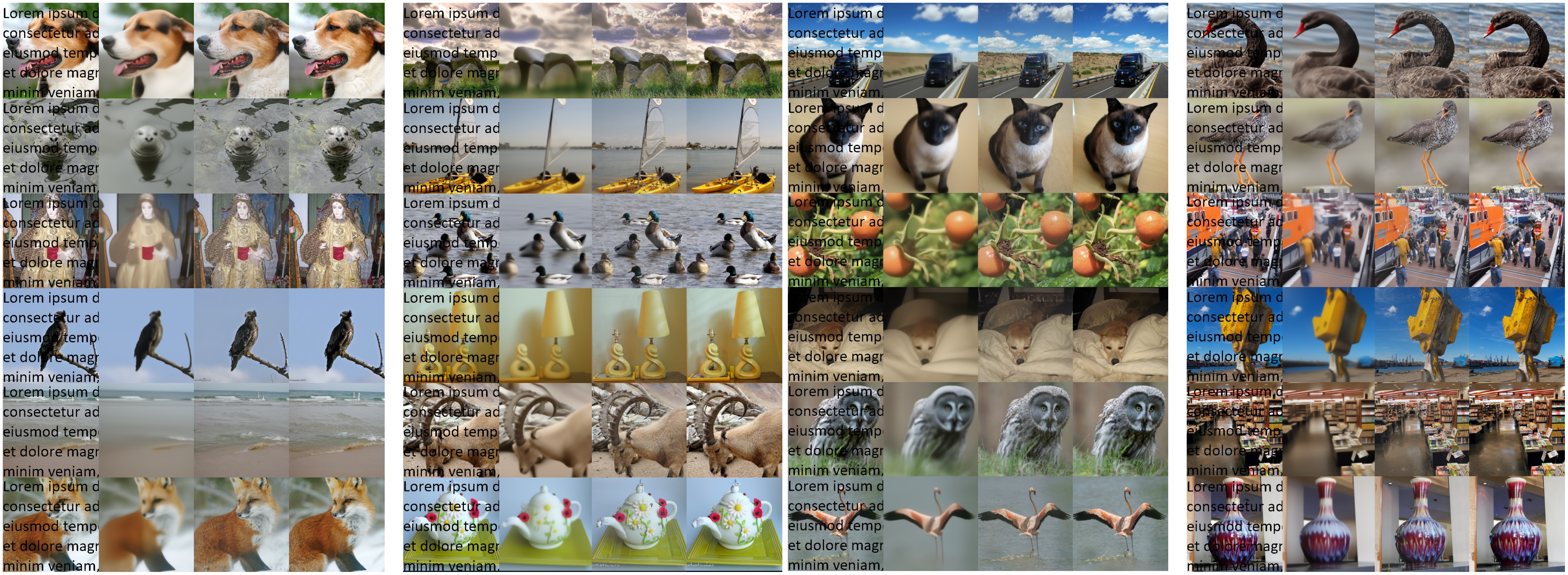

Original
Blurred
DDRM
GDP-x_t
GDP-x_0


Original
Blurred
GDP-x_t
GDP-x_0

25% Inpainting

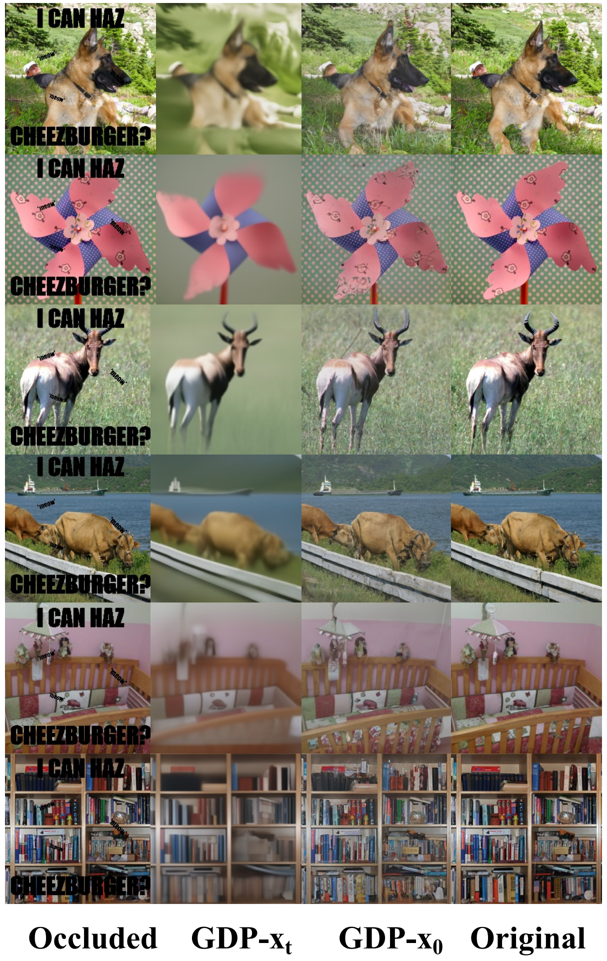


Inpainting-lorem



Occluded GDP-x_t GDP-x_0 Original | Occluded GDP-x_t GDP-x_0 Original | Occluded GDP-x_t GDP-x_0 Original | Occluded GDP-x_t GDP-x_0 Original

Inpainting-lolcat



Colorization

$$L_{\mathrm{col}} = \sum_{\forall (m,n) \in \varepsilon} (Y^m - Y^n)^2, \varepsilon = \{(R,G),(R,B),(G,B)\}$$

Gray Images **GDP-x_t** **GDP-x_0** **DGP** **DDRM**

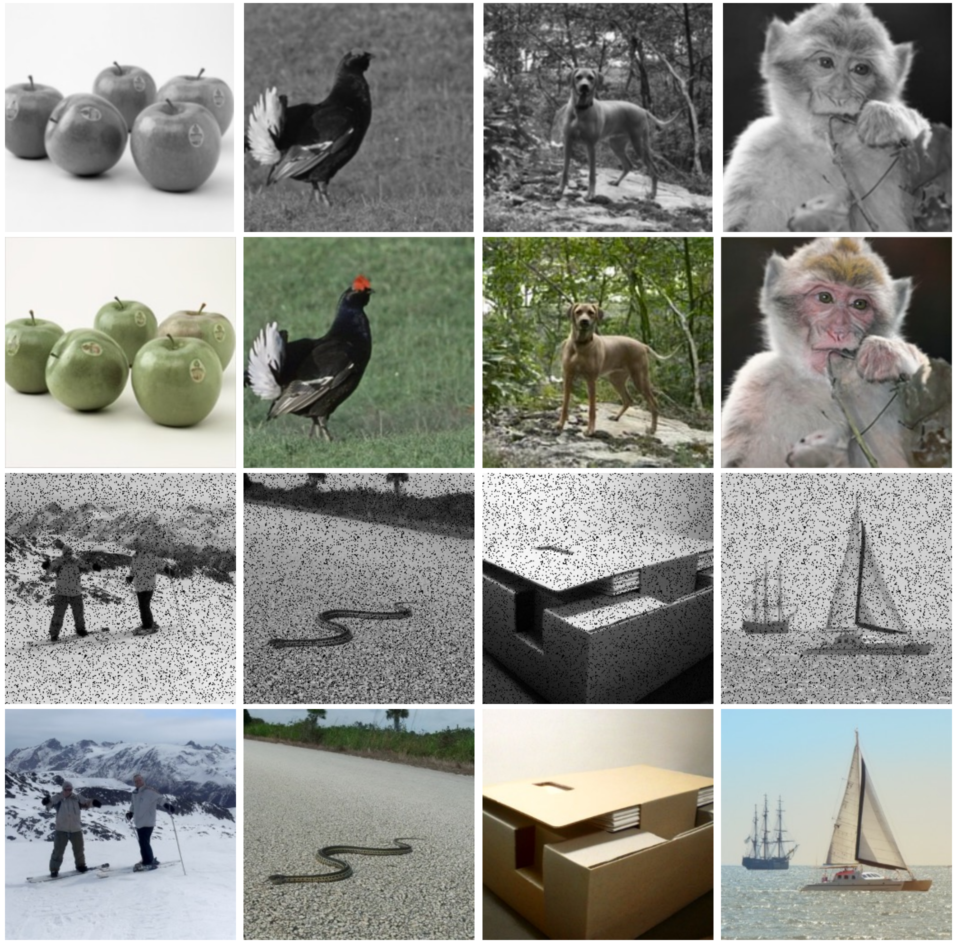
Multi-linear Degradation

Gray + Blur (3)

Output

Gray +
10 % inpainting

Output



Gray +
2x Super resolution

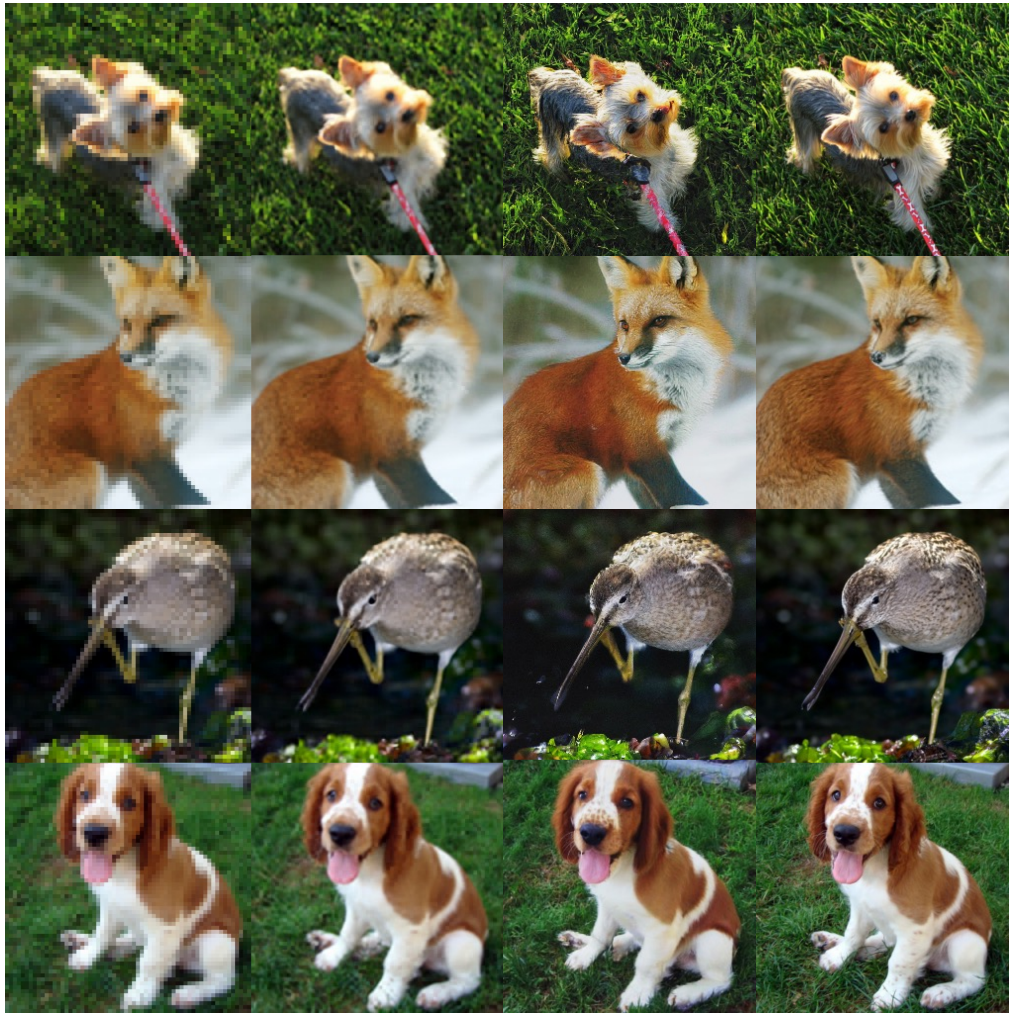
Output

Quantitative comparison

Method	4× Super-resolution				Deblur				25% Impainting				Colorization			
	PSNR ↑	SSIM ↑	Consistency ↓	FID ↓	PSNR ↑	SSIM ↑	Consistency↓	FID ↓	PSNR ↑	SSIM ↑	Consistency↓	FID ↓	PSNR ↑	SSIM ↑	Consistency ↓	FID ↓
DGP [57]	21.65	0.56	158.74	152.85	26.00	0.54	475.10	136.53	27.59	0.82	414.60	60.65	18.42	0.71	305.59	94.59
SNIPS [29]	22.38	0.66	21.38	154.43	24.73	0.69	60.11	17.11	17.55	0.74	587.90	103.50	-	-	-	-
RED [63]	24.18	0.71	27.57	98.30	21.30	0.58	63.20	69.55	-	-	-	-	-	-	-	-
DDRM [28]	**26.53**	**0.78**	19.39	40.75	**35.64**	**0.98**	50.24	4.78	34.28	0.95	**4.08**	24.09	**22.12**	0.91	37.33	47.05
GDP-x_t	24.27	0.67	80.32	64.67	25.86	0.75	54.08	5.00	31.06	0.93	8.80	20.24	21.30	0.86	75.24	66.43
GDP-x_0	24.42	0.68	**6.49**	**38.24**	25.98	0.75	**41.27**	**2.44**	**34.40**	**0.96**	5.29	**16.58**	21.41	**0.92**	**36.92**	**37.60**

- GDP-x_0 outperforms all baseline methods in Consistency and FID.
- Conventional automated evaluation measures (PSNR and SSIM) do not correlate well with human perception when the input resolution is low, and the magnification is large.

Accelerated by DDIM – 4x SR



Low-res | DDRM (20) | GDP-x_0 -DDIM (20) | Original

Low-res | DDRM (20) | GDP-x_0 -DDIM (20) | Original

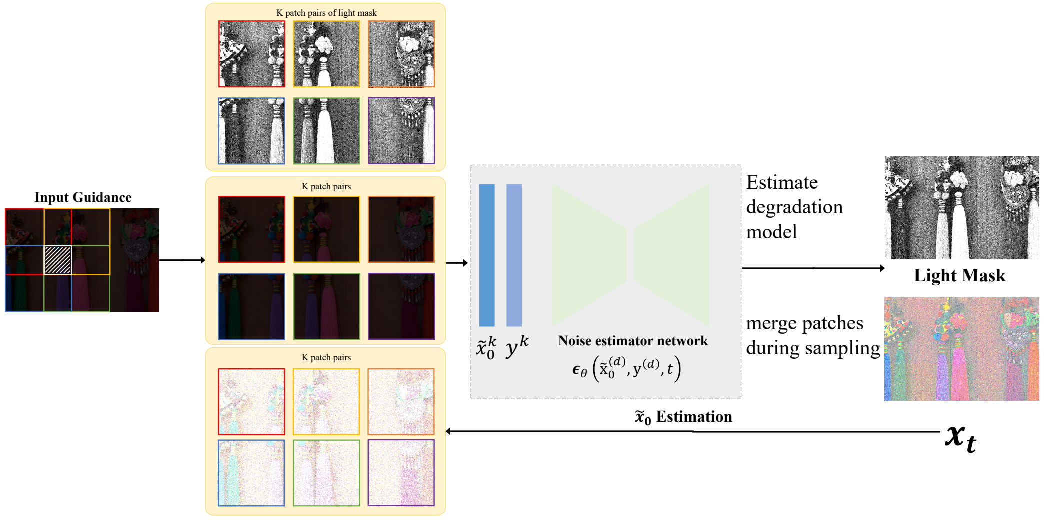
Accelerated by DDIM – Deblur

Blurred **DDRM (20)** **GDP-x_0 -DDIM (20)** **Original**

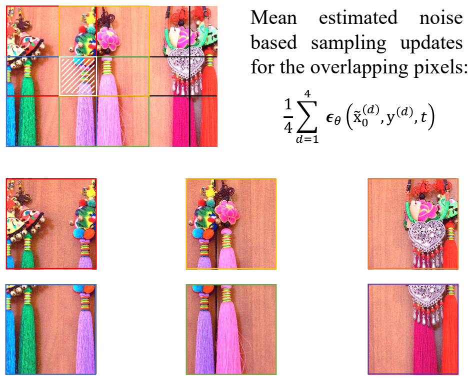
Blurred **DDRM (20)** **GDP-x_0 -DDIM (20)** **Original**

Task	4× super resolution				Deblur				25% Impainting				Colorization			
	PSNR	SSIM	Consistency	FID	PSNR	SSIM	Consistency	FID	PSNR	SSIM	Consistency	FID	PSNR	SSIM	Consistency	FID
DDRM(20) [6]	**26.53**	**0.784**	19.39	40.75	**35.64**	**0.978**	50.24	4.78	**34.28**	**0.958**	**4.08**	24.09	**22.12**	**0.924**	38.66	47.05
GDP-x_0-DDIM(20)	23.77	0.623	**9.24**	**39.46**	24.87	0.683	**44.39**	**3.66**	30.82	0.892	7.10	**19.70**	21.13	0.840	**37.33**	**41.38**

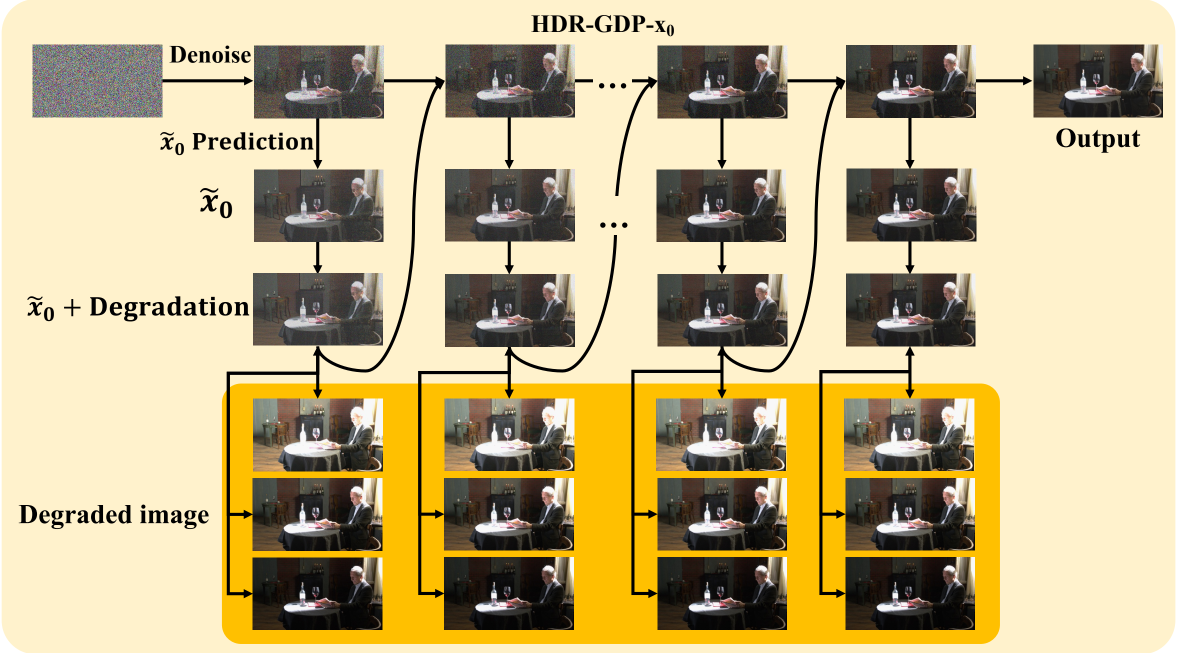
Non-linear and blind image restoration



(a) Patch-based diffusive image restoration



(b) Illustrating sampling for overlapping patches



$$\boldsymbol{y} = f\boldsymbol{x} + \boldsymbol{\mathcal{M}},$$

Algorithm 6: Restore Any-size Image

Input: Conditioner guided diffusion sampling on $\tilde{\boldsymbol{x}}_0$, given a diffusion model $(\mu_\theta(\boldsymbol{x}_t), \Sigma_\theta(\boldsymbol{x}_t))$, corrupted image conditioner $\boldsymbol{y}$, degradation model $\mathcal{D}_\phi : \boldsymbol{y} = f\boldsymbol{x} + \boldsymbol{\mathcal{M}}$ with randomly initiated parameters ϕ, learning rate l for optimizable degradation model. Dictionary of K overlapping patch locations, and a binary patch mask $\mathbf{P}^k$.

Output: Output image $\boldsymbol{x}_0$ conditioned on $\boldsymbol{y}$

Sample $\boldsymbol{x}_T$ from $\mathcal{N}(0, \mathbf{I})$

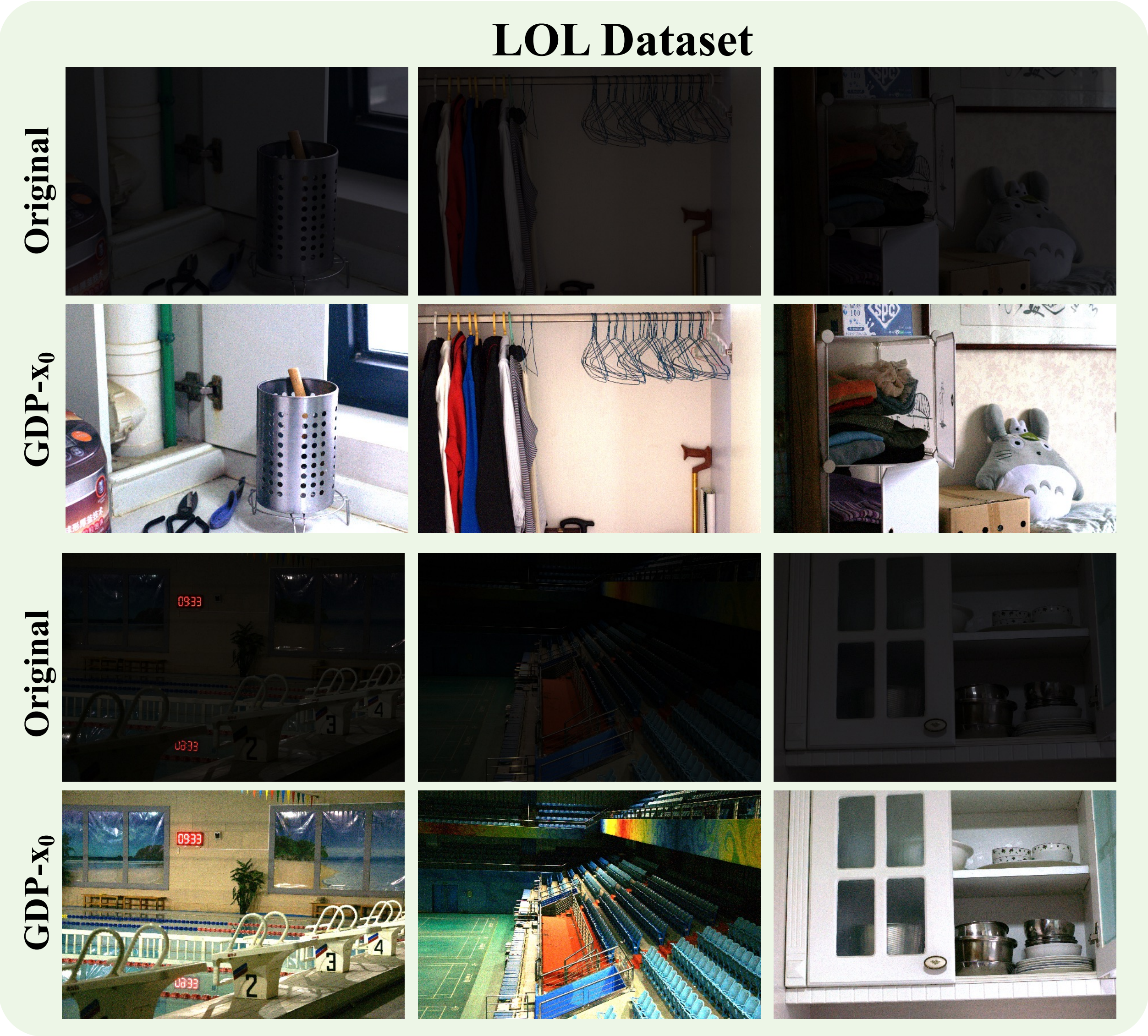
for t *from* T *to* 1 **do**

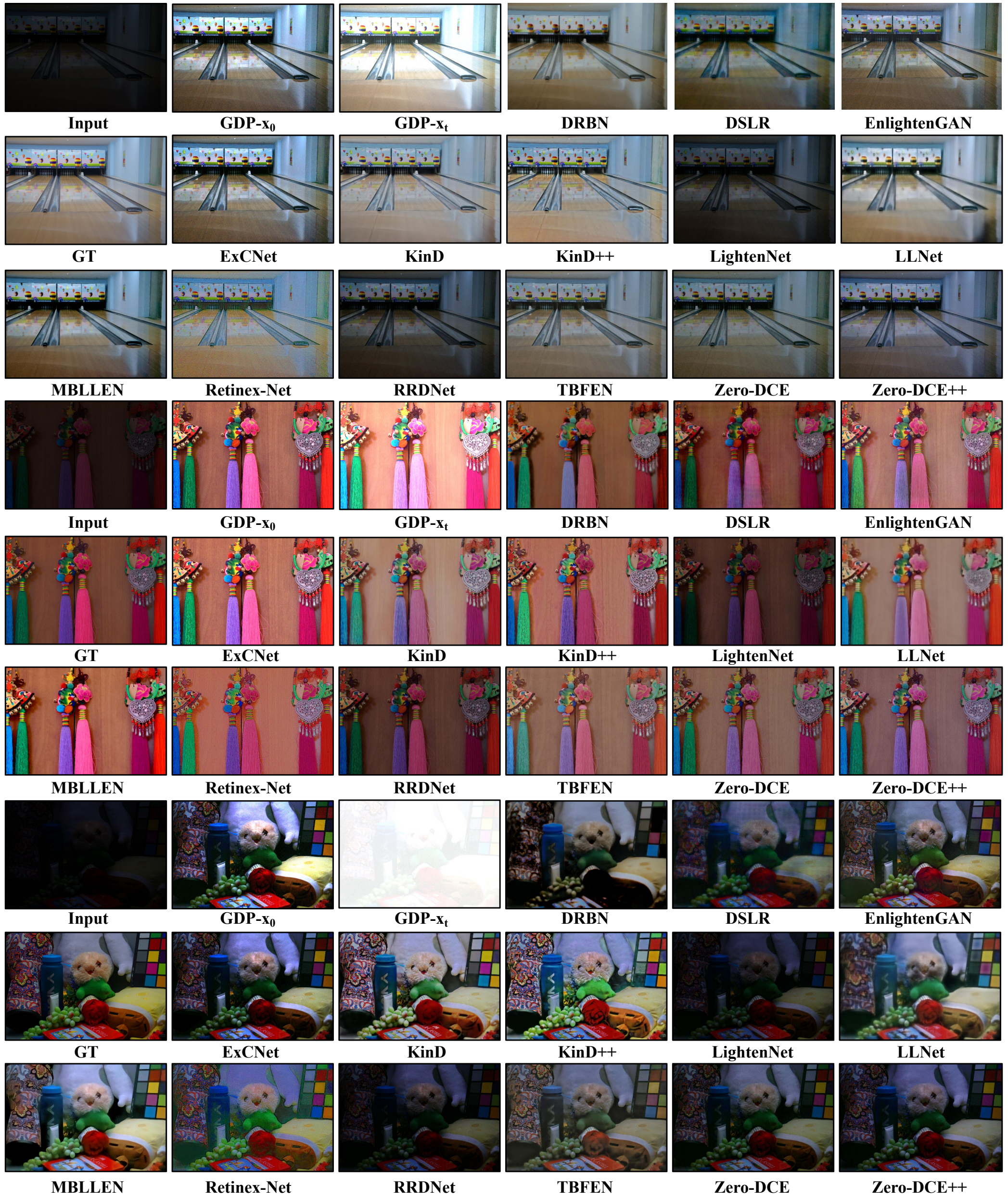
- $\mu, \Sigma = \mu_\theta(\boldsymbol{x}_t), \Sigma_\theta(\boldsymbol{x}_t)$
- Mean vector $\boldsymbol{\Omega}_t = \mathbf{0}$ and variance vector $\boldsymbol{\psi}_t = \mathbf{0}$ and weight vector $\mathbf{G} = \mathbf{0}$ and $f = \mathbf{0}$ and $\boldsymbol{\mathcal{M}} = \mathbf{0}$
- **for** $k = 1, \ldots, K$ **do**
  - $\boldsymbol{x}_t^k = \text{Crop}\left(\mathbf{P}^k \circ \boldsymbol{x}_t\right)$
  - $\boldsymbol{y}^k = \text{Crop}\left(\mathbf{P}^k \circ \boldsymbol{y}\right)$
  - $\boldsymbol{\mathcal{M}}^k = \text{Crop}\left(\mathbf{P}^k \circ \boldsymbol{\mathcal{M}}\right)$
  - $\tilde{\boldsymbol{x}}_0^k = \frac{\boldsymbol{x}_t^k}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta\left(\boldsymbol{x}_t^k, t\right)}{\sqrt{\bar{\alpha}_t}}$
  - $\mathcal{L}_{\phi, \tilde{\boldsymbol{x}}_0^k}^{total} = \mathcal{L}(\boldsymbol{y}^k, \mathcal{D}_\phi(\tilde{\boldsymbol{x}}_0^k)) + \mathcal{Q}(\tilde{\boldsymbol{x}}_0^k)$
  - $f^k \leftarrow f^k - l\nabla_{f^k}\mathcal{L}_{f^k, \tilde{\boldsymbol{x}}_0^k}^{total}$
  - $\boldsymbol{\mathcal{M}}^k \leftarrow \boldsymbol{\mathcal{M}}^k - l\nabla_{\boldsymbol{\mathcal{M}}^k}\mathcal{L}_{\boldsymbol{\mathcal{M}}^k, \tilde{\boldsymbol{x}}_0^k}^{total}$
  - $\mu^k = \mu + s\nabla_{\tilde{\boldsymbol{x}}_0^k}\mathcal{L}_{\phi, \tilde{\boldsymbol{x}}_0^k}^{total}$
  - $f = f + f^k$
  - $\boldsymbol{\Omega}_t = \boldsymbol{\Omega}_t + \mathbf{P}_k \cdot \mu^k$
  - $\boldsymbol{\psi}_t = \boldsymbol{\psi}_t + \mathbf{P}^k \cdot \sigma^k$
  - $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{M}} + \mathbf{P}^k \cdot \boldsymbol{\mathcal{M}}^k$
  - $\mathbf{G} = \mathbf{G} + \mathbf{P}^k$
- **end**
- $\boldsymbol{\Omega}_t = \boldsymbol{\Omega}_t \oslash \mathbf{G}$ $\quad$ $//\oslash$: element-wise division
- $\boldsymbol{\psi}_t = \boldsymbol{\psi}_t \oslash \mathbf{G}$
- $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{M}} \oslash \mathbf{G}$
- $f = f/K$
- Sample $\boldsymbol{x}_{t-1}$ by $\mathcal{N}(\boldsymbol{\Omega}_t, \boldsymbol{\psi}_t)$

end

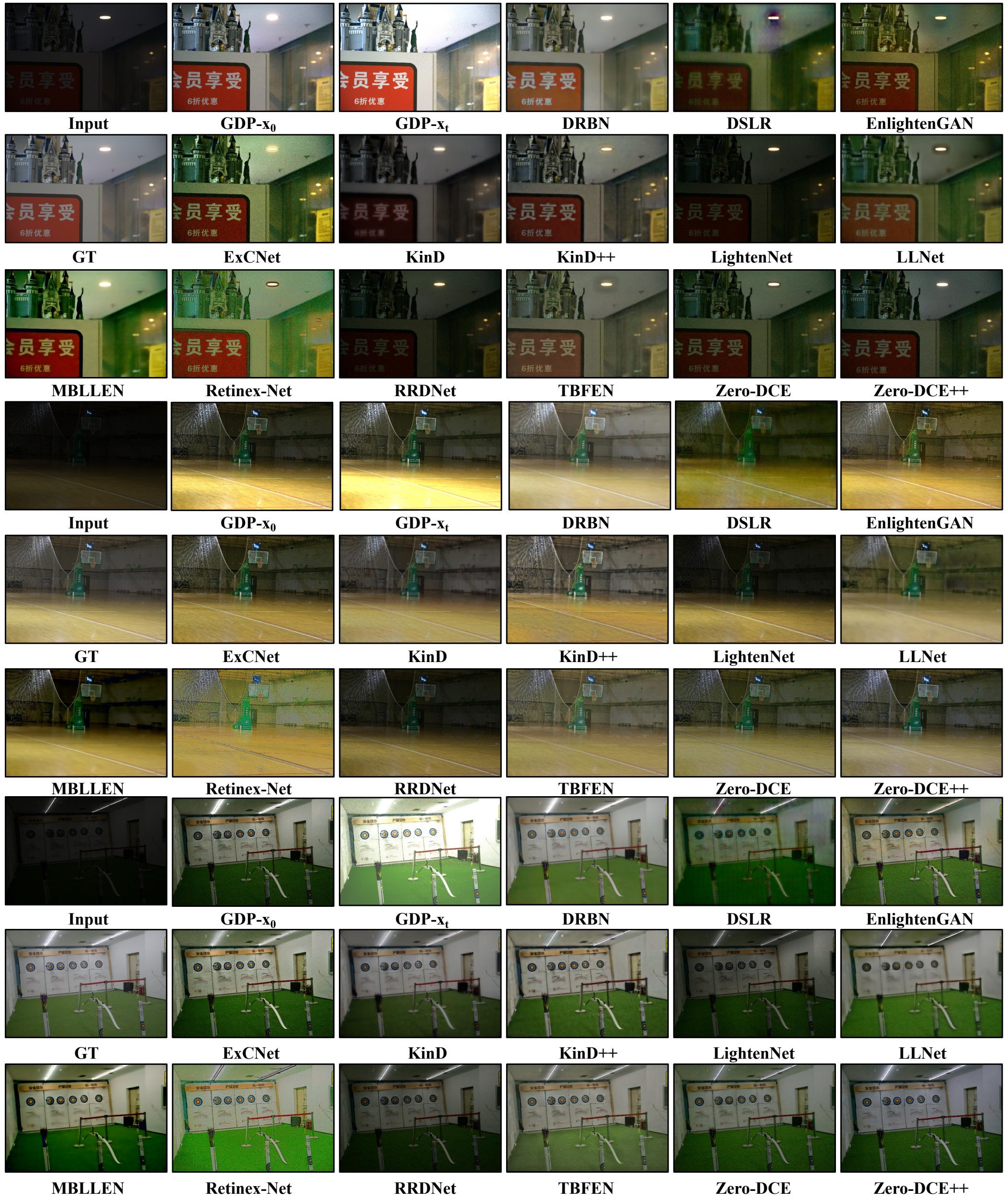
return Restored any-size image $\boldsymbol{x}_0$

Low-light enhancement-LOL

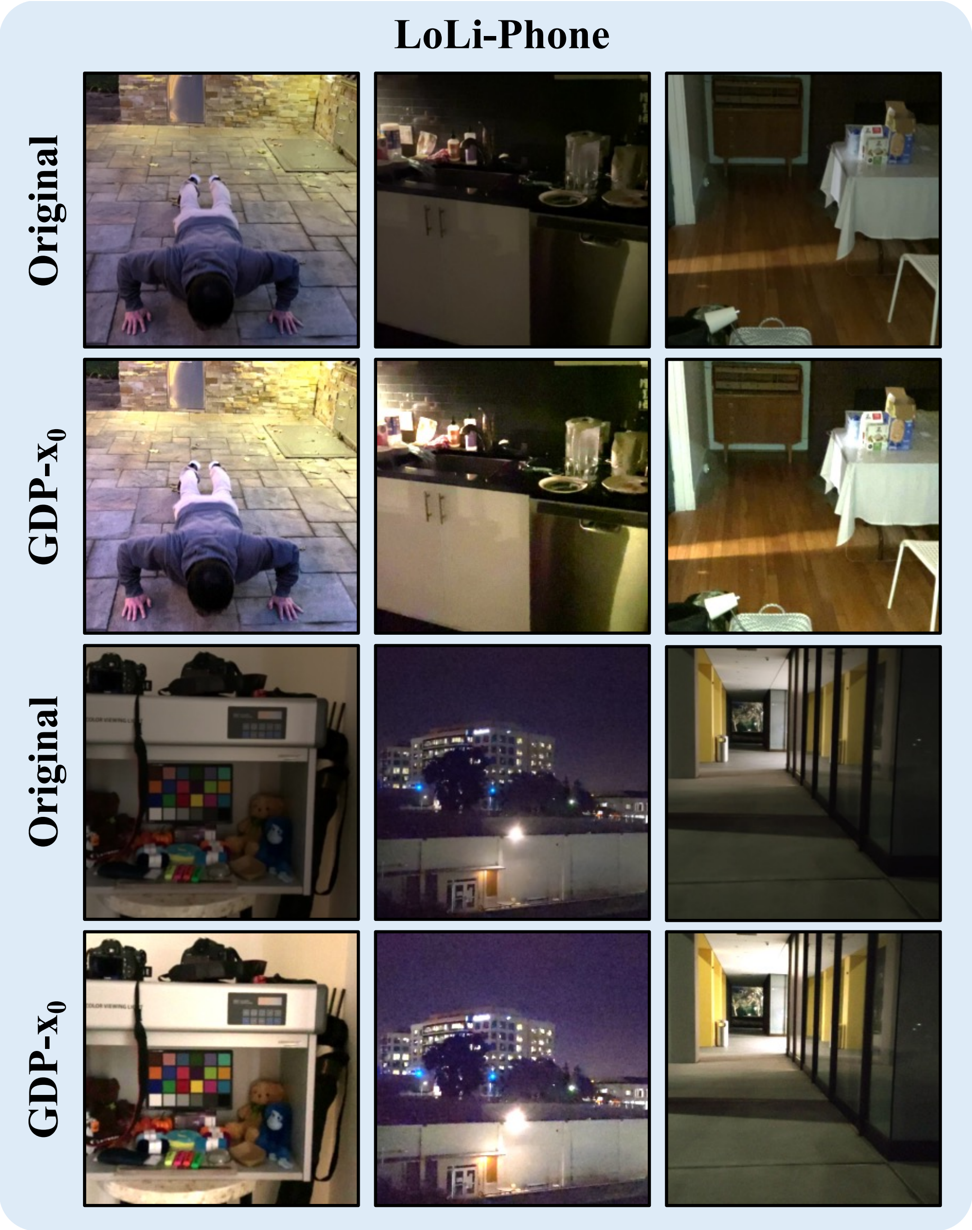


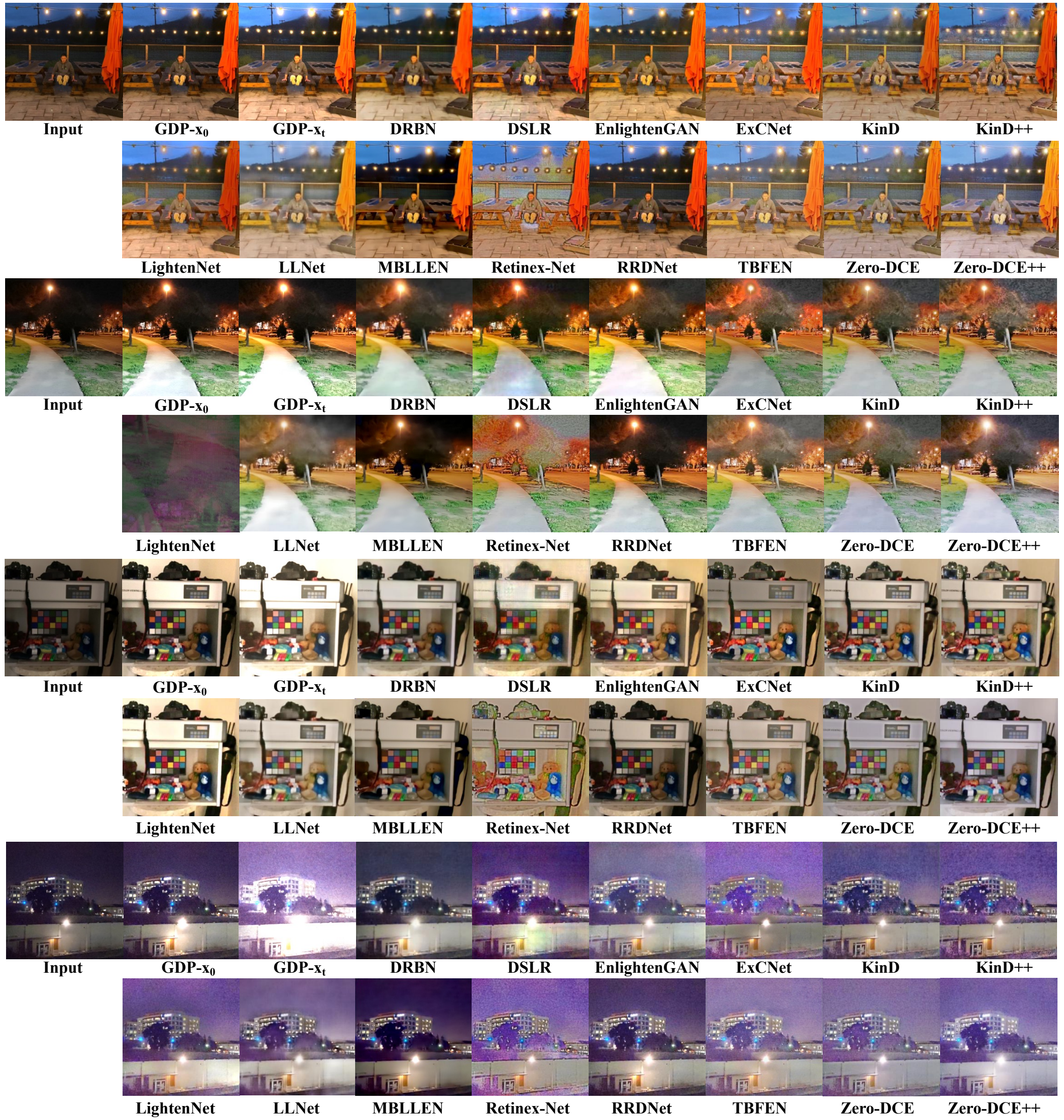


Low-light enhancement-VE-LOL

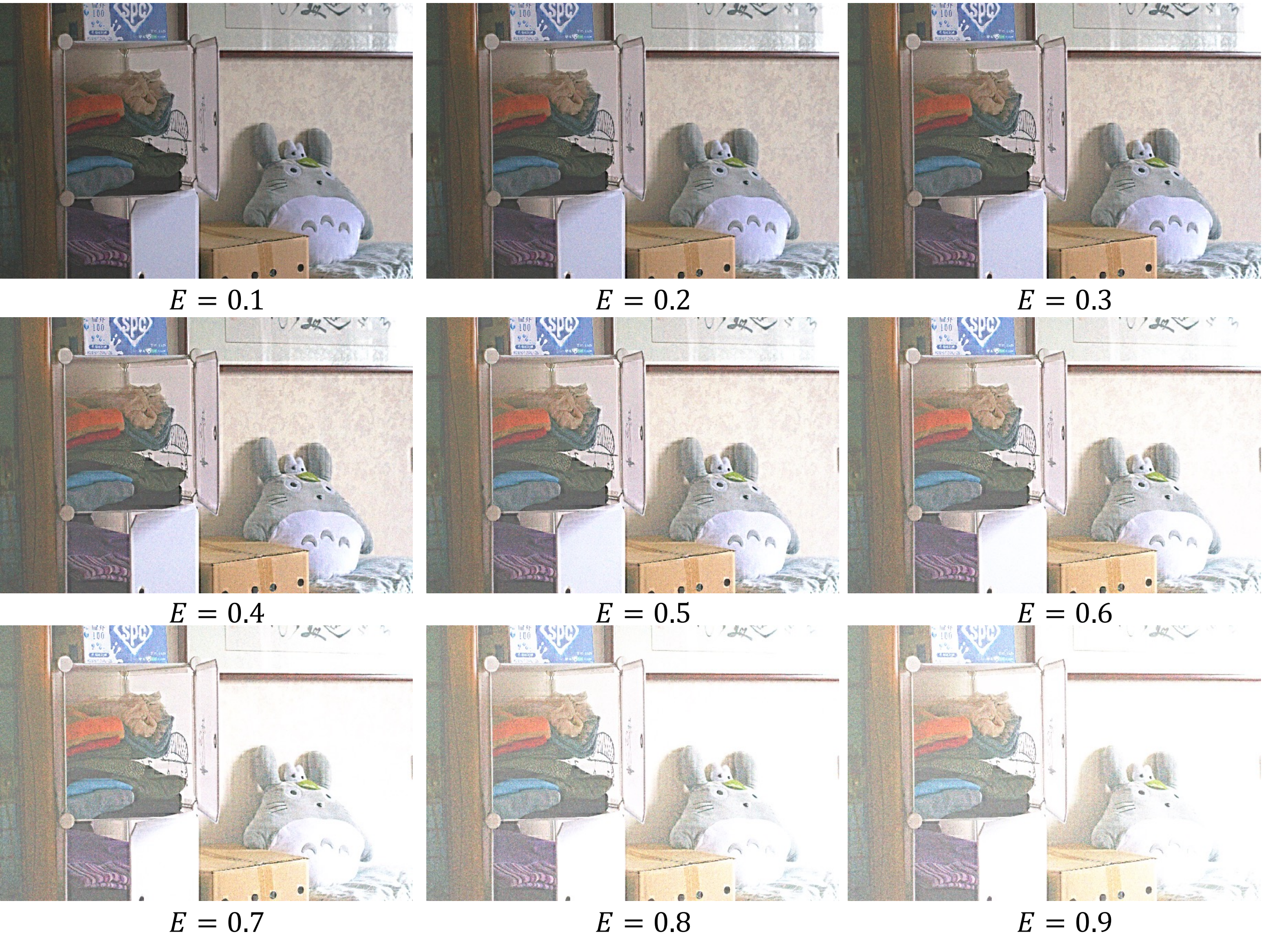


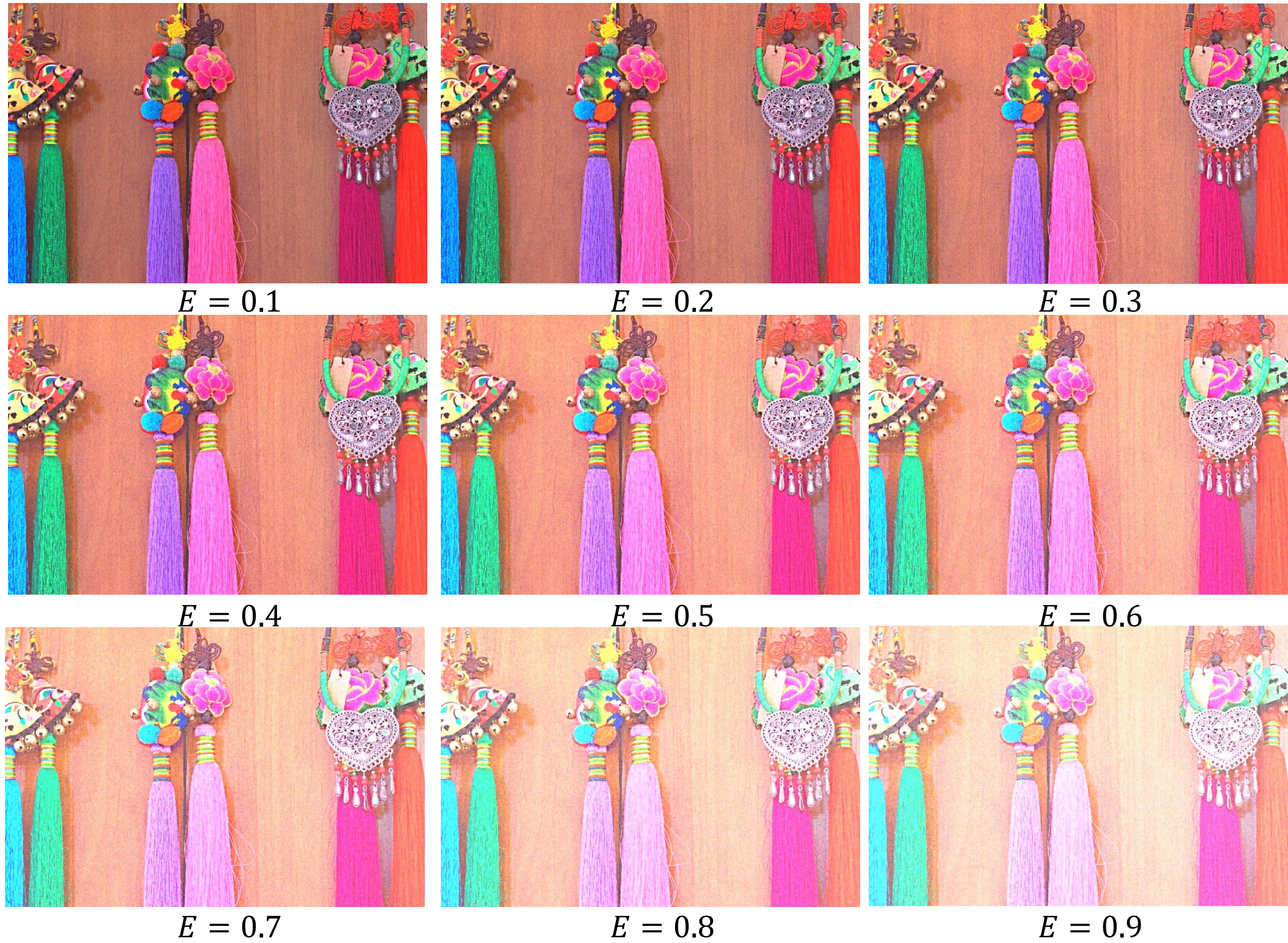
Low-light enhancement-LoLi-phone





Low-light enhancement-brightness control





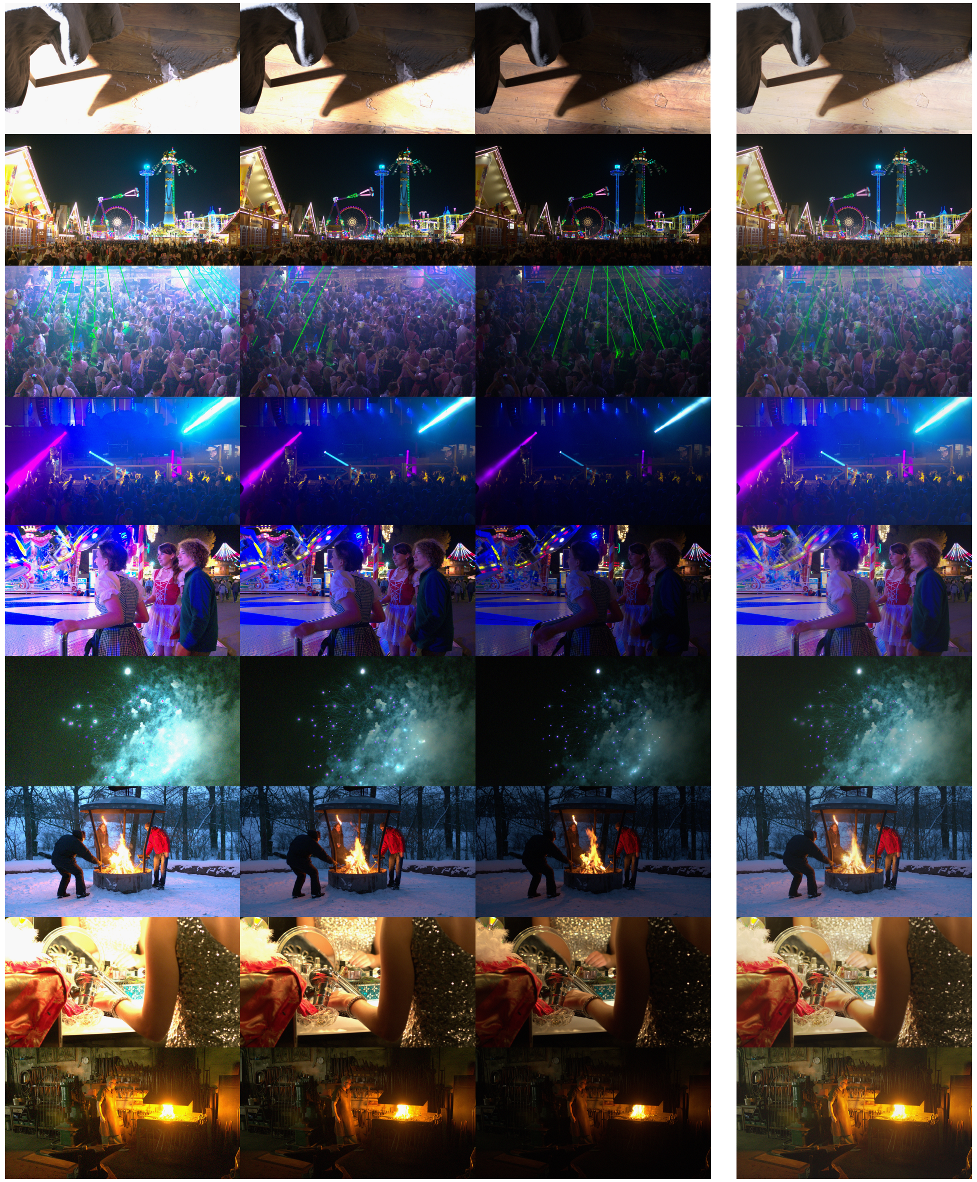
$$L_{\text{exp}} = \frac{1}{U} \sum_{k=1}^{U} |R_k - E| \,,$$

Low-light enhancement

Learning	Methods	LOL [82]					VE-LOL-L [43]					LoLi-Phone [37]	
		PSNR ↑	SSIM↑	FID↓	LOE↓	PI↓	PSNR↑	SSIM↑	FID ↓	LOE↓	PI↓	LOE↓	PI↓
Supervised learning	LLNet [45]	17.91	0.76	169.20	384.21	4.10	17.38	0.73	124.98	291.59	5.54	343.34	5.36
	LightenNet [39]	10.29	0.45	90.91	273.21	7.09	13.26	0.57	82.26	199.45	7.29	500.22	6.63
	Retinex-Net [82]	17.24	0.55	129.99	513.28	8.63	16.41	0.64	135.20	421.41	8.62	542.29	8.23
	MBLLEN [47]	17.90	0.77	122.69	175.10	8.39	15.95	0.70	105.74	114.91	7.45	137.34	6.46
	KinD [98]	17.57	0.82	74.52	377.59	7.41	18.07	0.78	80.12	253.79	7.51	265.47	6.84
	KinD++ [96]	17.60	0.80	100.15	712.12	7.96	16.80	0.74	101.23	421.97	7.98	382.51	7.71
	TBFEN [46]	17.25	0.83	90.59	367.66	8.29	18.91	0.81	91.30	276.65	8.02	214.30	7.34
	DSLR [42]	14.98	0.67	183.92	272.68	7.09	15.70	0.68	124.80	271.63	7.27	281.25	6.99
Unsupervised learning	EnlightenGAN [25]	17.44	0.74	82.60	379.23	8.78	17.45	0.75	86.51	311.85	8.27	373.41	7.26
Self-supervised learning	DRBN [88]	15.15	0.52	94.96	692.99	5.53	18.47	0.78	88.10	268.70	6.15	285.06	5.31
Zero-shot learning	ExCNet [94]	**16.04**	0.62	111.18	220.38	8.70	16.20	0.66	115.24	225.15	8.62	359.96	7.95
	Zero-DCE [20]	14.91	**0.70**	81.11	245.54	8.84	**17.84**	**0.73**	85.72	194.10	8.12	214.30	7.34
	Zero-DCE++ [38]	14.86	0.62	86.22	302.06	7.08	16.12	0.45	86.96	313.50	7.92	308.15	7.18
	RRDNet [100]	11.37	0.53	89.09	127.22	8.17	13.99	0.58	83.41	94.23	7.36	92.73	7.20
	GDP-x_t	7.32	0.57	238.92	364.15	8.26	9.45	0.50	152.68	194.49	7.12	508.73	8.06
	GDP-x_0	13.93	0.63	**75.16**	**110.39**	**6.47**	13.04	0.55	**78.74**	**79.08**	**6.47**	**75.29**	**6.35**

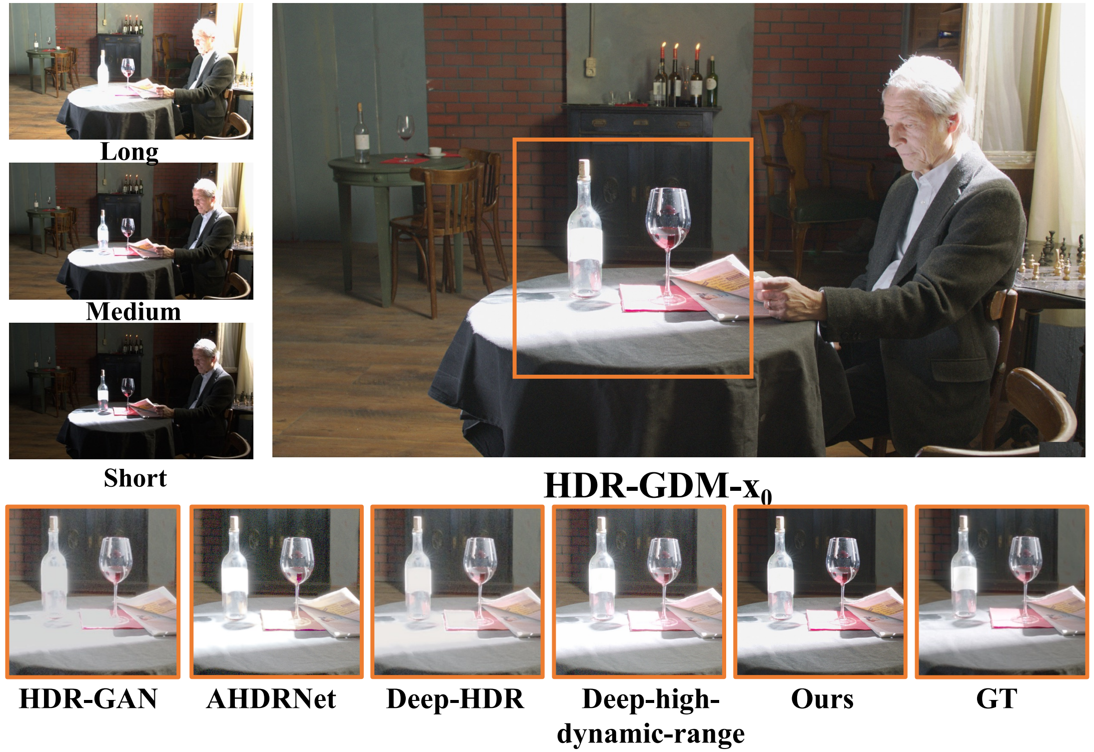
- GDP-x_0 fulfills the best FID, lightness order error (LOE), and perceptual index (PI) across all the zero-shot methods under three datasets.
- The lower LOE demonstrates better preservation for the naturalness of lightness, while the lower PI indicates better perceptual quality.

HDR image recovery



HDR image recovery

Methods	PSNR↑	SSIM↑	LPIPS ↓	FID↓
AHDRNet [86]	18.72	0.58	0.39	81.98
HDR-GAN [55]	21.67	0.74	0.26	52.71
Deep-HDR [84]	21.66	0.76	0.26	57.52
Deep-high-dynamic-range [26]	21.33	0.71	0.26	51.92
GDP-x_t	19.36	0.65	0.30	63.89
GDP-x_0	**24.88**	**0.86**	**0.13**	**50.05**



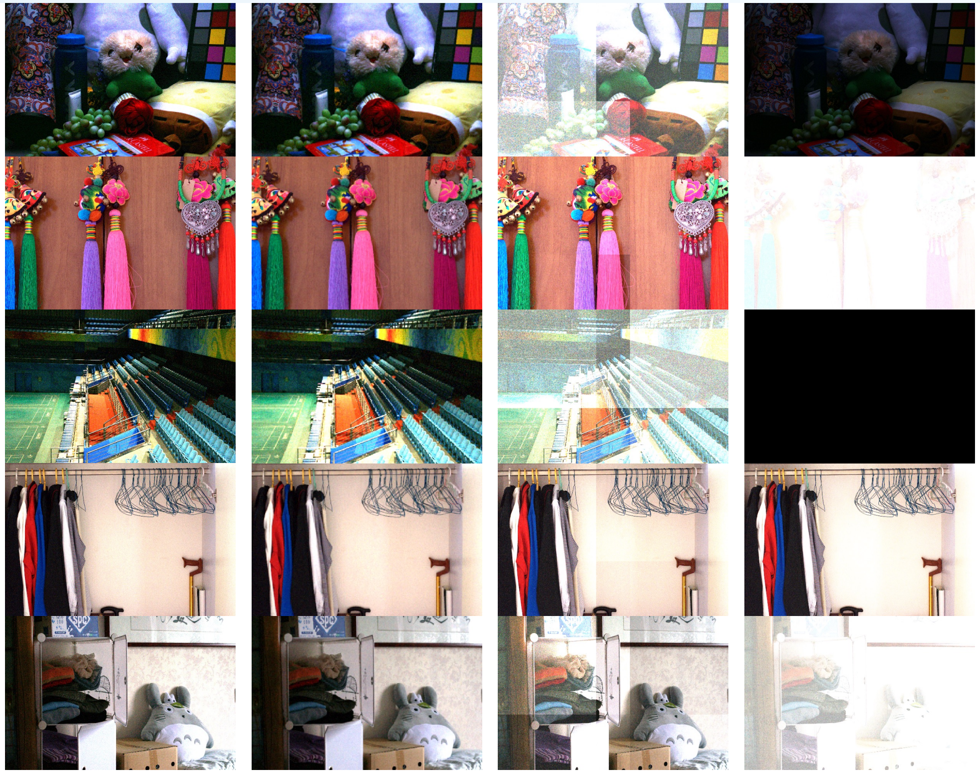
- HDR-GDP-x_0 exceeds the other methods in PSNR, SSIM, LPIPS, and FID.
- HDR-GDP-x_0 achieves a better quality of reconstructed images, where the low-light parts can be enhanced, and the over-exposure regions are adjusted.

Ablation Study

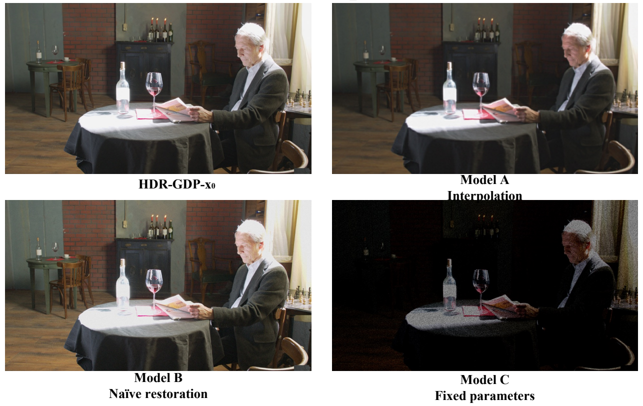
Task	4× Super resolution				Deblur			
	PSNR	SSIM	Consistency	FID	PSNR	SSIM	Consistency	FID
GDP -x_t with Σ	22.86	0.60	88.37	68.04	22.06	0.57	69.46	80.39
GDP -x_0 with Σ	22.09	0.58	93.19	41.22	23.49	0.65	68.67	50.29
GDP -x_t	24.27	0.67	80.32	64.67	25.86	0.73	54.08	5.00
GDP -x_0	**24.42**	**0.68**	**6.49**	**38.24**	**25.98**	**0.75**	**41.27**	**2.44**

Task	25% Inpainting				Colorization			
	PSNR	SSIM	Consistency	FID	PSNR	SSIM	Consistency	FID
GDP -x_t with Σ	25.28	0.70	171.44	73.32	17.67	0.70	246.26	145.20
GDP -x_0 with Σ	24.58	0.75	65.59	22.77	21.28	0.91	66.57	38.39
GDP -x_t	31.06	0.93	8.80	20.24	21.30	0.86	75.24	66.43
GDP -x_0	**34.40**	**0.96**	**5.29**	**16.58**	**21.41**	**0.92**	**36.92**	**37.60**

Methods	LOL					NTIRE			
	PSNR	SSIM	FID	LOE	PI	PSNR	SSIM	LPIPS	FID
Model A	11.05	0.49	156.51	707.57	8.61	24.12	0.67	0.32	86.69
Model B	9.01	0.37	355.99	969.89	9.04	9.83	0.04	1.02	253.11
GDP-x_t	7.32	0.57	238.92	364.15	8.26	19.36	0.65	0.30	63.89
GDP-x_0	**13.93**	**0.63**	**75.16**	**110.39**	**6.47**	**24.88**	**0.86**	**0.13**	**50.05**



GDP-x_0 **Model A Interpolation** **Model B Naïve restoration** **Model C Fixed parameters**



- Model A is devised to naively restore the images from patches and patches where the parameters are not related.
- Model B is designed with fixed parameters for all patches in the images.

Conclusion

(1) We introduce GDP, an effective and unsupervised posterior sampling method, for unified image restoration and enhancement.

(2) Our GDP is capable of optimizing the randomly initiated parameters of degradation that are unknown, resulting in a powerful GDP that can tackle any blind image restoration.

(3) Further, to achieve arbitrary size image generation, we propose hierarchical guidance and patch-based methods, greatly promoting the GDP on natural image enhancement.

(4) Moreover, the comprehensive experiments are carried out, different from the commonly utilized guidance way, where GDP directly predicts the temporary output given the noisy image in every step, which will be leveraged to guide the generation of images in the next step.

Thanks for listening!