南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

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
bfei21@m.fudan.edu.cn, wdyang@fudan.edu.cn, (lvzhaoyang,daibo)@pjlab.org.cn

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

1. Linear Inverse Image Restoration   Denoising Diffusion Restoration Models

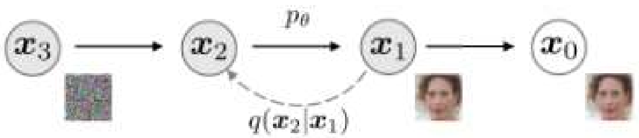
   image deblurring, super-resolution, and compressive sensing, inpainting, colorization

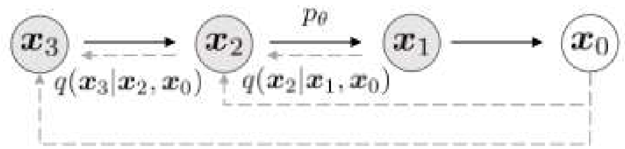
2. Non-linear Image Restoration.

   image low-light enhancement, HDR image recovery,
   JPEG Artifact Correction

3. Blind Image Restoration

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS





噪声采样 —去噪→ 生成图像

噪声采样 ←加噪— 生成图像

$$\boldsymbol{x}_t = \alpha_t \boldsymbol{x}_{t-1} + \beta_t \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

$$\boldsymbol{x}_t = \underbrace{(\alpha_t \cdots \alpha_1)}_{\text{记为}\bar{\alpha}_t} \boldsymbol{x}_0 + \underbrace{\sqrt{1 - (\alpha_t \cdots \alpha_1)^2}}_{\text{记为}\bar{\beta}_t} \bar{\boldsymbol{\varepsilon}}_t, \quad \bar{\boldsymbol{\varepsilon}}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

$$\alpha_t \bar{\beta}_{t-1} \bar{\boldsymbol{\varepsilon}}_{t-1} + \beta_t \boldsymbol{\varepsilon}_t$$

$$\beta_t \bar{\boldsymbol{\varepsilon}}_{t-1} - \alpha_t \bar{\beta}_{t-1} \boldsymbol{\varepsilon}_t$$

$$\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\omega}^\top] = \boldsymbol{0}$$

$$\boldsymbol{\varepsilon}_t = \frac{(\beta_t \boldsymbol{\varepsilon} - \alpha_t \bar{\beta}_{t-1} \boldsymbol{\omega}) \bar{\beta}_t}{\beta_t^2 + \alpha_t^2 \bar{\beta}_{t-1}^2} = \frac{\beta_t \boldsymbol{\varepsilon} - \alpha_t \bar{\beta}_{t-1} \boldsymbol{\omega}}{\bar{\beta}_t}$$

https://kexue.fm/archives/9119

$$\boldsymbol{\mu}(\boldsymbol{x}_t) = \frac{1}{\alpha_t}\left(\boldsymbol{x}_t - \beta_t \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\right)$$

$$\left\| \boldsymbol{\varepsilon}_t - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\bar{\alpha}_t \boldsymbol{x}_0 + \alpha_t \bar{\beta}_{t-1} \bar{\boldsymbol{\varepsilon}}_{t-1} + \beta_t \boldsymbol{\varepsilon}_t, t) \right\|^2$$

$$\begin{aligned} &\mathbb{E}_{\bar{\boldsymbol{\varepsilon}}_{t-1}, \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})}\left[\left\| \boldsymbol{\varepsilon}_t - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\bar{\alpha}_t \boldsymbol{x}_0 + \alpha_t \bar{\beta}_{t-1} \bar{\boldsymbol{\varepsilon}}_{t-1} + \beta_t \boldsymbol{\varepsilon}_t, t) \right\|^2\right] \\ =\,&\mathbb{E}_{\boldsymbol{\omega}, \boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})}\left[\left\| \frac{\beta_t \boldsymbol{\varepsilon} - \alpha_t \bar{\beta}_{t-1} \boldsymbol{\omega}}{\bar{\beta}_t} - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\bar{\alpha}_t \boldsymbol{x}_0 + \bar{\beta}_t \boldsymbol{\varepsilon}, t) \right\|^2\right] \end{aligned}$$

$$\left\| \boldsymbol{\varepsilon} - \frac{\bar{\beta}_t}{\beta_t} \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\bar{\alpha}_t \boldsymbol{x}_0 + \bar{\beta}_t \boldsymbol{\varepsilon}, t) \right\|^2$$

$$q\left(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T \mid \boldsymbol{x}_0\right)=\prod_{t=1}^T q\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}\right), \tag{1}$$

$$\boldsymbol{x}_t=\sqrt{\bar{\alpha}_t} \boldsymbol{x}_0+\sqrt{1-\bar{\alpha}_t} \boldsymbol{\epsilon}, \tag{2}$$

$$p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_0, \cdots, \boldsymbol{x}_{T-1} \mid \boldsymbol{x}_T\right)=\prod_{t=1}^T p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t\right), \quad p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t\right)=\mathcal{N}\left(\boldsymbol{x}_{t-1} ; \boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right), \Sigma_\theta \boldsymbol{I}\right) \tag{3}$$

$$\boldsymbol{\mu}_\theta\left(\boldsymbol{x}_t, t\right)=\frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta\left(\boldsymbol{x}_t, t\right)\right) \tag{4}$$

$$\tilde{\boldsymbol{x}}_0=\frac{\boldsymbol{x}_t}{\sqrt{\bar{\alpha}_t}}-\frac{\sqrt{1-\bar{\alpha}_t} \boldsymbol{\epsilon}_\theta\left(\boldsymbol{x}_t, t\right)}{\sqrt{\bar{\alpha}_t}} \tag{5}$$

$$q\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \tilde{\boldsymbol{x}}_0\right)=\mathcal{N}\left(\boldsymbol{x}_{t-1} ; \tilde{\boldsymbol{\mu}}_t\left(\boldsymbol{x}_t, \tilde{\boldsymbol{x}}_0\right), \tilde{\beta}_t \mathbf{I}\right),$$
$$\text{where} \quad \tilde{\boldsymbol{\mu}}_t\left(\boldsymbol{x}_t, \tilde{\boldsymbol{x}}_0\right)=\frac{\sqrt{\bar{\alpha}_{t-1}} \beta_t}{1-\bar{\alpha}_t} \tilde{\boldsymbol{x}}_0+\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t} \boldsymbol{x}_t$$
$$\text{and} \quad \tilde{\beta}_t=\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \beta_t \tag{6}$$

$$\boldsymbol{y}=\mathcal{D}(\boldsymbol{x}).$$

$$p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{t-1} | \boldsymbol{x}_t, \boldsymbol{y}\right).$$

$$\begin{aligned} \log p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{t-1} | \boldsymbol{x}_t, \boldsymbol{y}\right) & =\log \left(p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{t-1} | \boldsymbol{x}_t\right) p\left(\boldsymbol{y} | \boldsymbol{x}_t\right)\right)+K_1 \\ & \approx \log p(\boldsymbol{r})+K_2, \end{aligned} \tag{7}$$

$$\boldsymbol{r} \sim \mathcal{N}\left(\boldsymbol{r} ; \boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)+\Sigma \boldsymbol{g}, \Sigma\right)$$
$$\boldsymbol{g} = \nabla_{\boldsymbol{x}_t} \log p\left(\boldsymbol{y} \mid \boldsymbol{x}_t\right) \quad ; \Sigma=\Sigma_\theta\left(\boldsymbol{x}_t\right)$$

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

$$p(\boldsymbol{y} \mid \boldsymbol{x}_t) = \frac{1}{Z} \exp\left(-\left[s\mathcal{L}\left(\mathcal{D}(\boldsymbol{x}_t), \boldsymbol{y}\right) + \lambda\mathcal{Q}(\boldsymbol{x}_t)\right]\right) \quad (8)$$

a heuristic approximation

$\mathcal{L}$: image distance metric

Z : normalization factor

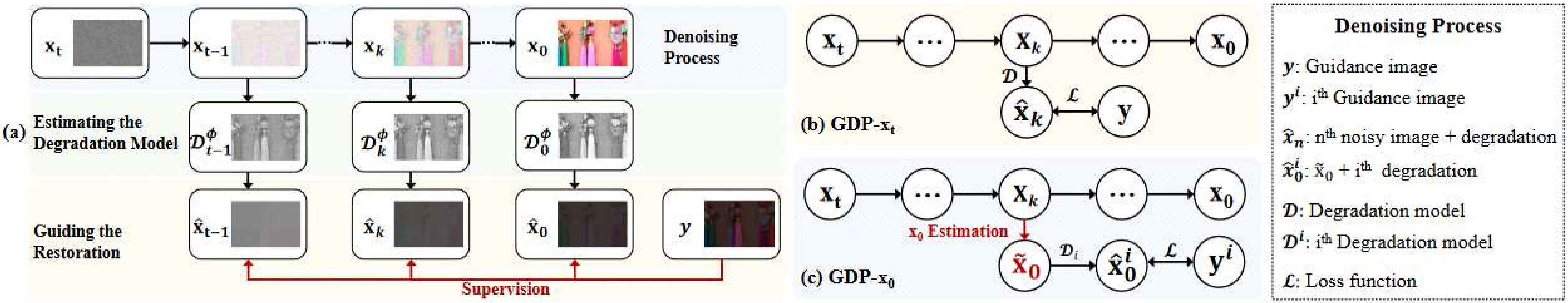
s : a scaling factor controlling the magnitude of guidance

$\mathcal{Q}$: the optional quality enhancement loss

λ : scale factor for adjusting the quality of images

$$\begin{aligned}\log p(\boldsymbol{y} \mid \boldsymbol{x}_t) &= -\log Z - s\mathcal{L}\left(\mathcal{D}(\boldsymbol{x}_t), \boldsymbol{y}\right) - \lambda\mathcal{Q}\left(\boldsymbol{x}_t\right) \\ \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y} \mid \boldsymbol{x}_t) &= -s\nabla_{\boldsymbol{x}_t}\mathcal{L}\left(\mathcal{D}(\boldsymbol{x}_t), \boldsymbol{y}\right) - \lambda\nabla_{\boldsymbol{x}_t}\mathcal{Q}\left(\boldsymbol{x}_t\right).\end{aligned} \quad (9)$$

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS



the way of adding guidance and the variance Σ negatively influence the reconstructed images.

The Influence of Variance Σ on the Guidance.

Guidance on xt

Algorithm 1: GDP-x_t with fixed degradation model: Conditioner guided diffusion sampling on x_t, given a diffusion model $(\mu_\theta(x_t), \Sigma_\theta(x_t))$, corrupted image conditioner y.

Input: Corrupted image 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 x_0 conditioned on y
Sample x_T from $\mathcal{N}(0, \mathbf{I})$
for *t from T to 1* **do**
  $\mu, \Sigma = \mu_\theta(x_t), \Sigma_\theta(x_t)$
  $\mathcal{L}_{x_t}^{total} = \mathcal{L}(y, \mathcal{D}(x_t)) + \mathcal{Q}(x_t)$
  Sample x_{t-1} by $\mathcal{N}(\mu + s\nabla_{x_t}\mathcal{L}_{x_t}^{total}, \Sigma)$
end
return x_0

Guidance on $\tilde{x}_0$.

Algorithm 2: GDP-x_0: Conditioner guided diffusion sampling on $\tilde{x}_0$, given a diffusion model $(\mu_\theta(x_t), \Sigma_\theta(x_t))$, corrupted image conditioner y.

Input: Corrupted image 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 x_0 conditioned on y
Sample x_T from $\mathcal{N}(0, \mathbf{I})$
for *t from T to 1* **do**
  $\mu, \Sigma = \mu_\theta(x_t), \Sigma_\theta(x_t)$
  $\tilde{x}_0 = \frac{x_t}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}}$
  $\mathcal{L}_{\phi,\tilde{x}_0}^{total} = \mathcal{L}(y, \mathcal{D}_\phi(\tilde{x}_0)) + \mathcal{Q}(\tilde{x}_0)$
  $\phi \leftarrow \phi - l\nabla_\phi\mathcal{L}_{\phi,\tilde{x}_0}^{total}$
  Sample x_{t-1} by $\mathcal{N}\left(\mu + s\nabla_{\tilde{x}_0}\mathcal{L}_{\phi,\tilde{x}_0}^{total}, \Sigma\right)$
end
return x_0

$$y = fx + \mathcal{M},$$

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Reconstruction Loss.	Quality Enhancement Loss.
MSE structural similarity index measure perceptual loss	Exposure Control Loss Color Constancy Loss Illumination Smoothness Loss

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

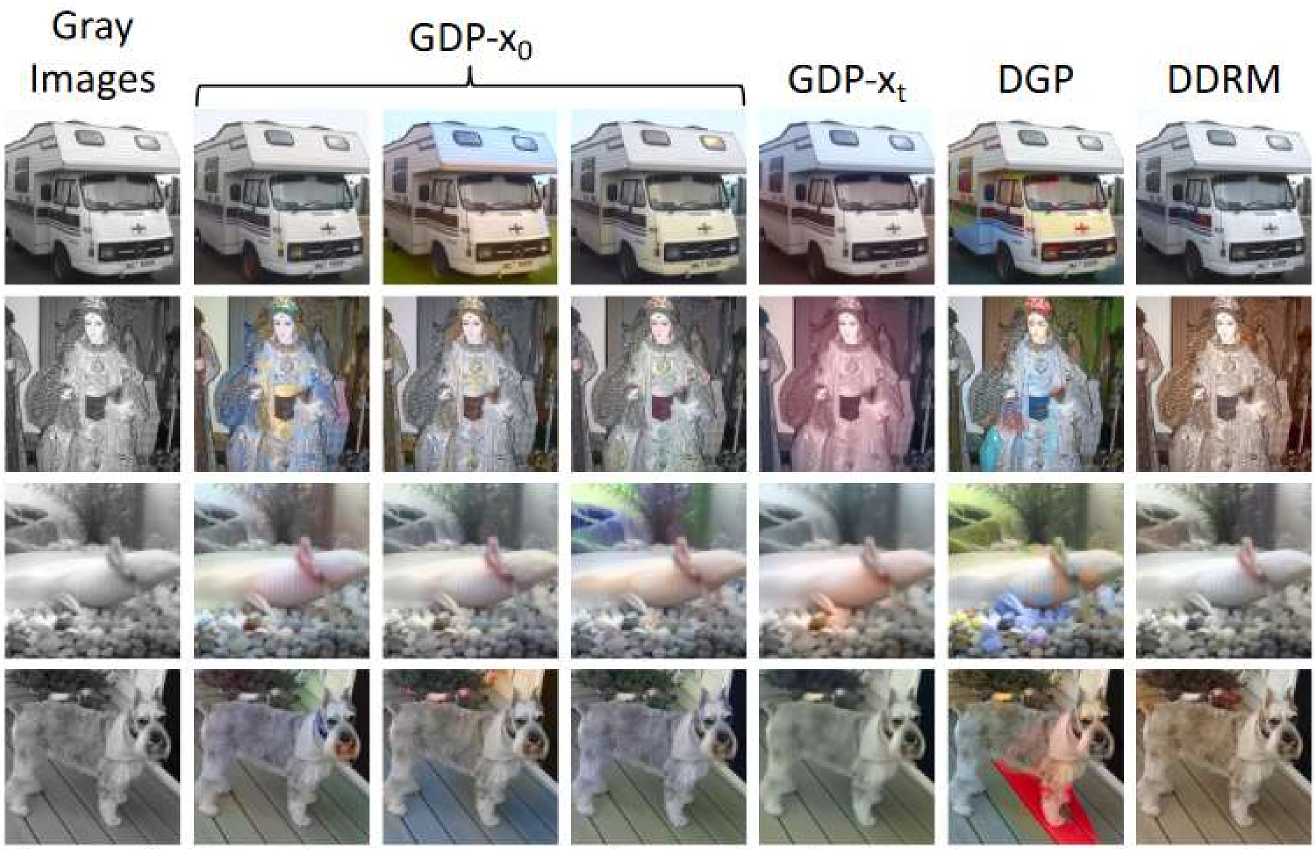
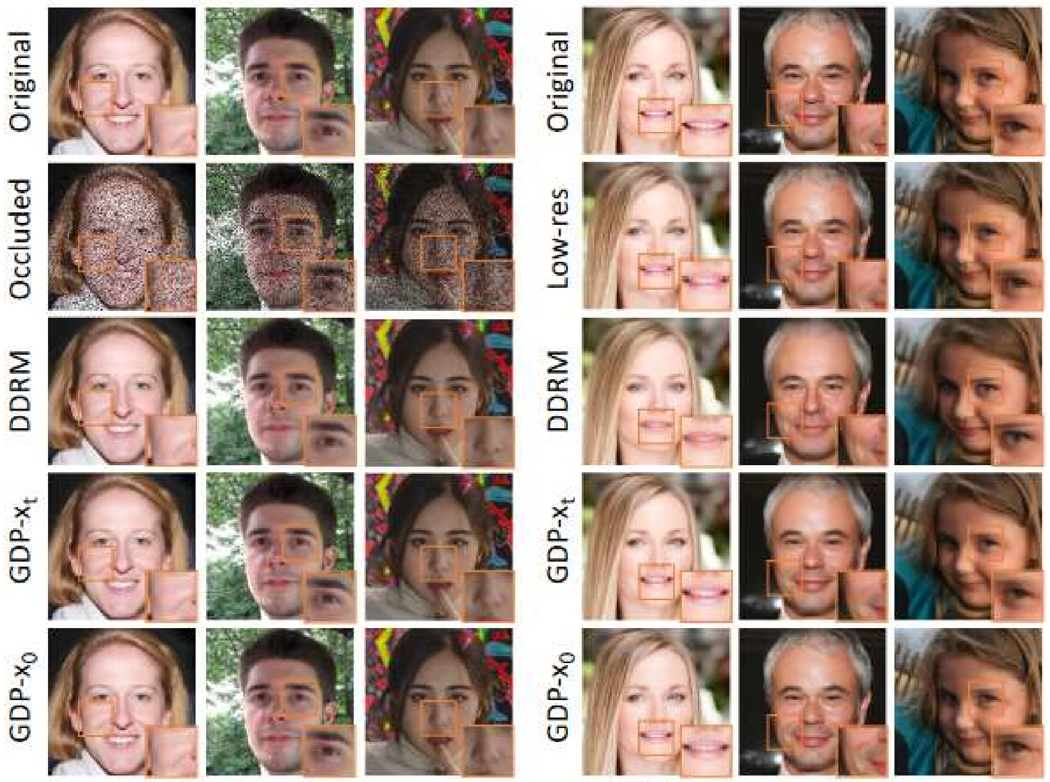


Figure 3. **Qualitative comparison of colorization results on ImageNet validation images**. GDP-x_0 generates various samples on the same input.

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Table 2. **Quantitative comparison of linear image restoration tasks on ImageNet 1k [62].** GDP outperforms other methods in terms of FID and Consistency across all tasks.

Method	4× Super-resolution				Deblur				25% Inpainting				Colorization			
	PSNR ↑	SSIM ↑	Consistency ↓	FID ↓	PSNR ↑	SSIM ↑	Consistency↓	FID ↓	PSNR ↑	SSIM ↑	Consistency↓	FID ↓	PSNR ↑	SSIM ↑	Consistency ↓	FID ↓
DGP [62]	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 [33]	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 [69]	24.18	0.71	27.57	98.30	21.30	0.58	63.20	69.55	-	-	-	-	-	-	-	-
DDRM [32]	**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**



(a) Inpainting (b) 4x super resolution

Figure 4. **Qualitative results of (a) 25 % inpainting and (b) 4× super-resolution on CelebA [31].**

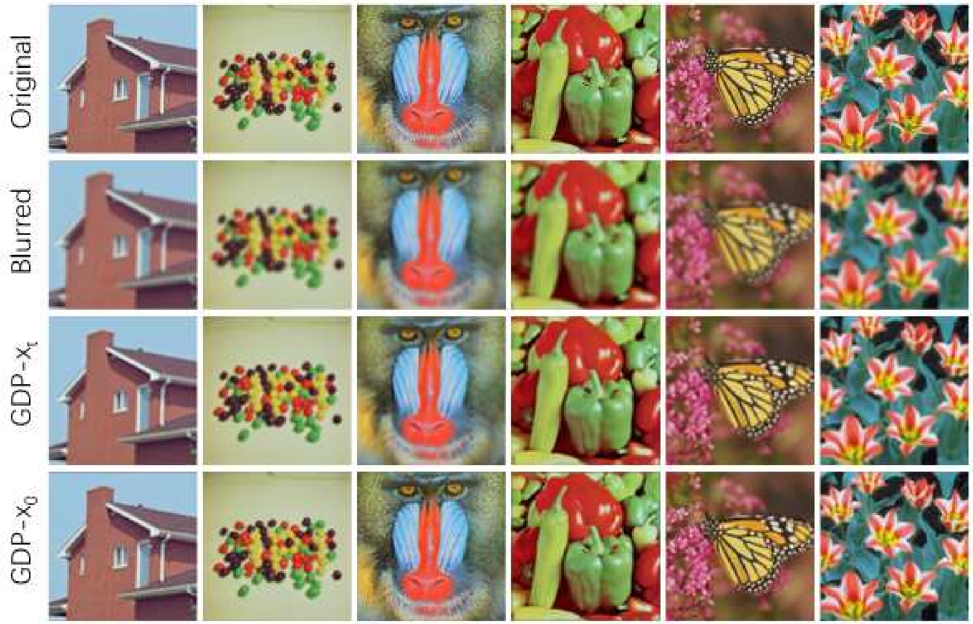


Figure 5. **Results of image deblurring task on 256 × 256 USC-SIPI images [87] using an ImageNet model.**

Figure 6. **Qualitative results of low-light enhancement on the LOL [88], VE-LOL [47], and LoLi-Phone [41] datasets.**

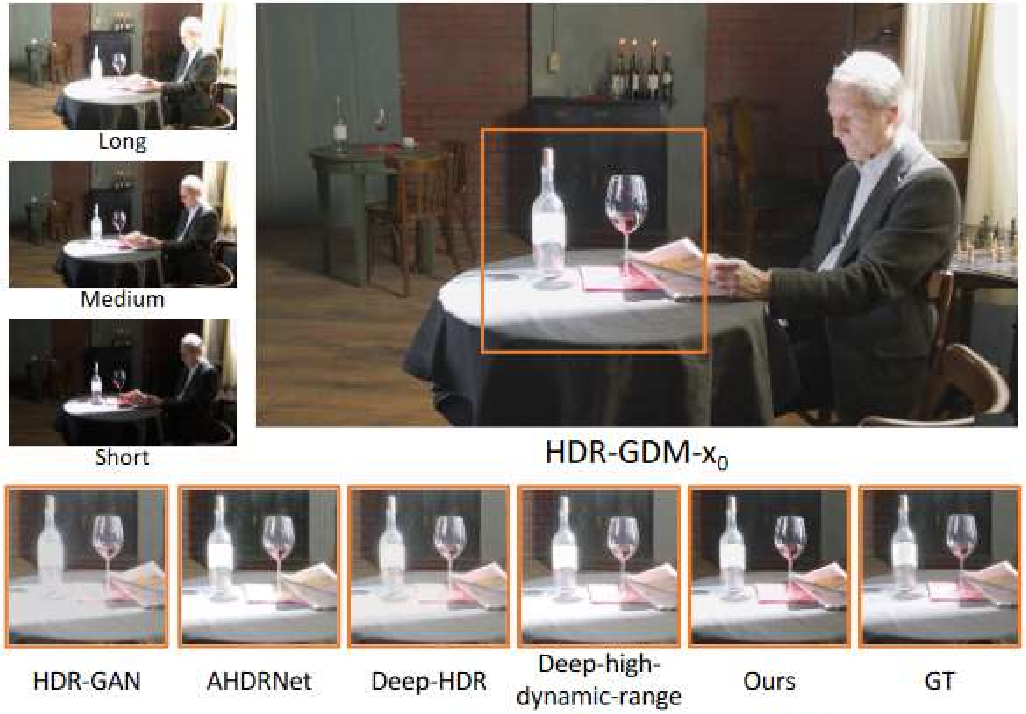


Figure 7. **Example from the NTIRE dataset [63].** We compare a set of patches cropped from the tone-mapped HDR images generated by state-of-the-art methods.

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Table 4. **Quantitative comparison on the NTIRE dataset [63].**

Methods	PSNR↑	SSIM↑	LPIPS ↓	FID↓
AHDRNet [91]	18.72	0.58	0.39	81.98
HDR-GAN [59]	21.67	0.74	0.26	52.71
Deep-HDR [90]	21.66	0.76	0.26	57.52
Deep-high-dynamic-range [30]	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**

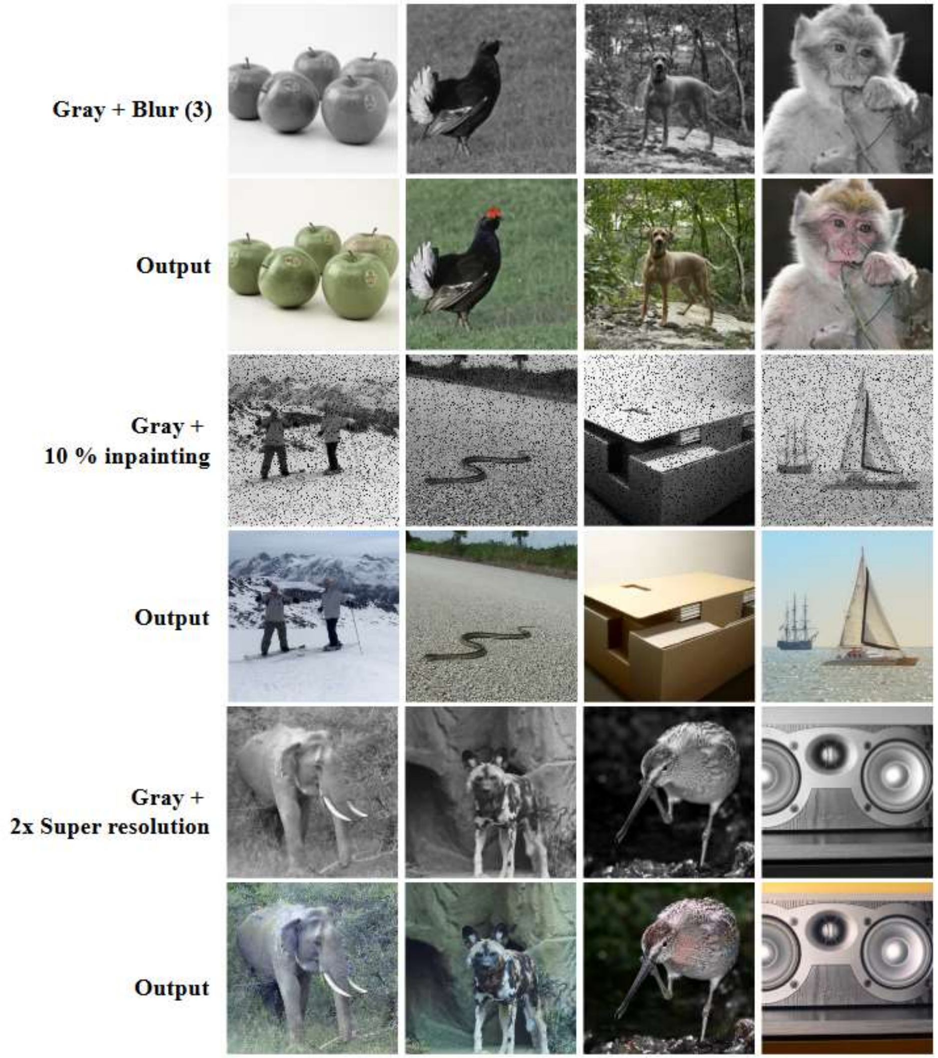
Table 5. **The ablation study on the variance Σ and the way of the guidance.**

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**

Table 6. **The ablation study on the optimizable degradation and patch-based tactic.**

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**


Gray + Blur (3)
Output
Gray +
10 % inpainting
Output
Gray +
2x Super resolution
Output

结　　束