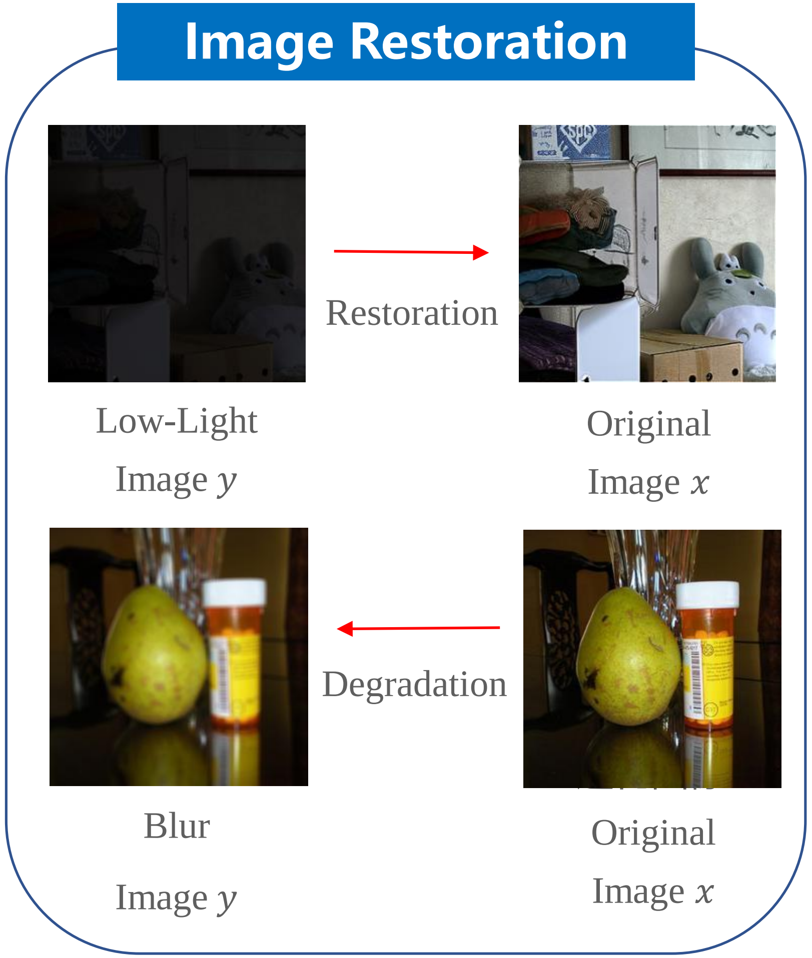
Taming Generative Diffusion Prior for Universal Blind Image Restoration

Siwei Tu[1], Weidong Yang[1,†], Ben Fei[2,†]

[1] Fudan University
[2] Chinese University of Hong Kong
[†] Corresponding author

➢ Research Background

- The original image x often undergoes quality degradation during capture, transmission, compression, and other processes, resulting in a degraded image y.
- The process of quality degradation can be expressed as $y = D(x)$, where D represents the degradation function.
- Image restoration tasks aim to restore the degraded image y to original x. It can be categorized into linear and blind problems.
- The latter necessitates the restoration of images when the degradation function remains unknown.



➢ Motivation-Setting Limitation of Guidance Scale

NEURAL INFORMATION PROCESSING SYSTEMS

Existing setting methods:

- Previous models, such as GDP[1], guidance scale is set as hyperparameter via manual specification.
- For diverse tasks, manual experimentation was required, and guidance scale are fixed across varying datasets and time steps.

Algorithm 2: GDP-$\boldsymbol{x}_0$: 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 $\boldsymbol{y}$.

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**

$\mu, \Sigma = \mu_\theta(\boldsymbol{x}_t), \Sigma_\theta(\boldsymbol{x}_t)$

$\tilde{\boldsymbol{x}}_0 = \frac{\boldsymbol{x}_t}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\boldsymbol{x}_t, t)}{\sqrt{\bar{\alpha}_t}}$

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

$\phi \leftarrow \phi - l\nabla_\phi \mathcal{L}_{\phi,\tilde{\boldsymbol{x}}_0}^{total}$

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

end

return $\boldsymbol{x}_0$

Limitation of setting as a fixed hyperparameter:

- The proper value varies from different modes of degradation, data, and diffusion steps;
- Larger values will result in the emergence of ore textures, whereas smaller values may lead to a loss of details.

Degraded Image Model Output Ground True

- **Fixed guidance scale value 80000 (Larger)**

Degraded Image Model Output Ground True

- **Fixed guidance scale value 80000 (Smaller)**

[1] Fei B, Lyu Z, Pan L, et al. Generative diffusion prior for unified image restoration and enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 9935-9946.

➢ Motivation-Setting Limitation of Degradation Modes

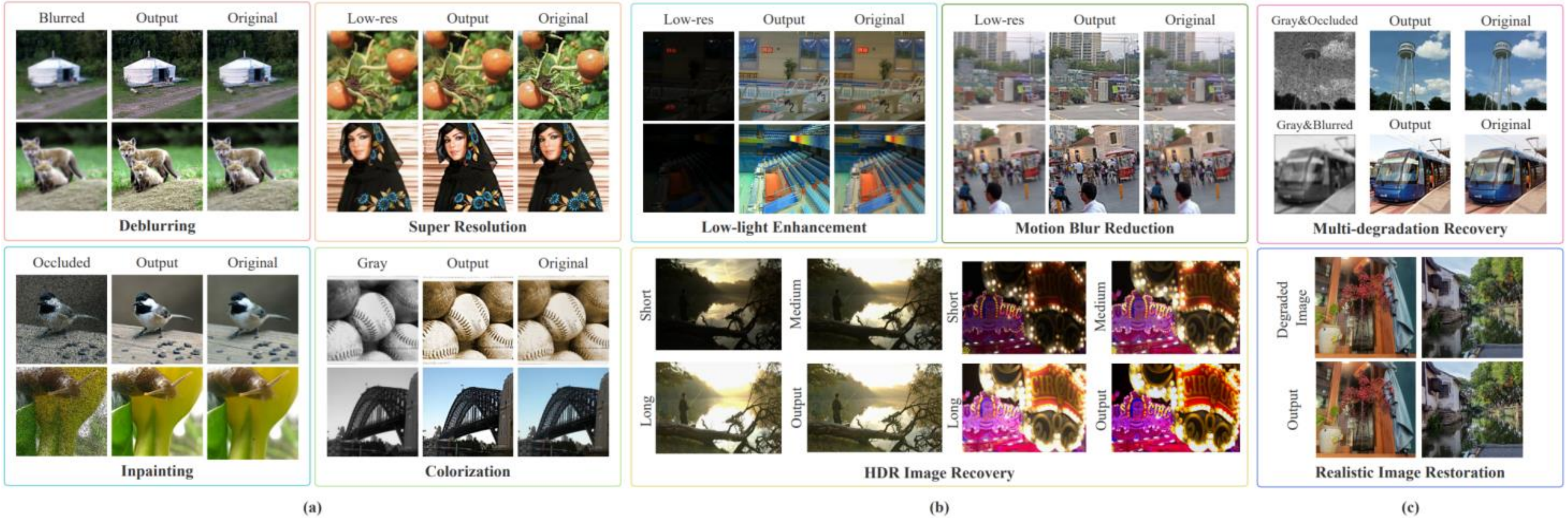
NEURAL INFORMATION PROCESSING SYSTEMS

- For linear reverse problems:
  - Can only address problems where the degenerate function is known, yet lacks the capability to simulate and construct it;

    For instance, in deblurring tasks, the blur kernel is required to be given rather than estimating it.

- For blind problems:
  - GDP tackles blind image restoration by only one parameter f, which limits its application in more complex degradation tasks.
  - When dealing with tasks involving mixed degradation, it can only resolve two types of degradation mixtures in a predetermined sequence.

BIR-D: Blind Image Restoration-Diffusion Model



➢ Empirical Formula of Guidance Scale

- A higher probability of $p_\theta\,(y|x_t)$ indicates that x_t is more consistent with the degraded image y;

- Performing Taylor expansion around x = μ and take the first two terms:

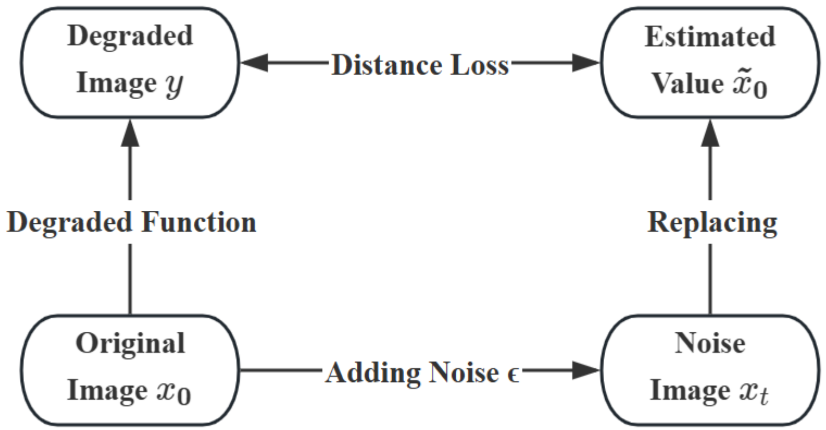
$$\log p_\theta\,(y|x_t) \;\approx \log p_\theta\,(y|x_t)|_{\,x_t=\mu} + (x_t-\mu)^T \nabla_{x_t}\log p_\theta\,(y|x_t)|_{x_t=\mu} = (x_t-\mu)^T \cdot g + C_1$$

- By combining the heuristic approximation formula: $\log p(y|x_{t-1}) = -\log p_\theta\,(y|x_t) \;- sL(D(\,x_t), y)$,
  it can be obtained that: （Taking K to replace $\log p_\theta\,(y|x_t)$ ）

$$(x_t-\mu)^T \cdot g + C_1 = -\log K - s\,L(D(x_t), y)\ ;$$

- Simplifying it, we can get： $s = f(x_t, y, D) = -\frac{(x_t-\mu)^T \cdot g + C_1 + \log K}{L(D(x_t),y)}$;
- Eliminating the noise introduced from x_t within the equation:

$$\tilde{x}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1-\bar{\alpha}_t} \cdot \varepsilon_t)$$



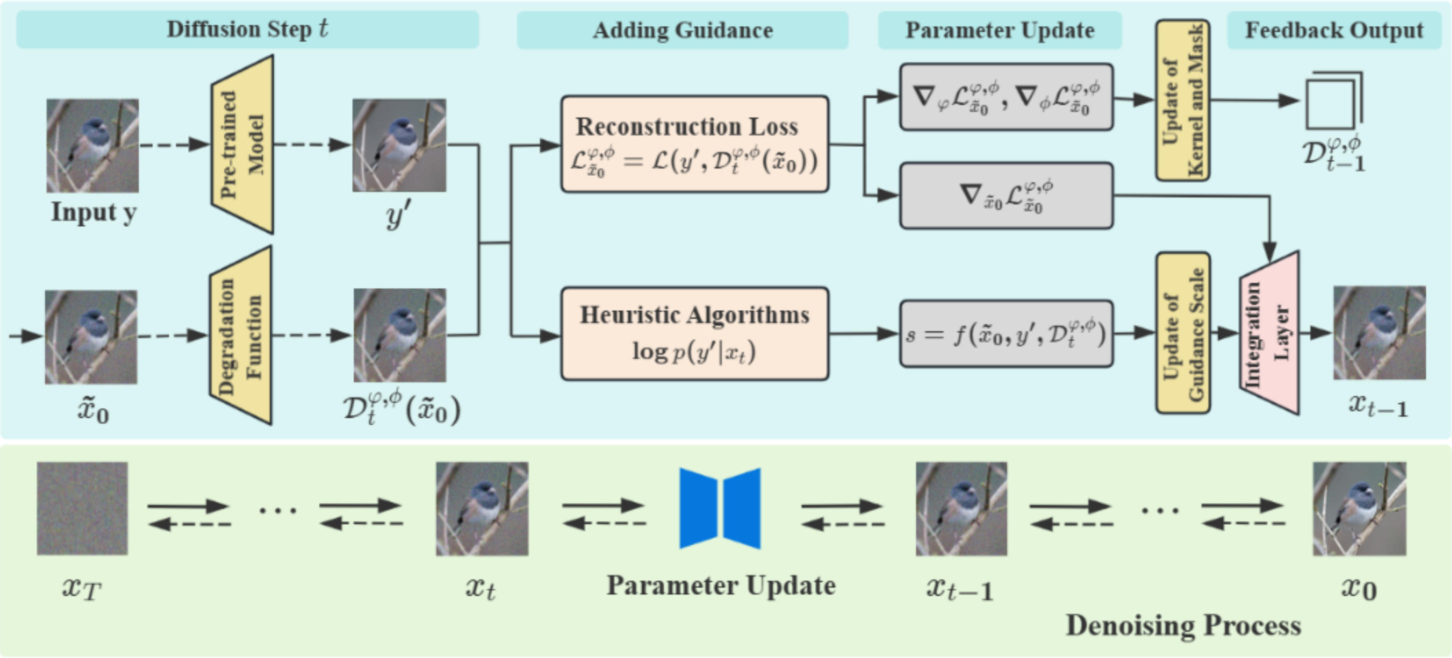
- Finally get the empirical formula of guidance scale $s = f(\tilde{x}_0, y, D) = -\frac{(x_t-\mu)^T \cdot g + C_1 + \log K}{L(D(\tilde{x}_0),y)}$;

- Guidance scale is adaptively updated at every reverse steps.

➢ Optimizable Convolutional Kernel and Reconstruction Loss

NEURAL INFORMATION PROCESSING SYSTEMS

- Utilizing optimizable convolutional kernel D_t^{φ} to simulate and construct the degradation function, the parameter φ is updated at every reverse steps.
- Reconstruction loss L is set with the purpose of measuring the distance between the $\tilde{x}_0$ after the degradation and degraded image y.
    - The gradient of the distance metric L with respect to $\tilde{x}_0$ is employed for sampling x_{t-1};
    - The gradient of the distance metric L with respect to parameter φ is employed for updating the convolutional kernel.



- **Flow-chart of BIR-D**

➢ Algorithm

Algorithm 1: Unconditional diffusion model with the guidance of degraded image y, given a diffusion model noise prediction function $\epsilon_\theta(x_t, t)$.

Input: Degraded image y, degradation function $\mathcal{D}$ composed of optimized convolutional kernels $\mathcal{K}$ with parameters φ and mask $\mathcal{M}$ with parameters ϕ, learning rate l, distant measure $\mathcal{L}$.
Output: Output image x_0 conditioned on y.

- Incorporate guidance from the degraded image y into μ and ∑ from the unconditional DDPM based on the distance metric L;
- The update of guidance scale is followed by the updates of μ and ∑ ;
- Update the parameter of convolutional kernel after sampling x_{t-1} to more closely align model output with the degraded image y.

Sample x_T from $\mathcal{N}(0, I)$
for t from T to 1 **do**

$$\tilde{x}_0 = \frac{x_t}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t,t)}{\sqrt{\bar{\alpha}_t}}$$

Estimate of Real-time Output

$$\mathcal{L}_{\varphi,\phi,\tilde{x}_0} = \mathcal{L}(y, \mathcal{D}^{\varphi,\phi}(\tilde{x}_0))$$

Distance Measure

$$s = -\frac{(x_t-\mu)^T g + C + \log N}{\mathcal{L}(\mathcal{D}^{\varphi,\phi}(\tilde{x}_0), y)}$$

Update of Guidance Scale

$$\tilde{x}_0 \leftarrow \tilde{x}_0 - \frac{s(1-\bar{\alpha}_t)}{\sqrt{\bar{\alpha}_{t-1}}\beta_t} \nabla_{\tilde{x}_0} \mathcal{L}_{\varphi,\phi,\tilde{x}_0}$$

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

$$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$$

Sample x_{t-1} from $\mathcal{N}(\tilde{\mu}_t, \tilde{\beta}_t I)$

$$\varphi \leftarrow \varphi - l\nabla_\varphi \mathcal{L}_{\varphi,\phi,\tilde{x}_0}$$

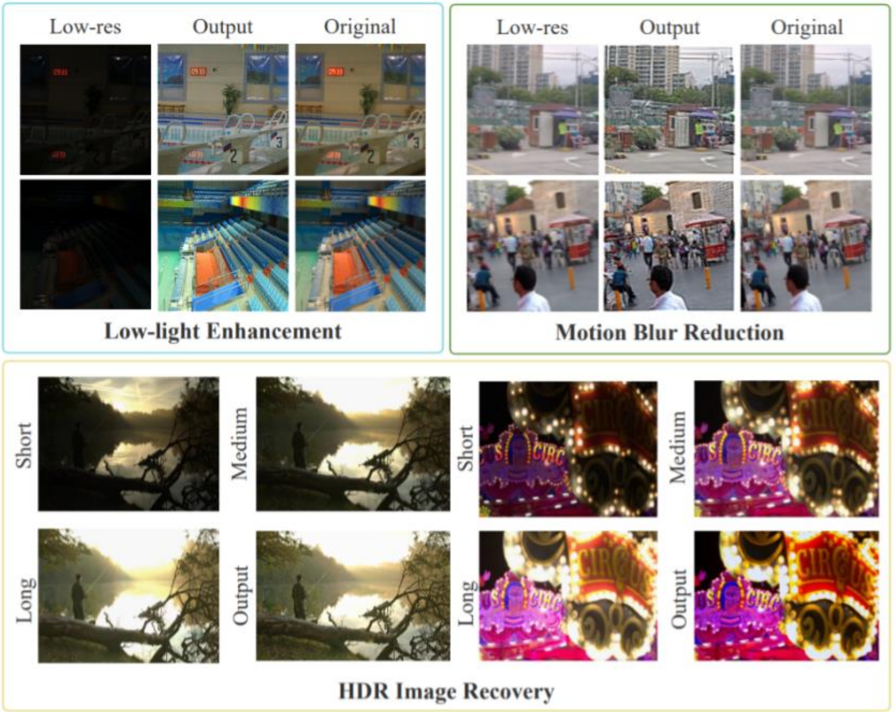
$$\phi \leftarrow \phi - l\nabla_\phi \mathcal{L}_{\varphi,\phi,\tilde{x}_0}$$

Parameter Update

return x_0

Results

The results of blind face restoration task on LFW and WIDER dataset.



Task Input, Output, and Ground Truth Comparison for Blind Image Restoration.

➢ Ablation Study

- To verify the effectiveness of optimizable convolutional kernel and adaptive guidance scale through metrics comparison on LOL and LoLi-Phone datasets.
- BIR-D outperformed other models in all indicators,

Methods	Dynamic Update		LOL					LoLi-Phone	
	Kernel	Guidance Scale	PSNR	SSIM	LOE	FID	PI	LOE	PI
Model A	✗	✗	8.96	0.46	210.88	113.36	8.24	110.05	8.36
Model B	✗	✓	9.58	0.48	203.83	102.47	7.90	102.55	8.25
Model C	✓	✗	14.35	0.54	113.56	82.14	5.23	75.34	7.94
BIR-D	✓	✓	**14.52**	**0.56**	**105.42**	**68.98**	**4.87**	**72.83**	**6.12**

Ablation Study

- The dynamic adjustment of optimizable parameters within the convolutional kernel equips the model with the capability to construct and simulate unknown degradation functions;
- The fixed convolutional kernel parameters prevents the model from providing accurate “guidance” during the diffusion steps.
- The adaptive guidance scale not only enhances the practicality of the model but also ensures greater stability in its outputs, thereby preventing the blurring of certain image results that may arise from biased guidance scale.

Degraded Image

Output of Fixed Kernel

BIR-D Output

Degraded Image

Output of Fixed Guidance Scale

BIR-D Output

Thanks For Listening

Siwei Tu[1], Weidong Yang[1,†], Ben Fei[2,†]

[1] Fudan University
[2] Chinese University of Hong Kong
[†] Corresponding author