Residual Denoising Diffusion Models

CVPR 2024

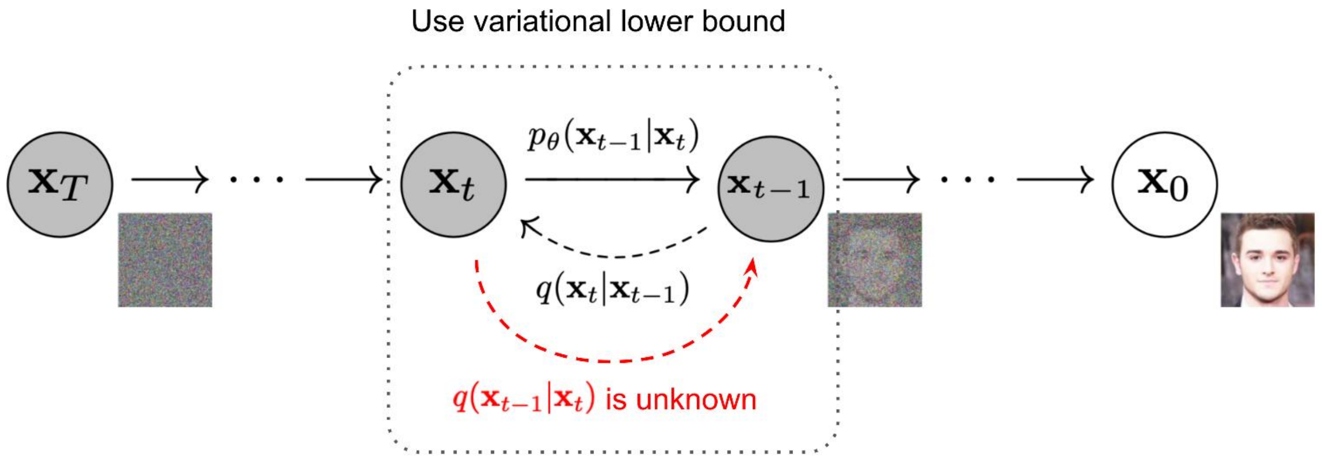
Presenter: Haofeng Huang

2024.09.22

Contents

- Authorship
- Background
- Method & Experiments
- Conclusion

Background: Denoising Diffusion Probabilistic Models



Algorithm 1 Training

repeat
  $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
  $t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  Take gradient descent step on
  $\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$
until converged

Algorithm 2 Sampling

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = T, \ldots, 1$ **do**
  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
  $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
end for
return $\mathbf{x}_0$

Ho et al. Denoising Diffusion Probabilistic Models. NIPS 2020.

Background: Denoising Diffusion Implicit Models

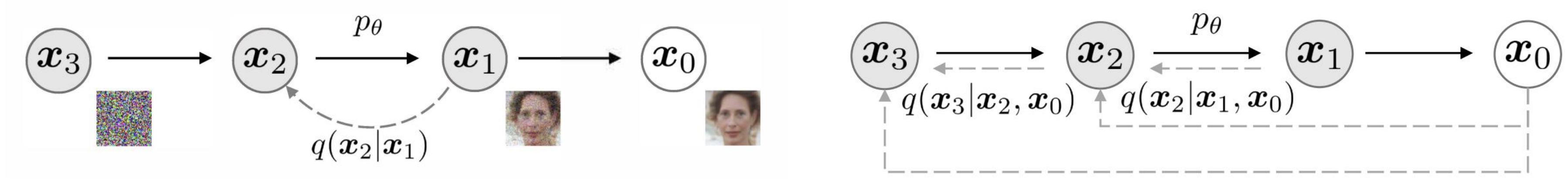


Figure 1: Graphical models for diffusion (left) and non-Markovian (right) inference models.

$$x_{t-1} = \sqrt{\alpha_{t-1}} \underbrace{\left(\frac{x_t - \sqrt{1-\alpha_t}\epsilon_\theta^{(t)}(x_t)}{\sqrt{\alpha_t}} \right)}_{\text{" predicted } x_0 \text{"}} + \underbrace{\sqrt{1-\alpha_{t-1}-\sigma_t^2} \cdot \epsilon_\theta^{(t)}(x_t)}_{\text{"direction pointing to } x_t \text{"}} + \underbrace{\sigma_t \epsilon_t}_{\text{random noise}}$$

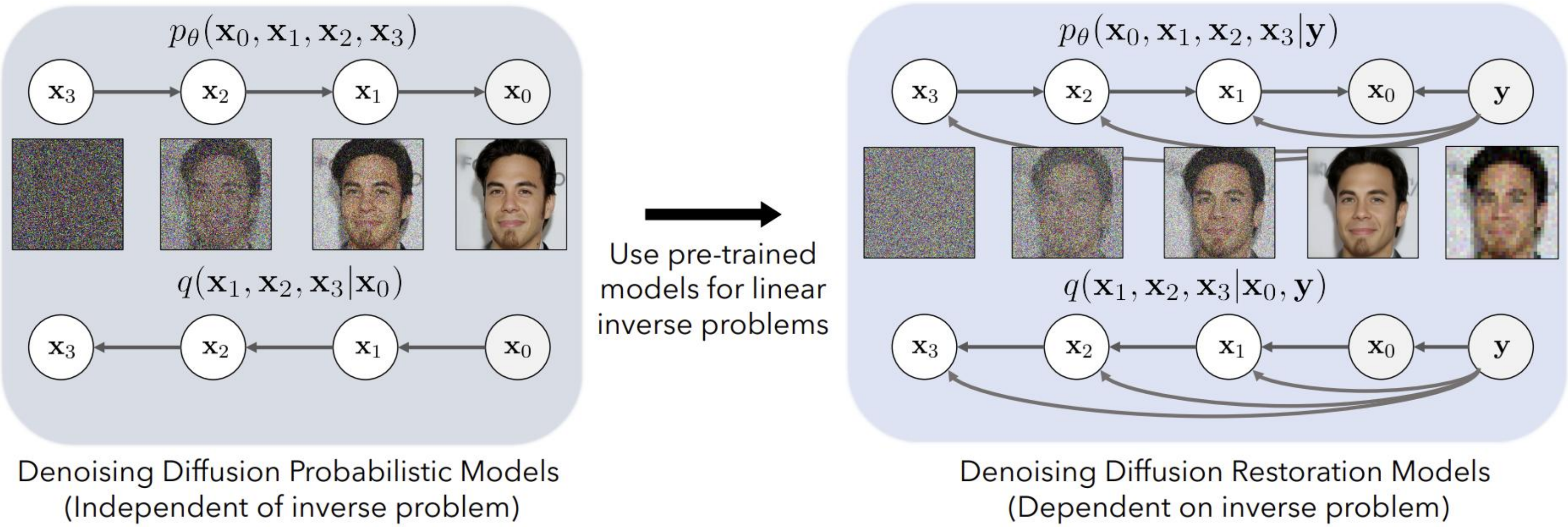
	S	CIFAR10 (32 × 32) 10	20	50	100	1000	CelebA (64 × 64) 10	20	50	100	1000
η	0.0	**13.36**	**6.84**	**4.67**	**4.16**	4.04	**17.33**	**13.73**	**9.17**	**6.53**	3.51
	0.2	14.04	7.11	4.77	4.25	4.09	17.66	14.11	9.51	6.79	3.64
	0.5	16.66	8.35	5.25	4.46	4.29	19.86	16.06	11.01	8.09	4.28
	1.0	41.07	18.36	8.01	5.78	4.73	33.12	26.03	18.48	13.93	5.98
$\hat{\sigma}$		367.43	133.37	32.72	9.99	**3.17**	299.71	183.83	71.71	45.20	**3.26**

Song et al. Denoising Diffusion Implicit Models. ICLR 2021.

Background: Denoising Diffusion Restoration Models

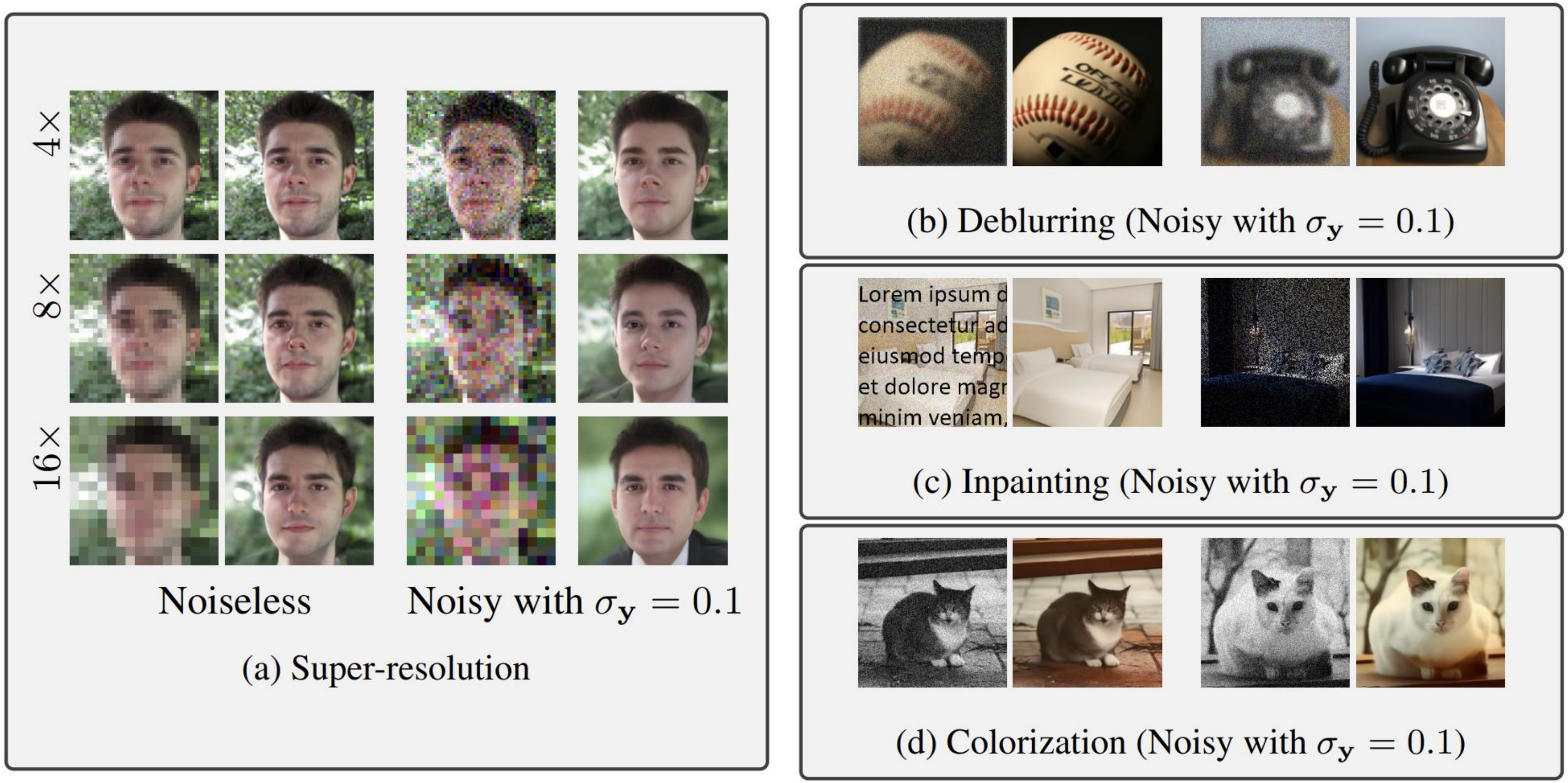
Linear Inverse Problems. A general linear inverse problem is posed as

$$\mathbf{y} = \boldsymbol{H}\mathbf{x} + \mathbf{z}, \quad (1)$$



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

Background: Denoising Diffusion Restoration Models



(a) Super-resolution

(b) Deblurring (Noisy with $\sigma_{\mathbf{y}} = 0.1$)

(c) Inpainting (Noisy with $\sigma_{\mathbf{y}} = 0.1$)

(d) Colorization (Noisy with $\sigma_{\mathbf{y}} = 0.1$)

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

Background: Denoising Diffusion Null-space Models

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2\sigma^2}||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2 + \lambda\mathcal{R}(\mathbf{x}).$$

$$\mathbf{x} \equiv \mathbf{A}^{\dagger}\mathbf{A}\mathbf{x} + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\mathbf{x}.$$

Algorithm 1 Sampling of DDNM

```
x_T ~ N(0, I)
for t = T, ..., 1 do
    x_{0|t} = (1/sqrt(ᾱ_t)) (x_t − Z_θ(x_t, t) sqrt(1 − ᾱ_t))
    x̂_{0|t} = A† y + (I − A† A) x_{0|t}
    x_{t−1} ~ p(x_{t−1} | x_t, x̂_{0|t})
return x_0
```

- $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- **for** $t = T, \ldots, 1$ **do**
  - $\mathbf{x}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathbf{x}_t - \mathcal{Z}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)\sqrt{1-\bar{\alpha}_t}\right)$
  - $\hat{\mathbf{x}}_{0|t} = \mathbf{A}^{\dagger}\mathbf{y} + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\mathbf{x}_{0|t}$
  - $\mathbf{x}_{t-1} \sim p(\mathbf{x}_{t-1}|\mathbf{x}_t, \hat{\mathbf{x}}_{0|t})$
- **return** $\mathbf{x}_0$

Algorithm 2 Sampling of DDNM^{+}

- $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- **for** $t = T, \ldots, 1$ **do**
  - $L = \min\{T - t, l\}$
  - $\mathbf{x}_{t+L} \sim q(\mathbf{x}_{t+L}|\mathbf{x}_t)$
  - **for** $j = L, \ldots, 0$ **do**
    - $\mathbf{x}_{0|t+j} = \frac{1}{\sqrt{\bar{\alpha}_{t+j}}}\left(\mathbf{x}_{t+j} - \mathcal{Z}_{\boldsymbol{\theta}}(\mathbf{x}_{t+j}, t+j)\sqrt{1-\bar{\alpha}_{t+j}}\right)$
    - $\hat{\mathbf{x}}_{0|t+j} = \mathbf{x}_{0|t+j} - \boldsymbol{\Sigma}_{t+j}\mathbf{A}^{\dagger}(\mathbf{A}\mathbf{x}_{0|t+j} - \mathbf{y})$
    - $\mathbf{x}_{t+j-1} \sim \hat{p}(\mathbf{x}_{t+j-1}|\mathbf{x}_{t+j}, \hat{\mathbf{x}}_{0|t+j})$
- **return** $\mathbf{x}_0$

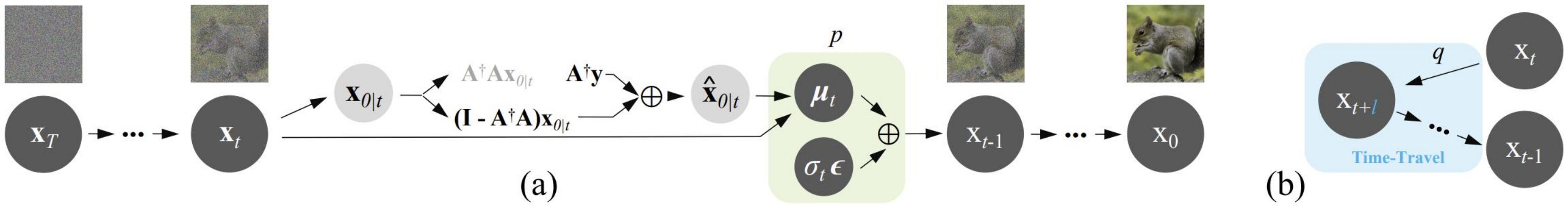


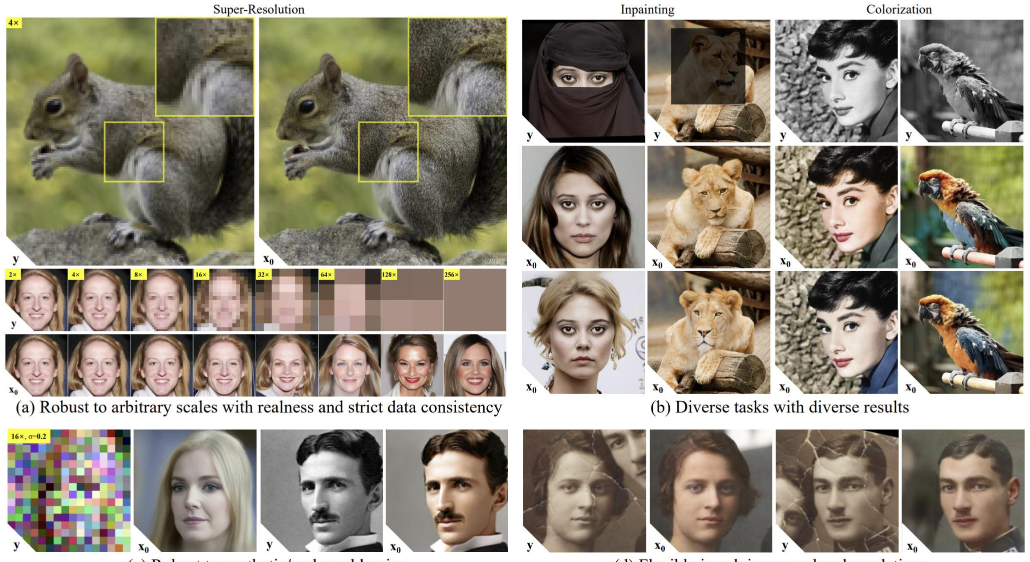
Figure 2: Illustration of (a) DDNM and (b) the time-travel trick.

Wang et al. Zero-shot Image Restoration Using Denoising Diffusion Null-space Model. ICLR 2022.

Background: Denoising Diffusion Null-space Models

ImageNet	4× SR	Deblurring	Colorization	CS 25%	Inpainting
Method	PSNR↑/SSIM↑/FID↓	PSNR↑/SSIM↑/FID↓	*Cons*↓/FID↓	PSNR↑/SSIM↑/FID↓	PSNR↑/SSIM↑/FID↓
$\mathbf{A}^{\dagger}\mathbf{y}$	24.26 / 0.684 / 134.4	18.56 / 0.6616 / 55.42	0.0 / 43.37	15.65 / 0.510 / 277.4	14.52 / 0.799 / 72.71
DGP	23.18 / 0.798 / 64.34	N/A	- / 69.54	N/A	N/A
ILVR	27.40 / **0.870** / 43.66	N/A	N/A	N/A	N/A
RePaint	N/A	N/A	N/A	N/A	31.87 / **0.968** / 12.31
DDRM	27.38 / 0.869 / 43.15	43.01 / 0.992 / 1.48	260.4 / 36.56	19.95 / 0.704 / 97.99	31.73 / 0.966 / 4.82
DDNM(ours)	**27.46 / 0.870/ 39.26**	**44.93 / 0.994 / 1.15**	**42.32 / 36.32**	**21.66 / 0.749 / 64.68**	**32.06 / 0.968 / 3.89**

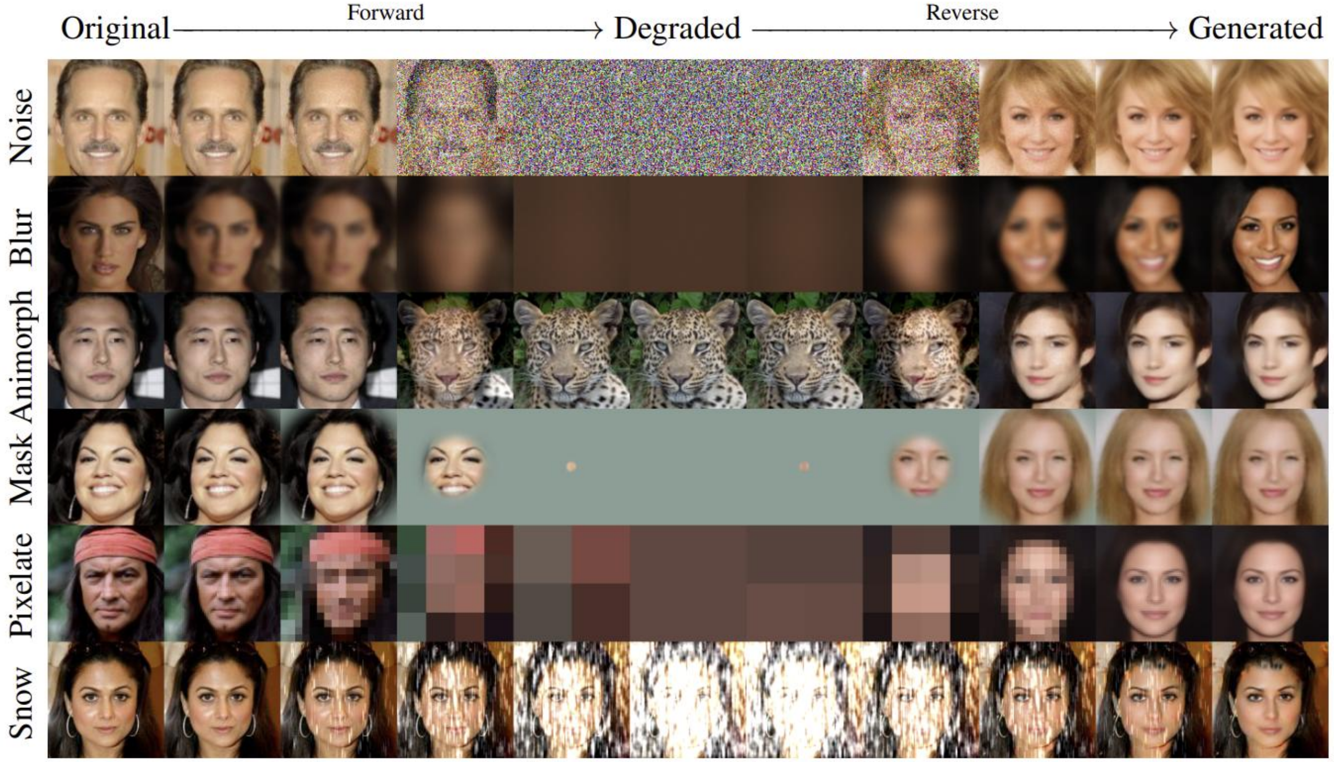
CelebA	4× SR	Deblurring	Colorization	CS 25%	Inpainting
Method	PSNR↑/SSIM↑/FID↓	PSNR↑/SSIM↑/FID↓	*Cons*↓/FID↓	PSNR↑/SSIM↑/FID↓	PSNR↑/SSIM↑/FID↓
$\mathbf{A}^{\dagger}\mathbf{y}$	27.27 / 0.782 / 103.3	18.85 / 0.741 / 54.31	0.0 / 68.81	15.09 / 0.583 / 377.7	15.57 / 0.809 / 181.56
PULSE	22.74 / 0.623 / 40.33	N/A	N/A	N/A	N/A
ILVR	31.59 / **0.945** / 29.82	N/A	N/A	N/A	N/A
RePaint	N/A	N/A	N/A	N/A	35.20 / 0.981 /14.19
DDRM	**31.63 / 0.945** / 31.04	43.07 / 0.993 / 6.24	455.9 / 31.26	24.86 / 0.876 / 46.77	34.79 / 0.978 /12.53
DDNM(ours)	**31.63 / 0.945 / 22.27**	**46.72 / 0.996 / 1.41**	**26.25 / 26.44**	**27.56 / 0.909 / 28.80**	**35.64 / 0.982 /4.54**



(a) Robust to arbitrary scales with realness and strict data consistency
(b) Diverse tasks with diverse results
(c) Robust to synthetic/real-world noise
(d) Flexible in solving complex degradations

Wang et al. Zero-shot Image Restoration Using Denoising Diffusion Null-space Model. ICLR 2022.

Background: Cold Diffusion



Algorithm 1 Naive Sampling (Eg. DDIM)

Input: A degraded sample x_t
for $s = t, t-1, \ldots, 1$ **do**
$\hat{x}_0 \leftarrow R(x_s, s)$
$x_{s-1} = D(\hat{x}_0, s-1)$
end for
Return: x_0

Algorithm 2 Transformation Agnostic Cold Sampling (TACoS)

Input: A degraded sample x_t
for $s = t, t-1, \ldots, 1$ **do**
$\hat{x}_0 \leftarrow R(x_s, s)$
$x_{s-1} = x_s - D(\hat{x}_0, s) + D(\hat{x}_0, s-1)$
end for

Bansal et al. Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise. NIPS 2023.

Background: Summary

北京大學 PEKING UNIVERSITY

- Generative Diffusion Models
  - DDPM
  - DDIM

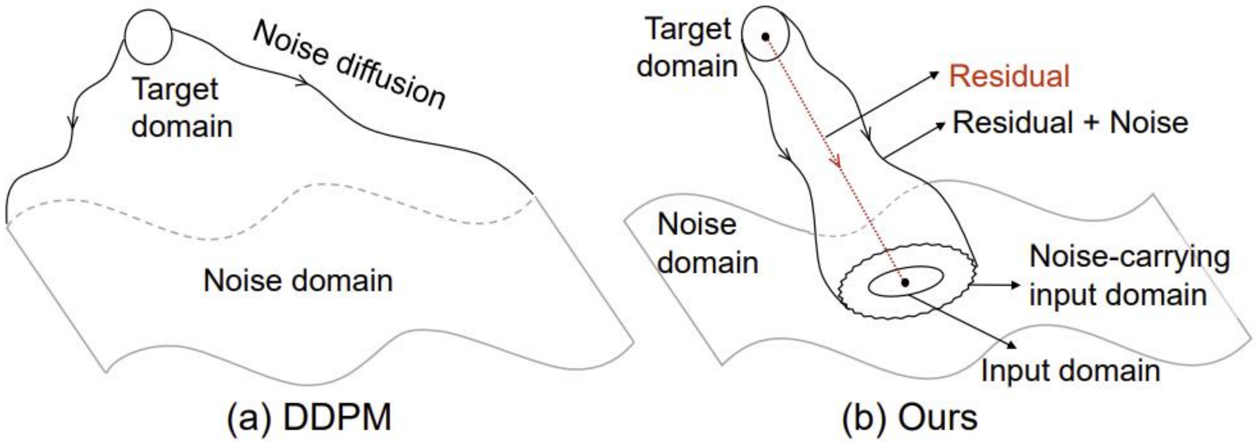
    Cons: only consider mapping between pure noise and natural image
- Restoration Diffusion Models
  - DDRM
  - DDNM

    Cons: only use the degraded image as the condition for generation, non-interpretability of forward process
  - Cold Diffusion

    Cons: lack of generality and theoretical justification

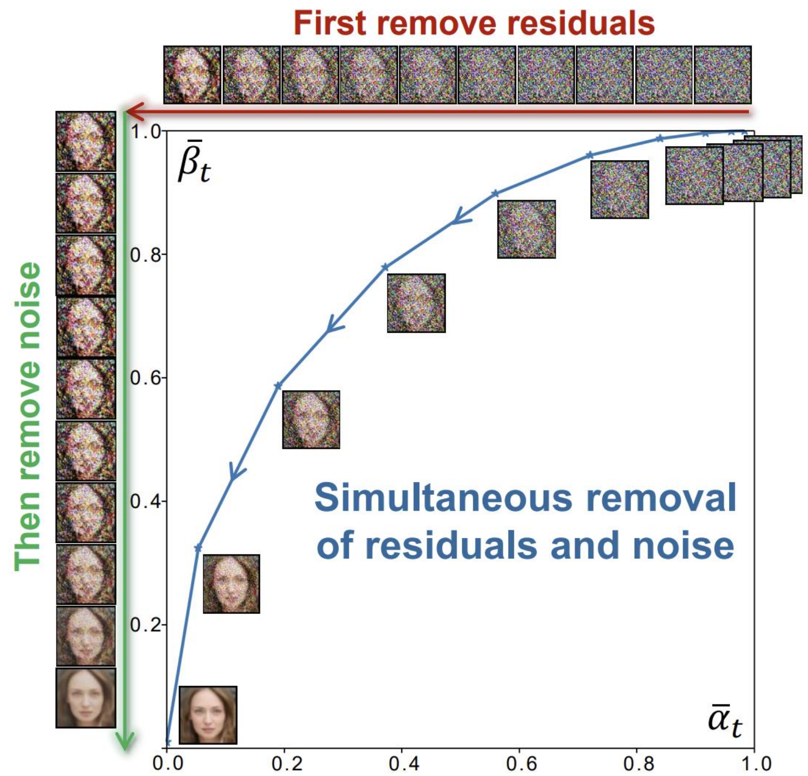
Method: Overview

- A novel dual diffusion process - Residual Denoising Diffusion Models, which decouples the previous diffusion process into *residual diffusion* and *noise diffusion*.
- *Residual diffusion* prioritizes certainty and represents a directional diffusion from the target image to the conditional input image, and *noise diffusion* emphasizes diversity and represents random perturbations in the diffusion process.



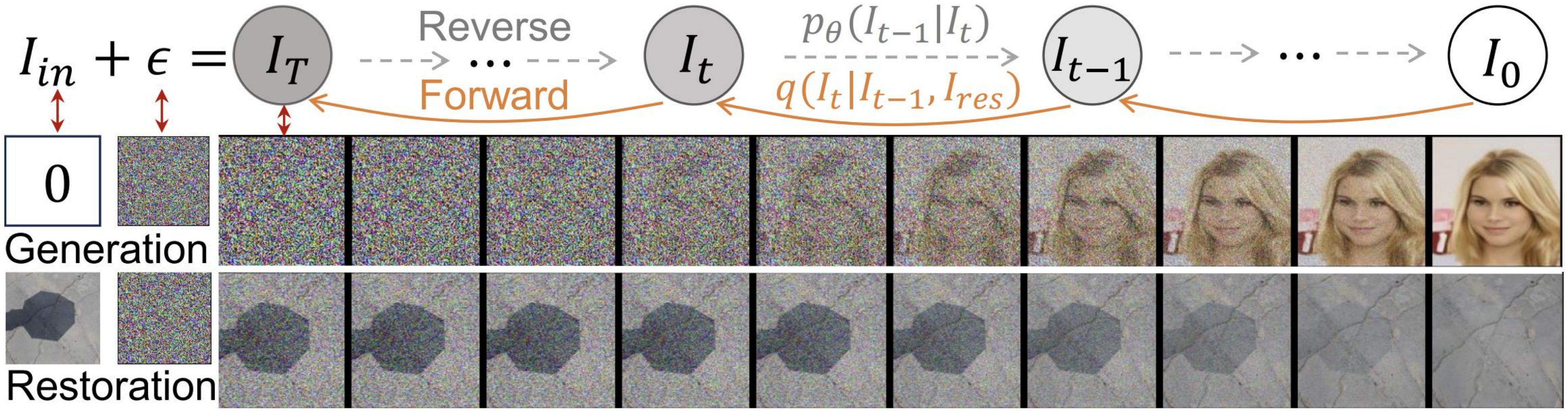
Method: Decomposition of diffusion step

- Unlike the previous denoising diffusion model, which uses one coefficient schedule to control the mixing ratio of noise and images, RDDM employs two independent coefficient schedules to control the diffusion speed of residuals and noise.



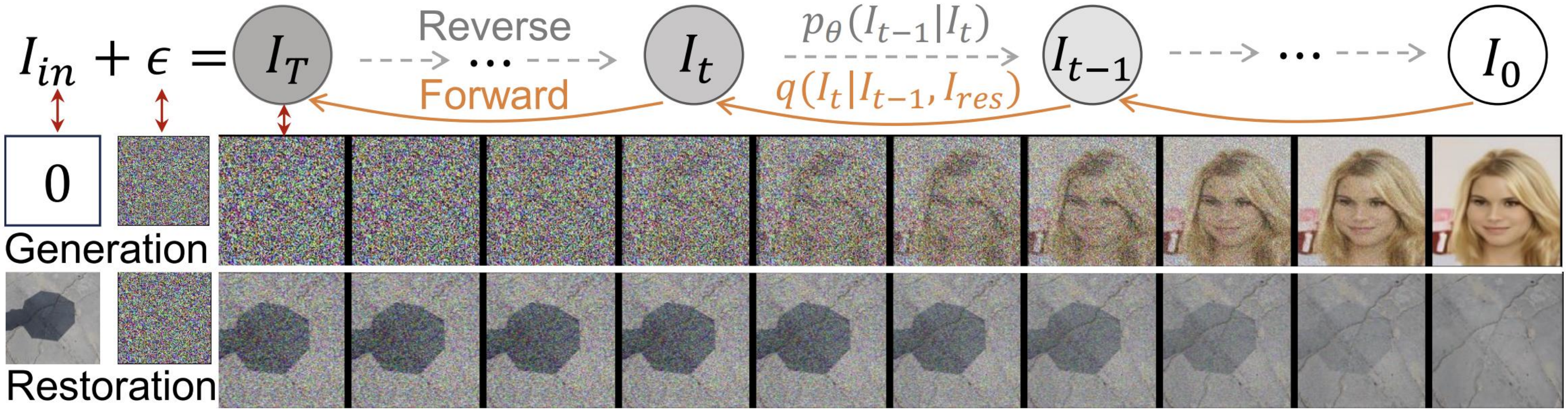
Method: Introduce Residual I_res

- $I_{res} = I_0 - I_{in}$



- $I_{in} = 0$ for generation, $I_{in} = I_{deg}$ for restoration.

Method: Introduce Residual I_res

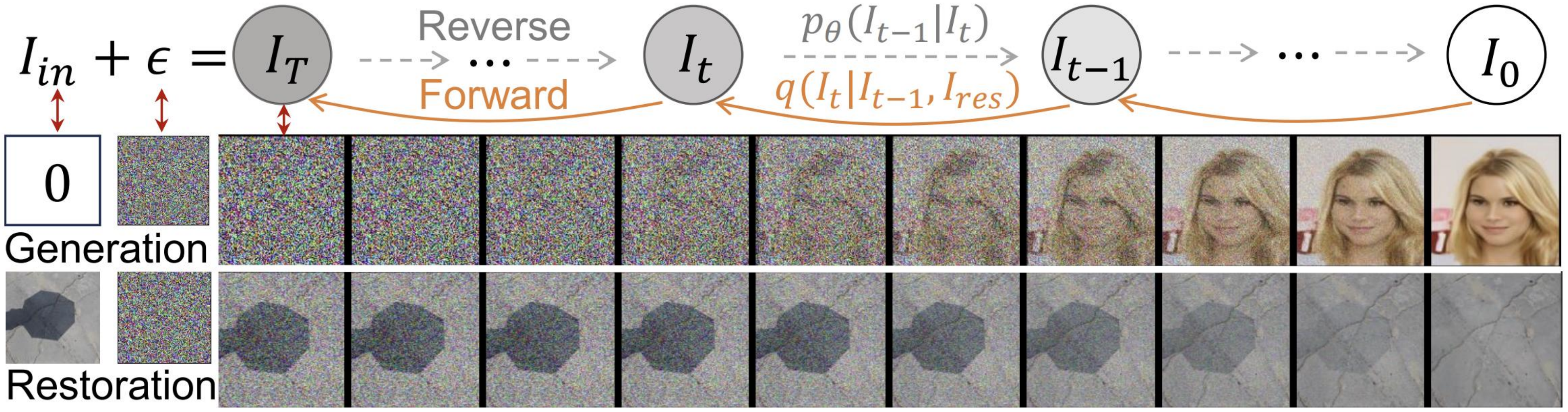


- Forward:

$$I_t = I_{t-1} + I_{res}^t, \qquad I_{res}^t \sim \mathcal{N}(\alpha_t I_{res}, \beta_t^2 \mathbf{I}),$$

$$\begin{aligned} I_t =& I_{t-1} + \alpha_t I_{res} + \beta_t \epsilon_{t-1}, \\ =& I_{t-2} + (\alpha_{t-1} + \alpha_t) I_{res} + (\sqrt{\beta_{t-1}^2 + \beta_t^2}) \epsilon_{t-2} \\ =& \dots \\ =& I_0 + \bar{\alpha}_t I_{res} + \bar{\beta}_t \epsilon, \end{aligned}$$

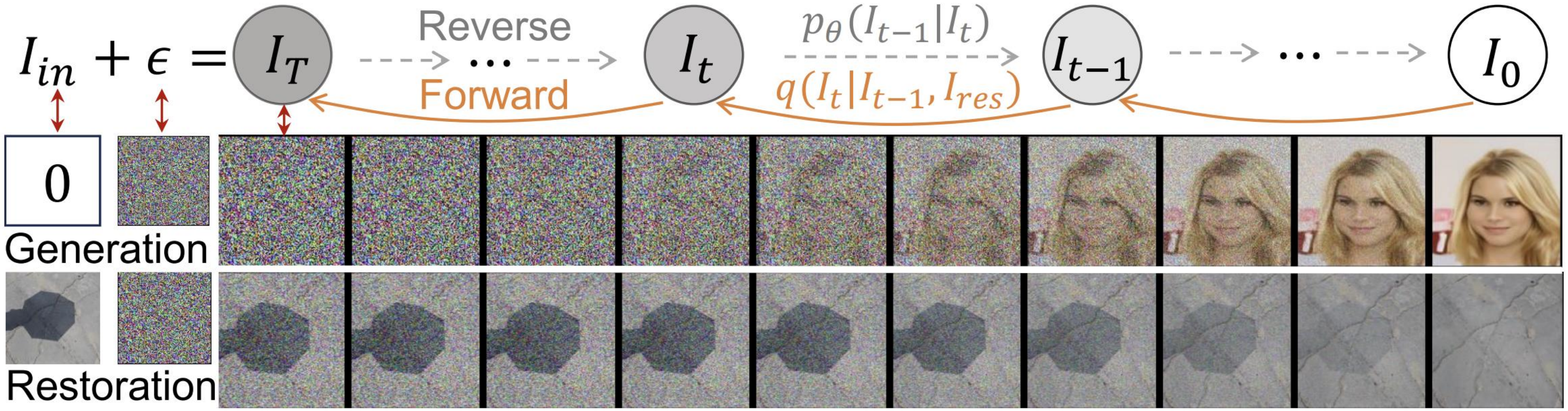
Method: Introduce Residual I_res



- Reverse: a network $I_{res}^{\theta}(I_t, t, I_{in})$ to predict residual, a network $\epsilon_\theta(I_t, t, I_{in})$ to estimate noise, then

$$I_0^\theta = I_t - \bar{\alpha}_t I_{res}^\theta - \bar{\beta}_t \epsilon_\theta.$$

Method: Introduce Residual I_res



- Reverse: therefore

$$I_{t-1} = I_t - (\bar{\alpha}_t - \bar{\alpha}_{t-1}) I_{res}^{\theta} - (\bar{\beta}_t - \sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2}) \epsilon_\theta + \sigma_t \epsilon_t$$

Method: Sampling Method

北京大學 PEKING UNIVERSITY

- Recall that $$I_t = I_0 + \bar{\alpha}_t I_{res} + \bar{\beta}_t \epsilon,$$

replace I_0 with I_{in},

$$I_t = I_{in} + (\bar{\alpha}_t - 1) I_{res} + \bar{\beta}_t \epsilon.$$

- For the generation process, we know I_{in}, I_t. It means that estimated I_{res} and ϵ can represent each other.

Method: Sampling Method

$$I_t = I_{in} + (\bar{\alpha}_t - 1)I_{res} + \bar{\beta}_t\epsilon.$$

- SM-Res: Predict *residual* and represent noise with residual.
- SM-N: Predict *noise* and represent residual with noise.
- SM-Res-N: Predict both *residual* and *noise*.

$$I_{t-1} = I_t - (\bar{\alpha}_t - \bar{\alpha}_{t-1})I^{\theta}_{res} - (\bar{\beta}_t - \sqrt{\bar{\beta}^2_{t-1} - \sigma^2_t})\epsilon_\theta + \sigma_t\epsilon_t$$

Method: Sampling Method

北京大學 PEKING UNIVERSITY

- SM-Res: Predict *residual* and represent noise with residual.
- SM-N: Predict *noise* and represent residual with noise.
- SM-Res-N: Predict both *residual* and *noise*.

Sampling Method	Generation (CelebA)		Shadow removal (ISTD)			Low-light (LOL)		Deraining (RainDrop)	
	FID (↓)	IS (↑)	MAE(↓)	PSNR(↑)	SSIM(↑)	PSNR(↑)	SSIM(↑)	PSNR(↑)	SSIM(↑)
SM-Res	31.47	1.73	4.76	30.72	0.959	**25.39**	**0.937**	31.96	0.9509
SM-N	**23.25**	**2.05**	81.01	11.34	0.175	16.30	0.649	19.15	0.7179
SM-Res-N	28.90	1.78	**4.67**	**30.91**	**0.962**	23.90	0.931	**32.51**	**0.9563**

Table 1. Sampling method analysis. The sampling steps are 10 on the CelebA 64 × 64 [36] dataset, 5 on the ISTD [57] dataset, 2 on the LOL [61] dataset, and 5 on the RainDrop [45] dataset.

- *"residual predictions prioritize certainty, whereas noise predictions emphasize diversity."*

Method: Automatic Objective Selection

- Joint loss function

$$L_{auto}(\theta) := \lambda_{res}^{\theta} E\left[\left\|I_{res} - I_{res}^{\theta}(I_t, t, I_{in})\right\|^2\right] + (1 - \lambda_{res}^{\theta}) E\left[\left\|\epsilon - \epsilon_{\theta}(I_t, t, I_{in})\right\|^2\right].$$

In the loss function, λ_{res}^{θ} is a learnable parameter.

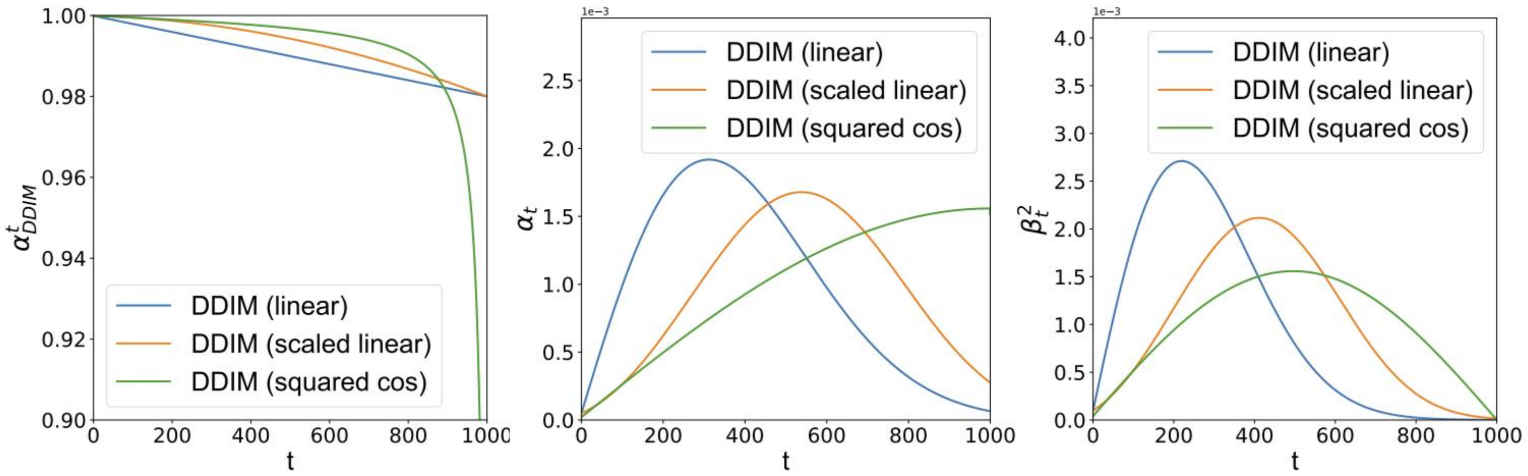
- When $|\lambda_{res}^{\theta} - 0.5|$ surpass a pre-defined threshold, switch the simultaneous training to sole training of $I_{res}^{\theta}(I_t, t, I_{in})$ or $\epsilon_{\theta}(I_t, t, I_{in})$.

Method: Partially Path Independent Generation

- For generation

$$I_{t-1} = I_t - (\bar{\alpha}_t - \bar{\alpha}_{t-1}) I^{\theta}_{res} - (\bar{\beta}_t - \sqrt{\bar{\beta}^2_{t-1} - \sigma^2_t})\epsilon_\theta + \sigma_t \epsilon_t$$

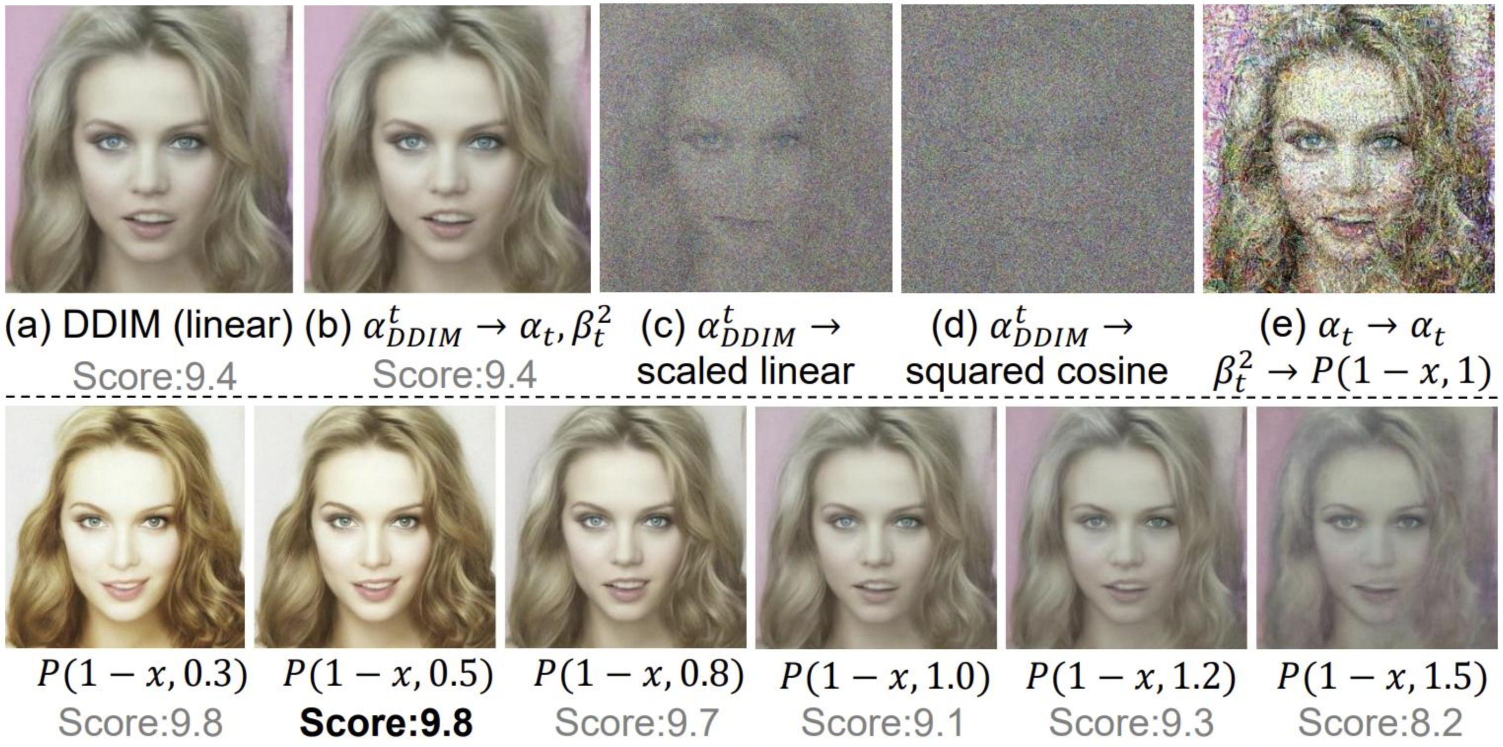
settings of $\bar{\alpha}_t$ and $\bar{\beta}_t$ will affect the results.



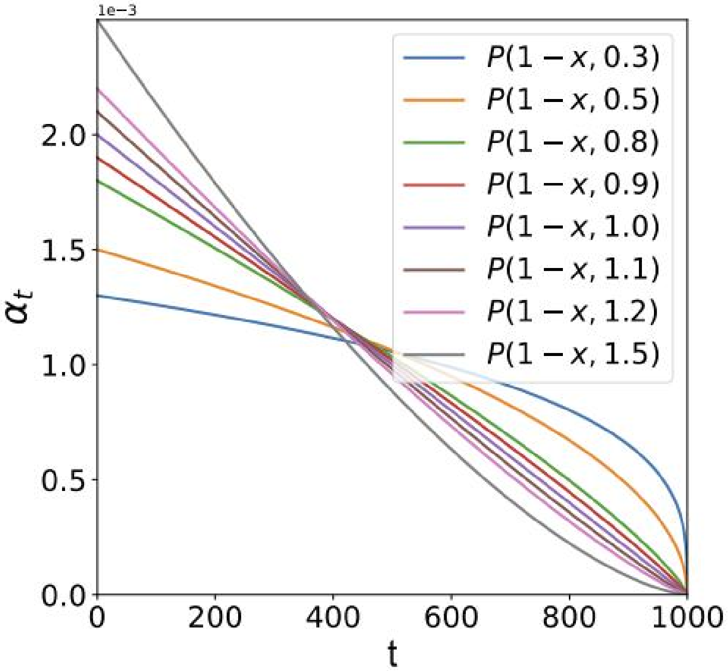
(a) α^t_{DDIM} (b) $\alpha^t_{DDIM} \rightarrow \alpha_t$ (c) $\alpha^t_{DDIM} \rightarrow \beta^2_t$

Method: Partially Path Independent Generation

- Directly changing schedule without retraining will fail.
- Solely readjust the $\bar{\alpha}_t$ may lead to a higher score; readjusting the $\bar{\beta}_t$ will fail.



(f) convert α_{DDIM}^t to α_t, β_t^2 and readjust the converted α_t without touching the β_t^2



(e) $P(x,a)$

Method: Partially Path Independent Generation

- Change the network input:

$$I^{\theta}_{res}(I_t, t, 0) \rightarrow I^{\theta}_{res}(I_t, \bar{\alpha}_t \cdot T, 0),$$
$$\epsilon_\theta(I_t, t, 0) \rightarrow \epsilon_\theta(I_t, \bar{\beta}_t \cdot T, 0).$$

- Compared with original model, if we *assumed* that the network is trained well and robust, there is

$$\frac{\partial I^{\theta}_{res}(I(t), \bar{\alpha}(t)\cdot T)}{\partial \bar{\beta}(t)} \approx 0, \frac{\partial \epsilon_\theta(I(t), \bar{\beta}(t)\cdot T)}{\partial \bar{\alpha}(t)} \approx 0.$$

- It means

$$I_t - I_{t-1} = (\bar{\alpha}_t - \bar{\alpha}_{t-1}) I^{\theta}_{res} + (\bar{\beta}_t - \bar{\beta}_{t-1})\epsilon_\theta,$$

or say

$$\mathrm{d}I(t) = I^{\theta}_{res}(I(t), \bar{\alpha}(t)\cdot T, 0)\mathrm{d}\bar{\alpha}(t) + \epsilon_\theta(I(t), \bar{\beta}(t)\cdot T, 0)\mathrm{d}\bar{\beta}(t),$$

is path independent.

Method: Partially Path Independent Generation

- Experimental evidence:

Experiment: Subjective Results

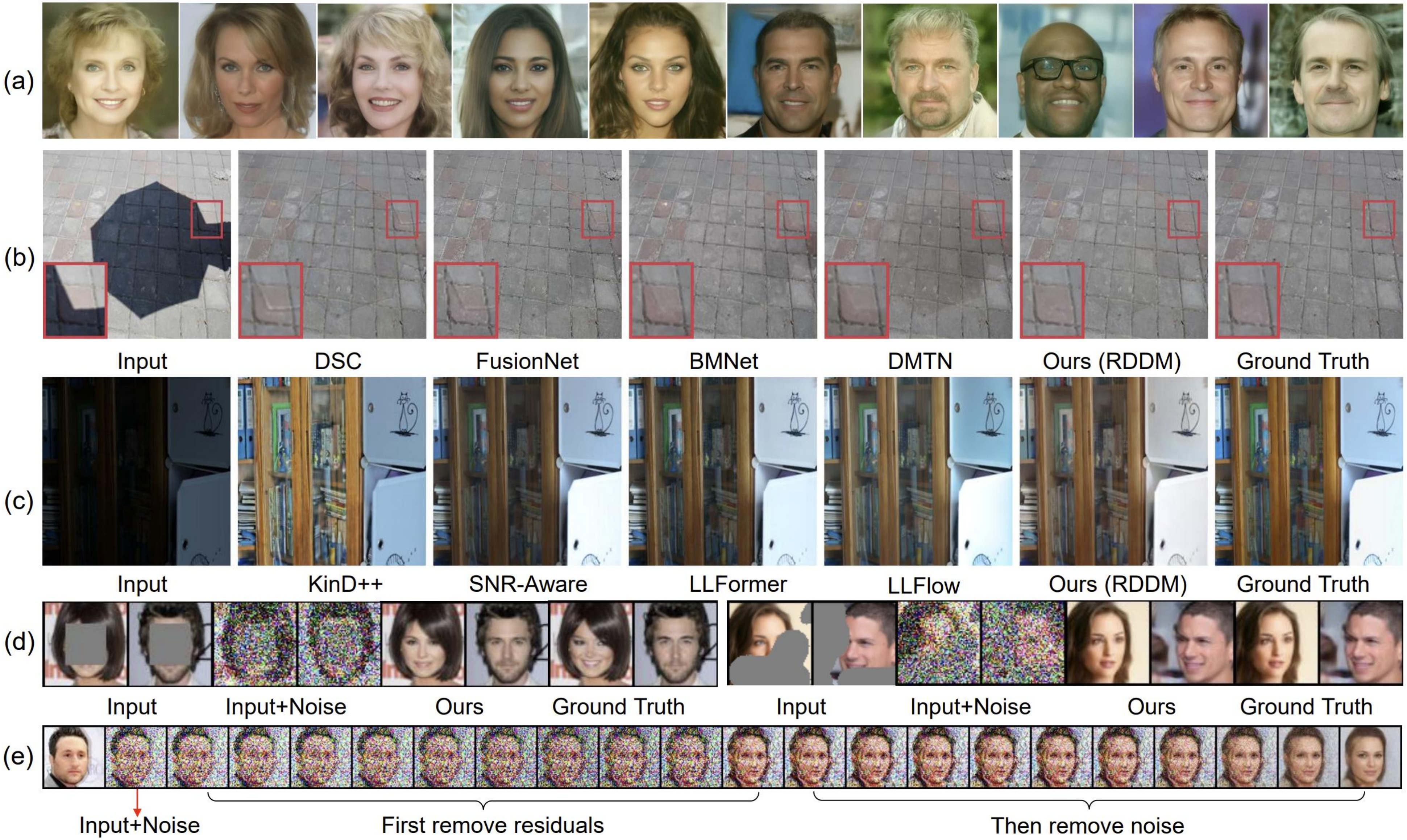
(a) CelebA (FID)	5 steps	10 steps	15 steps	20 steps	100 steps
DDIM	69.60	40.45	32.67	30.61	23.66
DDIM→RDDM	69.60	40.41	32.71	30.77	24.92

(b) Shadow Removal	MAE(↓)			SSIM(↑)			PSNR(↑)		
	S	NS	ALL	S	NS	ALL	S	NS	ALL
DSC [19] ¶	9.48	6.14	6.67	0.967	-	-	33.45	-	-
FusionNet [13]	7.77	5.56	5.92	0.975	0.880	0.945	34.71	28.61	27.19
BMNet [79]	7.60	4.59	5.02	0.988	0.976	0.959	35.61	32.80	30.28
DMTN [31]	7.00	4.28	4.72	**0.990**	**0.979**	**0.965**	35.83	33.01	30.42
Ours (RDDM)	**6.67**	**4.27**	**4.67**	0.988	**0.979**	0.962	**36.74**	**33.18**	**30.91**

(c) Low-light	PSNR(↑)	SSIM(↑)	LPIPS (↓)
KinD++ [76]	17.752	0.760	0.198
KinD++-SKF [68]	20.363	0.805	0.201
DCC-Net [77]	22.72	0.81	-
SNR-Aware [66]	24.608	0.840	0.151
LLFlow [59]	25.19	0.93	**0.11**
LLFormer [58]	23.649	0.816	0.169
Ours (RDDM)	**25.392**	**0.937**	0.116

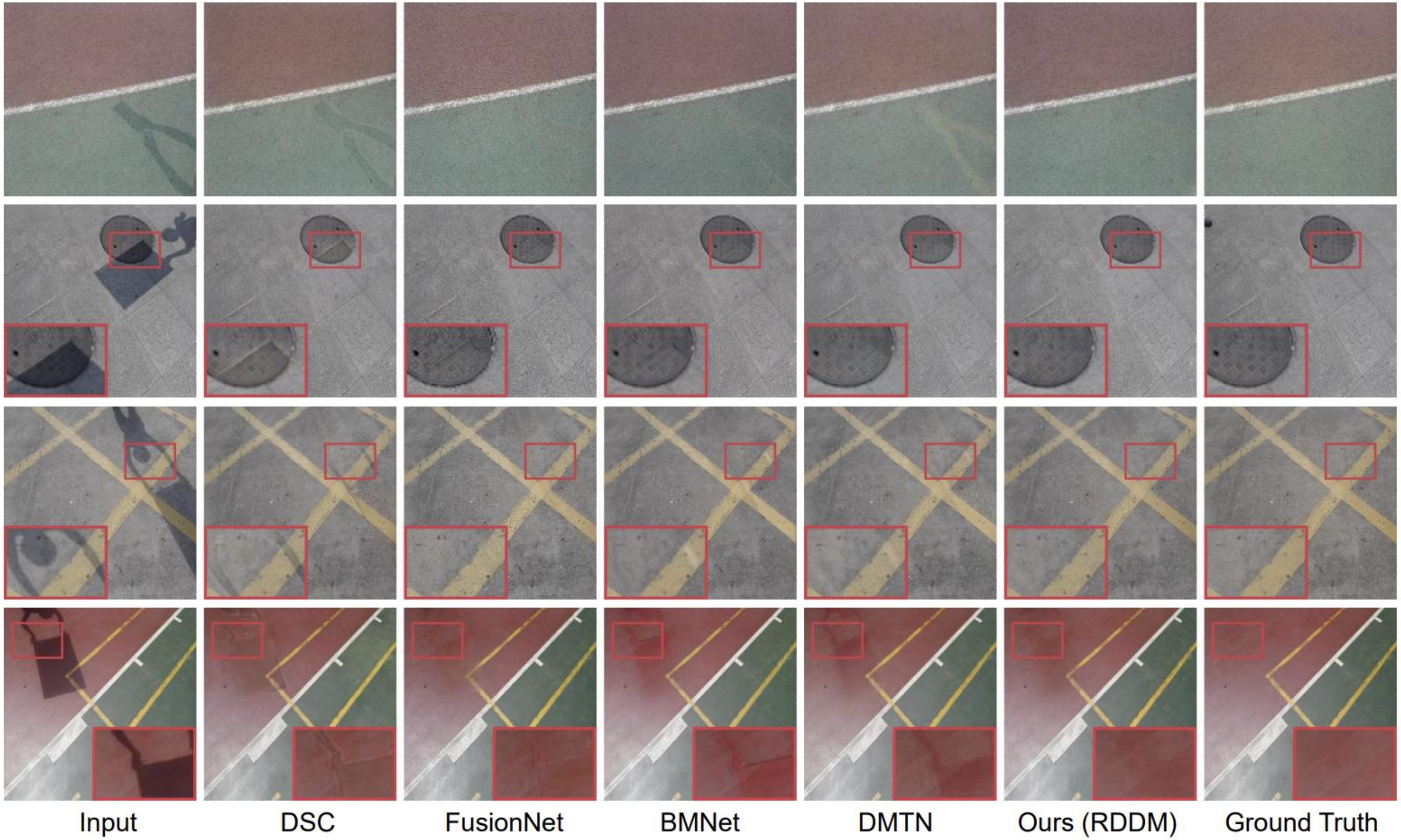
(d) Deraining	PSNR(↑)	SSIM(↑)
AttnGAN [45]	31.59	0.9170
DuRN [34]	31.24	0.9259
RainAttn [46]	31.44	0.9263
IDT [64]	31.87	0.9313
RainDiff64 [82]	32.29	0.9422
RainDiff128 [82]	32.43	0.9334
Ours (RDDM)	**32.51**	**0.9563**

Experiment: Objective Results



Experiment: More Results

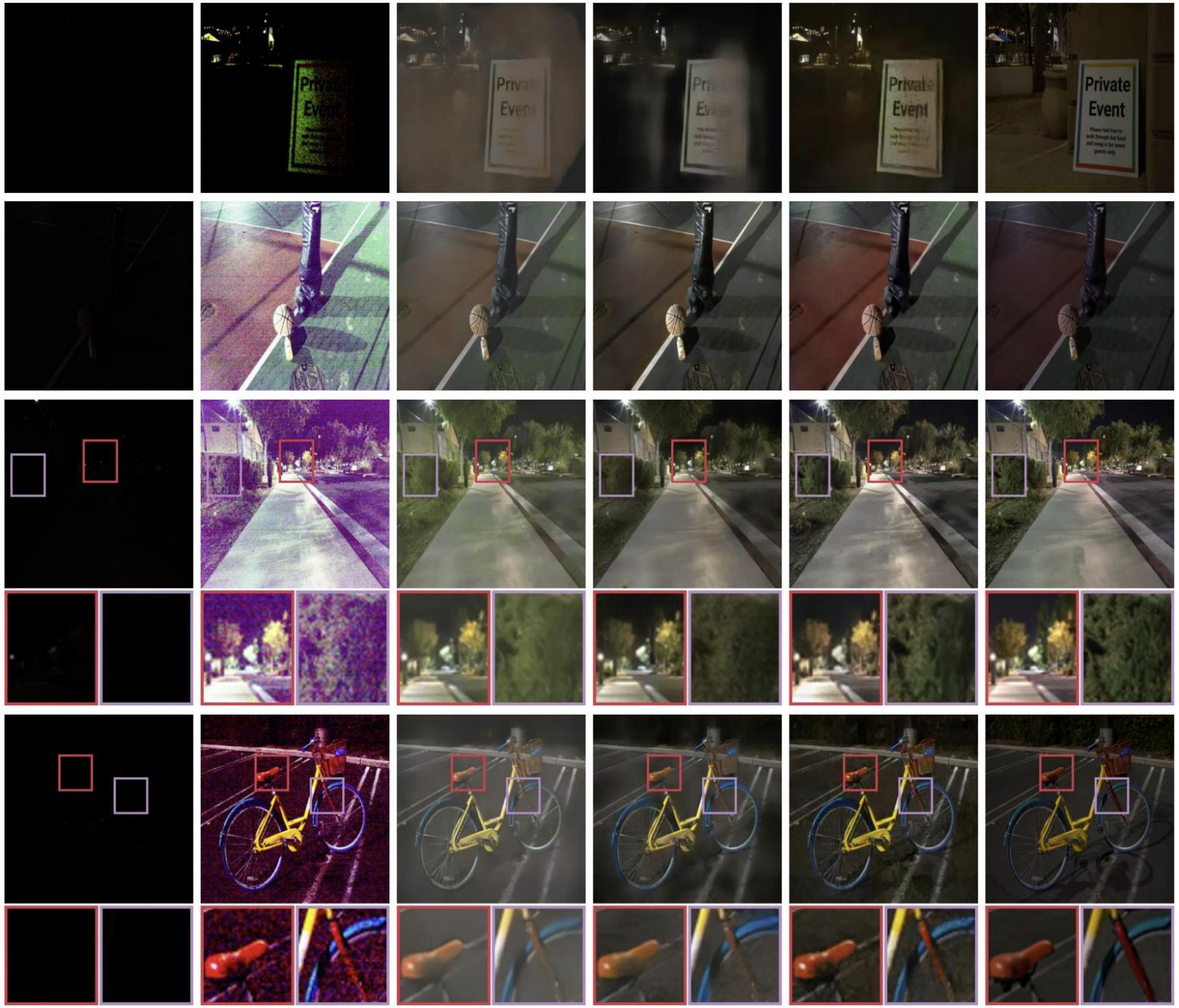
Experiment: More Results



Experiment: More Results

Input DRBN Zero-DCE++ KinD++ SNR-Aware Ours (RDDM) Ground Truth

Experiment: More Results



Input Histogram Equalization D&E UTVNet Ours (RDDM) Ground Truth

Experiment: More Results

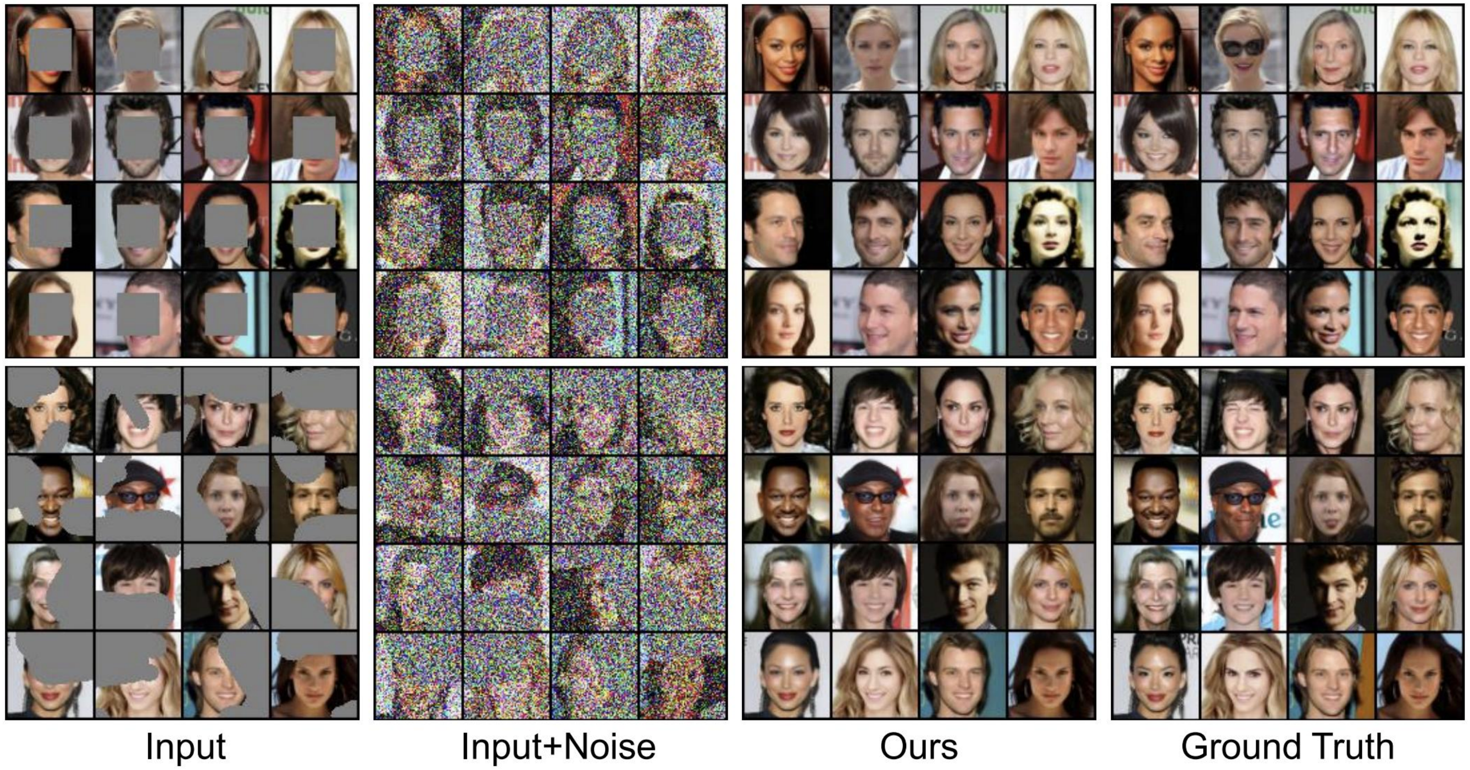
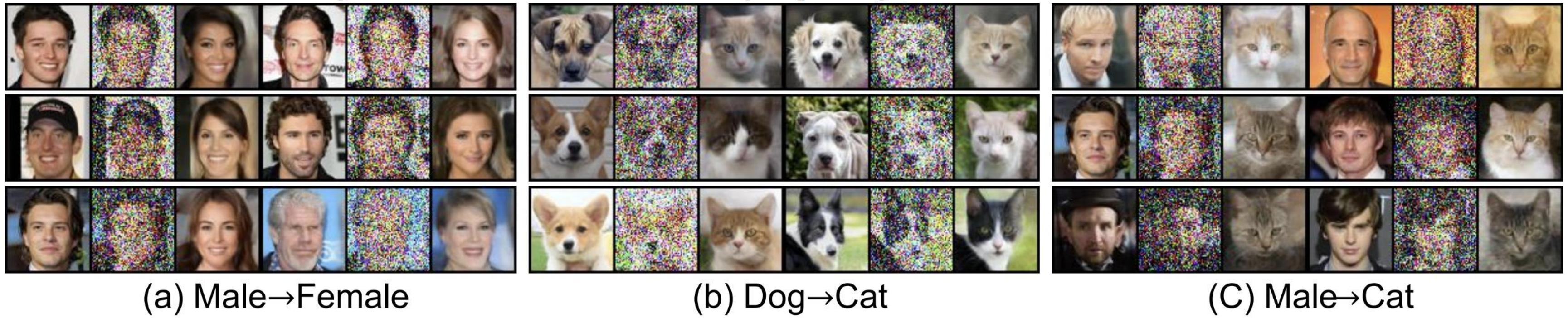


Figure 14. More visual results for image inpainting on the CelebA-HQ [23] dataset.



Conclusion

- A unified dual diffusion model for image restoration and image generation.
- A partially path-independent generation process.
- Limitations and discussion:
  - Specific model for specific task
  - Degenerated into generative model or restoration model
  - Comparison w/ cold diffusion, rectified flow, etc.

Thanks for listening!

Presenter: Haofeng Huang
2024.9.22