Appendix

In the appendix, we provide additional discussions and results complementing Sec. 4. In Sec. A, we conduct further ablation studies on initial condition and iterative refinement to show the control and conditioning effect these two mechanisms bring to the DR2 generative process. In Sec. B, we provide (1) our detailed settings of DR2 controlling parameters (N, τ) for each testing dataset, and (2) show more qualitative comparisons on each split of CelebA-Test dataset in this section to illustrate how our methods and previous state-of-the-art methods perform over variant levels of degradation.

A. More Ablation Studies

In this section, we explore the effect of initial condition and iterative refinement in DR2. To avoid the influence of the enhancement modules varying in structures, embedded facial priors, and training strategies, we only conduct experiments on DR2 outputs with no enhancement. To evaluate the degradation removal performance and fidelity of DR2 outputs, we use bicubic downsampled images as ground truth low-resolution (GT LR) image. This is intuitive as DR2 is targeted to produce clean but blurry middle results.

A.1. Conditioning Effect of Initial Condition with Iterative Refinement Enabled

During DR2 generative process, diffused low-quality inputs is provided through initial condition and iterative refinement. The latter one yields stronger control to the generative process because it is performed at each step, while initial condition only provides information in the beginning with heavy Gaussian noise attached. To quantitatively evaluate the effect of initial condition, we follow the settings of Sec4.4 by calculating the pixel-wise metric (PSNR) and identity distance (Deg) between DR2 outputs and ground truth low-resolution images on CelebA-Test ($8\times$, medium split) dataset. Quantitative results are shown in Tab. A1. We fix $(N, \tau) = (4, 300)$ and change the value of ω. When $\omega = 1000 = T$, no initial condition is provided because $\mathbf{y}_{1000}$ is pure Gaussian noise. As shown in the table, with iterative refinement providing strong control to DR2 generative process, the quality and fidelity of DR2 outputs are not evidently affected as ω varies.

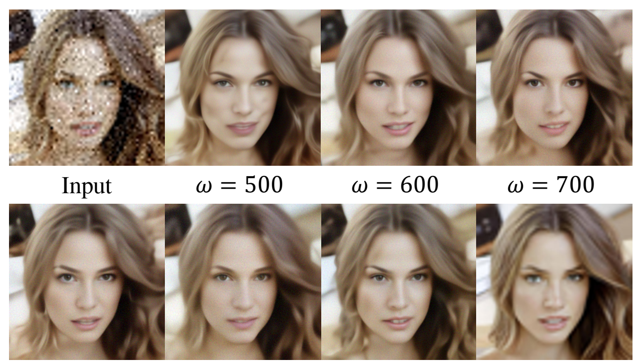
Qualitative results are provided in Fig. A1. With fixed iterative refinement controlling parameters, ω has little visual effect on DR2 outputs. Although the initial condition provides limited information compared with iterative refinement, it significantly reduces the total steps of DR2 denoising process.

ω	350	400	500	600	700	800	900	1000
PSNR↑	26.86	26.87	26.87	26.83	26.80	26.77	26.77	26.76
Deg ↓	56.15	56.02	56.03	56.94	57.38	57.86	57.67	57.78

Table A1. **Effect of different ω with iterative refinement enabled.** With iterative refinement, initial condition has little effect on DR2 output quality as long as $\omega - \tau$ is not two small.

τ	ω	300	400	500	600	700
300	PSNR↑	-	24.79	22.95	20.84	18.34
	Deg↓	-	61.87	68.53	74.84	79.84
150	PSNR↑	24.58	24.05	22.57	20.61	18.22
	Deg↓	63.07	64.24	69.64	75.04	80.51
0	PSNR↑	23.98	23.70	22.33	20.47	18.17
	Deg↓	64.70	64.29	69.71	74.83	80.50

Table A2. **Quantitative results with iterative refinement disabled.** Without iterative refinement, initial condition can only provide limited control to the generative process especially when ω is big.



ω = 800 ω = 900 ω = 1000 GT LR

Figure A1. **Qualitative effect of ω with iterative refinement enabled.** Testing image is from CelebA-Test ($8\times$, medium split) dataset with iterative refinement controlling parameters set as $(N, \tau) = (4, 300)$.

A.2. Conditioning Effect of Initial Condition with Iterative Refinement Disabled

We conduct experiments without iterative refinement in this section to show that generative results bear less fidelity to the input without it. Without iterative refinement, DR2 generative process relies solely on the initial condition to utilize information of low-quality inputs, and generate images through DDPM denoising steps stochastically from initial condition. ω now becomes an important controlling parameter determining how much conditioning information is provided. We also calculate PSNR and Deg between DR2 outputs and ground-truth low-resolution images on CelebA-Test ($8\times$, medium split) dataset. Quantitative results with different (ω, τ) are provided in Tab. A2. Note that PSNR

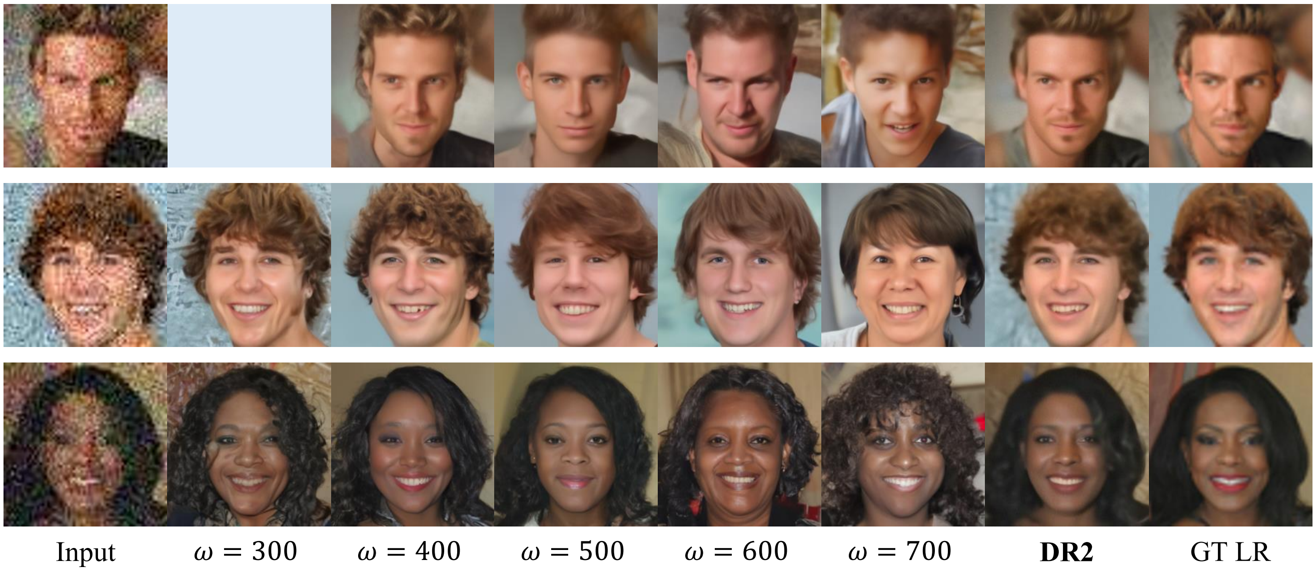


Figure A2. **Qualitative results without iterative refinement.** Testing images are from CelebA-Test ($8\times$, medium split) dataset. The three rows are sampled with $\tau = 300$, 150, and 0 respectively. When $\tau = 0$, truncated output is no longer needed. DR2 outputs are sampled with iterative refinement controlling parameters set as $(N, \tau) = (4, 300)$.

Methods	Split	CA $\times 16$		CA $\times 8$		CA $\times 4$		W-Cr		W-Nm		CelebC		LFW	
		N	τ	N	τ	N	τ	N	τ	N	τ	N	τ	N	τ
DR2 + SPAR	mild	8	220	4	200	4	80								
	medium	16	320	8	350	4	220	16	200	8	100	4	10	4	60
	severe	32	250	8	370	8	190								
DR2 + VQFR	mild	8	250	4	200	4	80								
	medium	16	300	4	350	4	180	8	250	8	100	4	30	4	60
	severe	32	250	8	370	8	190								

Table A3. **Controlling parameter settings** for **CelebA-Test** (CA), **WIDER-Critical** (W-Cr), **WIDER-Normal** (W-Nm), **CelebChild** (CelebC) and **LFW-Test** (LFW). For more severely degraded dataset, bigger N and τ are adopted and vice versa.

and Deg are all worse than those in Tab. A1, and have a negative correlation with ω because less information of inputs is used as ω increases.

Qualitative results are shown in Fig. A2. When $\omega \geqslant 400$, added noise in initial condition is strong enough to cover the degradation in inputs so the output tends to be smooth and clean. But as ω increases, the outputs become more irrelevant to the input because the initial conditions are weakened. Compared with results that were sampled with iterative refinement, the importance of it on preserving semantic information is obvious.

B. Detailed Settings and Comparisons

B.1. Controlling Parameter Settings

As introduced in Sec4.1, to evaluate the performance on different levels of degradation, we synthesize three splits (mild, medium, and severe) for each upsampling task ($16\times$, $8\times$, and $4\times$) together with four real-world datasets. During the experiment in Sec4.2, different controlling parameters (N, τ) are used for each dataset or split. Generally speaking, big N and τ are more effective to remove the degradation but lead to lower fidelity and vice versa. We provide detailed settings we employed in Tab. A3

B.2. More Qualitative Comparisons

For more comprehensive comparisons with previous methods on different levels of degraded dataset, we provide qualitative results on each split of CelebA-Test dataset under each upsampling factor in Figs. A1 to A3. As shown in the figures, for inputs with slight degradation, DR2 transformation is less necessary because previous methods can also be effective. But for severe degradation, previous methods fail since they never see such degradation during training. While our method shows great robustness even though no synthetic degraded images are employed for training.

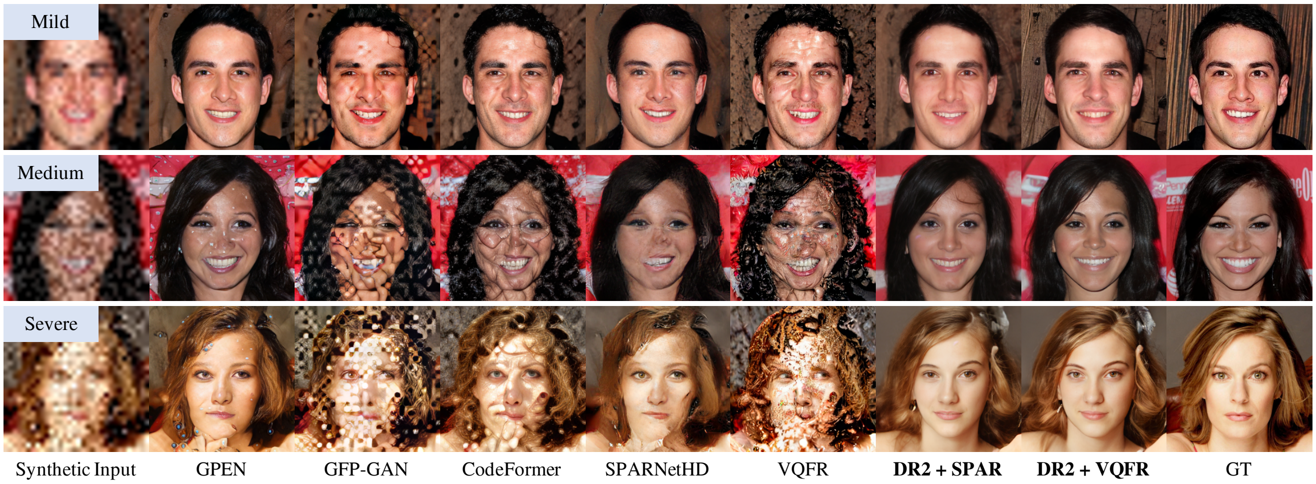


Figure A1. **Qualitative results on CelebA-Test** ($\times 16$)**.** Previous methods produce more artifacts when inputs are heavily degraded.

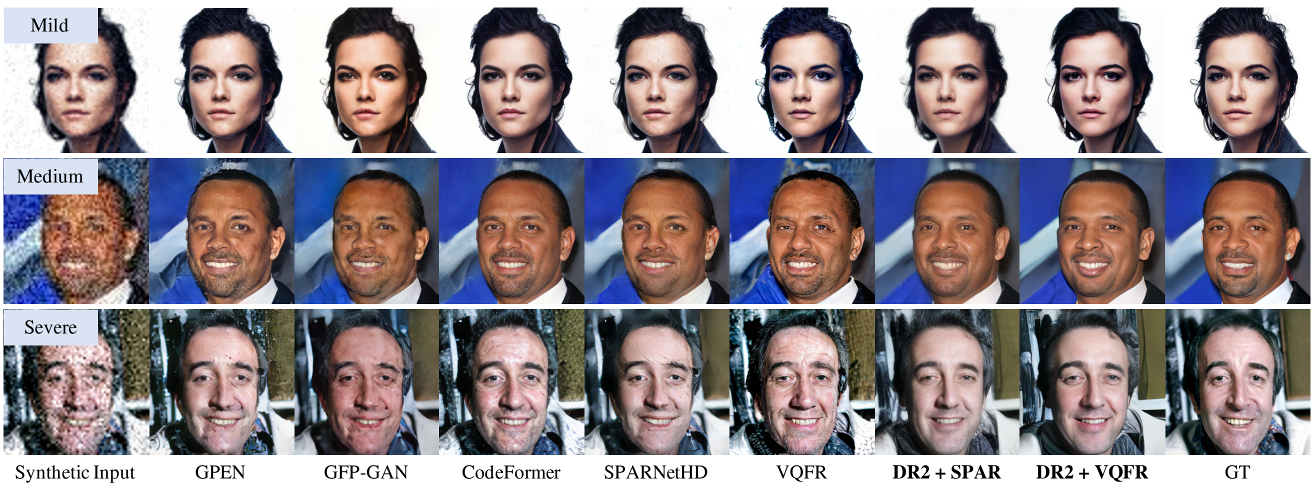


Figure A2. **Qualitative results on CelebA-Test** ($\times 8$)**.** Previous methods produce more artifacts when inputs are heavily degraded.

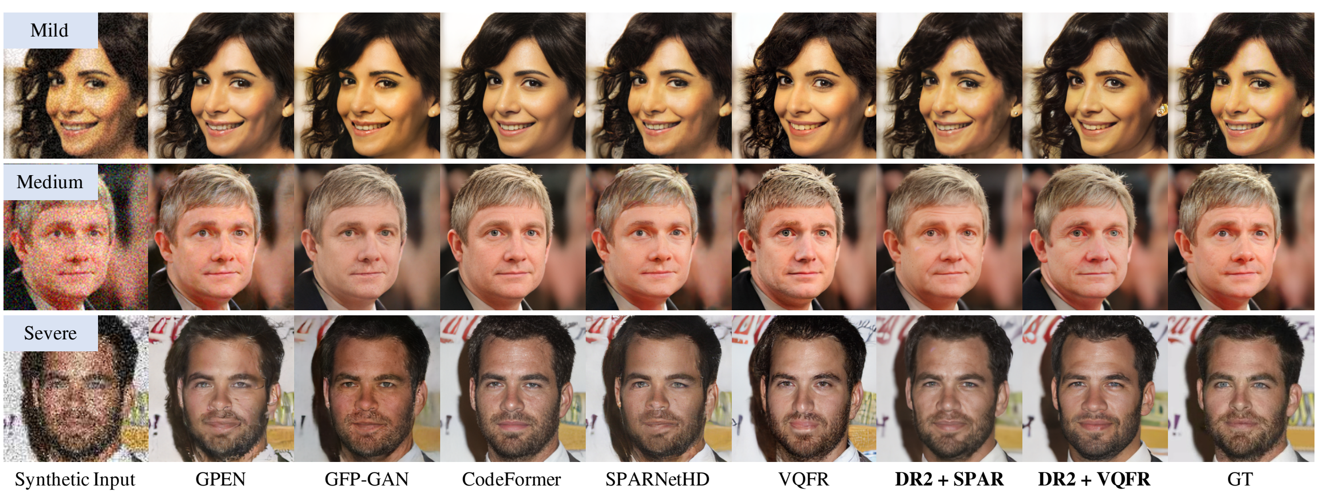


Figure A3. **Qualitative results on CelebA-Test** ($\times 4$)**.** Our methods produce comparable results with previous arts on mildly degraded data.