Accelerating Diffusion Sampling with Optimized Time Steps

Supplementary Material

8. Additional Results and Experiment Details

8.1. More Completed Results for CIFAR-10, ImageNet 64x64, and ImageNet 256x256

In this subsection, we present the completed quantitative results for CIFAR-10, ImageNet 64x64, and ImageNet 256x256 for the case of combining our optimized steps with DPM-Solver++ and UniPC. More specifically, more completed results for CIFAR-10 are presented in Table 3. More completed results for ImageNet 64x64 are presented in Table 4. Since the cosine schedule that leads to unbounded λ at $T = 1.0$ is used for the corresponding codebase, we set $T = 0.992$ instead of $T = 1.0$. More completed results for ImageNet 256x256 are presented in Table 5.

8.2. Additional Quantitative Results

In this subsection, we also present the quantitative results for FFHQ 64x64 and AFHQv2 64x64, as well as ImageNet 512x512. Since we observed from the quantitative results for CIFAR-10, ImageNet 64x64, and ImageNet 256x256 that for most cases, UniPC outperforms DPM-Solver++, throughout the following, we only present the results for UniPC. As mentioned in Section 6, for the EDM model, the time ranges from 0.002 to 80 during sampling rather than 0 to 1, which makes the results of the uniform-t scheme significantly worse than those of the other two baseline discretization schemes. Therefore, we do not include the results of the uniform-t scheme for FFHQ 64x64 and AFHQv264x64, for which we use the EDM unconditional model. Additionally, we observed from the results for ImageNet 256x256 that the EDM scheme does not perform well when the number of NFEs is small. As the generation for ImageNet 512x512 is time-consuming, we do not perform experiments for ImageNet 512x512 using the EDM scheme.

We also observed that for pixel-space generation, the optimized time steps initialized from the uniform-λ scheme often lead to the best generation performance, and for latent-space generation, the optimized time steps initialized from the uniform-t scheme often lead to the best generation performance. Therefore, for the pixel-space generation of images for FFHQ 64x64 and AFHQv264x64, we only present the quantitative results for the time steps initialized from the uniform-λ scheme. For the latent-space generation of images for ImageNet 512x512, we only present the quantitative results for the time steps initialized from the uniform-t scheme.

The quantitative results for FFHQ 64x64 and AFHQv2 64x64 are presented in Tables 6 and 7, respectively.

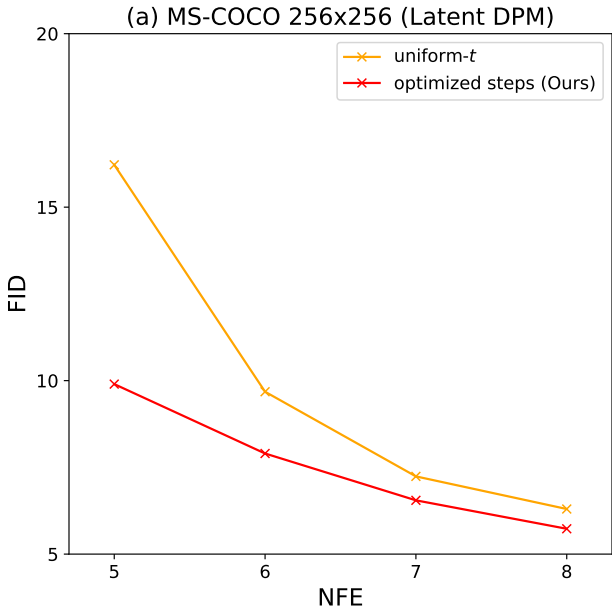

Figure 3. Sampling quality measured by FID (↓) of different discretization schemes of time steps for UniPC [55] with varying NFEs on MS-COCO 256x256 using PixArt-α-256 model [5] (with cfg scale $s = 2.5$).

The quantitative results for ImageNet 512x512 are presented in Table 8.

8.3. Text to Image Generation

We also evaluate our algorithm on text-to-image generation tasks using the PixArt-α model[8] [5]. The COCO [29] validation set is the standard benchmark for evaluating text-to-image models. We randomly draw 30K prompts from the validation set and report the FID score between model samples generated on these prompts and the reference samples from the full validation set following Imagen [41]. The results are shown in Fig. 3. The uniform-t schedule is a widely adopted sampling schedule for latent space text-to-image tasks. In our early experiments, we found the uniform-t schedule significantly outperforms the uniform-λ and EDM schedule on latent space text-to-image tasks. Thus we only report the results compared with the uniform-t schedule. Our proposed optimized sampling schedule consistently outperforms the baseline schedules. For example, we improved the FID from 16.22 to 9.90 with only 5NFEs. We also include some samples to demonstrate the effect of

[8]For FID scores, the checkpoint we use is a SAM pretrained model which is then finetuned on COCO.

Methods \NFEs	5	6	7	8	9	10	12	15
DPM-Solver++ with uniform-λ	29.22	13.28	7.18	5.12	4.40	4.03	3.45	3.17
DPM-Solver++ with uniform-t	28.16	19.63	15.29	12.58	11.18	10.15	8.50	7.10
DPM-Solver++ with EDM	40.48	25.10	15.68	10.22	7.42	6.18	4.85	3.49
DPM-Solver++ with uniform-λ-opt	12.91	8.35	5.44	4.74	3.81	3.51	3.24	3.15
DPM-Solver++ with uniform-t-opt	12.67	8.13	5.63	4.98	5.47	3.66	4.63	3.16
DPM-Solver++ with EDM-opt	12.93	8.04	5.90	8.63	5.14	4.72	4.12	3.16
UniPC with uniform-λ	23.22	10.33	6.18	4.80	4.19	3.87	3.34	3.17
UniPC with uniform-t	25.11	17.40	13.54	11.33	9.83	8.89	7.38	6.18
UniPC with EDM	38.24	23.79	14.62	8.95	6.60	5.59	4.18	3.16
UniPC with uniform-λ-opt	12.11	7.23	**4.96**	**4.46**	**3.75**	3.50	**3.19**	3.13
UniPC with uniform-t-opt	12.10	**7.01**	5.27	4.53	4.69	**3.25**	3.89	**2.78**
UniPC with EDM-opt	**11.91**	7.19	5.62	6.62	4.53	4.12	3.63	2.87

Table 3. Sampling quality measured by FID (↓) of different discretization schemes of time steps for DPM-Solver++ [32] and UniPC [55] with varying NFEs on CIFAR-10 (with $T = 1.0$ and $\epsilon = 0.0005$).

Methods \NFEs	5	6	7	8	9	10	12	15
DPM-Solver++ with uniform-λ	21.78	11.08	7.07	5.38	4.57	4.18	3.67	3.23
DPM-Solver++ with uniform-t	20.62	14.32	10.83	8.83	7.50	6.65	5.53	4.61
DPM-Solver++ with EDM	25.72	15.23	9.80	7.10	5.63	4.83	3.98	3.41
DPM-Solver++ with uniform-λ-opt	13.98	8.28	5.66	4.46	4.20	3.81	3.56	3.22
DPM-Solver++ with uniform-t-opt	13.98	8.28	5.51	**4.39**	**3.88**	3.59	3.71	3.76
DPM-Solver++ with EDM-opt	13.98	8.28	5.57	4.92	5.65	4.85	3.83	3.29
UniPC with uniform-λ	25.77	11.27	6.89	5.26	4.56	4.10	3.48	3.01
UniPC with uniform-t	12.36	9.14	7.85	7.13	6.59	6.20	5.57	4.91
UniPC with EDM	32.65	16.72	10.18	7.43	5.98	5.14	4.25	3.51
UniPC with uniform-λ-opt	**10.47**	**6.74**	**5.29**	4.53	3.99	**3.49**	**3.25**	**2.99**
UniPC with uniform-t-opt	**10.47**	**6.74**	5.60	4.66	3.96	3.91	3.88	3.59
UniPC with EDM-opt	**10.47**	**6.74**	5.39	4.98	5.02	4.30	3.53	3.18

Table 4. Sampling quality measured by FID (↓) of different discretization schemes of time steps for DPM-Solver++ [32] and UniPC [55] with varying NFEs on ImageNet 64x64 (with $T = 0.992$ and $\epsilon = 0.001$).

our algorithm on text-to-image tasks on Sec. 9.

9. Additional Samples

We include additional samples with only 5 NFEs in this section. In Fig. 4, we include the samples generated by DiT-XL-2 [38] on ImageNet 256x256. The samples generated by our method have more details and higher quality. In Fig. 5,6,7 and 8, we include the generated samples corresponding to text prompts generated by PixArt-α-512 model [5]. Our generated samples are clearer and more detailed.

10. Combining with SciRE-Solver

In this section, we provide experimental results for the case of combining our optimized steps with the recently proposed SciRE-Solver [28]. See Table 9.

Methods \NFEs	5	6	7	8	9	10	12	15
DPM-Solver++ with uniform-λ	38.04	20.96	14.69	11.09	8.32	6.47	4.50	3.33
DPM-Solver++ with uniform-t	31.32	14.36	7.62	4.93	3.77	3.23	2.78	2.51
DPM-Solver++ with EDM	65.82	25.19	11.17	7.50	6.98	12.46	6.54	4.03
DPM-Solver++ with uniform-λ-opt	12.53	5.44	3.58	7.54	5.97	4.12	3.61	3.36
DPM-Solver++ with uniform-t-opt	12.53	5.44	3.89	3.81	3.13	2.79	**2.55**	2.39
DPM-Solver++ with EDM-opt	12.53	5.44	3.95	3.79	3.30	3.14	2.91	2.44
UniPC with uniform-λ	41.89	30.51	19.72	12.94	8.49	6.13	4.14	2.98
UniPC with uniform-t	23.48	10.31	5.73	4.06	3.39	3.04	2.73	2.50
UniPC with EDM	45.89	21.24	15.52	14.38	14.24	12.98	8.62	4.10
UniPC with uniform-λ-opt	**8.66**	**4.46**	**3.57**	3.72	3.40	3.01	2.94	2.53
UniPC with uniform-t-opt	**8.66**	**4.46**	3.74	**3.29**	**3.01**	**2.74**	**2.55**	**2.36**
UniPC with EDM-opt	**8.66**	**4.46**	3.78	3.34	3.14	3.22	2.96	2.38

Table 5. Sampling quality measured by FID (↓) of different discretization schemes of time steps for DPM-Solver++ [32] and UniPC [55] with varying NFEs on ImageNet 256x256 (with $T = 1.0$ and $\epsilon = 0.001$).

Methods \NFEs	5	6	7	8	9	10	12	15
UniPC with uniform-λ	20.02	10.97	6.97	5.53	4.53	3.89	3.28	2.83
UniPC with EDM	26.54	15.07	11.20	11.65	10.91	8.89	5.43	3.40
UniPC with uniform-λ-opt	**13.66**	**8.41**	**6.49**	**4.84**	**3.82**	**3.41**	**3.00**	**2.82**

Table 6. Sampling quality measured by FID (↓) of different discretization schemes of time steps for UniPC with varying NFEs on FFHQ 64x64 (with $T = 80$ and $\epsilon = 0.002$).

Methods \NFEs	5	6	7	8	9	10	12	15
UniPC with uniform-λ	12.95	8.30	5.12	4.62	4.47	3.80	2.75	2.28
UniPC with EDM	15.83	10.30	8.46	7.83	6.78	6.38	5.25	3.09
UniPC with uniform-λ-opt	**12.11**	**7.49**	**5.05**	**3.86**	**3.27**	**2.74**	**2.51**	**2.17**

Table 7. Sampling quality measured by FID (↓) of different discretization schemes of time steps for UniPC with varying NFEs on AFHQv2 64x64 (with $T = 80$ and $\epsilon = 0.002$).

Methods \NFEs	5	6	7	8	9	10	12	15
UniPC with uniform-λ	41.14	19.81	13.01	9.83	8.31	7.01	5.30	4.00
UniPC with uniform-t	20.28	10.47	6.57	5.13	4.46	4.14	3.75	3.45
UniPC with uniform-t-opt	**11.40**	**5.95**	**4.64**	**4.36**	**4.05**	**3.81**	**3.54**	**3.28**

Table 8. Sampling quality measured by FID (↓) of different discretization schemes of time steps for UniPC with varying NFEs on ImageNet 512x512 (with $T = 1.0$ and $\epsilon = 0.001$).

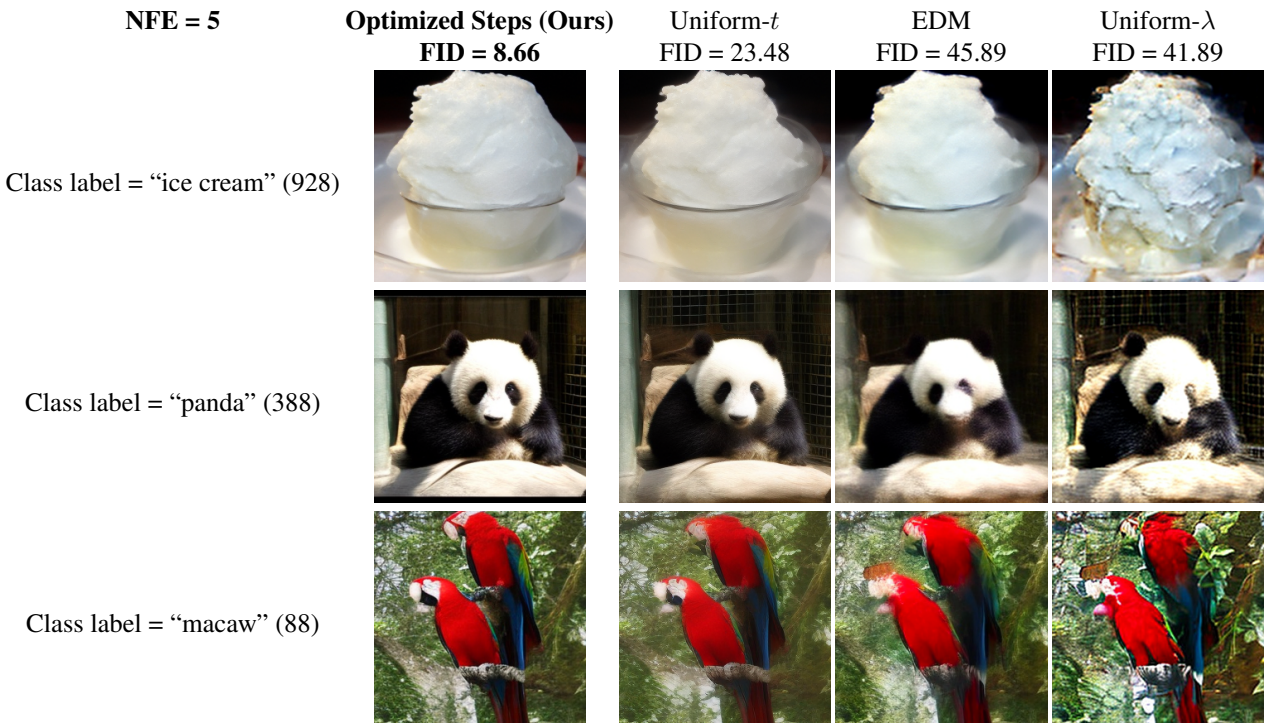

Figure 4. Generated images by UniPC [55] with only 5 NFEs for various discretization schemes of time steps from DiT-XL-2 ImageNet 256x256 model [38] (with cfg scale $s = 1.5$ and the same random seed).

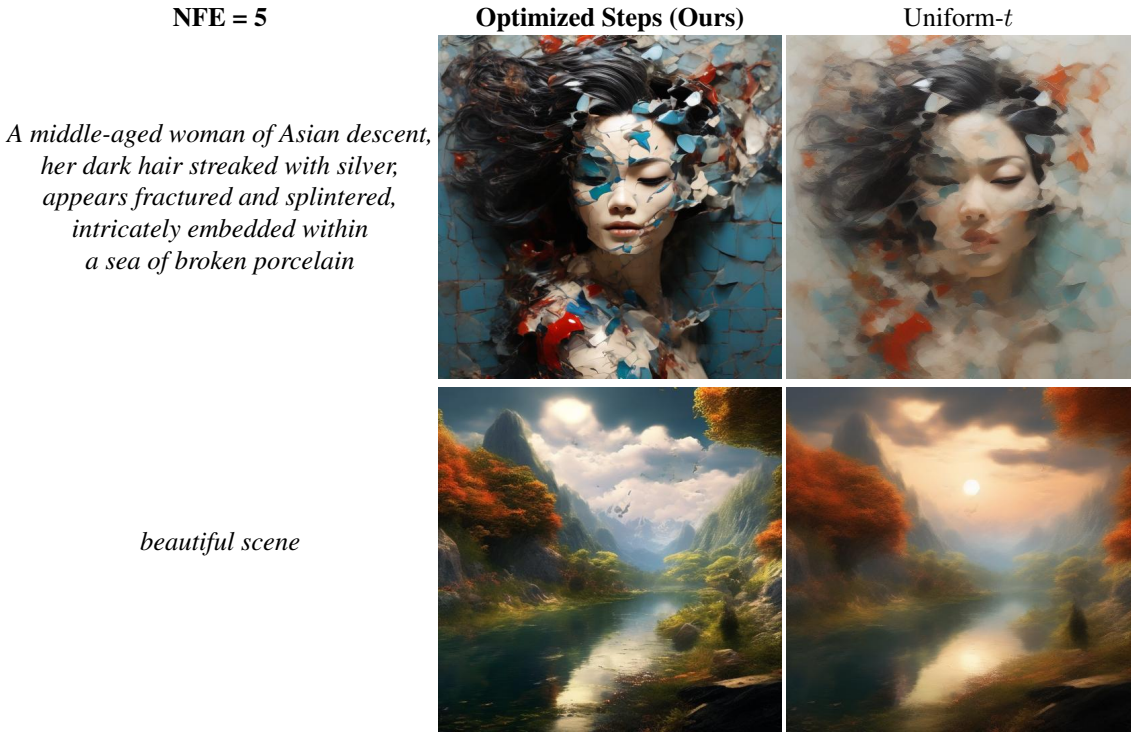

Figure 5. Generated images by UniPC [55] with only 5 NFEs for various discretization schemes of time steps from PixArt-α-512 model [5] (with cfg scale $s = 2.5$ and the same random seed).

Figure 6. Generated images by UniPC [55] with only 5 NFEs for various discretization schemes of time steps from PixArt-α-512 model [5] (with cfg scale $s = 2.5$ and the same random seed).

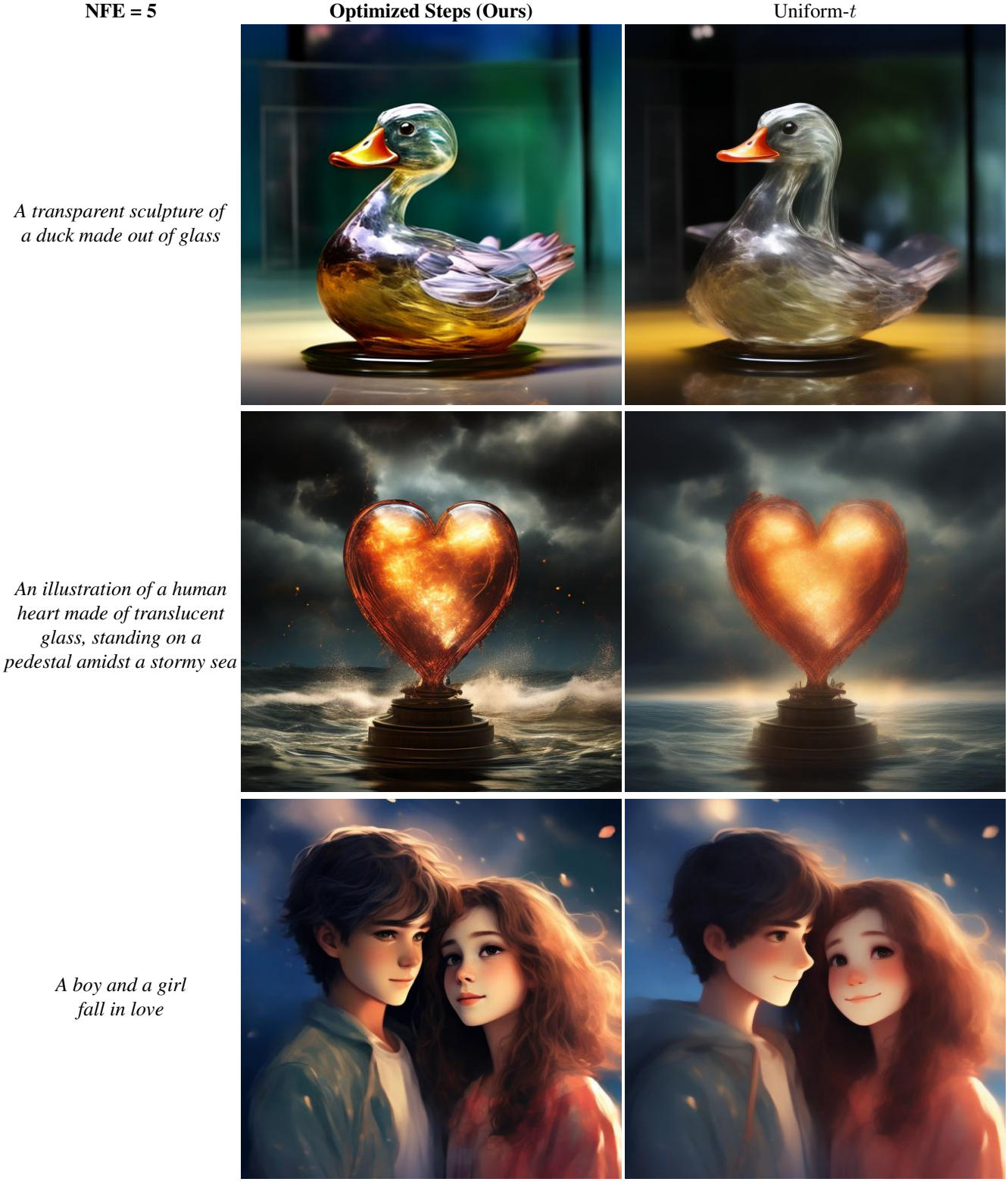

Figure 7. Generated images by UniPC [55] with only 5 NFEs for various discretization schemes of time steps from PixArt-α-512 model [5] (with cfg scale $s = 2.5$ and the same random seed).

Figure 8. Generated images by UniPC [55] with only 5 NFEs for various discretization schemes of time steps from PixArt-α-512 model [5] (with cfg scale $s = 2.5$ and the same random seed).

Methods \NFEs	5	6	7	8
Uniform-t	36.18	15.93	7.57	4.48
Uniform-λ	92.80	41.19	21.08	11.25
EDM	90.48	47.18	23.32	11.10
Optimized steps (Ours)	**16.07**	**6.31**	**4.12**	**3.66**

Table 9. Sampling quality measured by FID (↓) of different discretization schemes of time steps for SciRE-Solver with varying NFEs on ImageNet 256x256 (using the DiT-XL-2 model with cfg = 1.5).