Adv-CPG: A Customized Portrait Generation Framework with Facial Adversarial Attacks

Supplementary Material

Table 6. The prompts used in the experiment for Fig. 7 and Fig. 8. The face in the bottom right is the target of 5 original faces.

	
"A man, in a chef outfit"	"A woman, in the city"
	
"A woman, in a firefighter outfit"	"A woman, in the city"
	
"A woman, in the park"	**Target**
prompt = "cinematic photo," + prompt + "bokeh, 4k" negative prompt = "monochrome, lower, bad anatomy, worst quality, low quality, blurry"	

6. Additional Algorithm Showcase

Based on whether or not the augmented text is used, the generation process of Adv-CPG is divided into two stages: **1**) **Progressive Facial Privacy Protection Stage**: In this stage, unenhanced raw text prompt is utilized for initial context introduction. Since the fine-grained facial description is not used, the designed facial privacy protection modules (ID encryptor and encryption enhancer) play an important role in progressive protection, and meanwhile the original textual prompt enables effective background generation. **2**) **Fine-Grained Customized Portrait Generation Stage**: In this phase, fine-grained face generation is achieved by utilizing detailed facial features and integrating other modules. This accomplishes not only a balance of visual semantics between the human body and the background, but also a balance between facial privacy protection and portrait generation. The overall framework of the proposed Adv-CPG to perform facial privacy protection and customized portrait generation is shown in Fig. 2, and the algorithm of the generation process is summarized in Alg. 1.

7. Additional Ablation Study

The impact of T_2 to Adv-CPG without protection. In Figure 7, we provide visualization results to evaluate the impact of the Stage II start time T_2 on portrait generation (Adv-CPG W/O Protection). The term 'Stage II step' denotes the first time step in which the augmented text prompt is utilized. No protection means the ID encryptor and encryption enhancer are removed. In general, the influence of fine-grained facial features diminishes as the 'Stage II step' increases. As shown in Figure 7, if the 'Stage II step' is set to 0, it indicates that fine-grained facial features dominate the generation process, which may result in the loss of semantic information from the text prompt. Conversely, if the 'Stage II step' is set to 50, the guiding effect of the textual prompt is maximized, but there may be a lack of adherence to the semantic information of the facial description. In summary, the earlier the stage II begins, the better the balance of fine-grained faces and backgrounds.

The impact of T_2 to Adv-CPG. In Figure 8, the visualization results for evaluating the effect of Stage II start time T_2 on Adv-CPG are provided. The two protection modules work and function in both the I and II stages. In the experiment, we set T_2 to 30. In the stage I, as enhanced text is not used, facial privacy preservation and background introduction perform vital roles. In the stage II, all modules operate to facilitate the balance between facial privacy protection and fine-grained portrait generation.

Algorithm 1 Generation Process of the proposed Adv-CPG

Input: An original facial image, $\mathbf{I}_o$; A target facial image, $\mathbf{I}_t$; An original text prompt, $\mathcal{P}_o$; An augmented text prompt, $\mathcal{P}$; The pre-trained CLIP image (text) encoder, $\mathcal{I}$ ($\mathcal{T}$); The denoising time T; Stage II start time T_2;
Output: A generated adversarial example $\mathbf{I}_{adv}$;

Generating the visual semantic feature by $\mathbf{C}_i = \mathcal{I}(\mathbf{I}_o)$;
Exactoring the powerful target ID feature $\mathbf{C}_{id}$ by Eq. 5;
Generating the original text feature by $\mathbf{C}_t^o = \mathcal{T}(\mathcal{P}_o)$;
Generating the augmented text feature by $\mathbf{C}_t = \mathcal{T}(\mathcal{P})$;
A denoising step is denoted as $\mathbf{z}_{t-1} = DM(\mathbf{z}_t, t, ...)$;
for $t = T, T-1, ..., 1$ **do**
  if $t > T_2$ **then**
    $\mathbf{z}_{t-1} = DM(\mathbf{z}_t, t, \mathbf{C}_{id}, \mathbf{C}_i, \mathbf{C}_t^o)$;
  else
    $\mathbf{z}_{t-1} = DM(\mathbf{z}_t, t, \mathbf{C}_{id}, \mathbf{C}_i, \mathbf{C}_t)$;
  end if
  $\widetilde{\mathbf{z}}_0 = \frac{1}{\sqrt{\overline{a}_t}}(\mathbf{z}_{t-1} - \sqrt{1-\overline{a}_t}\epsilon_\theta(\mathbf{z}_{t-1}, t))$;
  $\mathbf{g}_t = \lambda_s \nabla_{\widetilde{\mathbf{z}}_0} F(\mathcal{D}(\widetilde{\mathbf{z}}_0), \mathbf{I}_t)$;
  $\hat{\mathbf{z}}_{t-1} := \mathbf{z}_{t-1} + \boldsymbol{\sigma}_t \mathbf{g}_t$;
end for
return $\mathbf{z}_0$;

Table 7. ASR (% ↑) of the six FR Models on LADN / FaceScrub.

Method	IR152	IRSE50	FaceNet	MF	MF1	TF
Clean	3.61 / 2.92	2.71 / 3.11	0.60 / 0.76	5.11 / 4.38	1.94 / 1.83	0.63 / 0.71
Adv-Makeup	10.03 / 15.54	29.64 / 42.64	0.97 / 18.62	22.38 / 26.53	20.78 / 21.95	19.73 / 26.49
FaceShifter[65]	49.12 / 55.13	80.41 / 83.57	52.13 / 46.62	72.43 / 50.71	45.85 / 48.92	45.01 / 48.72
Clip2Protect	53.31 / 57.32	91.57 / 91.89	47.91 / 52.46	77.94 / 79.67	67.82 / 66.17	47.85 / 56.15
Blendface[66]	63.04 / 62.73	89.64 / 76.96	52.48 / 56.04	75.16 / 64.19	59.05 / 57.14	52.78 / 51.46
DiffAM	69.48 / 72.36	90.26 / 91.25	68.16 / 73.44	72.68 / 75.23	88.33 / 85.46	57.82 / 61.35
Adv-CPG	**70.56 / 73.14**	91.46 / **93.18**	**70.42 / 74.57**	**79.54 / 81.32**	87.38 / **89.65**	**60.17 / 62.95**

Table 8. ASR of diverse backbones on FFHQ / CeleA-HQ.

Backbone	IR152	IRSE50	FaceNet	MF	MF1	TF
SD1.5	72.15 / 72.89	87.92 / 86.93	61.26 / 61.03	88.72 / 85.39	89.17 / 88.32	58.64 / 58.23
SD2.1	74.03 / 76.25	89.06 / 88.59	62.94 / 62.71	89.77 / 86.05	89.57 / 88.69	60.08 / 59.97
SDXL	**75.26 / 76.96**	**91.03 / 88.72**	**63.84 / 63.50**	**89.94 / 87.95**	**90.06 / 89.23**	**63.47 / 60.94**

In addition, the textual prompts used for the above two experiments are shown in Table 6. Among them, the last line is a generic prompt applied to each facial image, and the face in the bottom right is the target image.

Generalization of Adv-CPG. Adv-CPG has favorable generalization, which can be naturally extended to deverse T2I models. When replacing the T2I backbone, the only operation needed is to train the ID projector (in ID encryptor). As shown in Tab. 8, using SDXL as the T2I model is superior to using SD1.5 and SD2.1. In addition, SDXL is chosen for its consistency with classical portrait generation and its two-stage optimization for higher fidelity. Future work will extend Adv-CPG to models like SD3 and FLUX.

8. Additional Experiments

More Datasets and More FR models. In Tab. 7, we present attack results on the added datasets LADN [62] and FaceScrub [61] for more FR models (MF1 [63], TF [64]). Adv-CPG achieves optimal or suboptimal results on the protective effects of the six FR systems. Moreover, Adv-CPG exhibits solid robustness due to the sufficient diversity of FR systems. As shown in Tab. 9, we cover closed-source FR systems (Face++ and Aliyun in Fig. 4), Transformer-based models, and FR models used by dominant methods.

Table 9. Comparison of diverse FR Models used by Adv-CPG.

Classify	Face++	Aliyun	IR152	IRSE50	FaceNet	MF	MF1	TF
Open-source			✓	✓	✓	✓	✓	✓
Transformer-based								✓
DiffAM/Adv-Diffusion Used	✓	✓	✓	✓	✓	✓		

Comparison with face-swapping tasks. In Tab. 7, [65] and [66] can be used for image protection after the original and target images are exchanged. But their results are far less effective than Adv-CPG, which proves the effectiveness of our pipeline. Symbol bugs and figure optimizations will be updated in the final version.

References

[61] Hong-Wei Ng and Stefan Winkler. A data-driven approach to cleaning large face datasets. In *ICIP*, pages 343–347, 2014.

[62] Qiao Gu, Guanzhi Wang, Mang Tik Chiu, Yu-Wing Tai, and Chi-Keung Tang. Ladn: Local adversarial disentangling network for facial makeup and de-makeup. In *ICCV*, pages 10481–10490, 2019.

[63] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *CVPR*, pages 14220–14229, 2021.

[64] Jun Dan, Yang Liu, Haoyu Xie, Jiankang Deng, Haoran Xie, Xuansong Xie, and Baigui Sun. Transface: Calibrating transformer training for face recognition from a data-centric perspective. In *ICCV*, pages 20585–20596, 2023

[65] Lingzhi Li, Jianmin Bao, Hao Yang, Dong Chen, and Fang Wen. Advancing high fidelity identity swapping for forgery detection. In *CVPR*, pages 5073–5082, 2020.

[66] Kaede Shiohara, Xingchao Yang, and Takafumi Taketomi. Blendface: Re-designing identity encoders for face-swapping. In *ICCV*, pages 7600–7610, 2023.

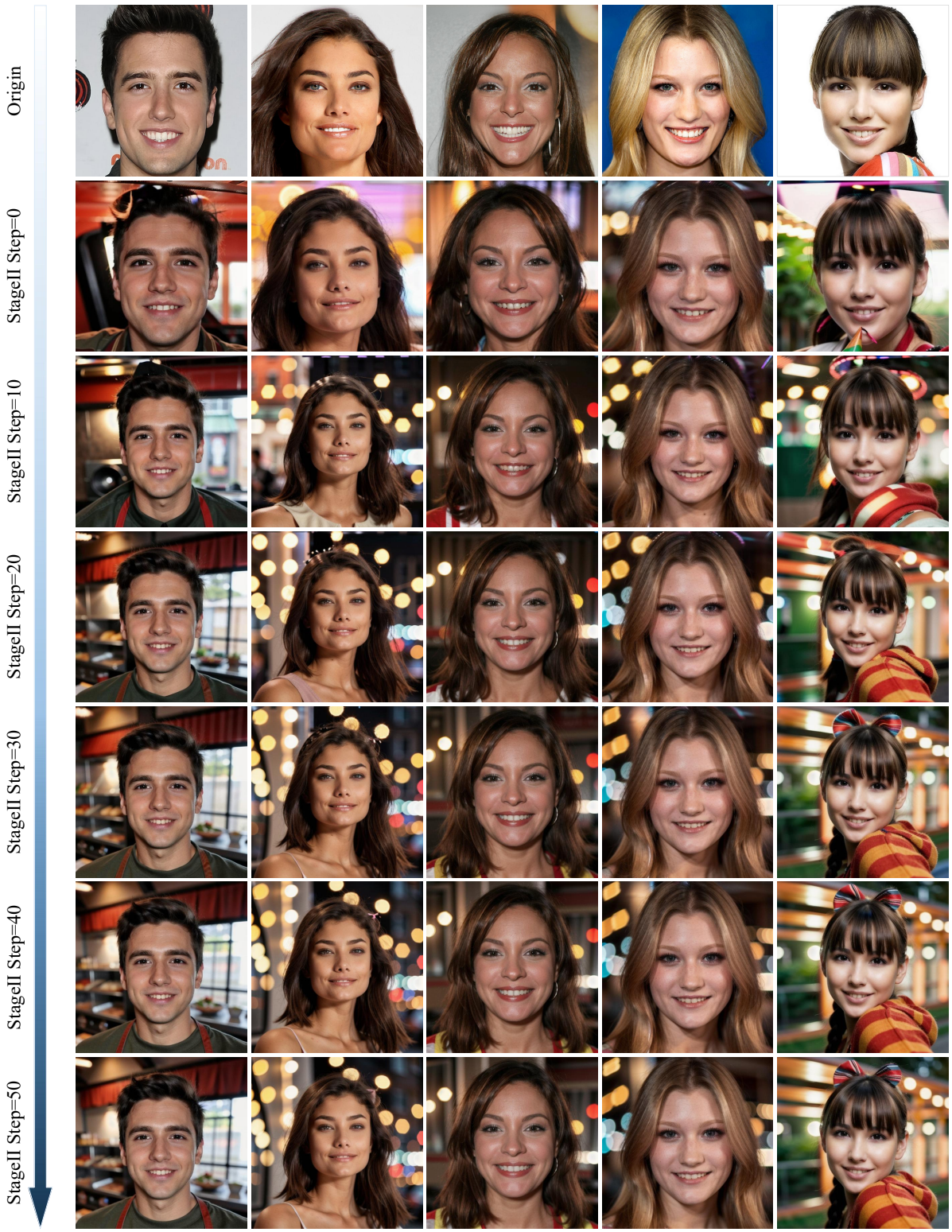

Figure 7. visualized results of Adv-CPG (W/O Protection) under different Stage II start time T_2. The term 'W/O Protection' indicates the ID encryptor and the encryptor enhancer are removed. The term 'Stage II step' denotes the first time step in which the augmented text prompt is utilized.

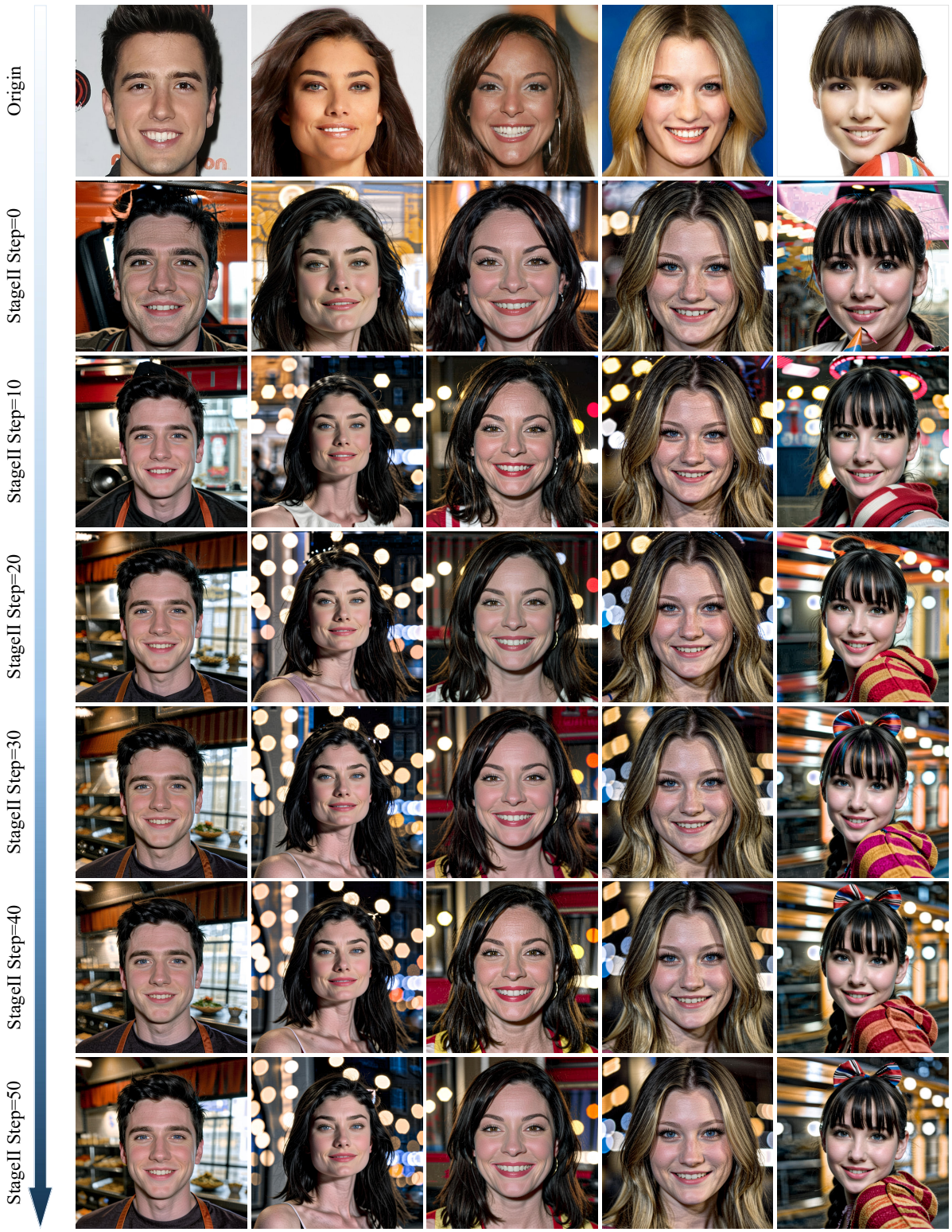

Figure 8. visualized results under different Stage II start time T_2. The term 'Stage II step' denotes the first time step in which the augmented text prompt is utilized.