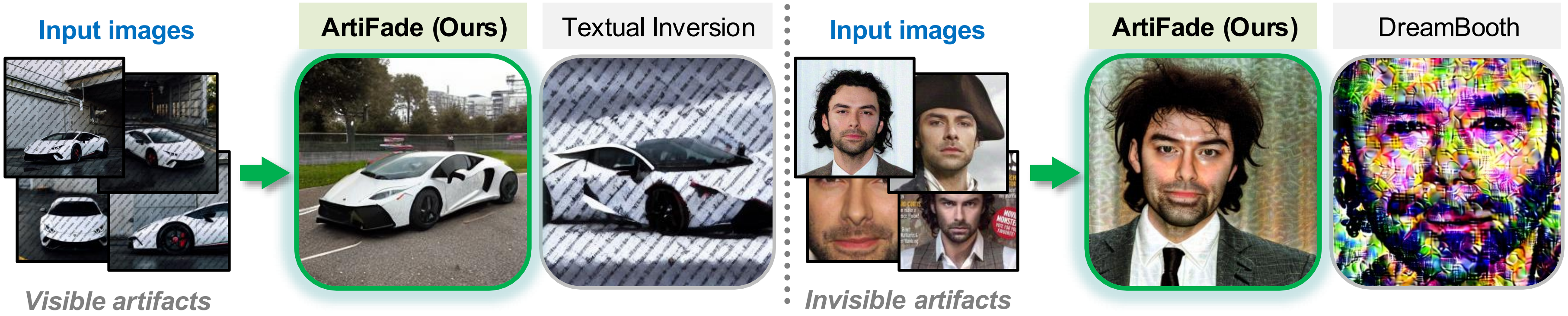
ArtiFade: Learning to Generate High-quality Subject from Blemished Images

Shuya Yang Shaozhe Hao Yukang Cao Kwan-Yee K. Wong
The University of Hong Kong



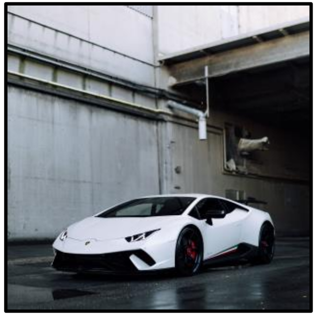
Outline

- Motivation
- Methodology
- **Experiments**
- Applications

Subject-driven image generation

Input images

Textual Inversion



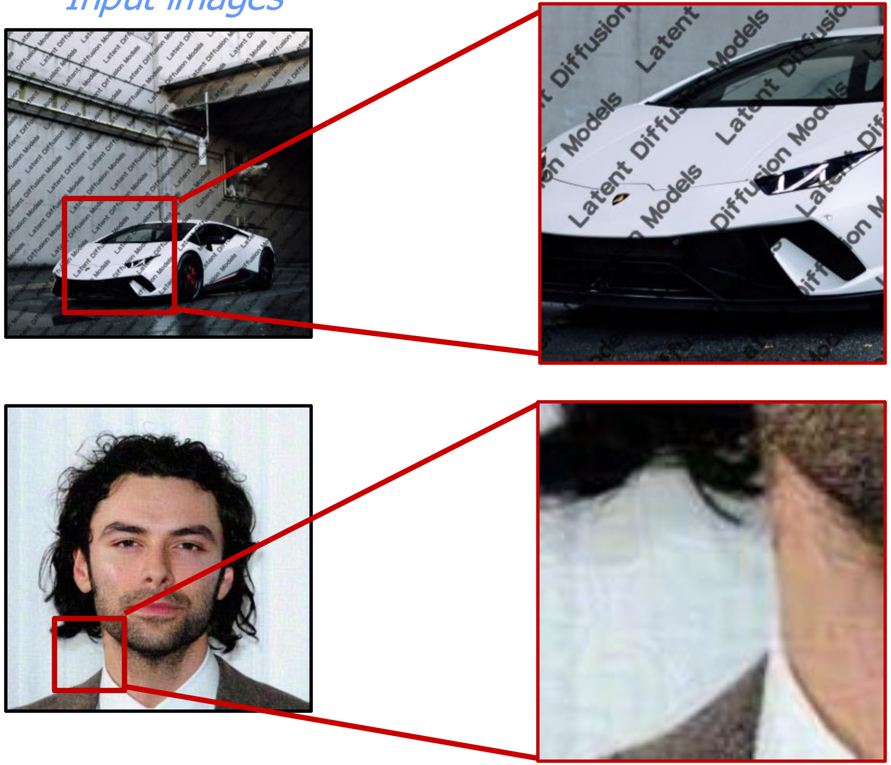
DreamBooth

Subject-driven image generation

Input images

Textual Inversion

DreamBooth



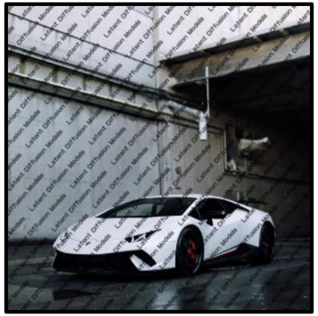
Visible Artifacts

Invisible Artifacts

Blemished Subject-driven image generation

Input images

Textual Inversion



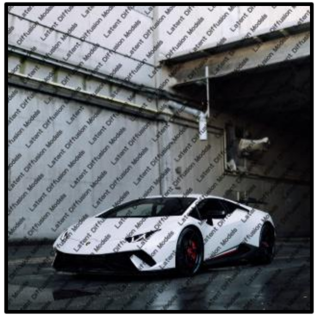
DreamBooth

Blemished Subject-driven image generation

Input images

ArtiFade

Textual Inversion



DreamBooth

Contribution

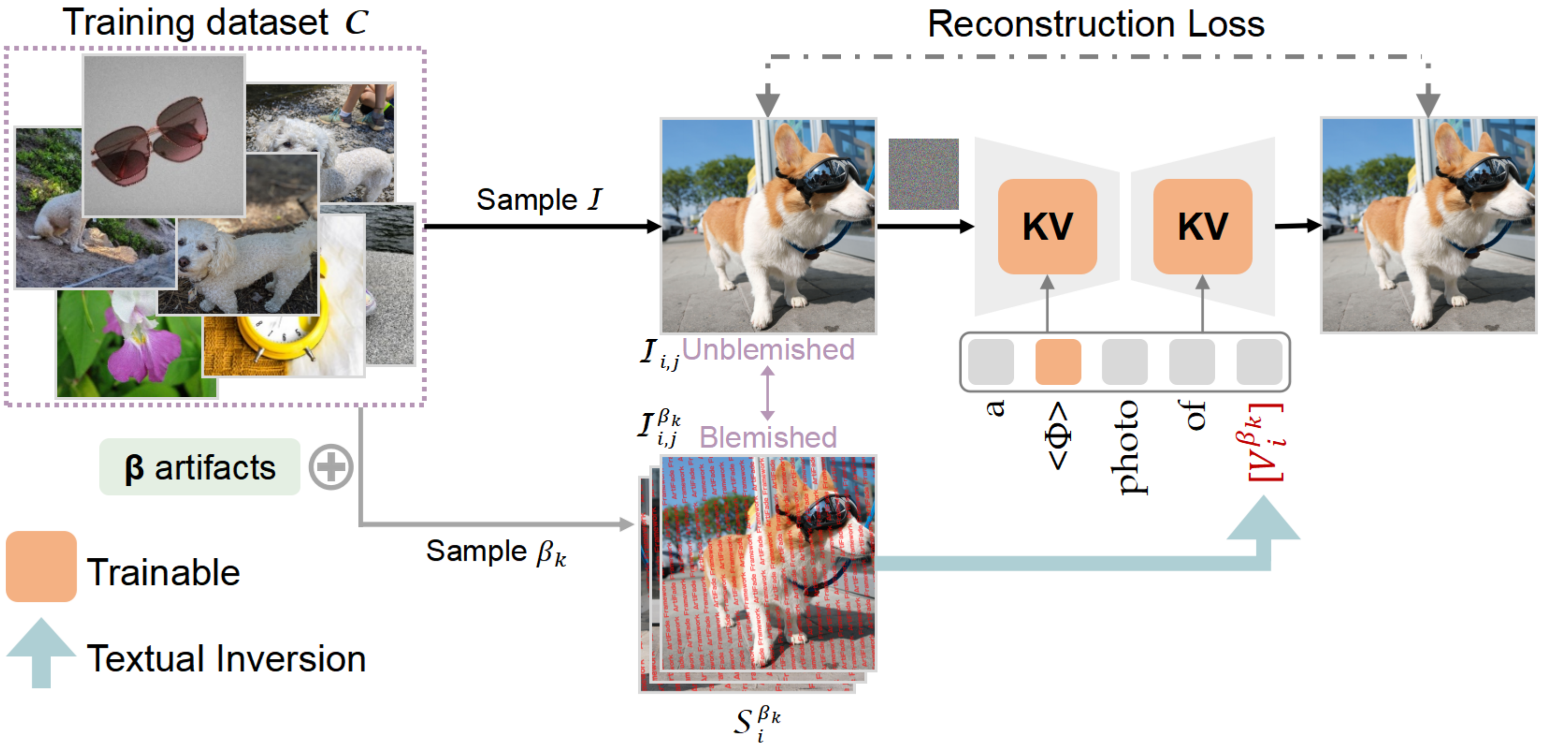
- **Novel challenge:** The first work to tackle the problem of blemished subject-driven generation.
- **Proposed method:** Introduce ArtiFade, which fine-tunes diffusion models to align unblemished and blemished data.
- **Benchmark:** Establish a new benchmark for evaluating blemished subject-driven generation.
- **Generalizability:** Demonstrate strong generalizability, effective on both in-distribution and out-of-distribution artifacts.

Method
of
ArtiFade

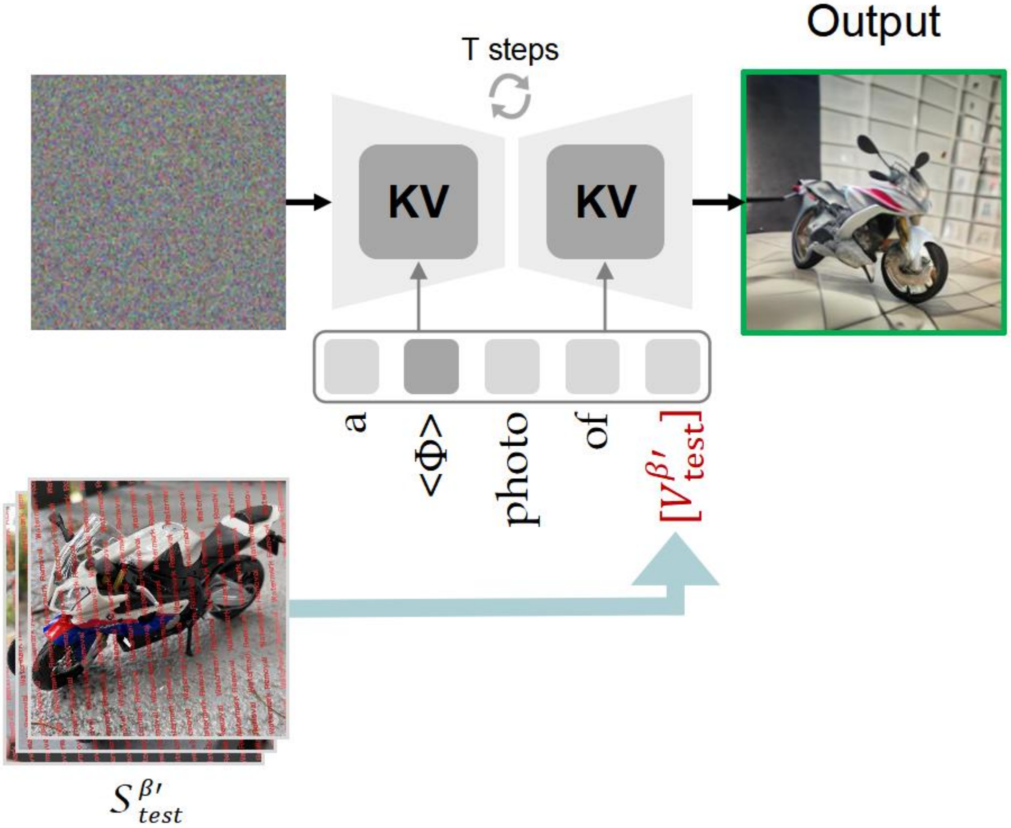
I. Data preparation Step

$$S_i^{\beta_k} \xrightarrow{\text{Textual Inversion}} [\mathtt{V}_{\mathtt{i}}^{\beta_k}], \quad i = 1, 2, \ldots, N; \quad k = 1, 2, \ldots, L$$

2. Fine-tuning Step



3. Inference Step



Experiments

ArtiFade with Textual Inversion - Quantitative

1. I^{CLIP}:= CLIP similarities between the generated images and the corresponding unblemished subsets
2. I^{DINO} := DINO similarities between the generated images and the corresponding unblemished subsets
3. T^{CLIP}:= CLIP similarity between the generated images and the text prompt
4. $R^{CLIP} = I^{CLIP} / I_{\beta}^{CLIP}$
5. $R^{DINO} = I^{DINO} / I_{\beta}^{DINO}$

ArtiFade with Textual Inversion - Quantitative

In-distribution

Method	WM-model on WM-ID-test				
	$I^{DINO}\uparrow$	$R^{DINO}\uparrow$	$I^{CLIP}\uparrow$	$R^{CLIP}\uparrow$	$T^{CLIP}\uparrow$
TI (unblemished)	0.488	1.349	0.730	1.070	0.283
TI (blemished)	0.217	0.852	0.576	0.909	0.263
Ours	**0.337**	**1.300**	**0.649**	**1.020**	**0.282**

Out-of-distribution

Method	WM-model on WM-OOD-test				
	$I^{DINO}\uparrow$	$R^{DINO}\uparrow$	$I^{CLIP}\uparrow$	$R^{CLIP}\uparrow$	$T^{CLIP}\uparrow$
TI (unblemished)	0.488	1.278	0.730	1.136	0.283
TI (blemished)	0.229	0.858	0.575	0.929	0.262
Ours	**0.356**	**1.237**	**0.654**	**1.079**	**0.282**

ArtiFade with Textual Inversion - Qualitative (In-distribution)

Input images | **Ours** | Textual Inversion | **Ours** | Textual Inversion





$[V_{test}^{\beta'}]$ in the street

$[V_{test}^{\beta'}]$ in the snow





$[V_{test}^{\beta'}]$ with a mountain in the background

$[V_{test}^{\beta'}]$ with a city in the background

ArtiFade with Textual Inversion - Qualitative (Out-of-distribution)

Input images | **Ours** | Textual Inversion | **Ours** | Textual Inversion





$[V^{\beta'}_{test}]$ on top of a wooden floor

$[V^{\beta'}_{test}]$ with a city in the background





$[V^{\beta'}_{test}]$ in the movie theater

$[V^{\beta'}_{test}]$ in a luxurious interior living room

ArtiFade with DreamBooth - Quantitative

In-distribution

Method	WM-ID-test				
	$I^{DINO}\uparrow$	$R^{DINO}\uparrow$	$I^{CLIP}\uparrow$	$R^{CLIP}\uparrow$	$T^{CLIP}\uparrow$
TI (unblemished)	0.488	1.349	0.730	1.070	0.283
TI (blemished)	0.217	0.852	0.576	0.909	0.263
DB (blemished)	0.503	0.874	0.738	0.939	0.272
Ours (TI-based)	0.337	1.300	0.649	1.020	0.282
Ours (DB-based)	**0.589**	**1.308**	**0.795**	**1.083**	**0.284**

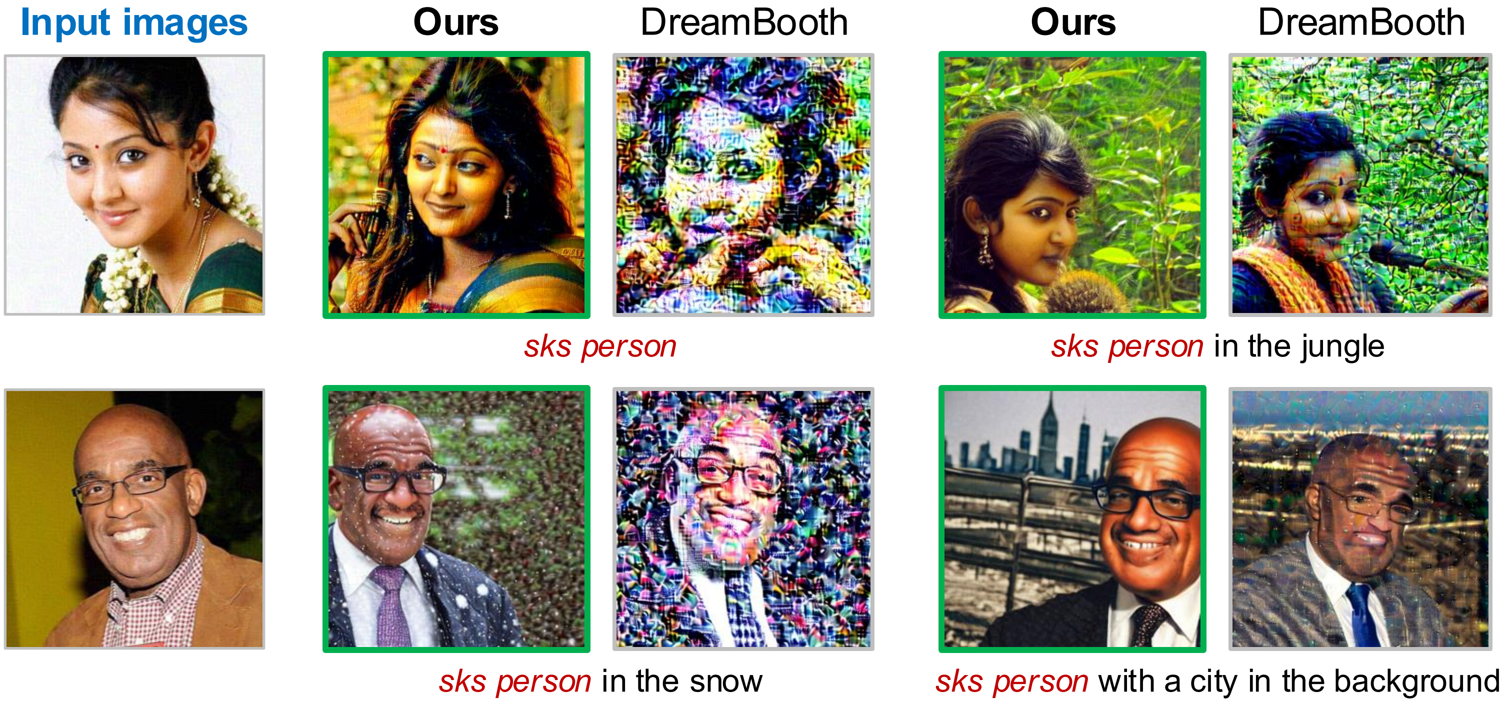
ArtiFade with DreamBooth - Qualitative (In-distribution)

Input images | **Ours** | DreamBooth | **Ours** | DreamBooth

sks cat with a beautiful sunset *sks cat* with the Eiffel tower in the background

sks motorbike in the snow *sks motorbike* in the jungle

ArtiFade with DreamBooth - Qualitative (Invisible artifacts)



Ablation Study

Method	W^{kv}	W^{q}	$\langle\Phi\rangle$	$\mathrm{I}^{\mathrm{DINO}}$	$\mathrm{R}^{\mathrm{DINO}}$	$\mathrm{I}^{\mathrm{CLIP}}$	$\mathrm{R}^{\mathrm{CLIP}}$	$\mathrm{T}^{\mathrm{CLIP}}$
$\mathtt{Var}_{\mathtt{A}}$			✓	0.154	**1.412**	0.566	0.984	0.265
$\mathtt{Var}_{\mathtt{B}}$		✓	✓	0.283	1.230	0.617	0.978	0.277
$\mathtt{Var}_{\mathtt{C}}$	✓			**0.342**	1.292	**0.652**	1.019	0.280
Ours	✓		✓	0.337	1.300	0.649	**1.020**	**0.282**

Applications

- stickers removal

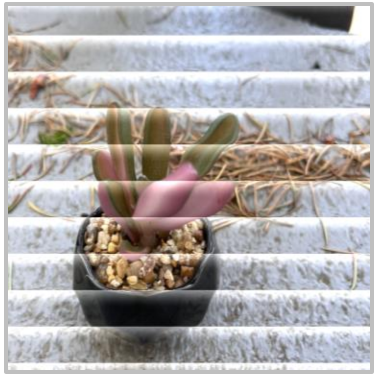
& glass effect removal

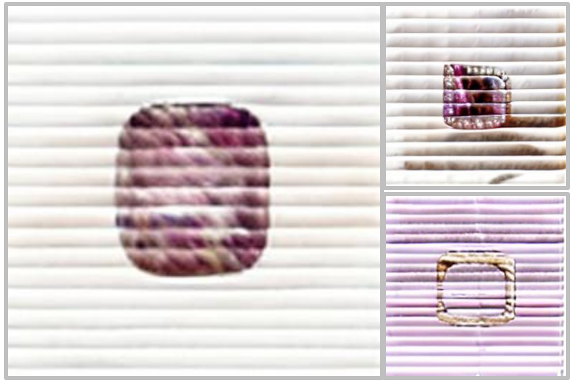
Input sample | **Ours** | Textual Inversion






Input sample | **Ours** | Textual Inversion





Thank you!