CVPR JUNE 3-7, 2026 DENVER COLORADO

See and Fix the Flaws :

Enabling VLMs and Diffusion Models to Comprehend Visual Artifacts via Agentic Data Synthesis

Jaehyun Park, Minyoung Ahn, Minkyu Kim, Jonghyun Lee, Jae-Gil Lee, Dongmin Park

KAIST

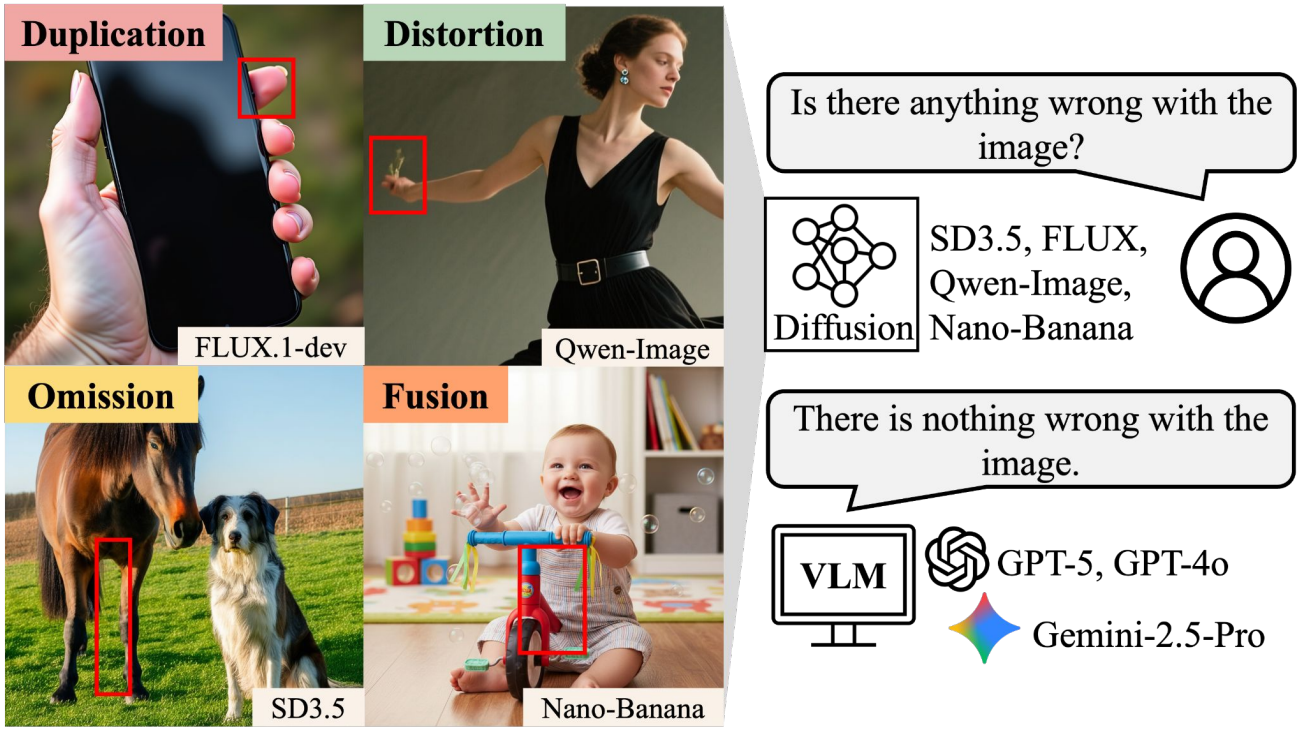
KRAFTON

VERI LUX TAS MEA

Visual Artifacts

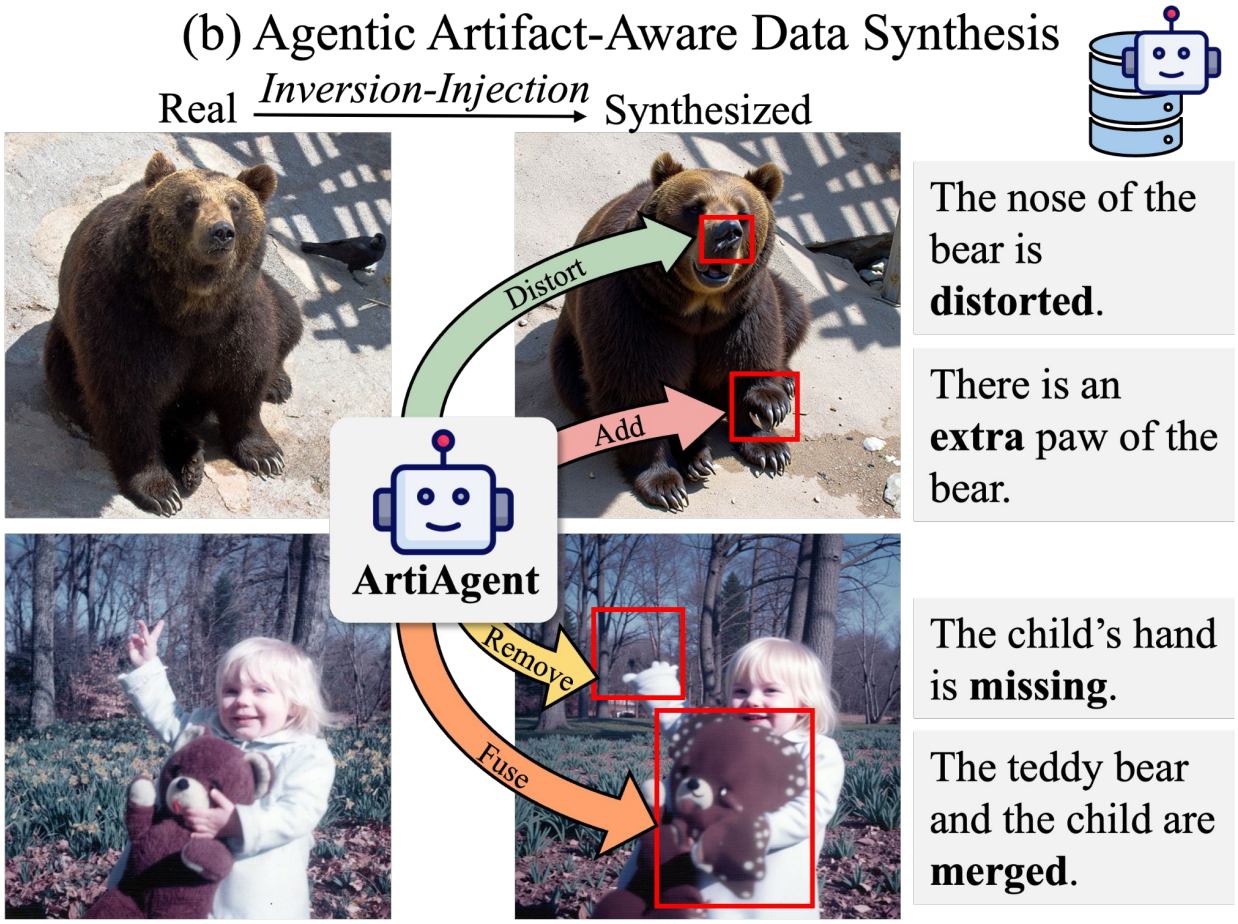
- SOTA diffusion models still generate images with **structural artifacts**, while VLMs do not successfully understand them.
- Prior works target simple artifacts generated from outdated models, while relying heavily on human-annotation.

(a) Limitations of SOTA Diffusions & VLMs



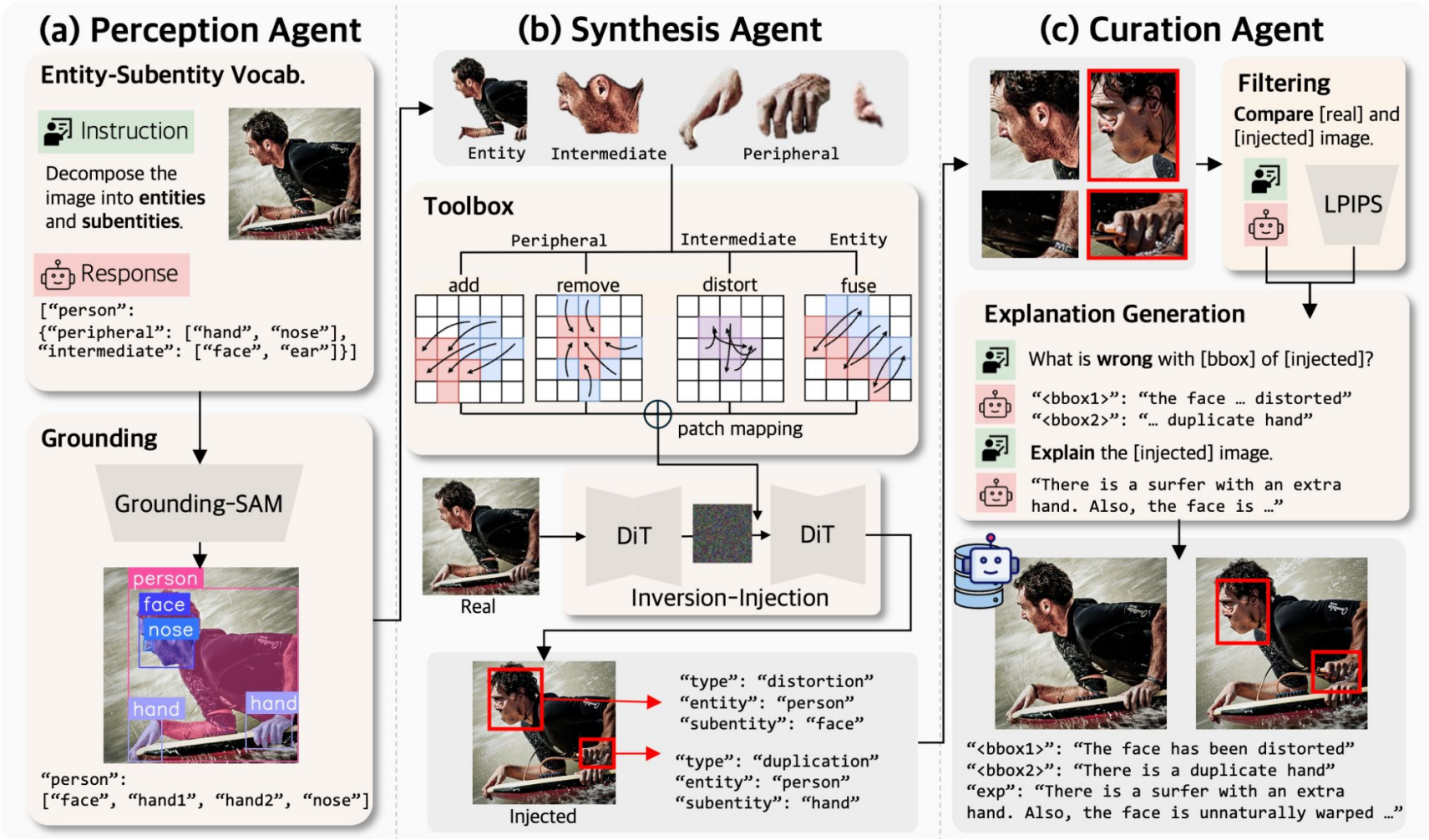
Method: ArtiAgent

- ArtiAgent automatically **injects artifacts** into real images, achieving a scalable pairwise dataset generation.



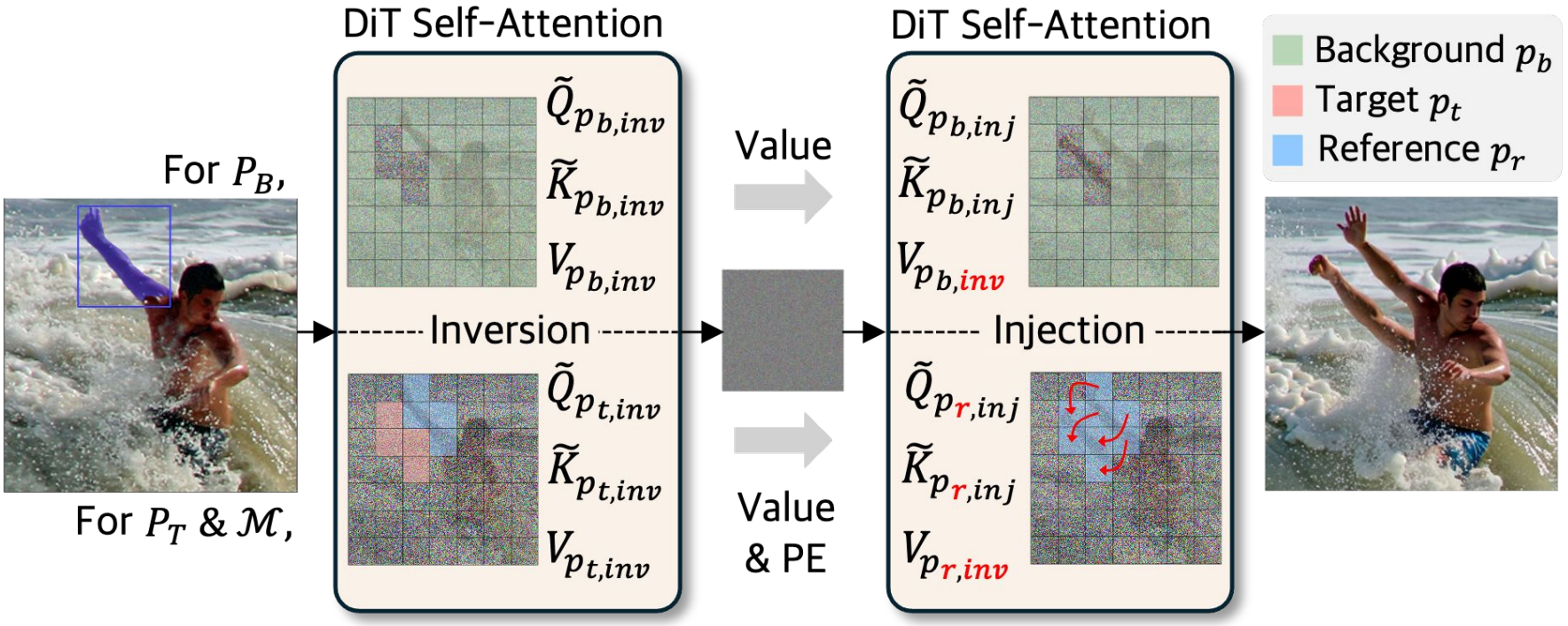
Method: ArtiAgent

- We generate a 100K paired dataset via our 3-step agentic pipeline.



Inversion-Injection Method

- We manipulate **positional information** in self-attention layers of the DiT while denoising.



ArtiBench

- We propose a 1K benchmark on **recent diffusion models**, focused on **structural artifacts** only.
- Human annotation + curation process
- Covers three tasks : binary detection / localization / explanation

Table 4. **Comparison of artifact benchmark datasets, their generative sources, and evaluation tasks.** Highlighted entries denote dataset sources that were reused by subsequent benchmarks.

Benchmark	Generative Sources	Sample	Bin.	Loc.	Exp.
RichHF [27]	SD2.1 [42], SDXL [38], Dreamlike Photoreal [10]	955		✓	
LOKI [51]	FLUX [5], SD1.4–2.1 [42], Midjourney [30], StyleGAN [22], pix2pix [18], CUT [36]	229		✓	✓
SynthScars [20]	RichHF [27], Chameleon [50], Midjourney [30], DALL·E 3 [32], SD1.x [42]	1K		✓	✓
ArtiBench	SD3.5 [11], FLUX [5], Qwen-Image [49], Nano-Banana [13]	1K	✓	✓	✓

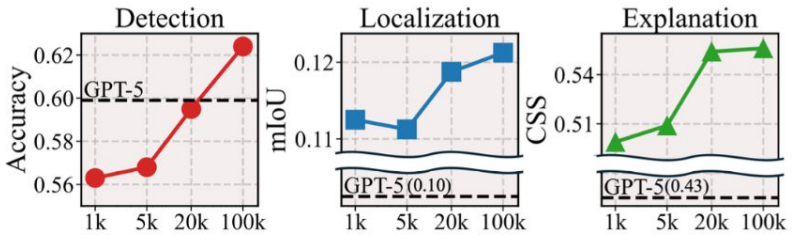
Main Results

Artifact Understanding Enhancement on VLMS

- We achieve SOTA performance on detecting, localizing, and explaining artifacts via 7B/8B models finetuned on ArtiAgent generated data.

METHOD	(A) DETECTION		(B) LOCALIZATION		(C) EXPLANATION	
	ACC	F1	MIOU	F1	R	CSS
PAL	—	—	.040	.066	—	—
DiffDoctor	—	—	.081	.137	—	—
LEGION	—	—	.062	.099	.143	.332
GPT-4o	.619	.601	.049	.084	.143	.433
Gemini-2.5-Pro	.582	.575	.095	.147	.159	.420
GPT-5	.599	.577	.061	.099	.145	.434
Qwen2.5-VL-7B	.501	.336	.010	.014	.117	.263
+ ArtiAgent	.627	**.627**	.111	.168	**.233**	**.643**
InternVL3.5-8B	.498	.357	.010	.015	.126	.256
+ ArtiAgent	**.630**	.620	**.119**	**.176**	.226	.625

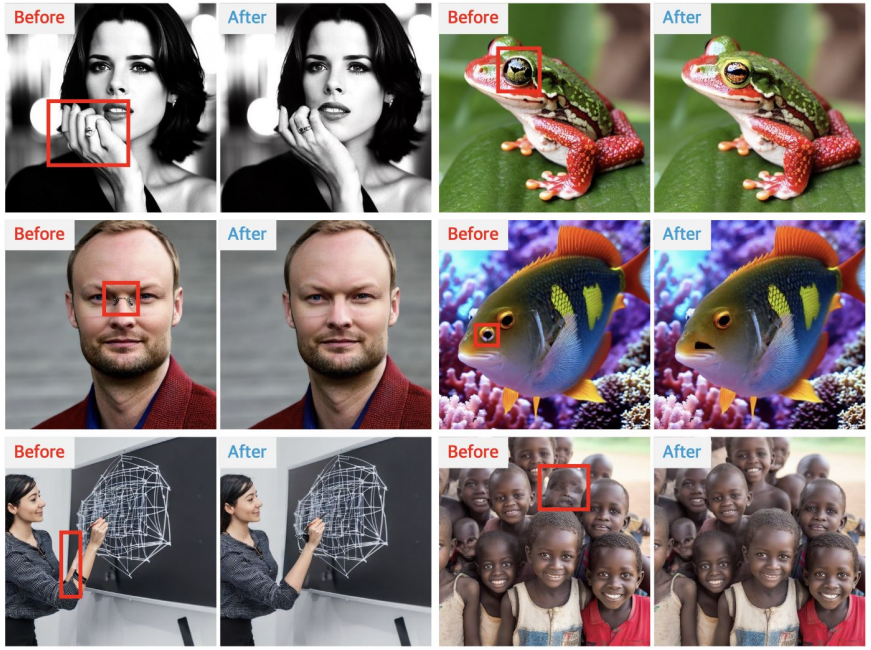
SCALES WITH ARTIAGENT TRAINING DATA



Qwen2.5-VL-7B fine-tuned on 1K→100K ArtiAgent samples crosses **GPT-5** on all 3 tasks.

Downstream Tasks

- **Reward-guided generation** : Reward model trained on an ArtiAgent generated dataset allows generation of artifact-free images
- **Image correction** : VLM trained on ArtiAgent guides image inpainting models for artifact correction



Thank you!

Paper

Dataset

https://github.com/krafton-ai/ArtiAgent