Where and When Identity Forms: Identity-Vital Attention Redistribution for Training-Free Subject-Driven Generation

LY Corporation

Luan Thanh Trinh, Kenji Doi, Atsuki Osanai (LY Corporation) {ttrinh, kedoi, atsuki.osanai}@lycorp.co.jp

NEURAL INFORMATION PROCESSING SYSTEMS

Problem: Identity Gets Lost

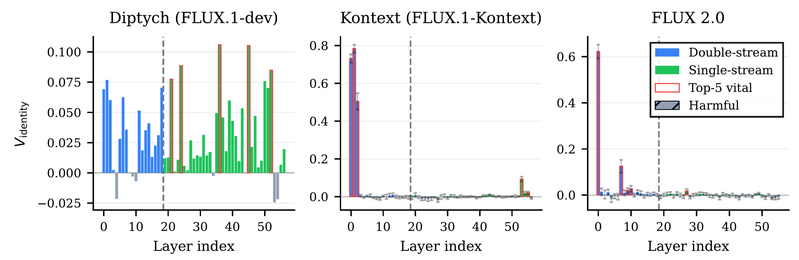
Subject-driven DiTs keep the object but lose its identity.

Reference influence is applied **globally**, but identity evidence is **local, sparse, time-dependent**. **VitalScore fixes it**: training-free, same model, same seed.

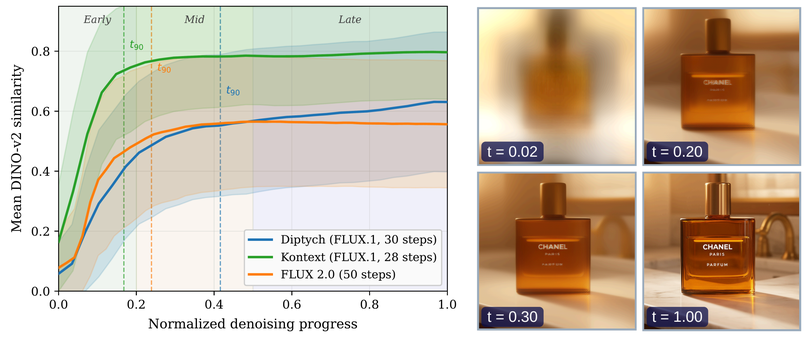


Where, When & What Identity Forms

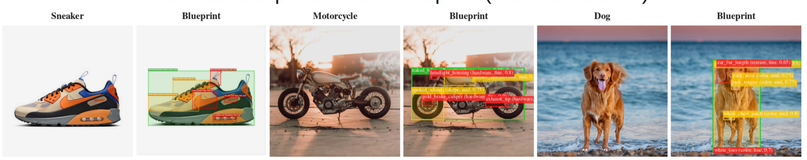
WHERE few layers, backbone-specific; ≠ edit-vital (Jaccard 0.18)



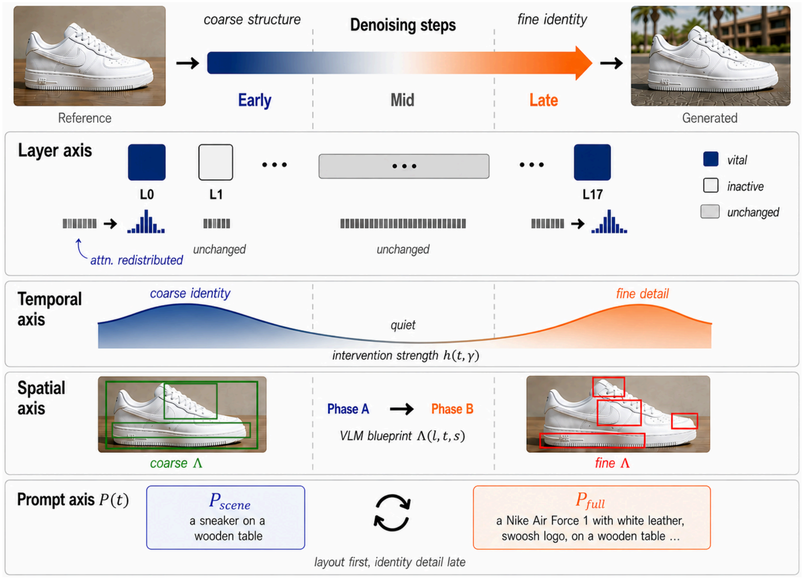
WHEN coarse identity early, fine details crystallize late



WHAT human-calibrated 2-phase VLM blueprint (feature F1 0.87)



VitalScore: Redistribute, Don't Amplify



$$\Lambda(l,t,s) = \lambda_o + g_l \cdot h(t, \gamma(s)) \cdot (M_s(s) - \lambda_o)$$

WHERE layer gate **WHEN** temporal window **WHAT** spatial blueprint

mean-normalized: same total attention budget, nothing amplified

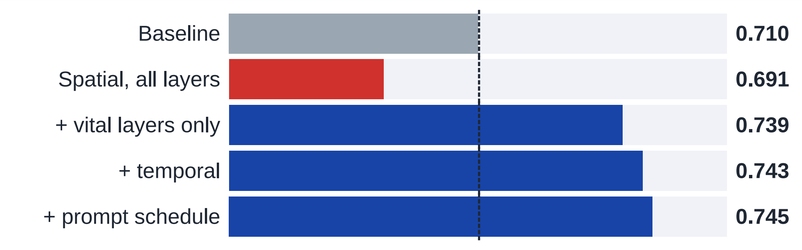
$$P(t) = P_{scene} \rightarrow P_{full} \text{ once } \rho(t) \geq \tau_p$$

scene & layout first → identity details later

1 Blueprint	2 Map	3 Schedule	4 Generate
VLM finds identity regions	boxes → latent tokens	build Λ(l,t,s), P(t)	one pass, frozen model

Training-free · **1.07–1.11× wall-clock** · **No parameter updates**

Ablation: Discovery Is What Matters



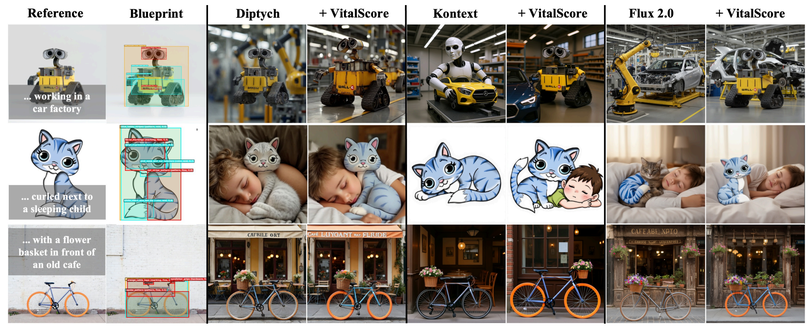
FLUX 2.0, product benchmark, DINO-v2 (dashed = baseline).

- Spatial mask on **all layers hurts** (0.710 → 0.691); on **vital layers it helps** (0.739).
- Robust: window shifts ±0.10 change DINO ≤ 0.001; ±20 box jitter keeps 81% of the gain.
- Local VLM works: Qwen3.5-9B 0.745 vs. Gemini 3.1 Pro 0.748.

Results: Consistent Gains on 5 Backbones

+15%	**+16.5%**	**80.5%**
DINO-v2 identity DreamBench++ (Diptych)	**hardest prompt axis** action prompts, product set	**human preference** overall, blinded A/B

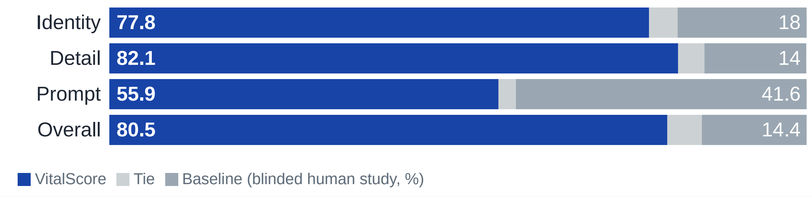
DreamBench++ 520 pairs · humans / animals / objects



Product identity benchmark 240 pairs · logos / text / texture / design



DINO-v2	DreamBench++	Product
Diptych	0.594 → **0.683** +15.0%	0.674 → **0.701** +4.0%
Kontext	0.668 → **0.759** +13.6%	0.794 → **0.850** +7.1%
FLUX 2.0	0.585 → **0.620** +6.0%	0.710 → **0.745** +4.9%
Qwen-Edit	0.655 → **0.708** +8.1%	0.689 → **0.712** +3.3%
FIBO-Edit	0.891 → **0.923** +3.6%	0.743 → **0.771** +3.8%



Probe first, then intervene.

- Identity is sparse in layers, time and space.
- Discover it once per backbone (~16 GPU-h), then redistribute.

[1] Stable Flow, CVPR'25 · [2] Diptych Prompting, CVPR'25 · [3] FLUX.1 Kontext, 2025 · [4] DreamBench++, ICLR'25

Paper