

Common Inpainted Objects In-N-Out of Context

Tianze Yang* · Tyson Jordan* · Ruitong Sun* · Ninghao Liu · Jin Sun

University of Georgia

* Equal contribution

CVPR 2026 · June 3–7, 2026 · Denver, Colorado

Code & Dataset: <https://co-in-co.github.io/>

1. Why Context Matters?

Context: a powerful but under-used cue

Humans instinctively spot objects that **don't belong** in a scene — wrong size, wrong location, or impossible co-occurrence.

But **out-of-context examples are rare** in real datasets — COCO, LVIS, COCO-Stuff all show objects in their natural settings.

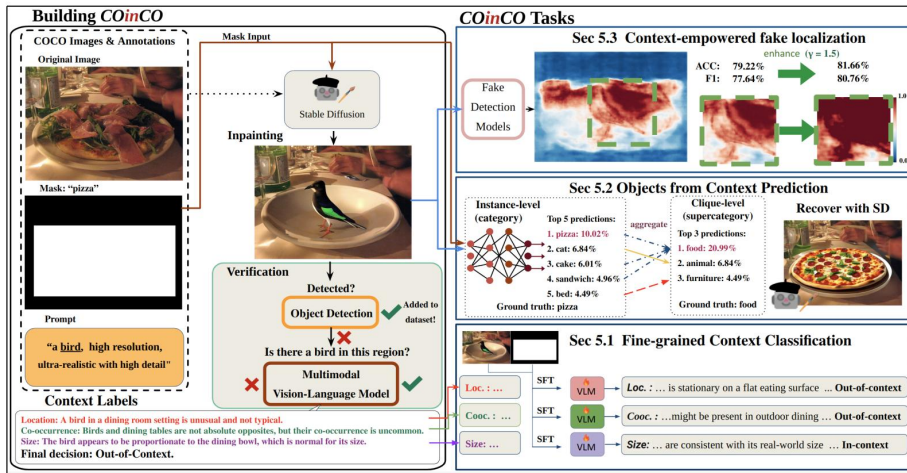
How do we **teach context at scale** when contextual violations are inherently uncommon?



Painted out-of-context (top 2 rows) and in-context (bottom row) objects.

2. Building the COinCO Dataset

Pipeline: Inpaint → Verify → Reason



1. Inpaint

Stable Diffusion replaces ONE object per COCO image with one of 80 random target classes.

2. Verify

Object Detector checks whether the new object is present; failed cases are regenerated.

3. Reason

Three LVLMs (Molmo-72B, Qwen2.5-VL-72B, InternVL3.5-38B) judge size, location, and co-occurrence.

Building COinCO: COCO image → Stable Diffusion inpainting → YOLO verification → multi-LVLM context reasoning.

Result: 97,722 successfully-inpainted images, each labelled in- or out-of-context.

2. Building the COinCO Dataset

Three criteria · Multi-LVLM consensus · Human-verified

Criteria for in / out-of-context

Location — is the object placed in a plausible spot?

Size — is its scale appropriate for the scene?

Co-occurrence — should it appear together with the surrounding objects?

Unanimous consensus = label kept

- All three LVLMs must agree on the final decision (in / out)
- 73,929 images with full criterion-level reasoning
- 1,000-image manual study → 95.10% agreement with human annotators

How COinCO differs

	COCO	COCOglide	Ours
Size	123K	512	97,722
Replacement classes	--	same	80
Out-of-context imgs	--	--	✓
Context reasoning	--	--	✓
Inpainting	--	✓	✓

97,722

inpainted images · 73,929 with full context labels

3. Tasks Enabled by COinCO

Fine-grained Context Classification

Question: *'Does this object belong in this context?'*

Distill 72B Qwen2.5-VL teacher → three **3B student models** (one per criterion).

Each student outputs a final decision *and* a natural-language explanation — modular, interpretable, deployable.

On real (un-manipulated) COCO too: **+37% avg.** on Location & Size — true context understanding, not artifact memorisation.

Accuracy (%) on COinCO test set

Criterion	Before	After	Δ
Size	55.4	65.9	+10.5
Location	67.6	75.4	+7.8
Co-occurrence	75.5	79.9	+4.4
Average	66.2	73.7	+7.6



"A cow should not be in the sky. It is impossible for a cow to appear together with the moon."
Final decision: Out-of-context.

Example explanation — a cow in the sky is flagged as out-of-context.

3. Tasks Enabled by COinCO

Object-from-Context Prediction & Context-Empowered Fake Detection

What objects fit this context?

Predict possible objects at **instance** (80 classes) and **clique** (12 super-categories) levels.

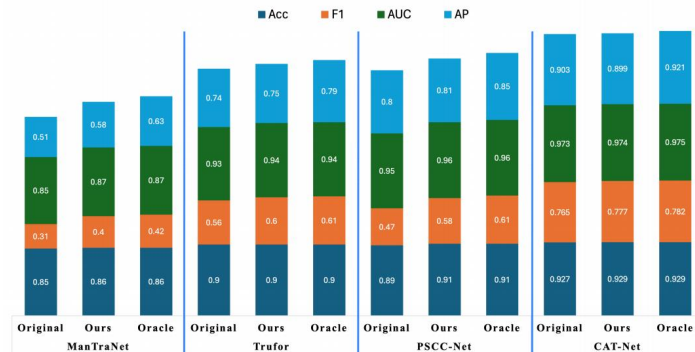
Top-1 instance: **16.32%** vs. Random 1.25 %, Co-occurrence 1.54 %.



Instance		Clique		Instance		Clique		Instance		Clique	
skis	31.6	sports	79.2	kite	22.9	sports	75.1	horse	5.0	animal	37.2
snowboard	24.3	accessory	6.6	skis	22.1	accessory	6.1	zebra	5.0	vehicle	14.9
kite	14.8	person	4.9	snowboard	19.6	vehicle	5.7	giraffe	4.9	sports	12.1
surfboard	7.0	vehicle	4.2	surfboard	8.6	person	5.5	person	4.6	outdoor	10.7
backpack	5.9	animal	3.0	person	5.5	animal	4.2	elephant	4.4	accessory	6.8

Context boosts SOTA fake detectors — without fine-tuning

- Use LVM to flag suspicious regions, then multiply prediction by $\gamma = 5$ inside those regions.
- Improves F1 / AP of CAT-Net, TruFor, PSCC-Net, and ManTra-Net — context is a free, complementary signal.



Thank You!

COinCO: a large-scale testbed for context-aware visual understanding & forensics.



Funding

National Science Foundation (SES-2521631)

University of Georgia Presidential Interdisciplinary Seed Grant

Code & Dataset: <https://co-in-co.github.io/>

CVPR 2026 · June 3–7, 2026 · Denver, Colorado