STABLE: Simulation-Ready Tabletop Layout Generation via a Semantics–Physics Dual System

Zhen Luo*[1 2], Yixuan Yang*[1 3], Xudong Xu†[3], Jinkun Hao[4], Zhaoyang Lyu[3], Feng Zheng†[1 5], Jiangmiao Pang[3], Yanwei Fu[2 6]

[1]SUSTech [2]SII [3]Shanghai AI Laboratory [4]SJTU [5]Spatialtemporal AI [6]FDU

Information

WeChat

Project Page

✓Email: luoz2024@mail.sustech.edu.cn

Background

➢ Task-Driven Scene Generation for Embodied AI

Synthetic 3D scenes provide scalable environments for training and evaluating embodied agents. A useful tabletop scene should satisfy:

- ✓ Task alignment: the scene should contain the correct objects and spatial relations specified by the instruction.
- ✓ Simulation-Ready: objects should be physically plausible, without collisions, interpenetration, or floating.

➢ Challenge

LLM-Based Generation: LLMs understand task instructions but struggle with precise continuous 3D poses.

❌ Collision Floating Interpenetration

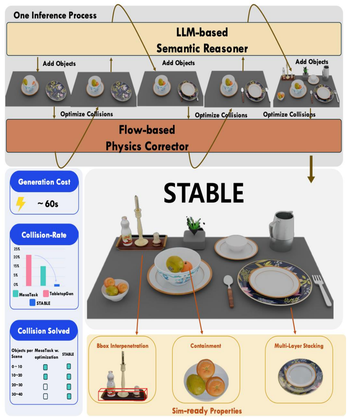
Post-hoc Optimization: Iterative optimization can reduce collisions, but may be slow, fail on severe cases, or break task-critical relations.

❌ High Cost Optimization Failure Task Misalignment

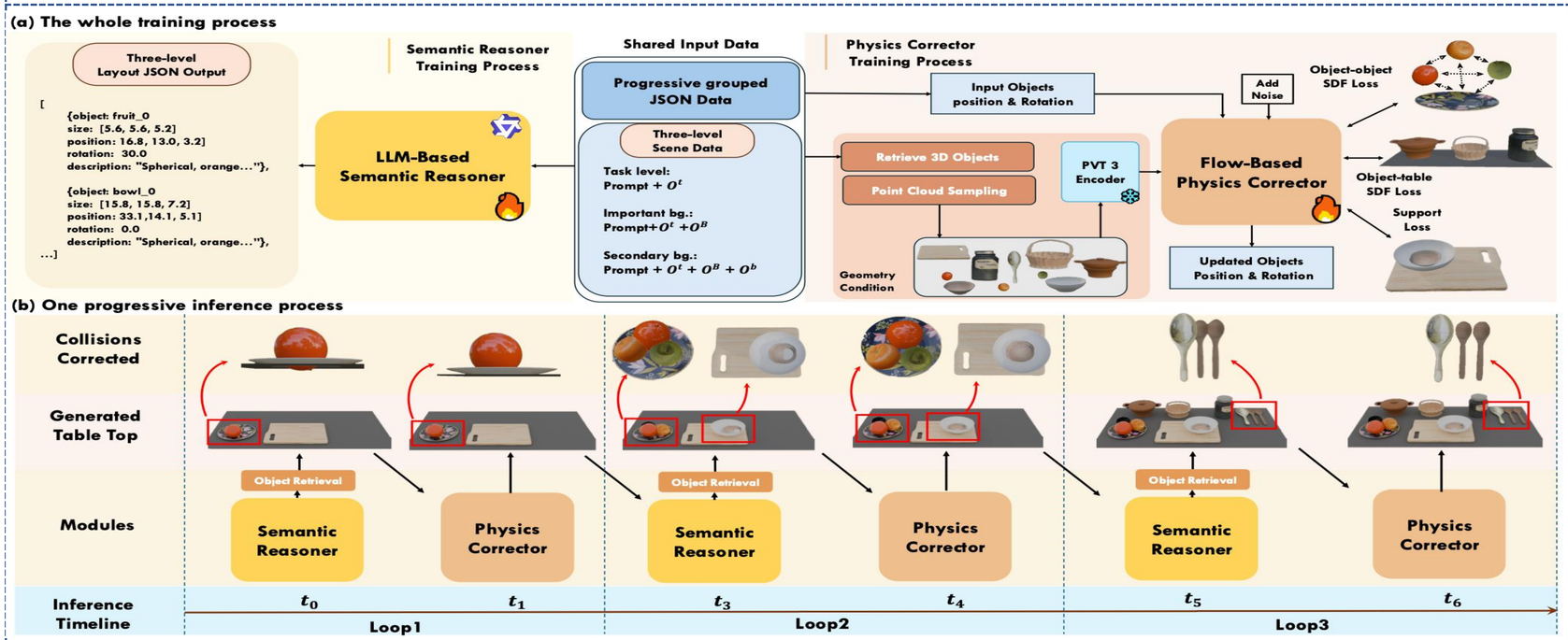
How can we preserve both task semantics and physical validity?

➢ Key Contributions

- ✓ We decouple task-aware semantic generation from physics-aware pose correction to directly generate simulation-ready tabletop layouts.
- ✓ Our Physics Corrector uses object geometry and mesh-level physical constraints to handle collision, containment, stacking, and support.
- ✓ Alternating semantic expansion and physical correction preserves task alignment while achieving collision-free, simulation-ready scenes.



STABLE Framework



Semantic Reasoner

Generates structured, task-aligned layouts from language instructions.

✓ **Better task grounding**
✓ **Fewer missing critical objects**

Physics Corrector

Refines object positions and rotations in while keeping object identity and semantics unchanged.

✓**Object–Object Collision**
✓**Object–Table Collision**
✓**Support Contact**

Progressive Dual-System Inference

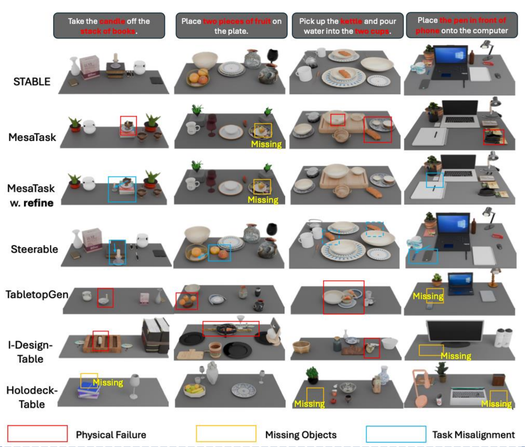
Semantic expansion and physical correction alternate at every stage.

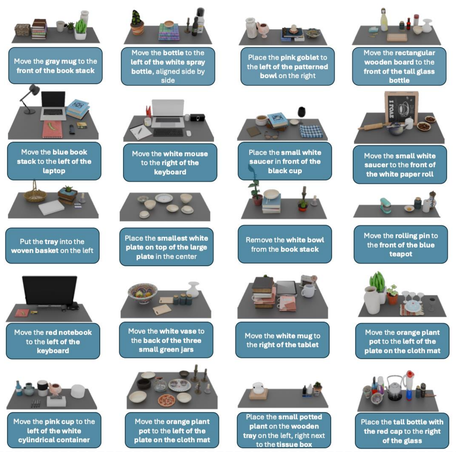
Prevent geometric errors from accumulating.

STABLE decouples semantic reasoning from physics correction, while tightly coupling them during inference.

Experiment

Results.





Quantitative comparison.

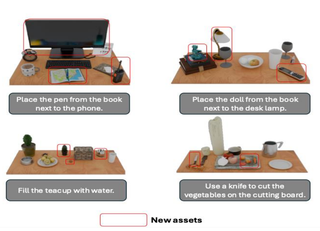
Method	FID↓	GPT Score						OC	AwT(%)	AwS(%)
		CwT	OSR	PPI	LCR	OV	Avg.			
GPT-4o	85.5	6.0	8.0	8.4	6.2	6.9	7.1	25.4	25.2	18.5
Holodeck-Table	90.2	4.2	7.0	7.8	5.0	8.5	6.5	0	18.4	11.2
I-Design-Table	94.8	5.0	8.0	8.4	5.4	7.7	6.9	10.3	16.2	9.6
MesaTask	40.6	7.8	9.2	9.4	8.6	9.0	8.8	15.6	90.2	81.5
TabletopGen	54.8	8.6	9.3	9.5	8.9	8.7	9.0	23.5	91.7	83.6
Steerable	43.7	7.5	9.2	9.7	8.7	8.8	8.8	0	99.4	91.1
Ours	38.6	9.0	9.4	9.6	9.1	8.9	9.2	0	99.4	99.0

Comprehensive Ablation Study.

Variant	OC ↓	Float ↓	AwT ↑	Distractor Rate ↑
(a) Ablation on Physical Constraints (Physics Corrector)				
w/o $\mathcal{L}_{\text{sup}}$	4.7	9.8	-	-
w/o $\mathcal{L}_{\text{obj-table}}$	13.6	5.4	-	-
w/o $\mathcal{L}_{\text{obj-obj}}$	11.9	15.8	-	-
Full (Ours)	0	0	-	-
(b) Ablation on Progressive Scene Construction (Semantic Reasoner)				
One-shot SR	-	-	89.9	78.6
Progressive SR (Ours)	-	-	99.4	86.1

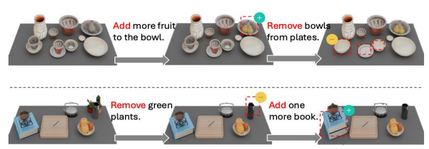
Generalization Results.

Generalization to unseen tabletop types.



Generalization to unseen object assets.

Editing Results.



Rearrange Results.