PhyScene: Physically Interactable 3D Scene Synthesis for Embodied AI

Yandan Yang*   Baoxiong Jia*   Peiyuan Zhi   Siyuan Huang†

State Key Laboratory of General Artificial Intelligence, Beijing Institute for General Artificial Intelligence (BIGAI)

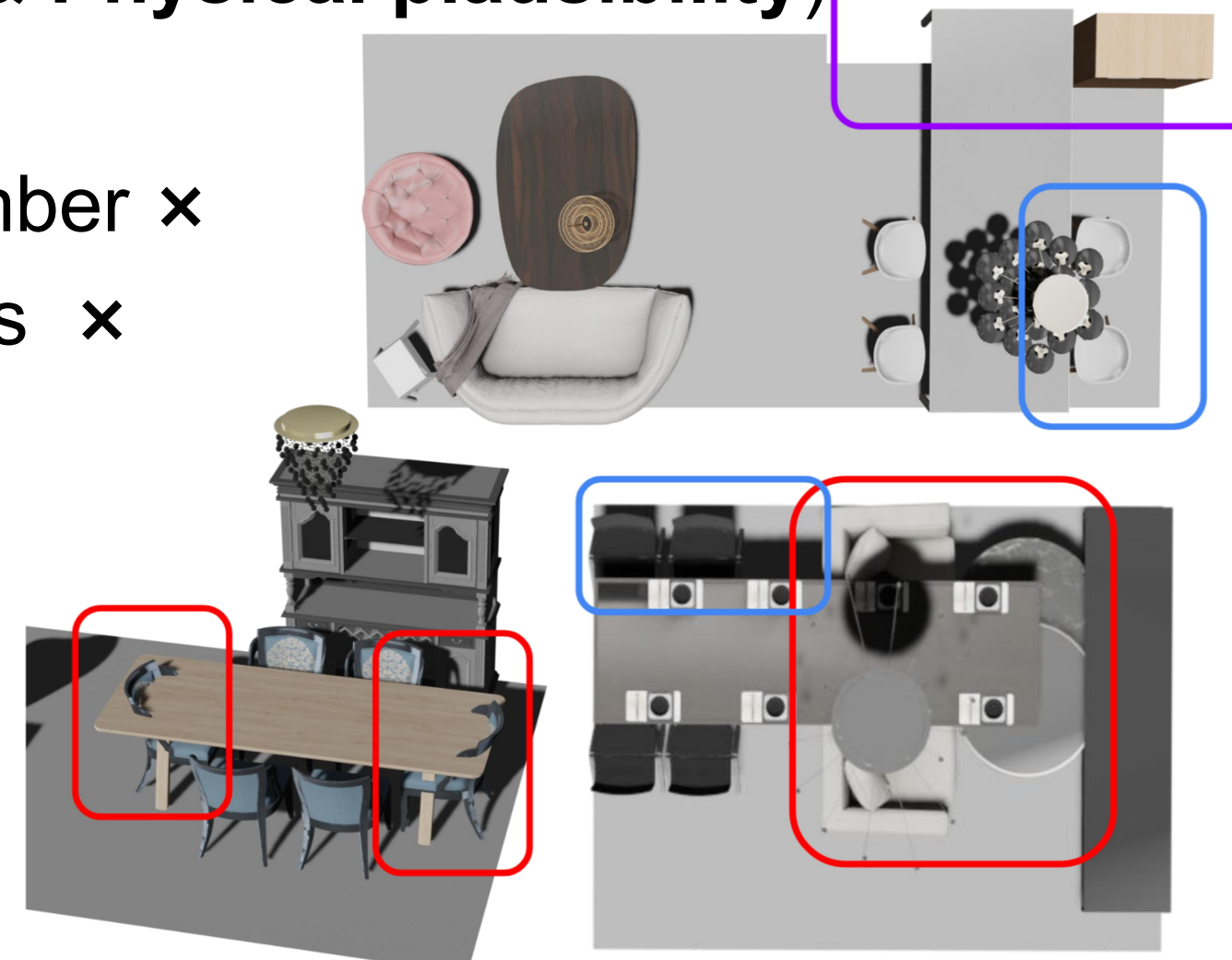* indicates equal contribution   † indicates corresponding author

CVPR
JUNE 17-21, 2024
SEATTLE, WA

https://physcene.github.io/

Overview

Motivation: Interactive 3D scene synthesis for embodied agents

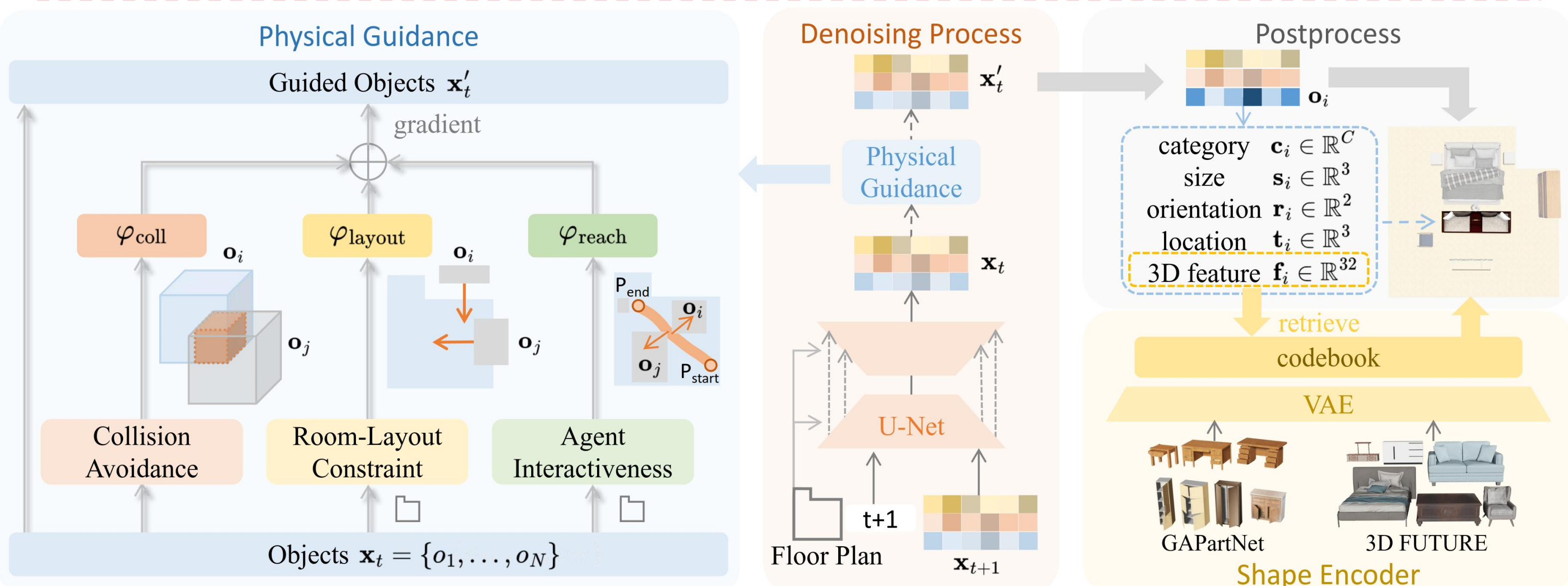**Challenge:** Lack of scene data (**Realism & Physical plausibility**)

➤ Procedurally created scenes: Unreal ✕
➤ Manually designed scenes: Limited Number ✕
➤ Scene synthesis: No physical constraints ✕

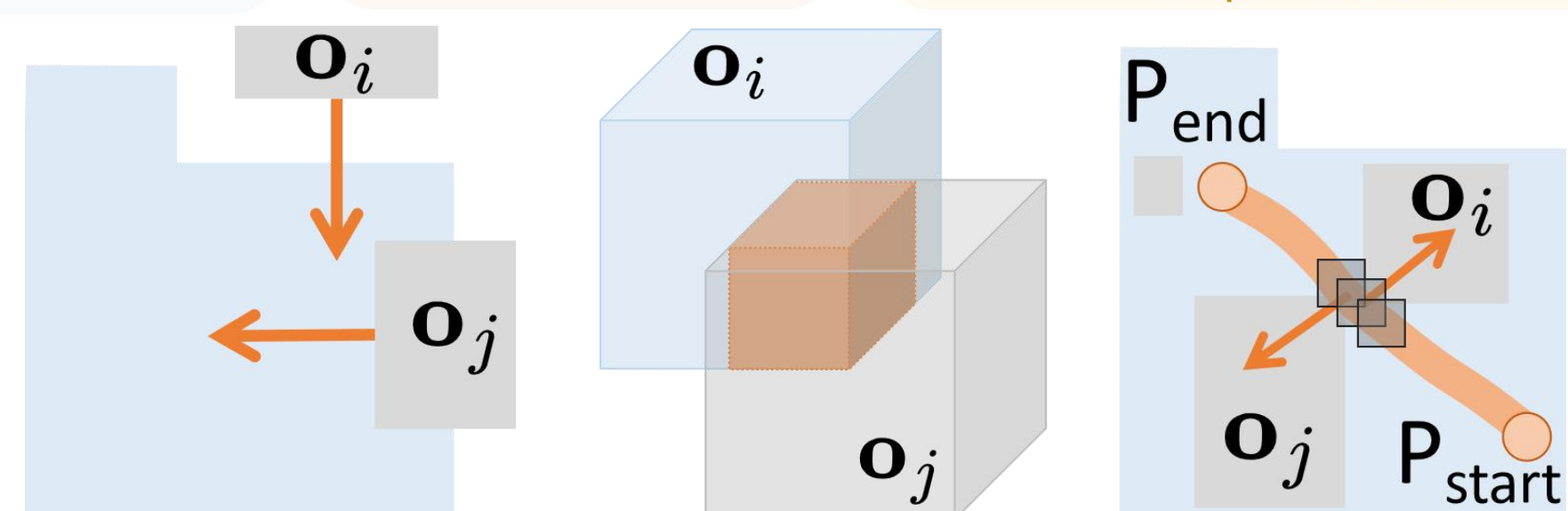| Data	Bedroom	Livingroom	Diningroom
Col_{obj} ↓	0.214	0.206	0.209
Col_{scene} ↓	0.42	0.625	0.57
R_{out} ↓	0.201	0.0584	0.159
R_{reach} ↑	0.850	0.841	0.876
$R_{walkable}$ ↑	0.749	0.828	0.807

Contributions:

1. **PhyScene** is the *first* framework for *physically interactable scene synthesis*.
2. Diffusion model + novel *physical guidance* + *retrieve articulated objects*.
3. Outperform SOTAs on *visual metrics* & our designed *physical metrics*.

Physical Guidance

➤ Collision Avoidance
➤ Room-Layout Guidance
➤ Reachability Guidance

Data

➤ 3D-FRONT for scene layout
➤ 3D-FUTURE for rigid objects
➤ GAPartNet for **articulated objects**
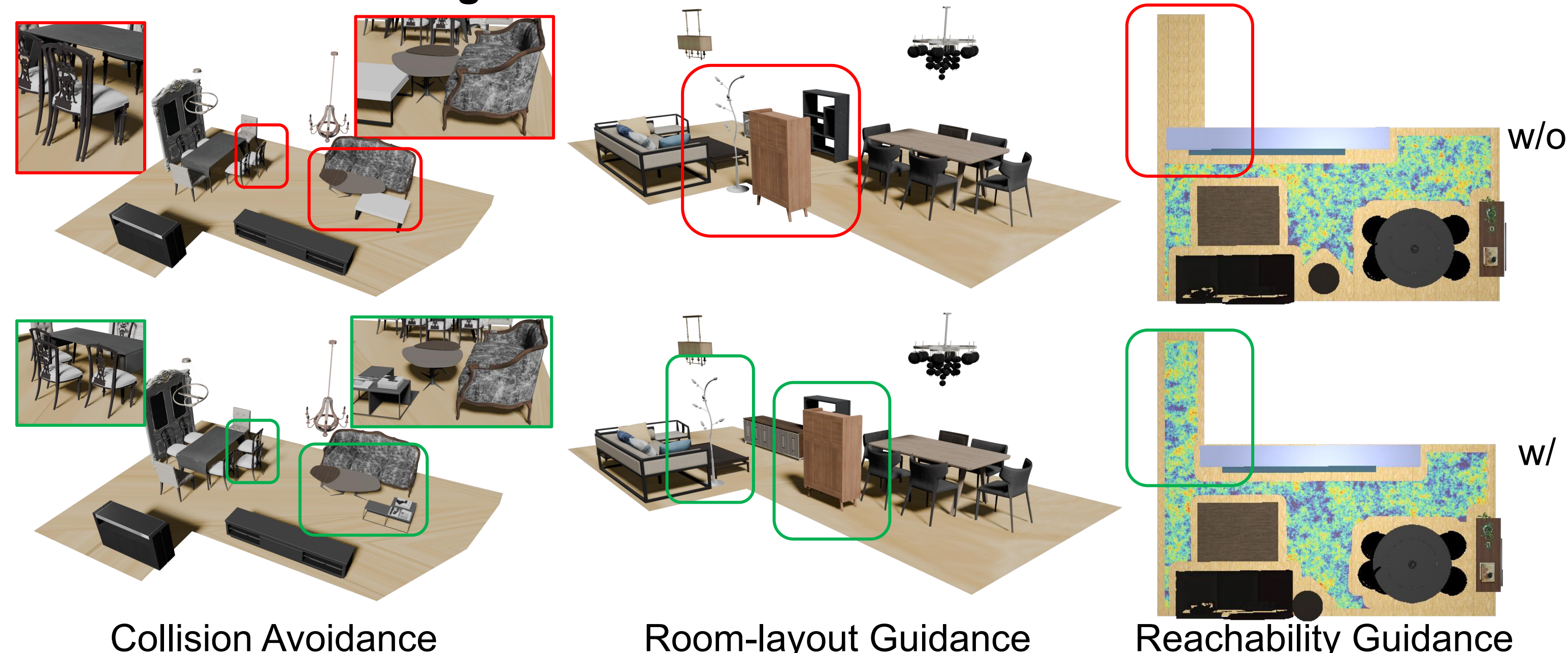

Metric

➤ **Visual Metric** (realism & diversity)
  • FID, KID, SCA, CKL
➤ **Physical Metric**
  • Collision rate: Col_{obj}, Col_{scene}
  • Objects outside the floor plan: R_{out}
  • Reachable rate of objects: R_{reach}
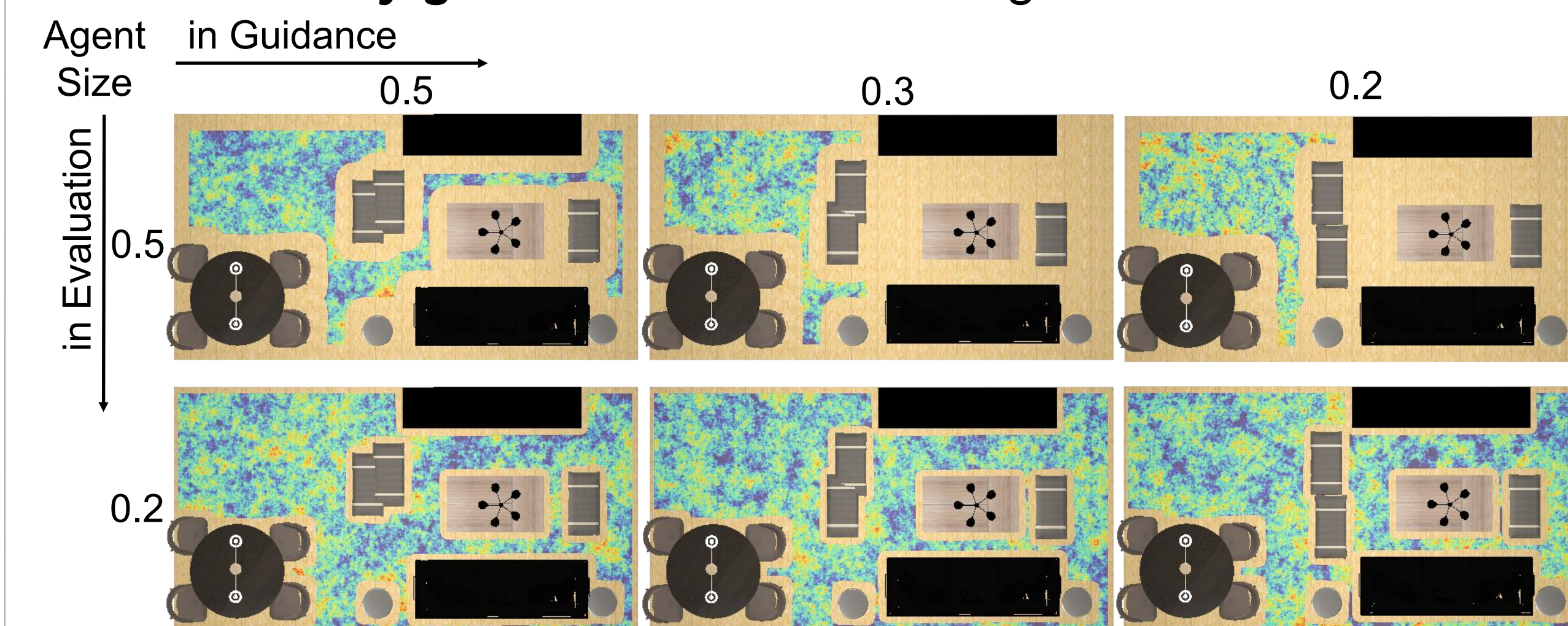 • Ratio of walkable area: $R_{walkable}$

Experiments

➤ visual metrics and physical metrics ☆means better than training dataset

Room Type	Method	FID ↓	KID ↓	SCA ↓	CKL ↓	Col_{obj} ↓	Col_{scene} ↓	R_{out} ↓	$R_{walkable}$ ↑	R_{reach} ↑
Bedroom	ATISS	30.19	0.0010	49.14	0.0028	0.248	0.46	0.286	0.839	0.736
	DiffuScene	**25.00**	**0.0004**	51.78	0.0031	0.228	0.43	0.272	0.827	0.755
	PhyScene (Ours)	25.52	0.0006	**50.10**	**0.0025**	**0.187** ☆	**0.36** ☆	**0.245**	**0.865** ☆	**0.762** ☆
Living Room	ATISS	45.66	0.0035	**51.64**	0.0016	0.316	0.85	**0.136**	0.814	**0.791**
	DiffuScene	**38.69**	**0.0012**	54.06	0.0017	0.198	0.69	0.238	0.790	0.756
	PhyScene (Ours)	43.33	0.0031	53.50	**0.0015**	**0.191** ☆	**0.63**	0.219	**0.815**	0.771
Dining Room	ATISS	41.66	0.0039	64.57	0.0040	0.591	0.96	**0.132**	**0.874**	**0.848**
	DiffuScene	**38.31**	**0.0020**	60.19	**0.0013**	0.160	0.55	0.244	0.787	0.847
	PhyScene (Ours)	39.90	0.0026	**60.00**	**0.0013**	**0.151** ☆	**0.53** ☆	0.217	0.852	0.789

➤ Effectiveness of *three guidances*

Collision Avoidance   Room-layout Guidance   Reachability Guidance

Reachability guidance with different agent sizes.

Agent Size — in Guidance → 0.5   0.3   0.2 / in Evaluation 0.5, 0.2

PhyScene in Simulation

➤ Embed articulated object
➤ Import scenes in Isaac Sim
➤ Evaluate with agent

Future Work

• Diversity. Generate more room types & add small objects.
• Interaction. Evaluate on specific planning & interaction tasks.
• Representation. Find a better/general object representation.