

From Gaussian Splats to Interactive Worlds: Automated Physics-Ready Scene Construction for Robot Manipulation

Motivation

Training robust manipulation policies in simulation requires interactive environments where the robot physically grasps, pushes, and collides with objects. Current approaches either hand-author these scenes (expensive, weeks per environment) or use Gaussian Splatting reconstructions as visual backdrops without physics interaction (SplatSim, RoboSplat). The result: visually rich but physically inert worlds that cannot train contact-rich skills. Meanwhile, the individual tools to bridge this gap -- semantic GS segmentation, differentiable mesh extraction, VLM-based physics estimation, and generative 3D asset creation -- now exist but have never been connected into a single automated pipeline for robotics.

Research Question

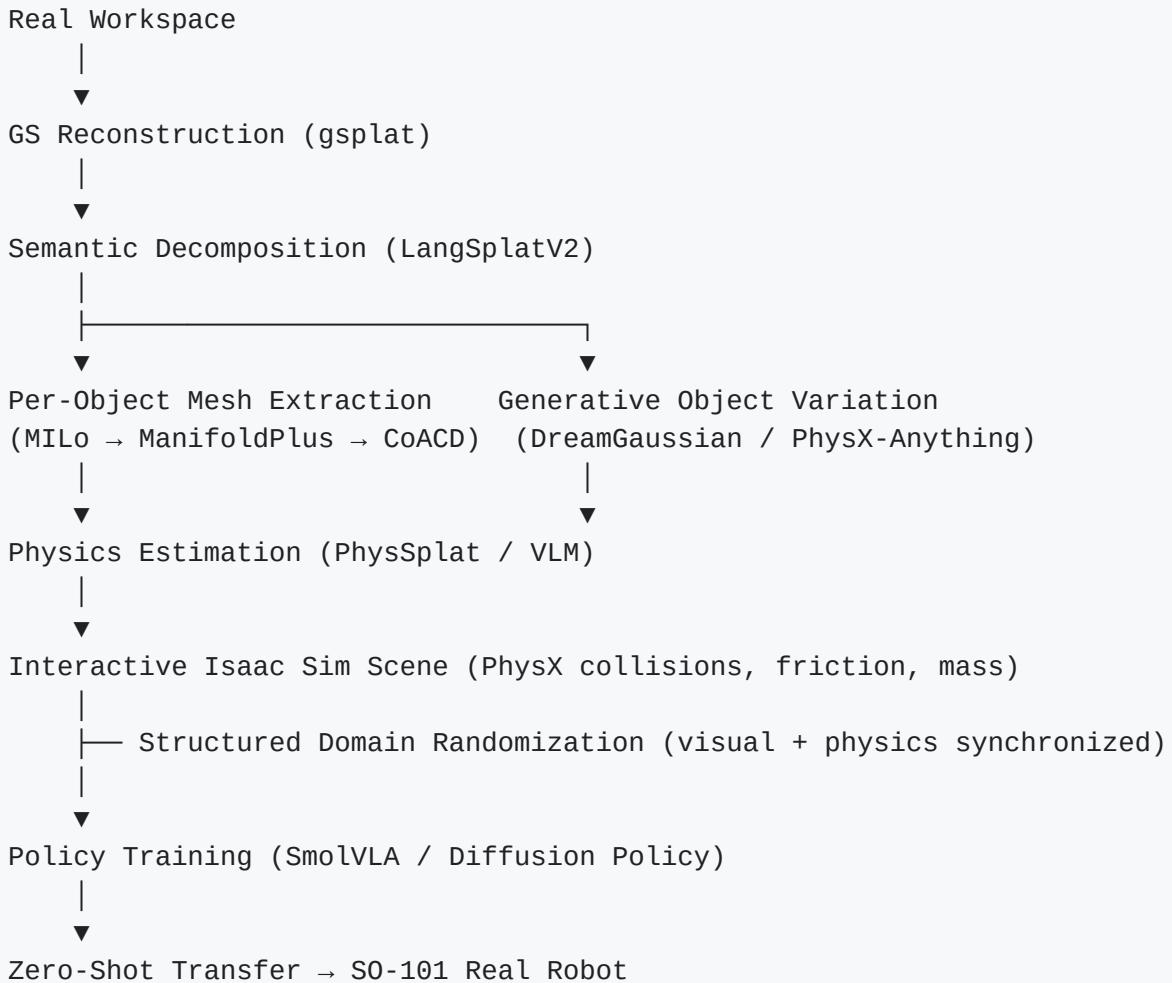
Can a Gaussian Splatting scan of a real workspace be automatically transformed into a fully interactive simulation environment -- with per-object collision meshes, estimated physics properties, and generative scene variations -- that produces manipulation policies transferable to the real world?

The Gap

What exists	What's missing
GS visual augmentation achieves 86-88% sim2real (SplatSim, RoboSplat RSS'25)	Visual-only -- no collision, no grasping, no force feedback
LangSplatV2 segments GS scenes at 384 FPS (NeurIPS'25)	Never used for per-object mesh extraction in robotics
MILO extracts meshes 10x more efficiently during GS training (SIGGRAPH Asia'25)	Never evaluated for physics simulation quality
PhysSplat estimates material properties via VLM in zero-shot (ICCV'25)	Uses custom MPM solver, not PhysX; accuracy unvalidated
PhysX-Anything generates sim-ready URDF from images (Nov'25)	No grasp success evaluation; image-only, no text input
World Labs Marble exports collider meshes to Isaac Sim (Jan'26)	No per-object physics properties; coarse geometry
RoboSimGS builds hybrid GS+mesh scenes with VLM physics (Oct'25)	Slow pipeline, no structured DR framework, no scaling analysis

The central finding: every component exists in isolation. Nobody has connected them end-to-end and measured whether the result supports contact-rich robot learning.

Proposed Pipeline



Key Technical Sub-Questions

- 1. Mesh quality threshold.** What is the minimum collision mesh quality from GS extraction/generation for stable grasp simulation in PhysX? MILo and SuGaR produce rendering-quality meshes, but nobody has measured whether they survive convex decomposition and contact simulation. This is the gate -- if meshes fail, the pipeline fails.
- 2. Automated physics estimation accuracy.** Can VLM-inferred physics properties (mass, friction, stiffness) from GS renders replace manual specification for manipulation training? Phys2Real showed 100% vs 60% over domain randomization on pushing, but only tested 2 simple objects. We need validation on diverse grasping tasks.
- 3. Generative variation for training diversity.** Does inserting parametrically varied generated objects (size, color, shape, deformability) into a GS-reconstructed scene improve policy transfer compared to standard domain randomization alone? Agriculture -- where every tomato differs in size, ripeness, and firmness -- is the ideal stress test.

Expected Contributions

1. **First empirical study** of GS-extracted mesh quality for contact simulation in Isaac Sim/PhysX
2. **First automated end-to-end pipeline** from GS scan to interactive training environment for manipulation
3. **Quantified comparison** of automated GS-based scenes vs hand-authored simulation for policy transfer
4. **Scaling analysis** of generative scene variation vs policy robustness for agricultural manipulation

Timeline

Phase	Weeks	Deliverable
Mesh quality threshold study	1-4	Go/no-go on mesh extraction approach
Full pipeline integration + DR comparison	5-10	Pipeline vs hand-authored baseline
Generative variation experiment	11-14	Parametric diversity ablation
Paper writing	14-16	Paper

Key References

- SplatSim (CMU, 2024) -- GS rendering for sim2real, 86.25% zero-shot
- RoboSplat (RSS 2025) -- GS augmentation, 87.8% one-shot, visual-only
- RoboSimGS (Oct 2025) -- hybrid GS+mesh with VLM physics, closest prior work
- PhysSplat (ICCV 2025) -- MLLM zero-shot physics estimation
- MILO (SIGGRAPH Asia 2025) -- differentiable mesh extraction from GS
- PhysX-Anything (Nov 2025) -- single image to sim-ready URDF
- PhysXGen (NeurIPS 2025 Spotlight) -- physics-grounded 3D generation
- LangSplatV2 (NeurIPS 2025) -- 384 FPS semantic GS segmentation
- World Labs Marble (Jan 2026) -- generative 3D worlds with Isaac Sim integration