SimuScene: Simulation-Ready Compositional 3D Scene Reconstruction from a Single Image

Inhee Lee* Sangwon Baik* Sungjoo Kim Hyeonwoo Kim Hyunsoo Cha Hanbyul Joo†

Seoul National University

Reconstructing 3D from an image

Recent 3D reconstruction approaches



SAM3D [1]



VGGT [2]

[1] S. 3D Team et al., “SAM 3D: 3Dfy Anything in Images,” in *preprint*, 2025

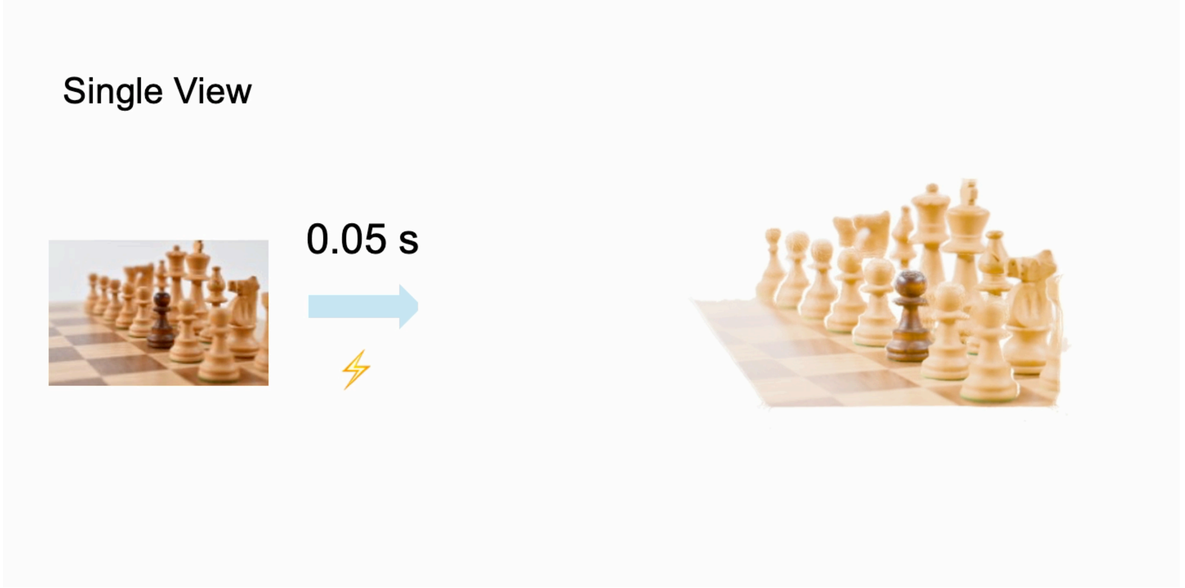
[2] J. Wang, M. Chen, N. Karaev, A. Vedaldi, C. Rupprecht, and D. Novotny, “Vggt: Visual geometry grounded transformer,” in CVPR, 2025

Reconstructing 3D from an image

Recent 3D reconstruction approaches



SAM3D



VGGT

These are not simulation-ready, and has limitation on downstream application

Inaccurate Pose



Input Image

Inaccurate Shape



Inaccurate Pose



SAM3D Output

Inaccurate Shape



Inaccurate Pose



SAM3D Result

Inaccurate Shape



Inaccurate Pose



SAM3D Result

Inaccurate Shape



Inaccurate Pose



SAM3D Result

Inaccurate Shape



Inaccurate Pose



SAM3D Result

Inaccurate Shape



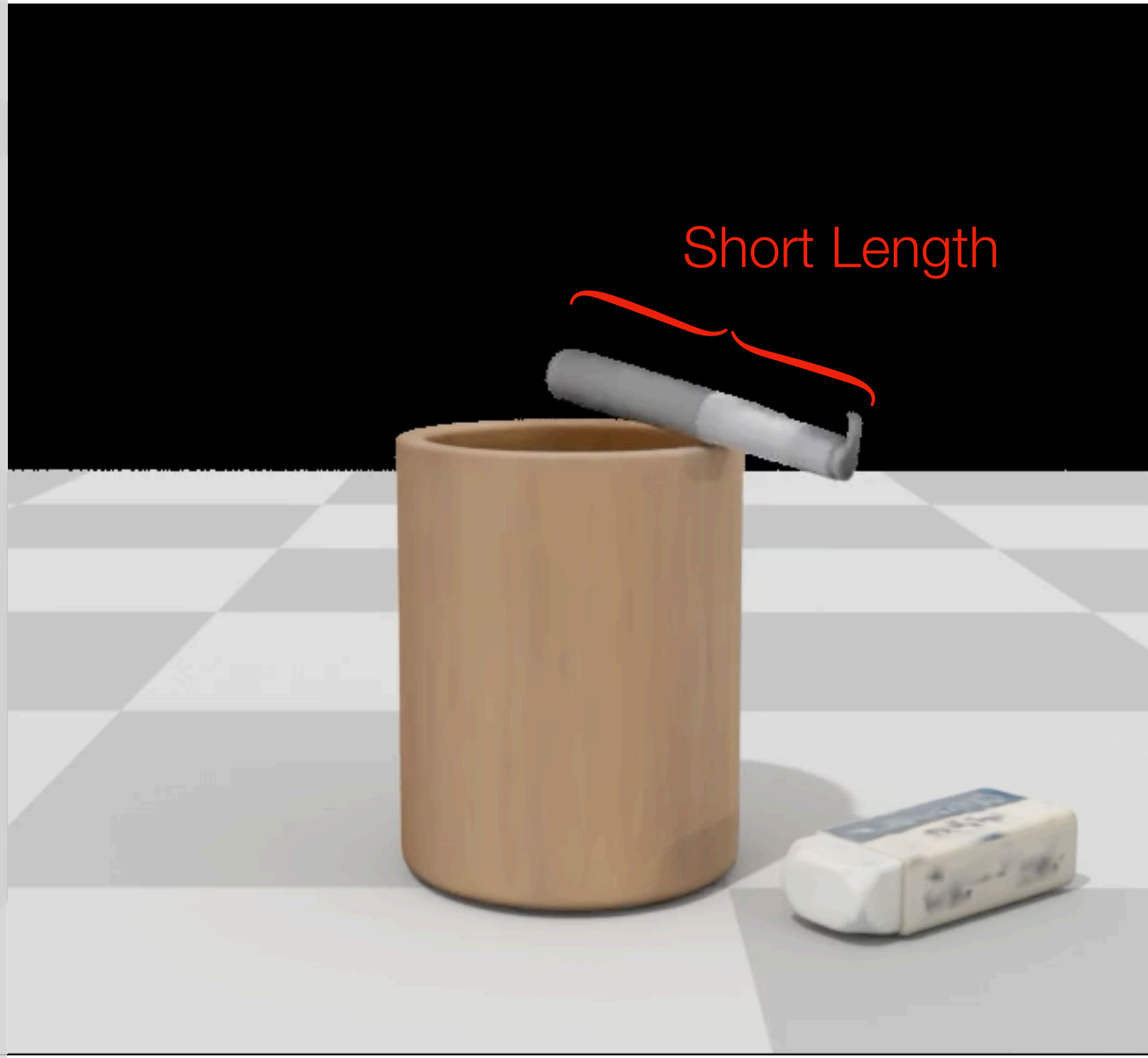
SAM3D Result

Inaccurate Pose



SAM3D Result

Inaccurate Shape



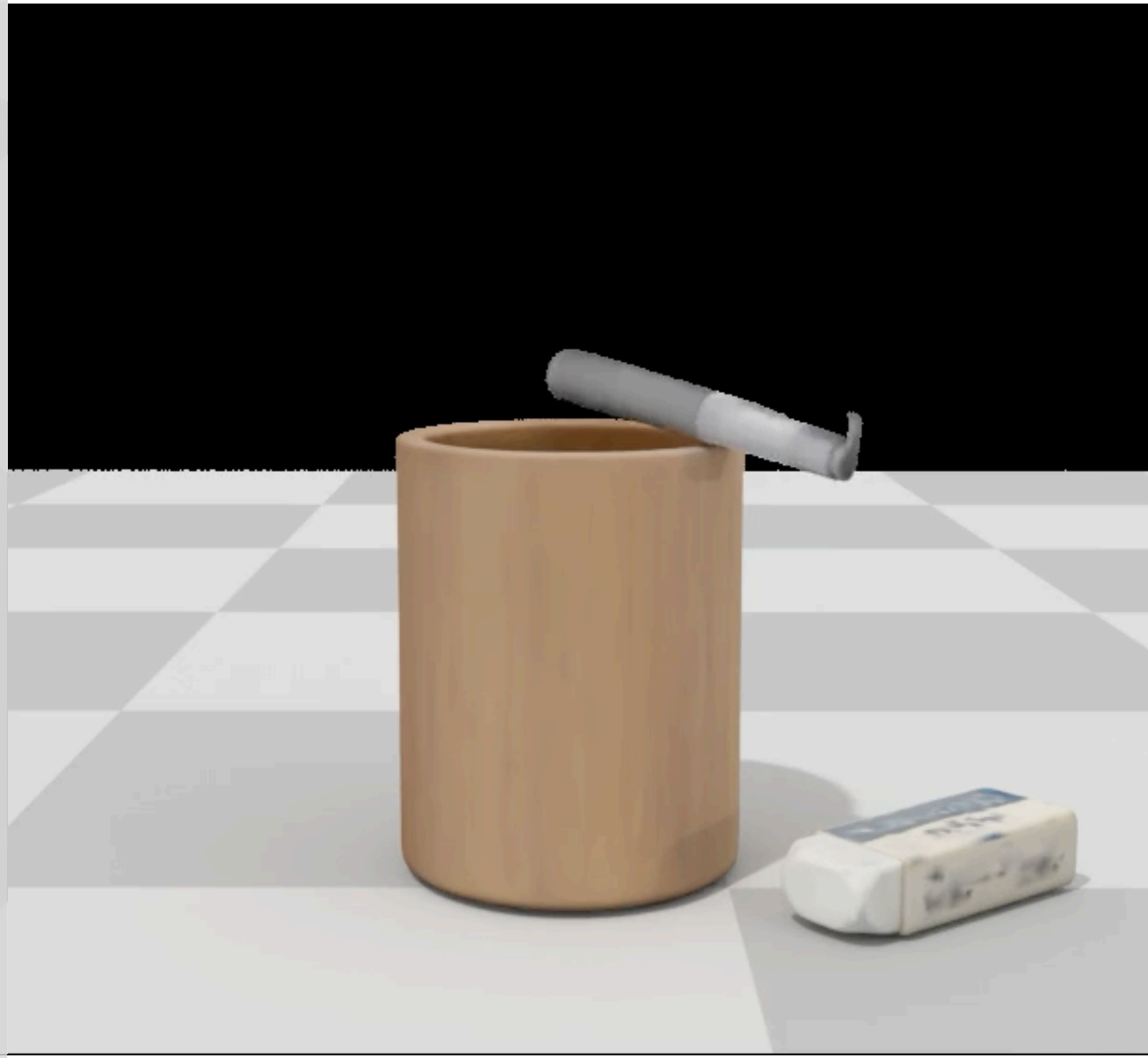
SAM3D Result

Inaccurate Pose



SAM3D Result

Inaccurate Shape



SAM3D Result

Inaccurate Pose

Inaccurate Shape

What we need is…

- **Pose Refinement**

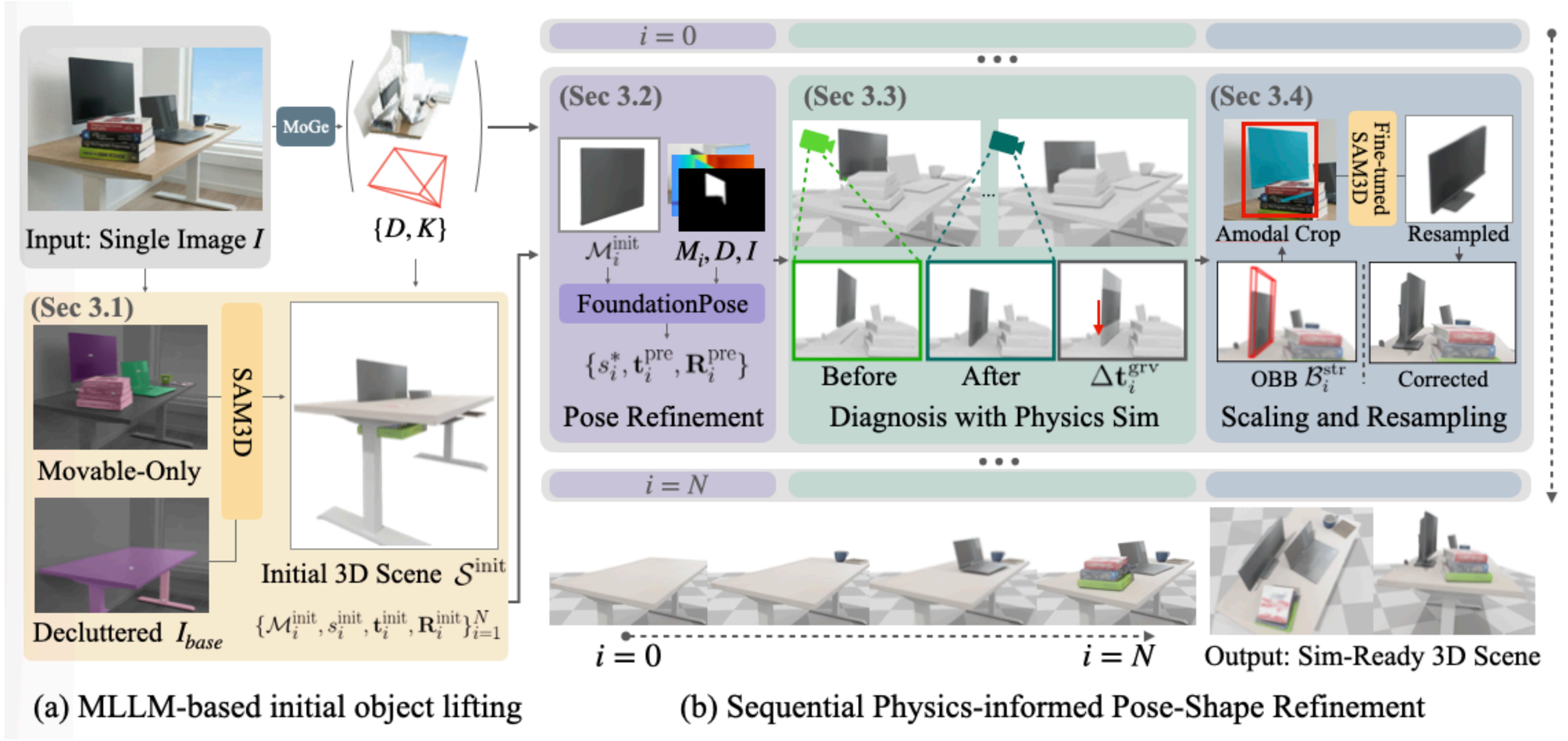
- **Shape Refinement**

SAM3D Result

SAM3D Result

Our Approach

Integrate collision + Diagnosis Implausible Pose by Physics



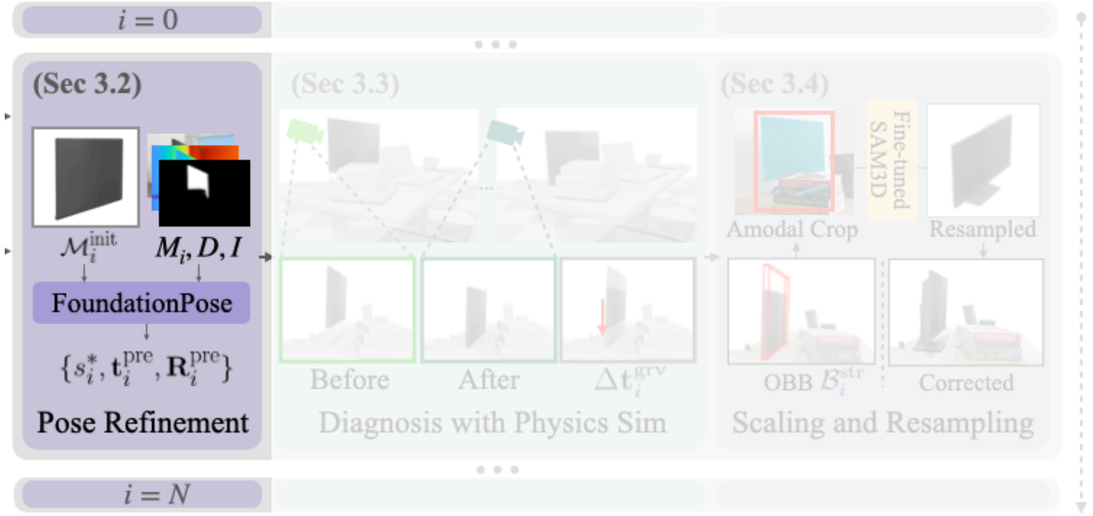
(a) MLLM-based initial object lifting

(b) Sequential Physics-informed Pose-Shape Refinement

Our Approach

Integrate collision + Diagnosis Implausible Pose by Physics



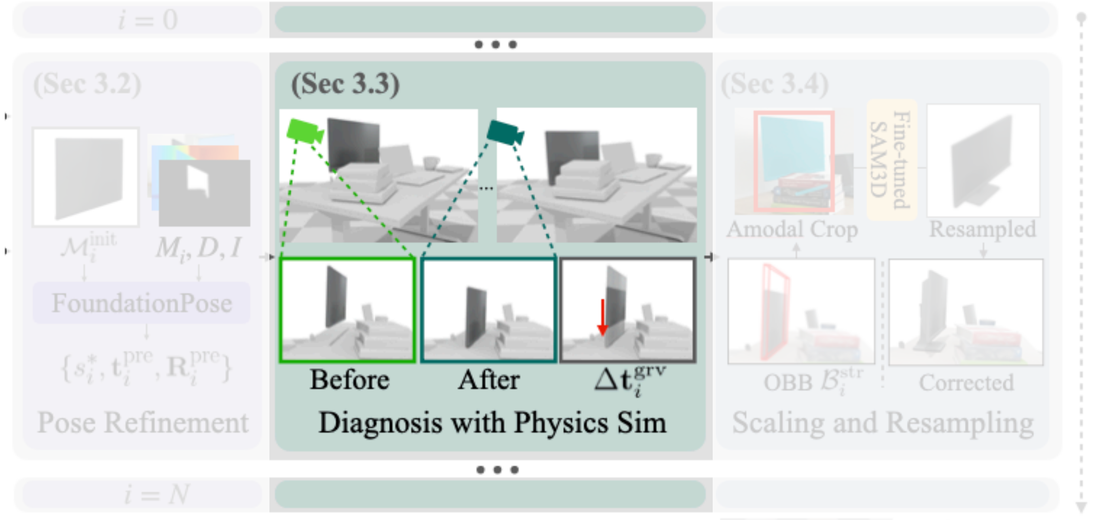


1. Add object in scene one by one w/ pose-refinement

Our Approach

Integrate collision + Diagnosis Implausible Pose by Physics



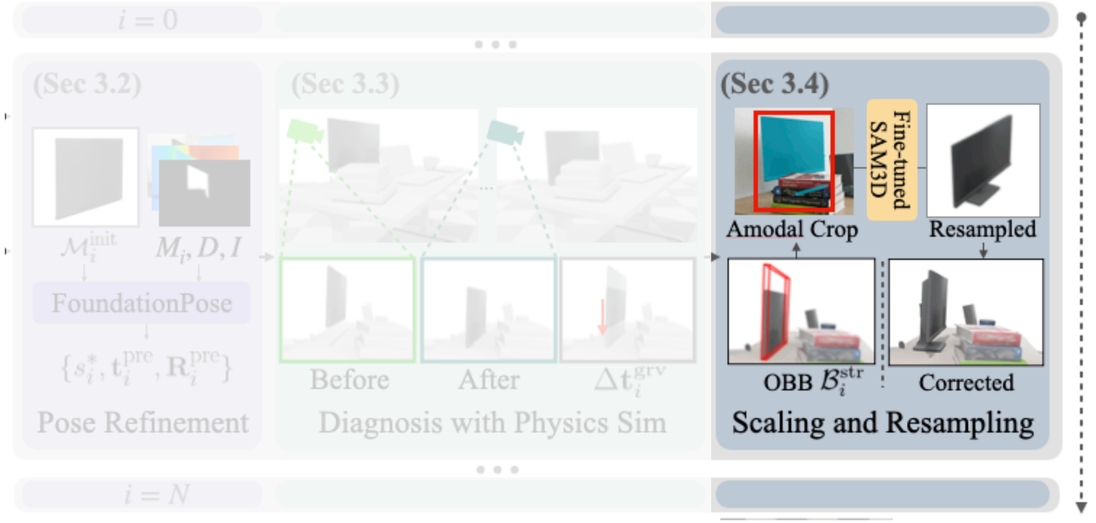


2. Apply Diagnostic Physics Simulation (include penetration resolve)

Our Approach

Integrate collision + Diagnosis Implausible Pose by Physics



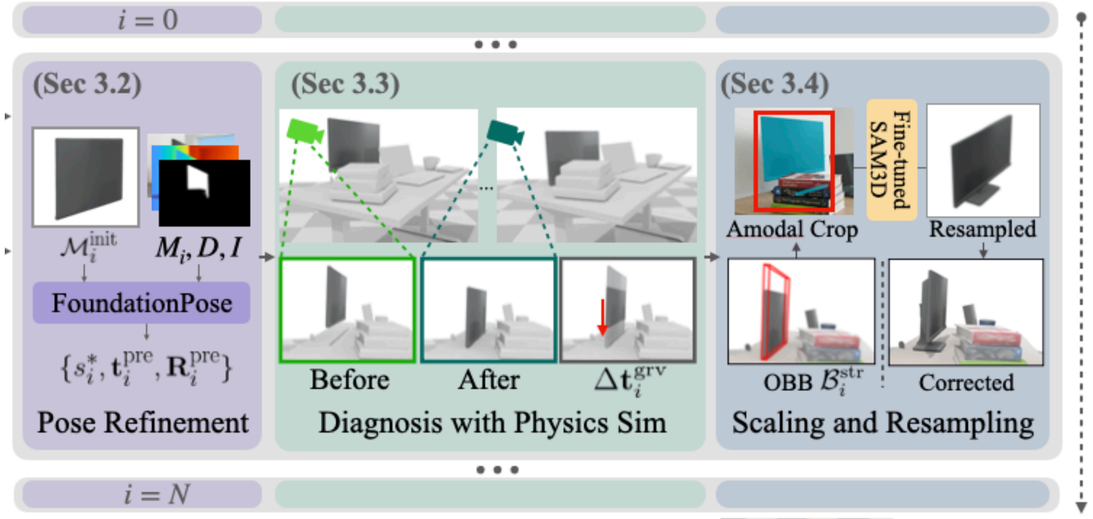


3. Adjust shape by scaling

Our Approach

Integrate collision + Diagnosis Implausible Pose by Physics



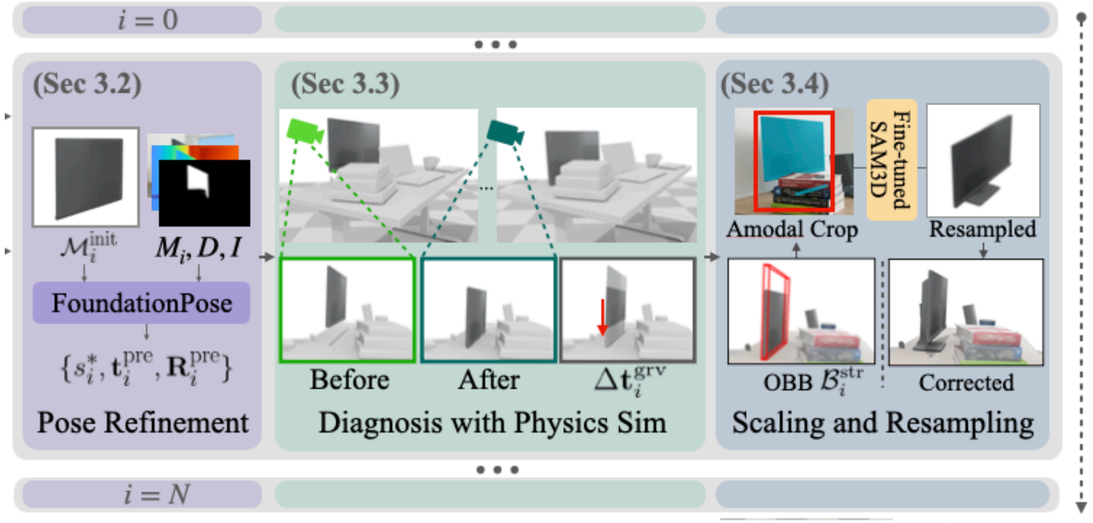


Process remaining objects sequentially

Our Approach

Integrate collision + Diagnosis Implausible Pose by Physics





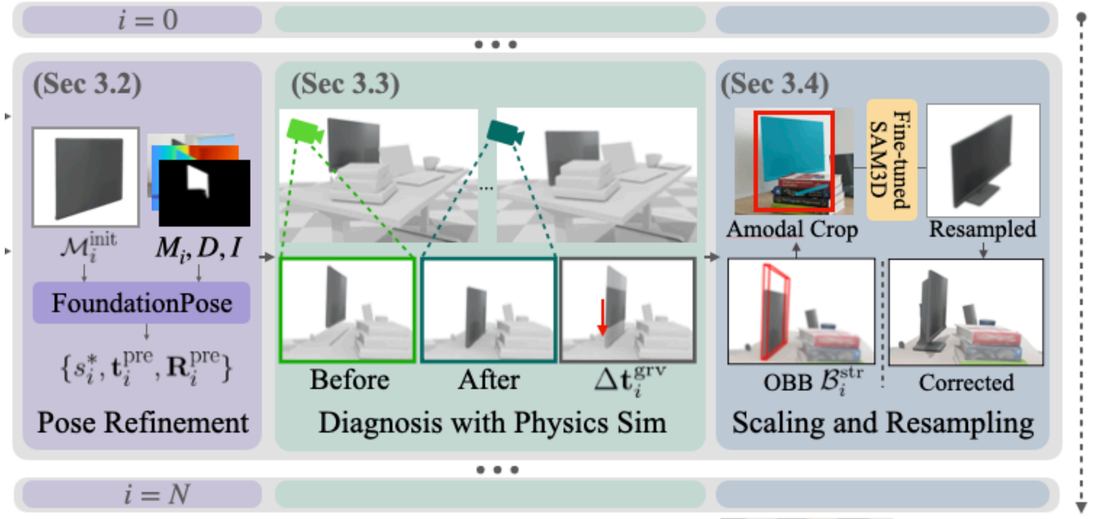
Having inaccurate “shape” need resampling

Our Approach

Integrate collision + Diagnosis Implausible Pose by Physics

Monitor doesn’t have “stand"





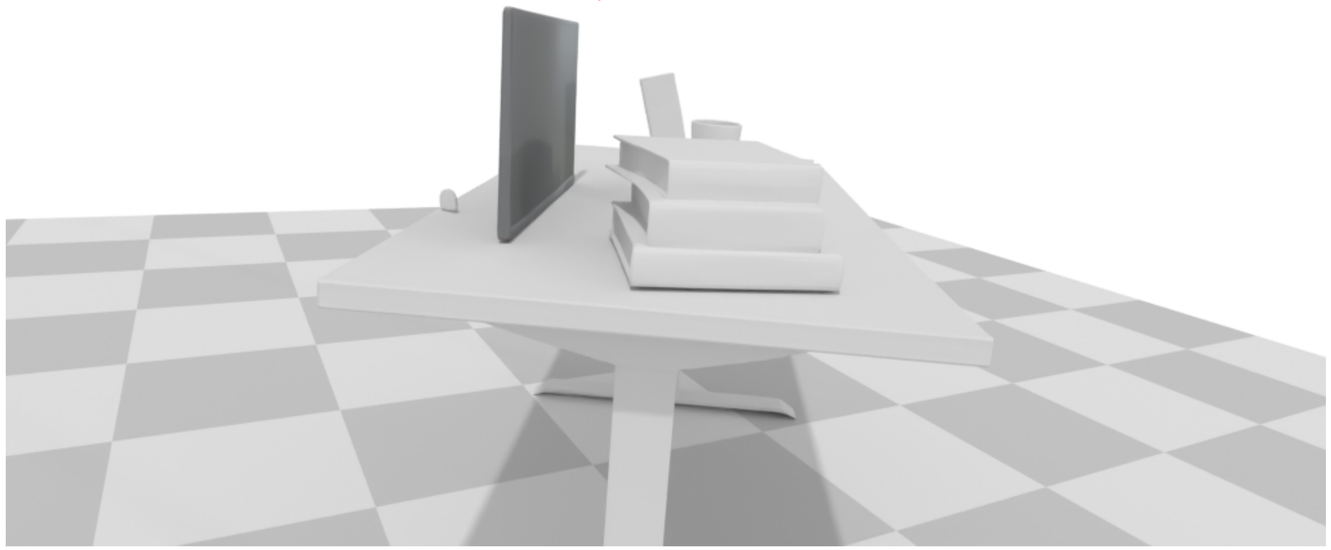
Large displacement -> Having inaccurate shape

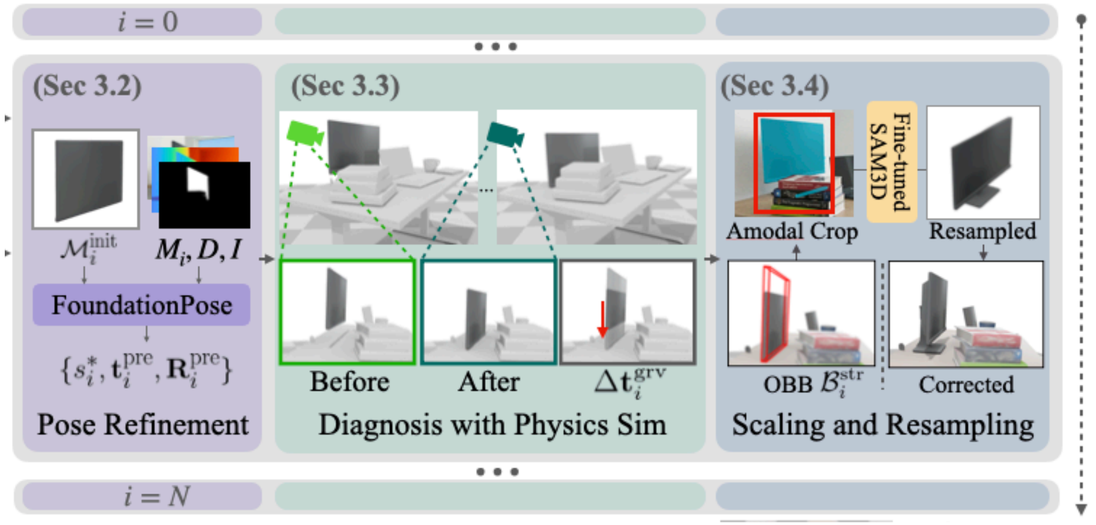
Our Approach

Integrate collision + Diagnosis Implausible Pose by Physics

Monitor doesn’t have “stand"

Side View

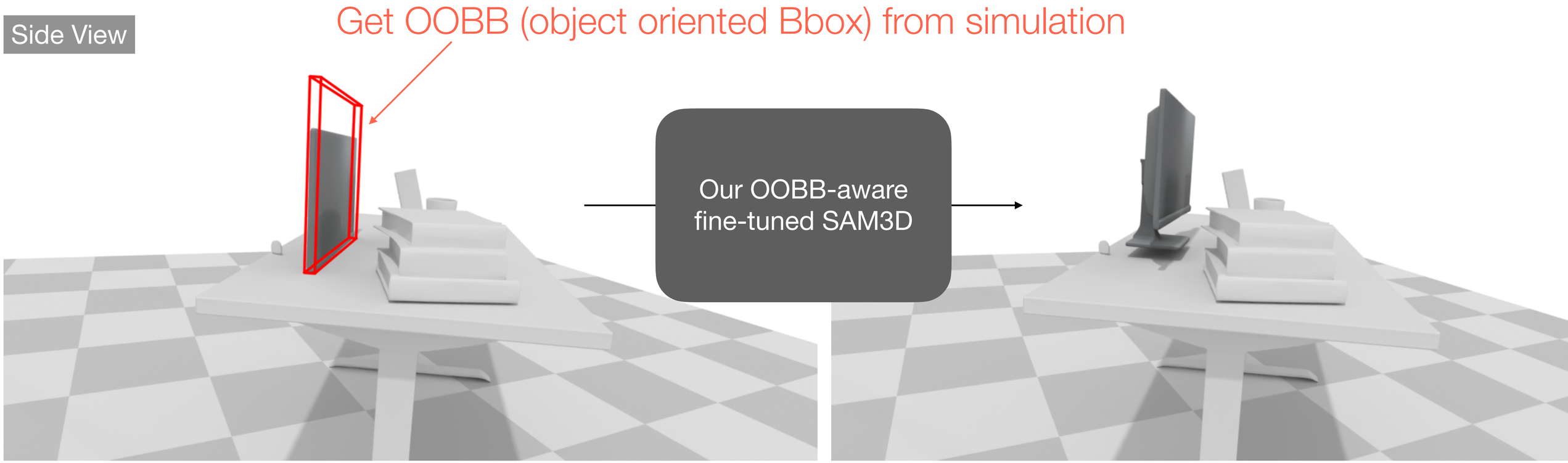




Large displacement -> Having inaccurate shape

Our Approach

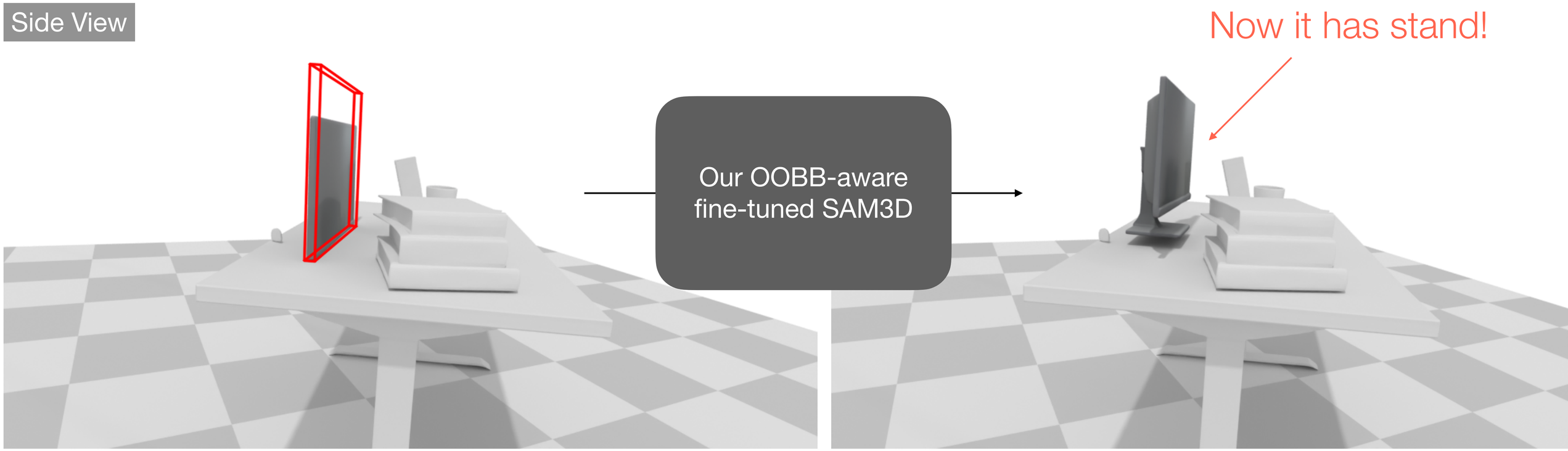
Integrate collision + Diagnosis Implausible Pose by Physics



4. If displacement is huge, we apply “resampling” with OOBB from simulation instead of stretching

Our Approach

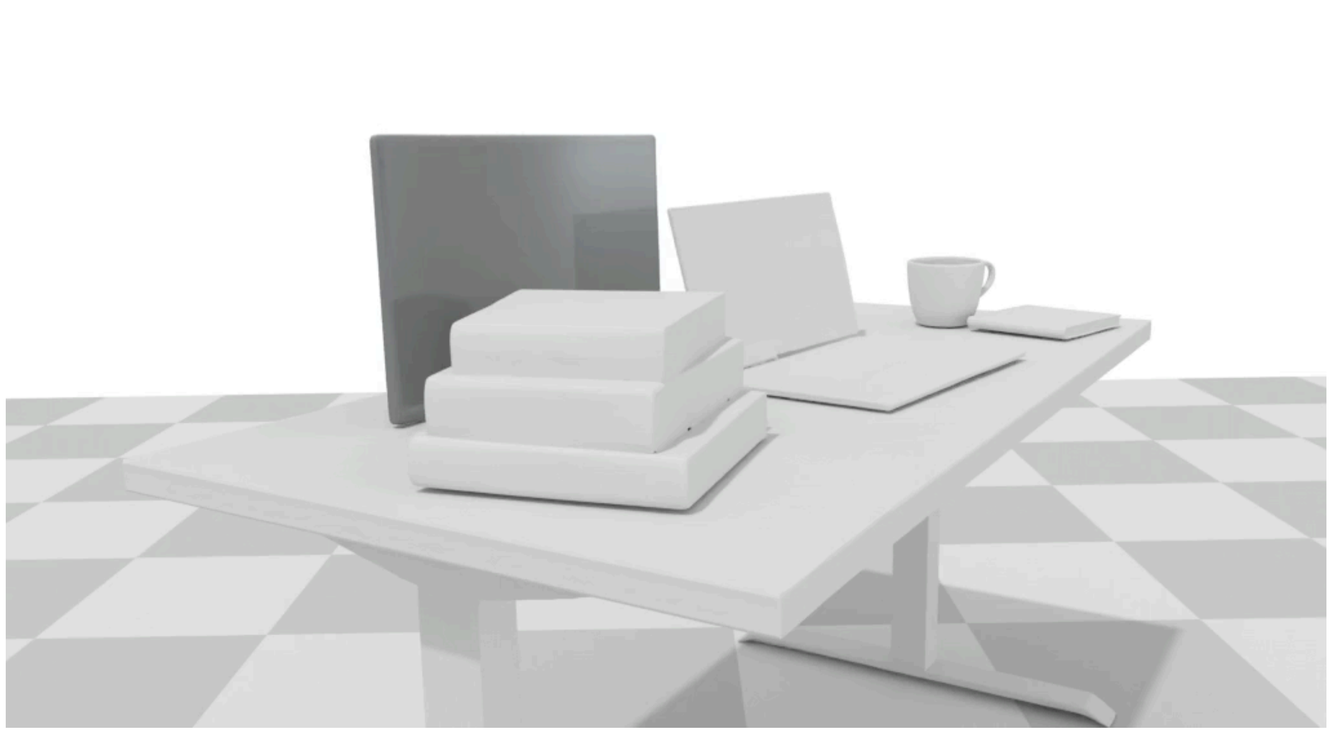
Integrate collision + Diagnosis Implausible Pose by Physics



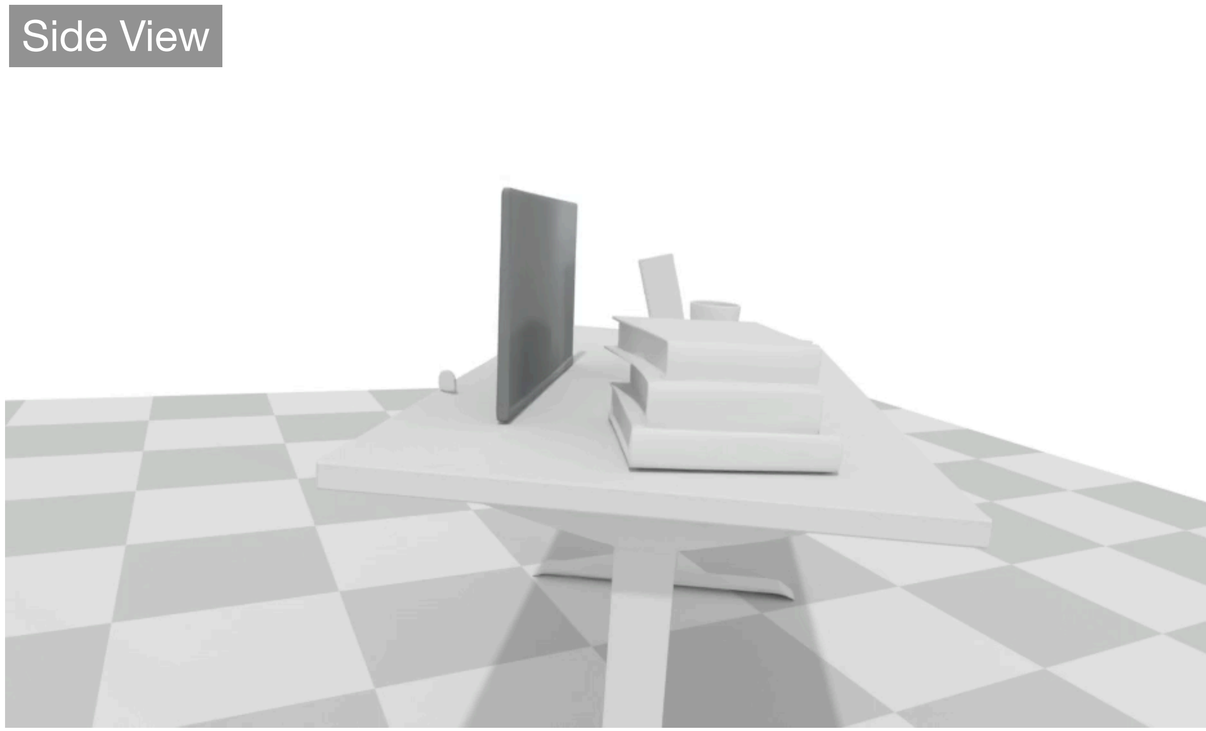
4. If displacement is huge, we apply “resampling” with OOBB from simulation instead of stretching

Our Approach

Integrate collision + Diagnosis Implausible Pose by Physics



Side View



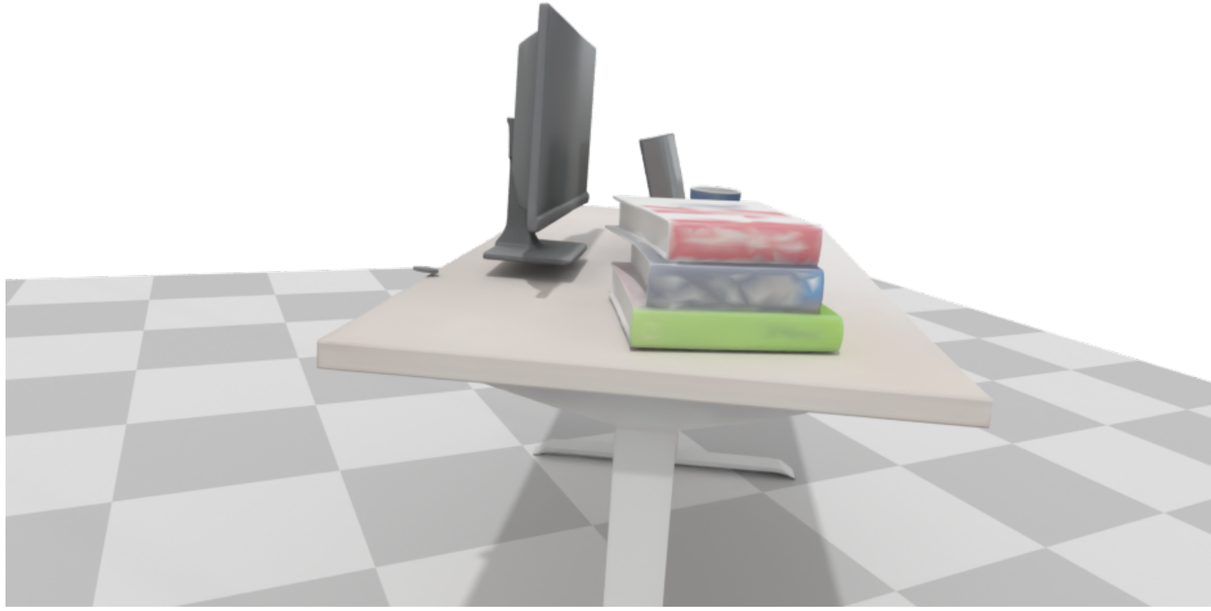
4. If displacement is huge, we apply “resampling” with OOBB from simulation instead of stretching

Our Approach

Integrate collision + Diagnosis Implausible Pose by Physics



Side View



Finally we get simulation-ready & image-aligned 3D scene!

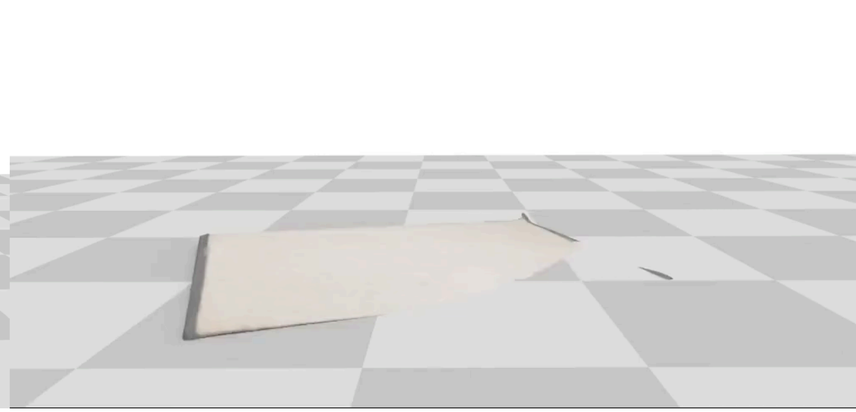
Our Approach

Integrate collision + Diagnosis Implausible Pose by Physics



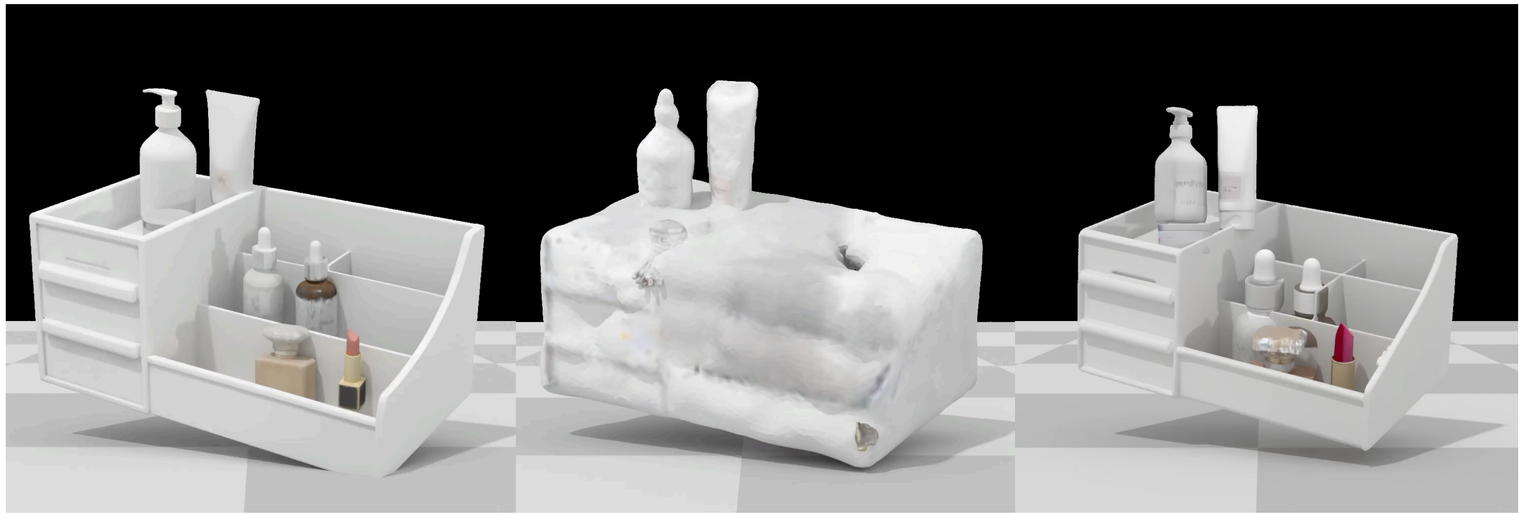




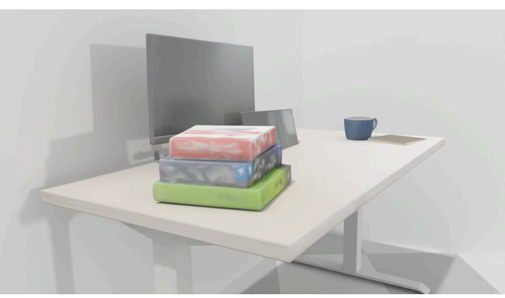



Results

Vs. Baselines



Input | SAM3D | Gen3DSR [1] | 3D-RE-GEN [2] | Ours

[1] A. Ardelean, M. Özer, and B. Egger, “Gen3dsr: Generalizable 3d scene reconstruction via divide and conquer from a single view,” in 3DV, 2025

[2] T. Sautter, J.-N. Dihlmann, and H. P. A Lensch, “3d-re-gen: 3d reconstruction of indoor scenes with a generative framework,” in CVPR, 2026

Results

Vs. Baselines



Input SAM3D Gen3DSR 3D-RE-GEN Ours

Results

Vs. Baselines



Input | SAM3D | Gen3DSR | 3D-RE-GEN | Ours

Results

Aria Digital Twin Dataset



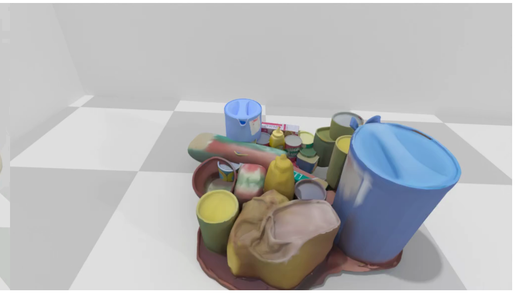
Input | SAM3D | Gen3DSR | 3D-RE-GEN | Ours

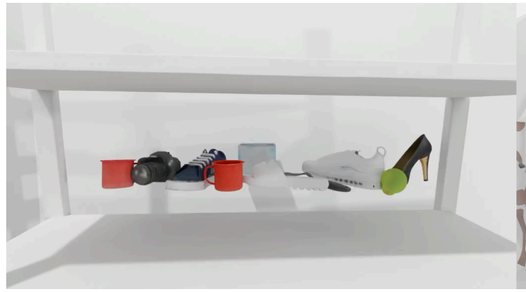
Our method shows more stable and input-aligned in complex room scale images

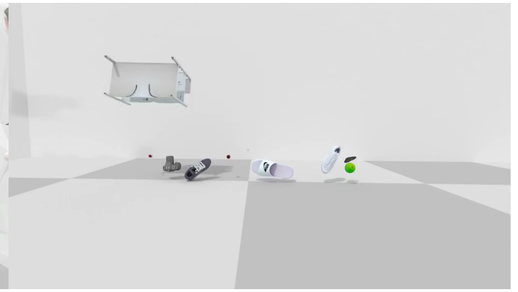
Results

GraspClutter6D



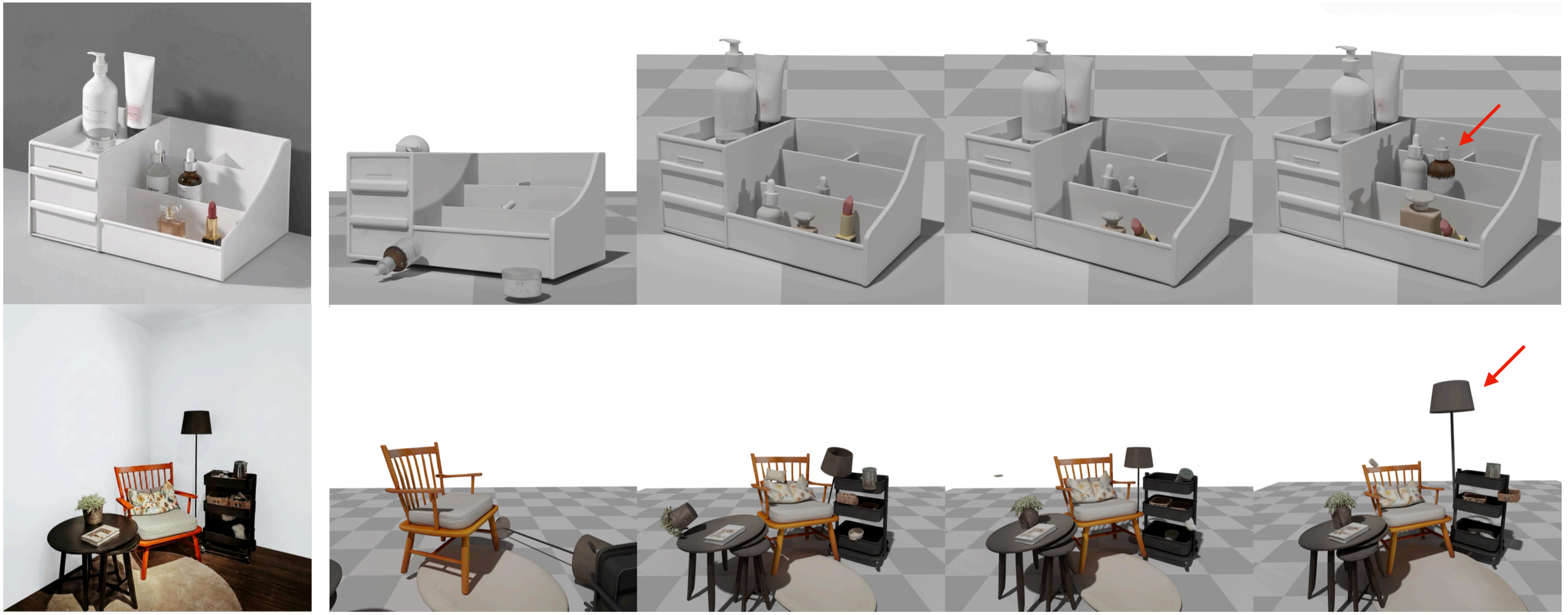



Input | SAM3D | Gen3DSR | 3D-RE-GEN | Ours

Our method more stable and input-aligned than baselines
in cluttered tabletop-scale images

Results

Ablation Study



Input SAM3D +Align. +Pen.+Grv. + Re. (Full)

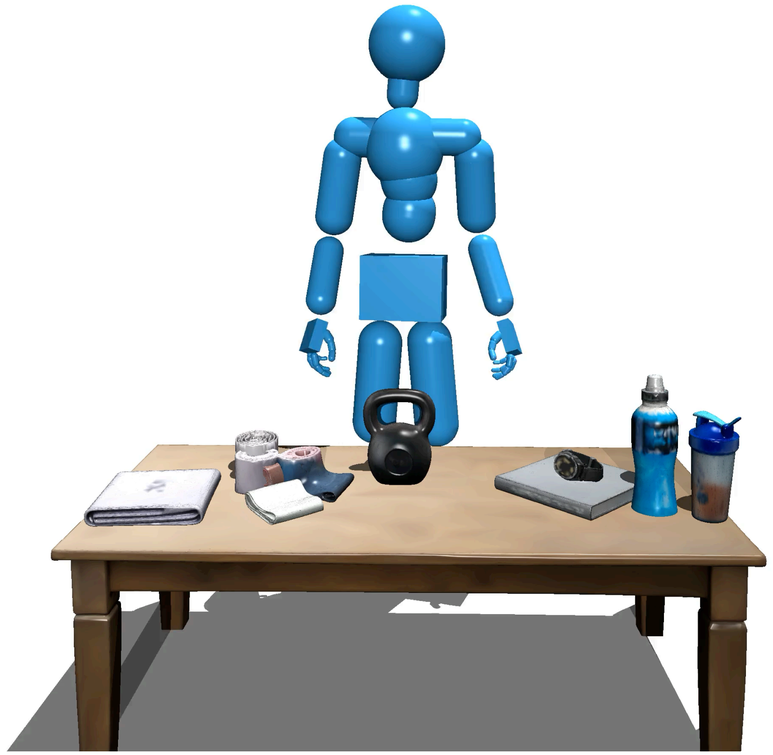
Pose alignment, Physics Simulation and Resampling improves alignment on simulator.

Applications

Humanoid Policy Learning in simulator



Input Image



Results of DeVI [1] in Our reconstruction

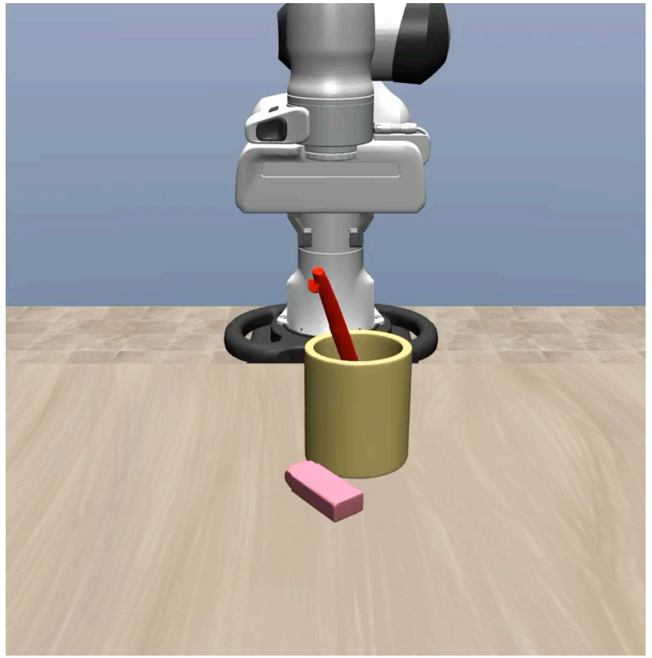
[1] H. Kim, J. Kim, K. Cho, and H. Joo, “DeVI: Physics-based Dexterous Human-Object Interaction via Synthetic Video Imitation,” in *preprint*, 2026

Applications

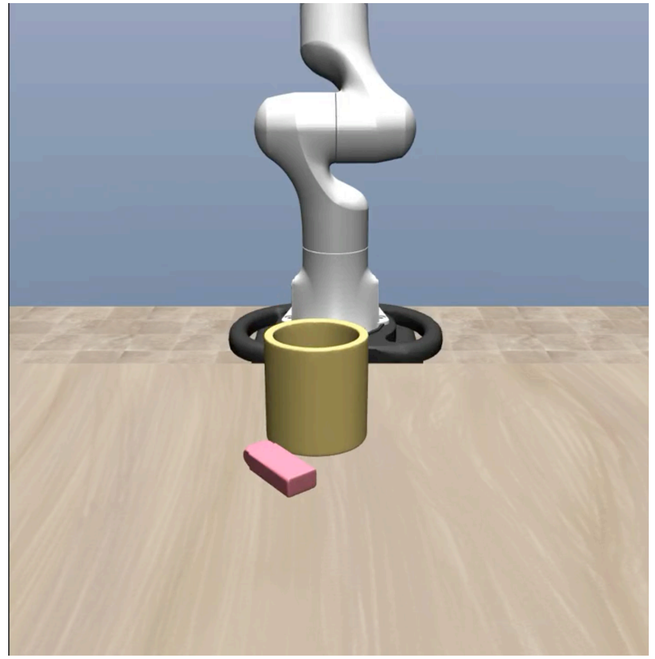
Robot Manipulation Test



Input Image



Ours



SAM3D

Robot manipulation in simulator using VLMPose [1] on reconstructed 3D scene

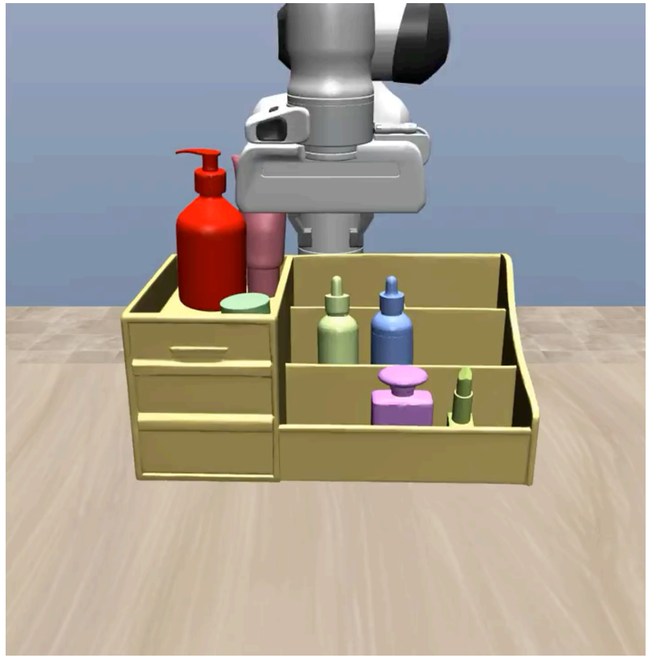
[1] S. Baik, G. Kim, M. Choi, and H. Joo, “Text-Guided 6D Object Pose Rearrangement via Closed-Loop VLM Agents,” in *preprint*, 2026

Applications

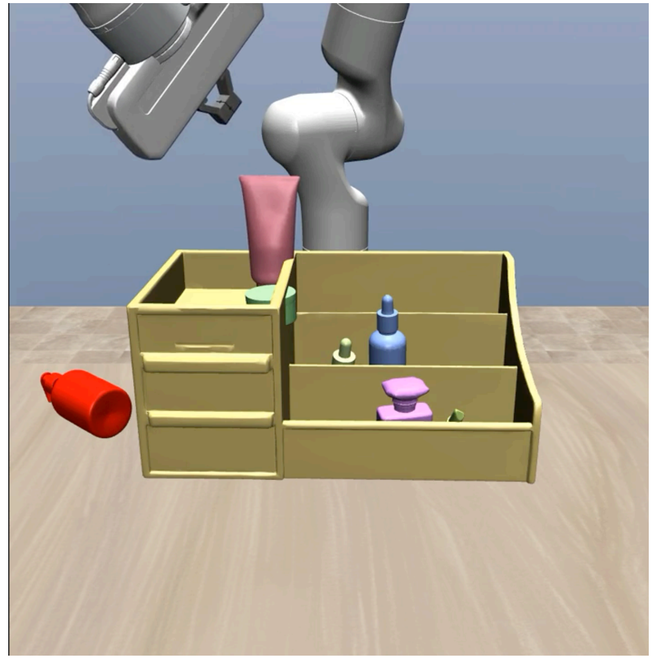
Robot Manipulation Test



Input Image



Ours



SAM3D

Robot manipulation in simulator using VLMPose [1] on reconstructed 3D scene

[1] S. Baik, G. Kim, M. Choi, and H. Joo, “Text-Guided 6D Object Pose Rearrangement via Closed-Loop VLM Agents,” in *preprint*, 2026

Thank you

Project Page

Paper