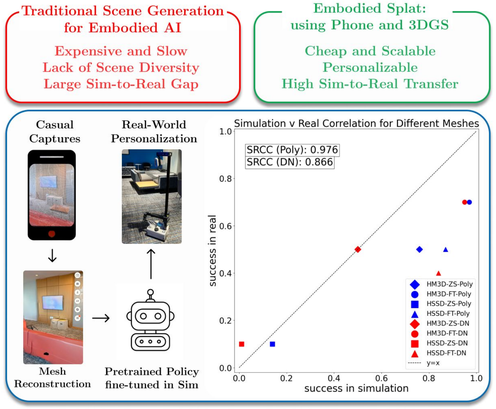
EmbodiedSplat: Personalized Real-to-Sim-to-Real Navigation with Gaussian Splats from a Mobile Device

Gunjan Chhablani[1], Xiaomeng Ye[1], Rynaa Grover[1], Muhammad Zubair Irshad[2], Zsolt Kira[1]

TL;DR



Problem ⚠️: Sim-to-real transfer is limited:

- Synthetic scenes (e.g., HSSD) are unrealistic.
- Real datasets (e.g., HM3D) are expensive.

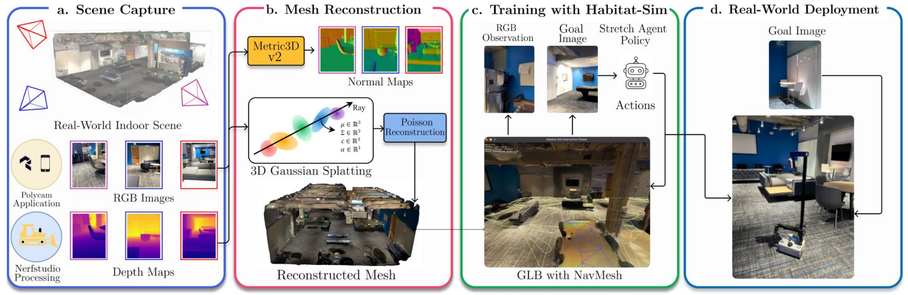
Our Solution 🧠: EmbodiedSplat

- Captures real scenes using a mobile phone + 3D Gaussian Splatting (3DGS)/Polycam.
- Enables realistic fine-tuning in Habitat-Sim [2].

Key Results 🏆: +20–40% success in real-world ImageNav over pre-trained policies. High SRCC: 0.87–0.97.

Impact 🎯: Democratizes realistic policy training with fast, low-cost scene capture, scaling diverse scene collection.

Method



Real-world scenes → 3DGS reconstruction → Habitat-Sim integration + Policy fine-tuning → Real-world deployment

Scene Capture 📱: Use iPhone 13 Pro Max + Polycam to capture RGB-D data in a university setting (out-of-distribution from pre-training). Process 1K frames using nerfstudio.

Mesh Reconstruction 🛠️: Train 3DGS (DN-Splatter) and generate mesh. Reconstruction **takes 20–30 min capture + 1–2 hrs training**. Much faster and cheaper than Matterport-style pipelines. Use Polycam raw mesh as an alternative.

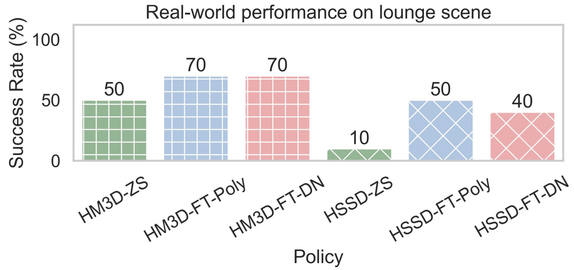
Simulation & Training 🧪: Load into Habitat-Sim, generate ImageNav and train an LSTM policy end-to-end with DD-PPO and VC-1 visual encoder.

Real-World Deployment 🤖: Deploy trained policy on a Stretch robot in real-world.

Policy Training and Evaluation Details:

- **Pretraining**: HM3D policy: 83.1% val SR after 600M steps. HSSD policy: 63.2% val SR after 1200M steps.
- **Scene-Level Fine-Tuning**: for 20M steps on DN and Polycam meshes.
- **Evaluation:** On the real-world scene over 10 episodes.

Results



Zero-Shot vs Fine-Tuned:

- HM3D Zero-Shot: 50% SR → up to 70% with fine-tuning.
- HSSD Zero-Shot: 10% SR → 50% with Polycam, 40% with DN-Splatter.

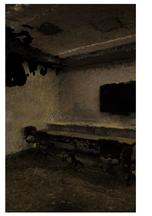
Polycam vs DN-Splatter:

- Polycam > DN-Splatter due to **better visual fidelity**.
- Sneak Peek 👀

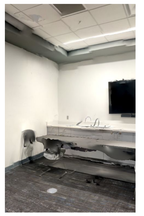
  Overfitted policy:

  with Polycam → 50% SR.

  with DN → only 10% SR.



DN-Splatter



Polycam

Pre-training and visual fidelity crucial!

Conclusion and Future Work 🔎🔮

Achieved: bridged sim-to-real gap, enabled fast and low-cost low-effort scene collection, supported high-fidelity personalized training in realistic environments. 🎉

Next: complex tasks, large environments, diverse scenes