

---

# **CGI-Scene: Controllable Generation of Open-Vocabulary Interactive Scenes**

Liu Dai, Fanbo Xiang, Minghua Liu, Hao Su

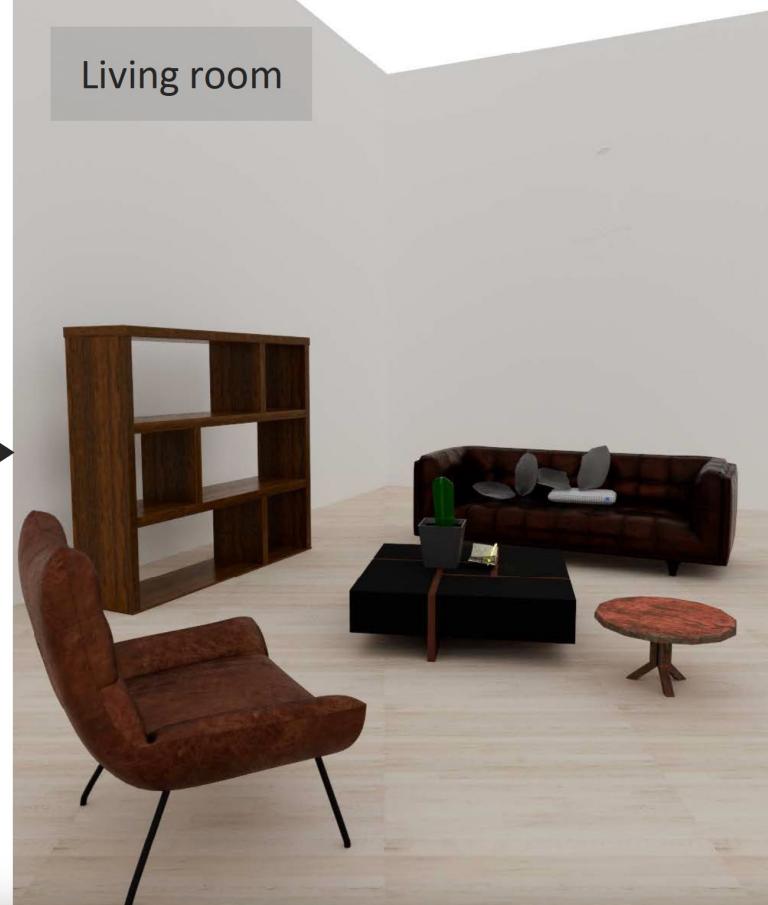
---

# What did we do?

2D Image



3D Scene



# What can we do with it?

---

Text Prompt

"A car and an arm chair, interior scene."



Sketch



Real Photo

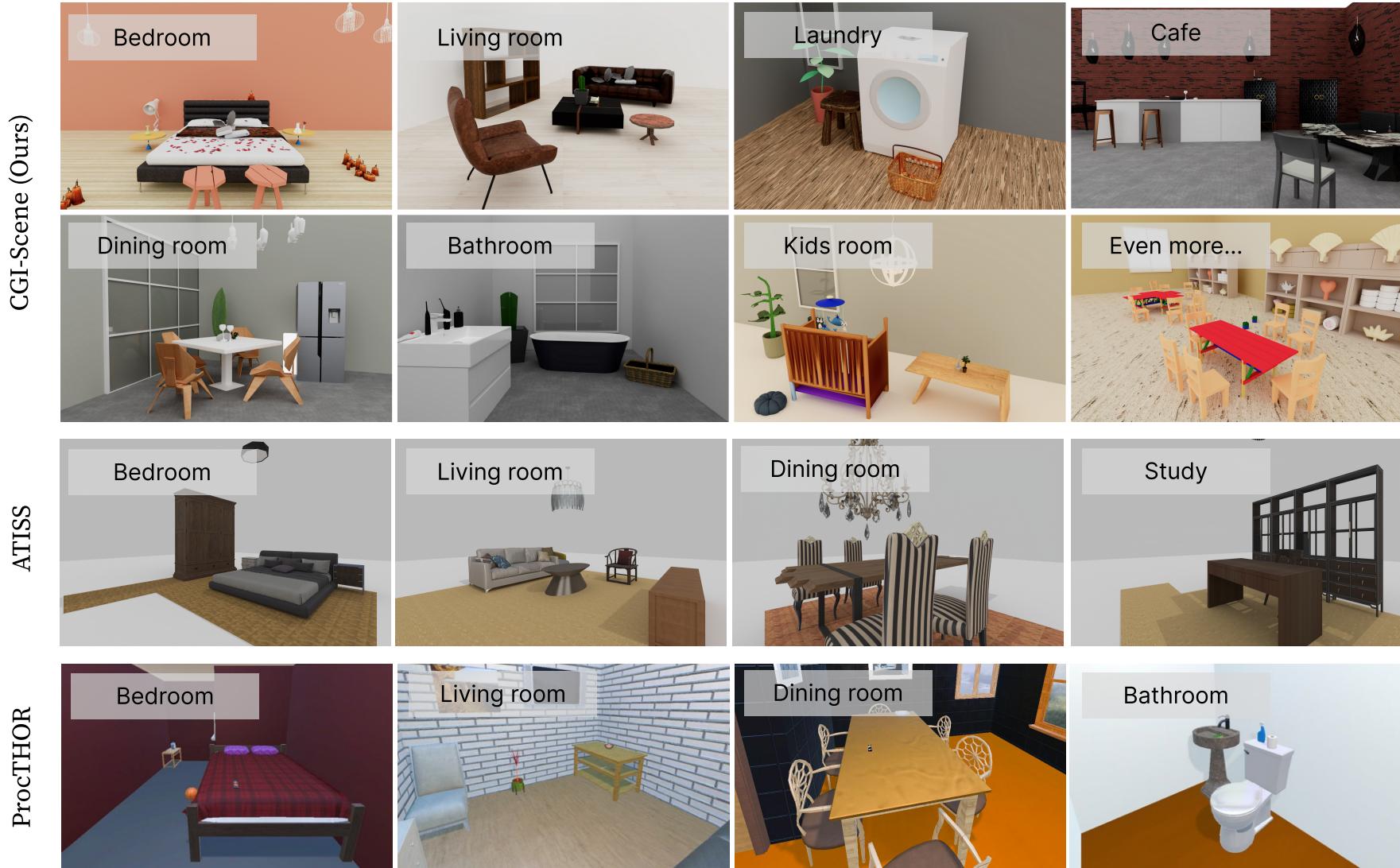


"A toilet and a washing machine, interior scene."



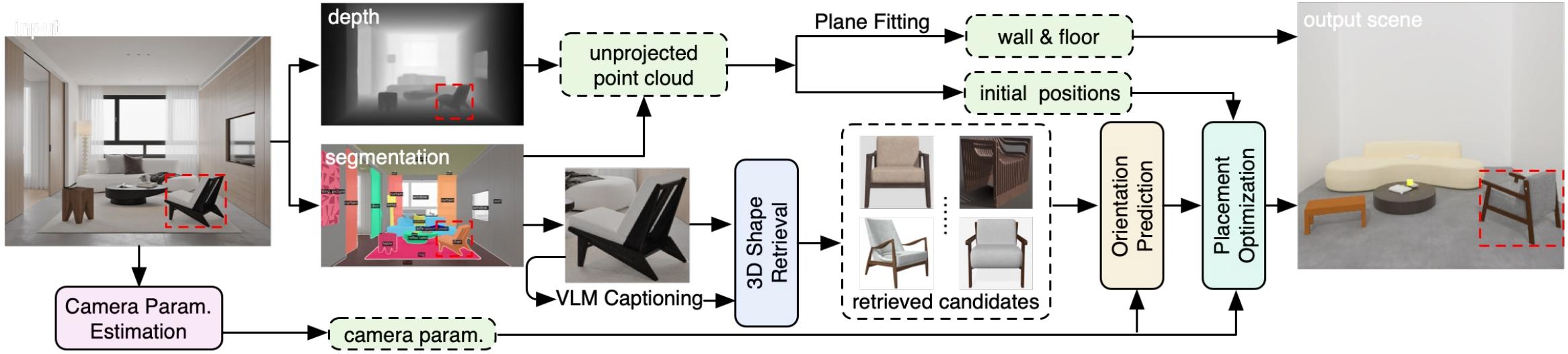
# What can we do with it?

---



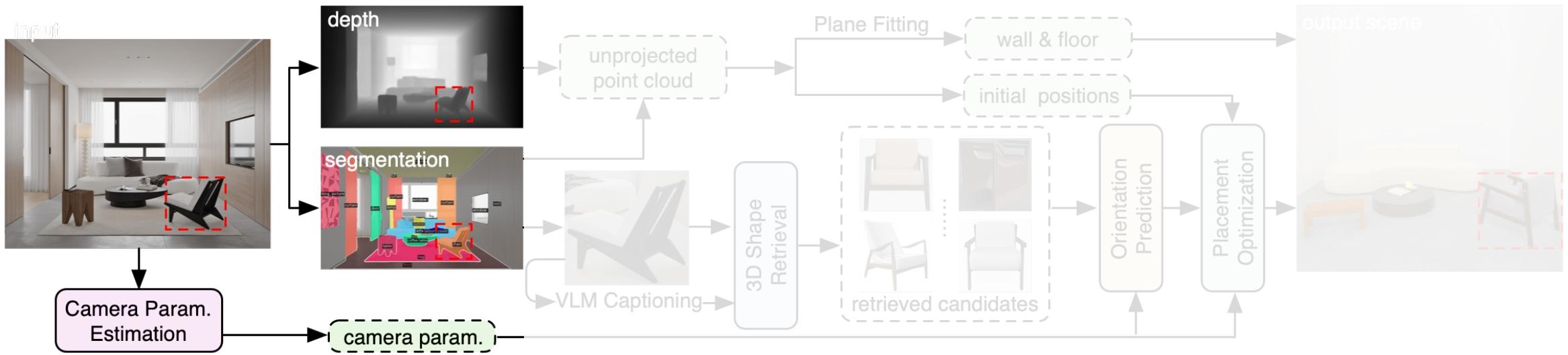
# How did we do it?

---



# How did we do it?

---

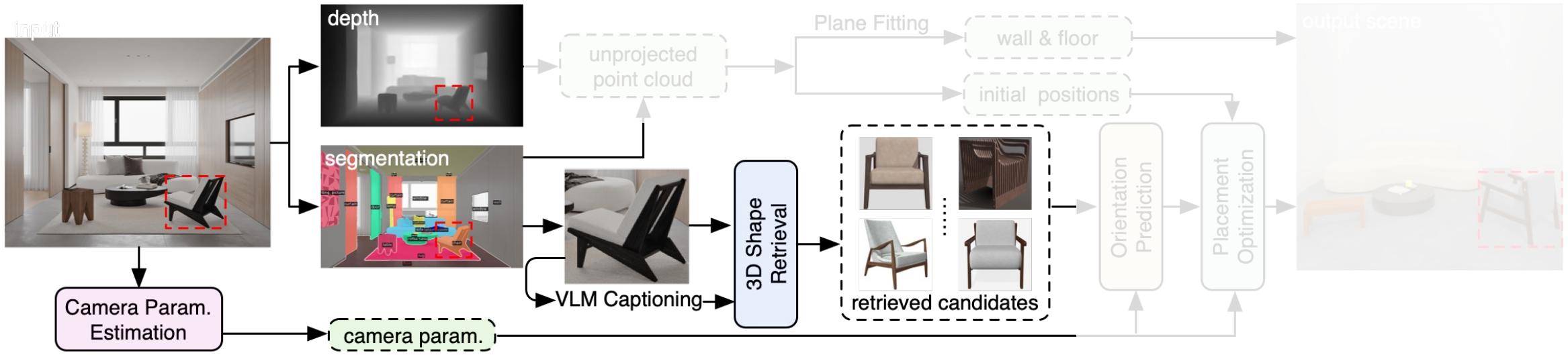


**Step 1:** Extract basic information from reference image

---

# How did we do it?

---

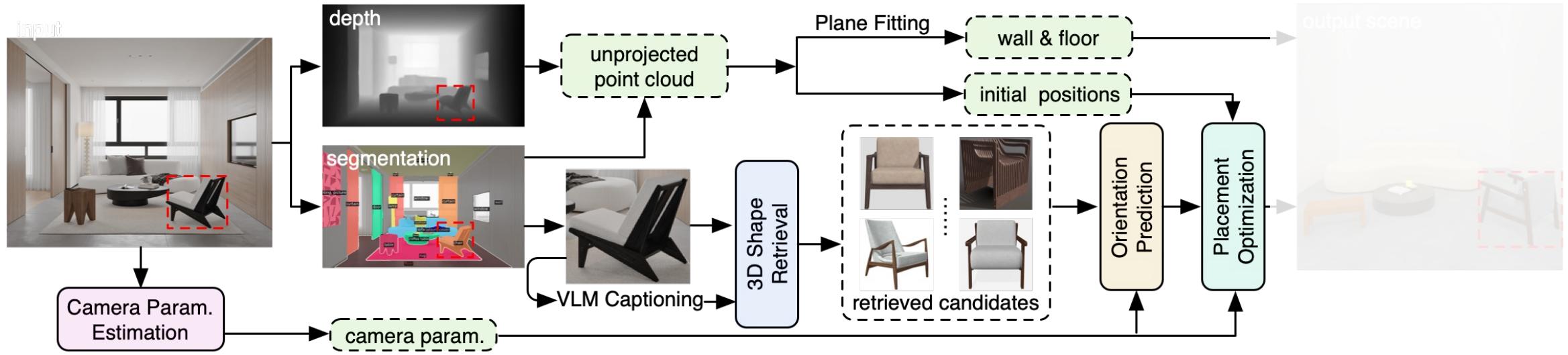


**Step1:** Extract basic information from reference image

**Step2:** Retrieve candidate objects from Objaverse

# How did we do it?

---



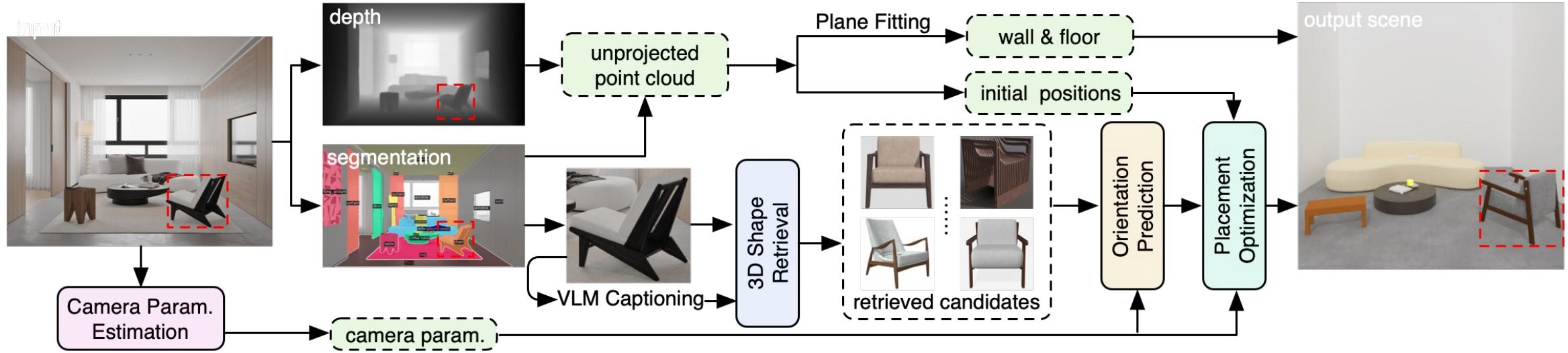
**Step1:** Extract basic information from reference image

**Step2:** Retrieve candidate objects from Objaverse

**Step3:** Place each candidate with optimal pose and scale

# How did we do it?

---



**Step1:** Extract basic information from reference image

**Step2:** Retrieve candidate objects from Objaverse

**Step3:** Place each candidate with optimal pose and scale

**Step4:** Finalize the objects and the complete scene

---