**Raw Models (6.3M)**
GitHub
Sketchfab

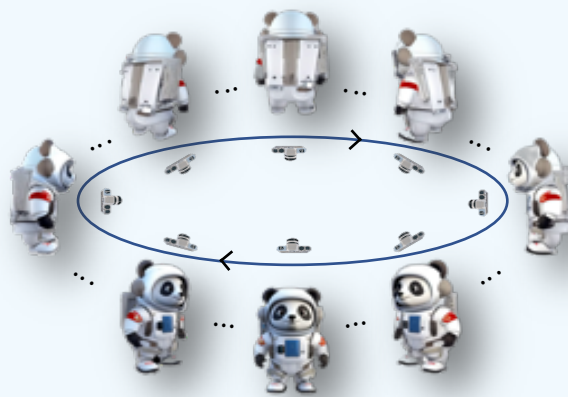**Rendering**

S1 **Coarse**-Rendering [*pyrender*]
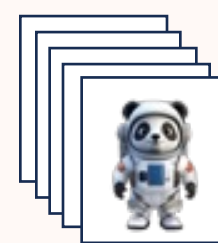
S2 **Fine**-Rendering [*BlenderProc*]

**Metric Filtering**
- aesthetic score > 4.0
- image quality score > 4.0

❌ ✅ ❌ ✅

**Caption Generation**
VLM ❄
LoRA
GRPO
Brief text description
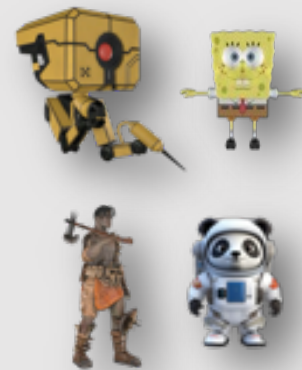Align detailed Caption

**Droplet3D-4M (4M)**

The video is filmed from a frontal perspective, starting by showing the robot's ......rotation, ...... details from all angles.