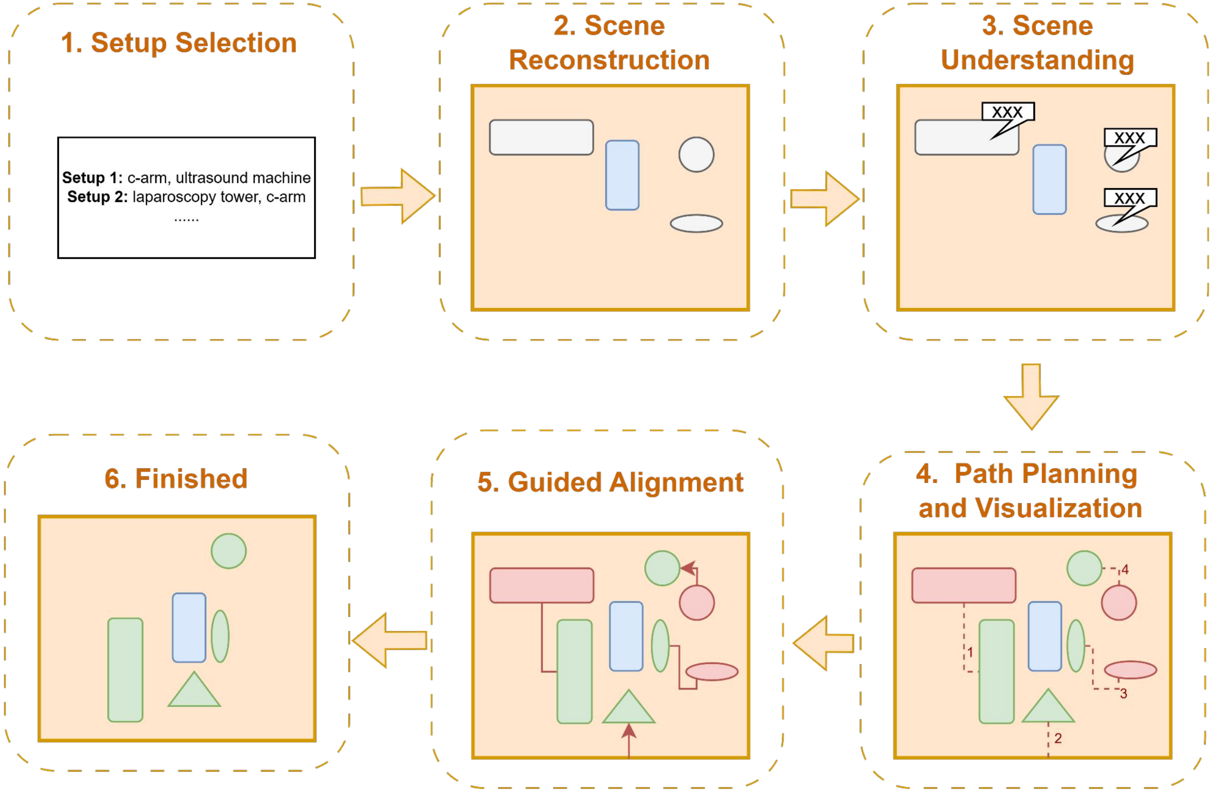
Pipeline Overview



1. Setup Selection

Setup 1: c-arm, ultrasound machine
Setup 2: laparoscopy tower, c-arm
......

2. Scene Reconstruction

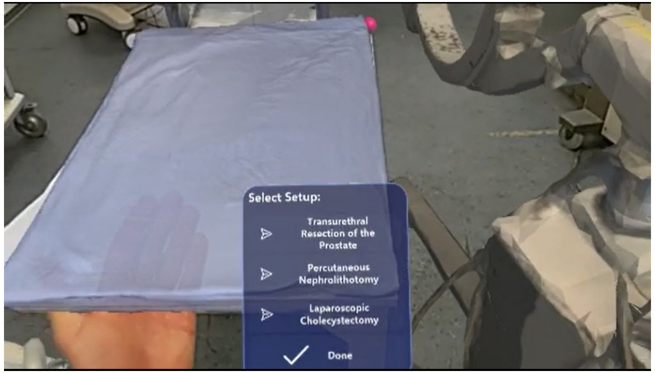
3. Scene Understanding

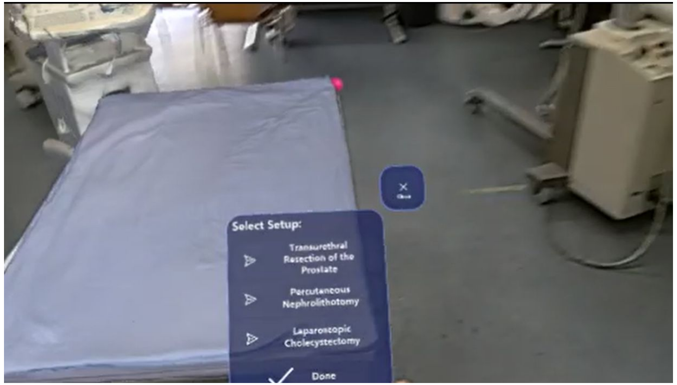
4. Path Planning and Visualization

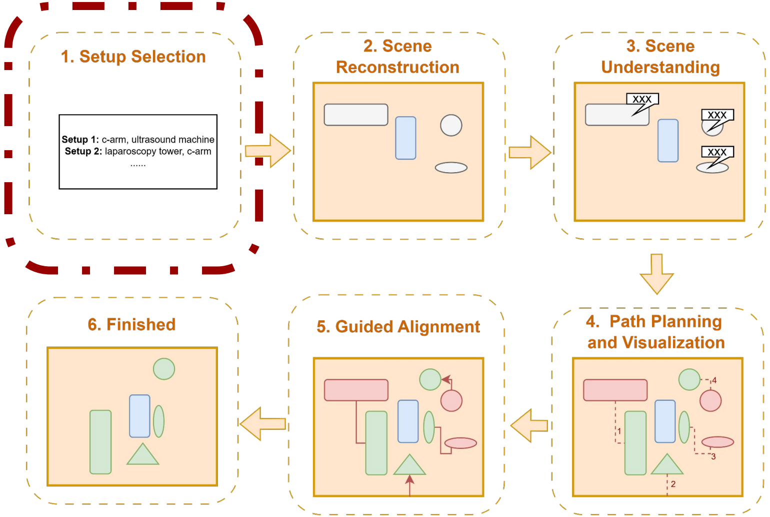
5. Guided Alignment

6. Finished

1. Setup Selection


Select Setup:
Transurethral Resection of the Prostate
Percutaneous Nephrolithotomy
Laparoscopic Cholecystectomy
Done


×
Select Setup:
Transurethral Resection of the Prostate
Percutaneous Nephrolithotomy
Laparoscopic Cholecystectomy
Done



1. Setup Selection

Setup 1: c-arm, ultrasound machine
Setup 2: laparoscopy tower, c-arm
......

2. Scene Reconstruction

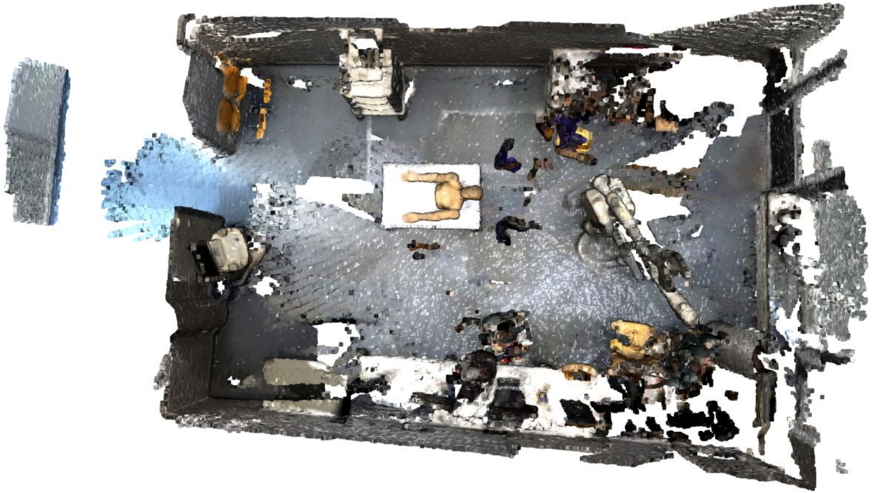
3. Scene Understanding

4. Path Planning and Visualization

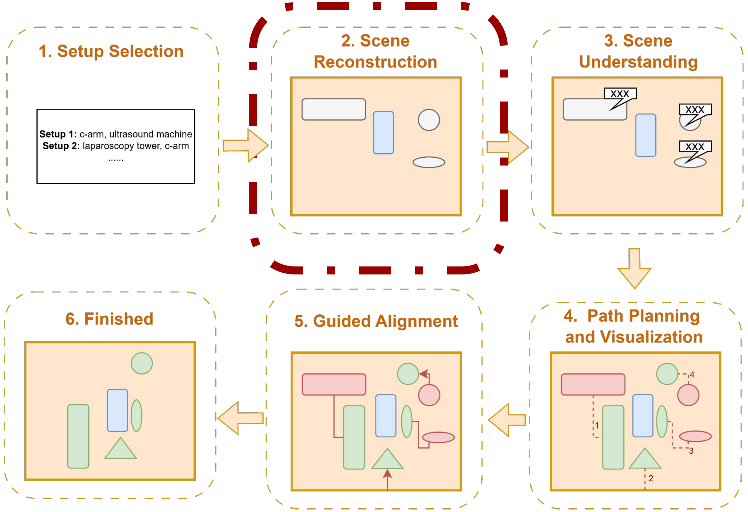
5. Guided Alignment

6. Finished

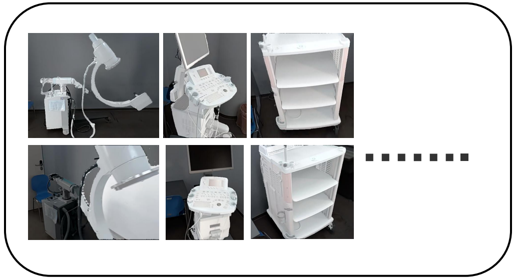
2. Scene Reconstruction



Reconstructed Pointcloud

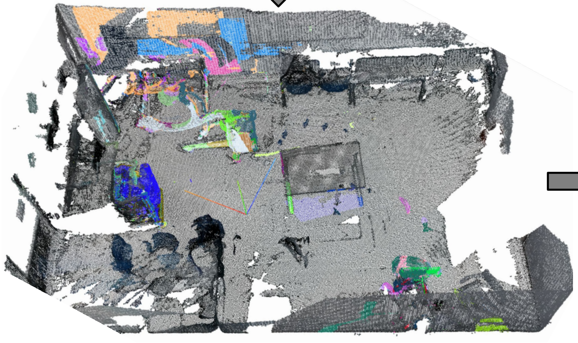


3. Scene Understanding

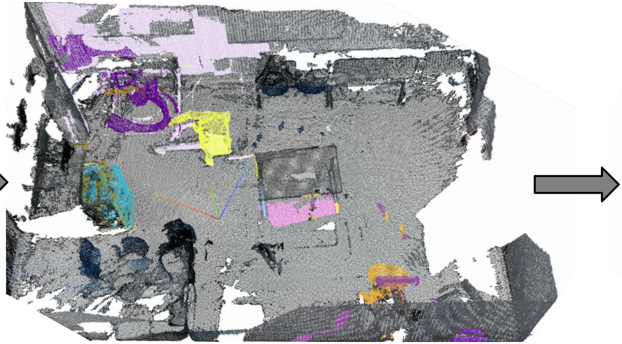


2D masks

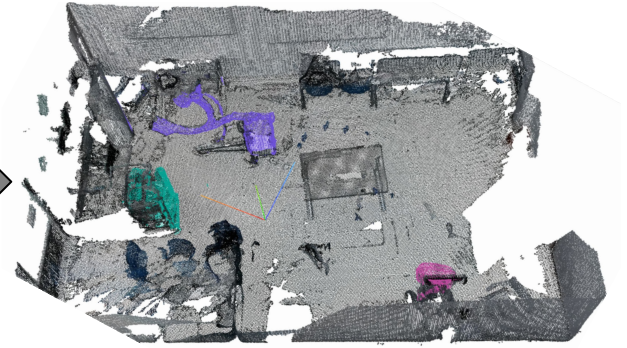
Grounding DINO[1]
(Detection)
+
CLIP[3] Verification
+
SAM[2]
(Segmentation)



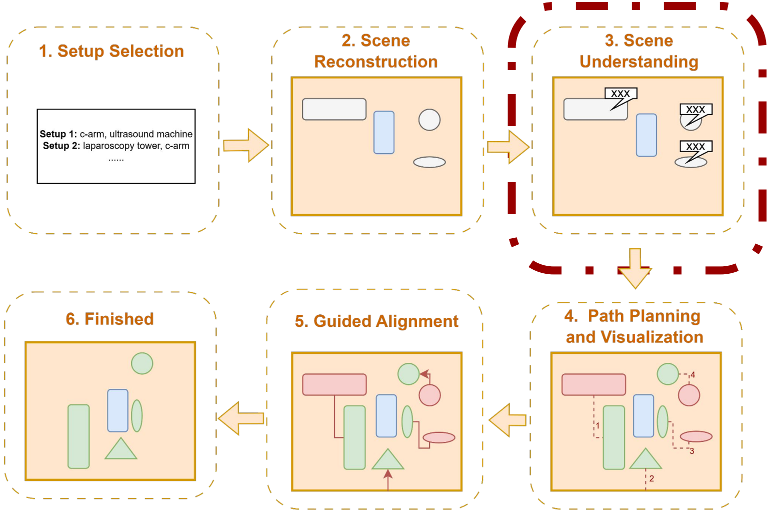
Projected masks



Aggregated masks



Filtered masks



1. Setup Selection

Setup 1: c-arm, ultrasound machine
Setup 2: laparoscopy tower, c-arm
......

2. Scene Reconstruction

3. Scene Understanding

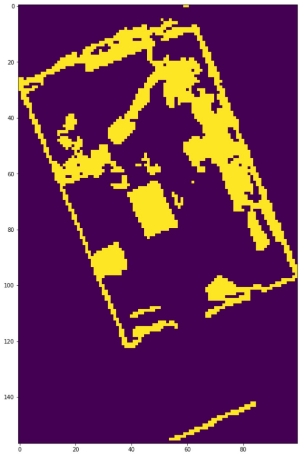
4. Path Planning and Visualization

5. Guided Alignment

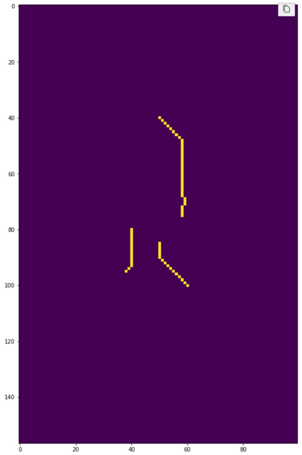
6. Finished

[1] Liu, Shilong, et al. "Grounding dino: Marrying dino with grounded pre-training for open-set object detection." *arXiv preprint arXiv:2303.05499* (2023).
[2] Kirillov, Alexander, et al. "Segment anything." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.
[3] Menon, Sachit, and Carl Vondrick. 'Visual Classification via Description from Large Language Models'. arXiv, 1 December 2022. https://doi.org/10.48550/arXiv.2210.07183.

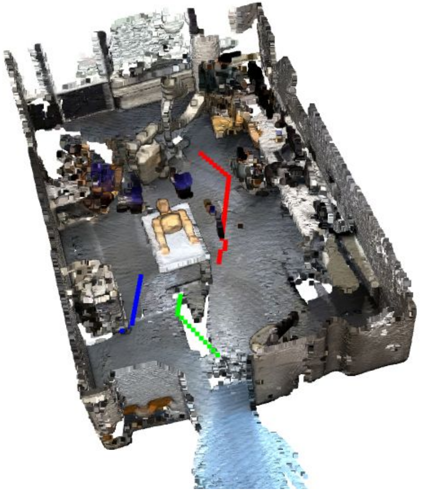
4. Path Planning



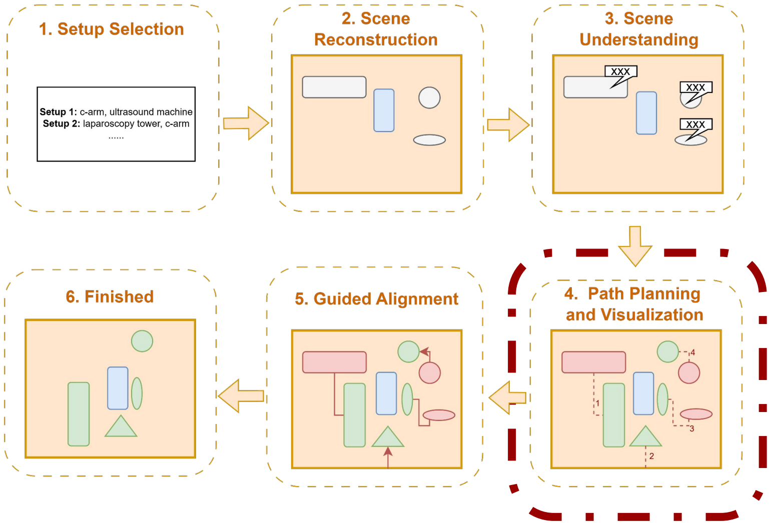
obstacles projected
on ground



2D Path



3D Path



1. Setup Selection

Setup 1: c-arm, ultrasound machine
Setup 2: laparoscopy tower, c-arm
......

2. Scene Reconstruction

3. Scene Understanding

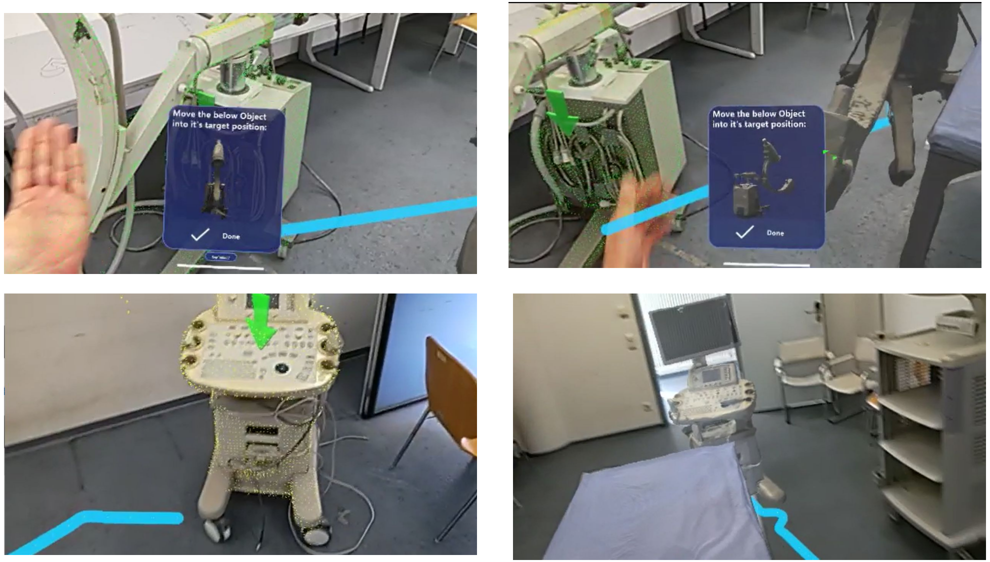
4. Path Planning and Visualization

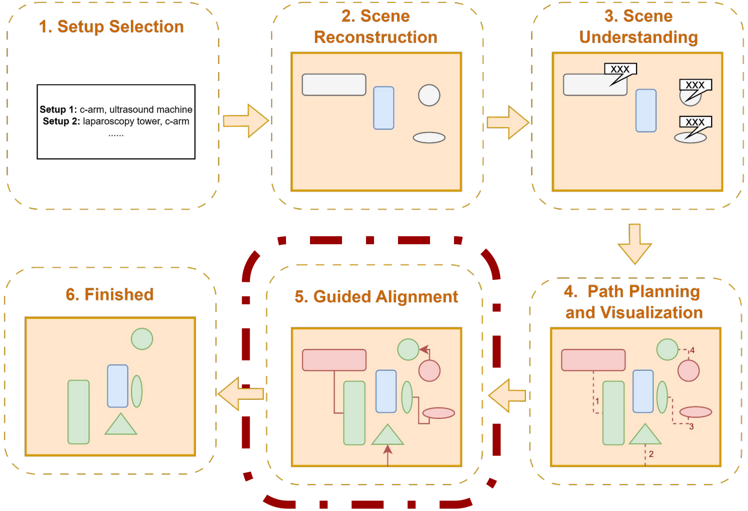
5. Guided Alignment

6. Finished

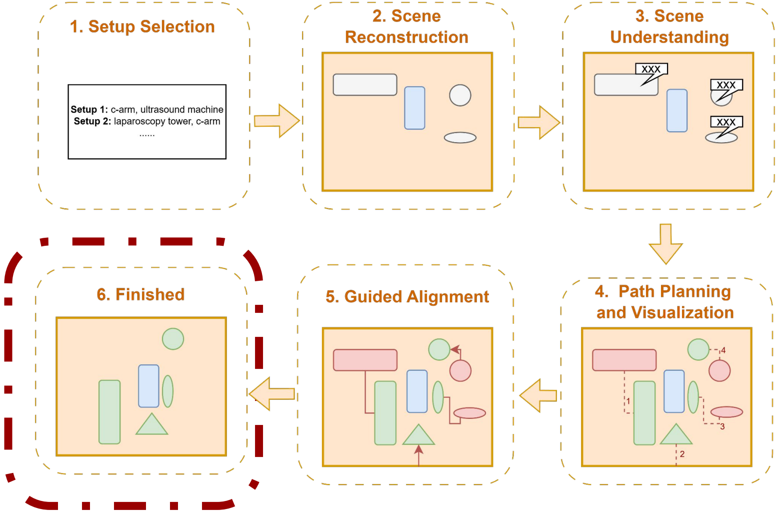
5. Guided Alignment

The user **follow paths on ground** and **aligns the devices with virtual objects**


Move the below Object into it's target position:



6. Finished



1. Setup Selection

Setup 1: c-arm, ultrasound machine
Setup 2: laparoscopy tower, c-arm
......

2. Scene Reconstruction

3. Scene Understanding

XXX

XXX

XXX

4. Path Planning and Visualization

1

2

3

4

5. Guided Alignment

6. Finished