3D Scene Reconstruction for Autonomous Robot Navigation

Subramaniya Siva T S
Om Sai Krishna Madhav Lella
Thejesh Mallidi
Sai Mohan Gajapaka

College of Natural Science
MICHIGAN STATE UNIVERSITY

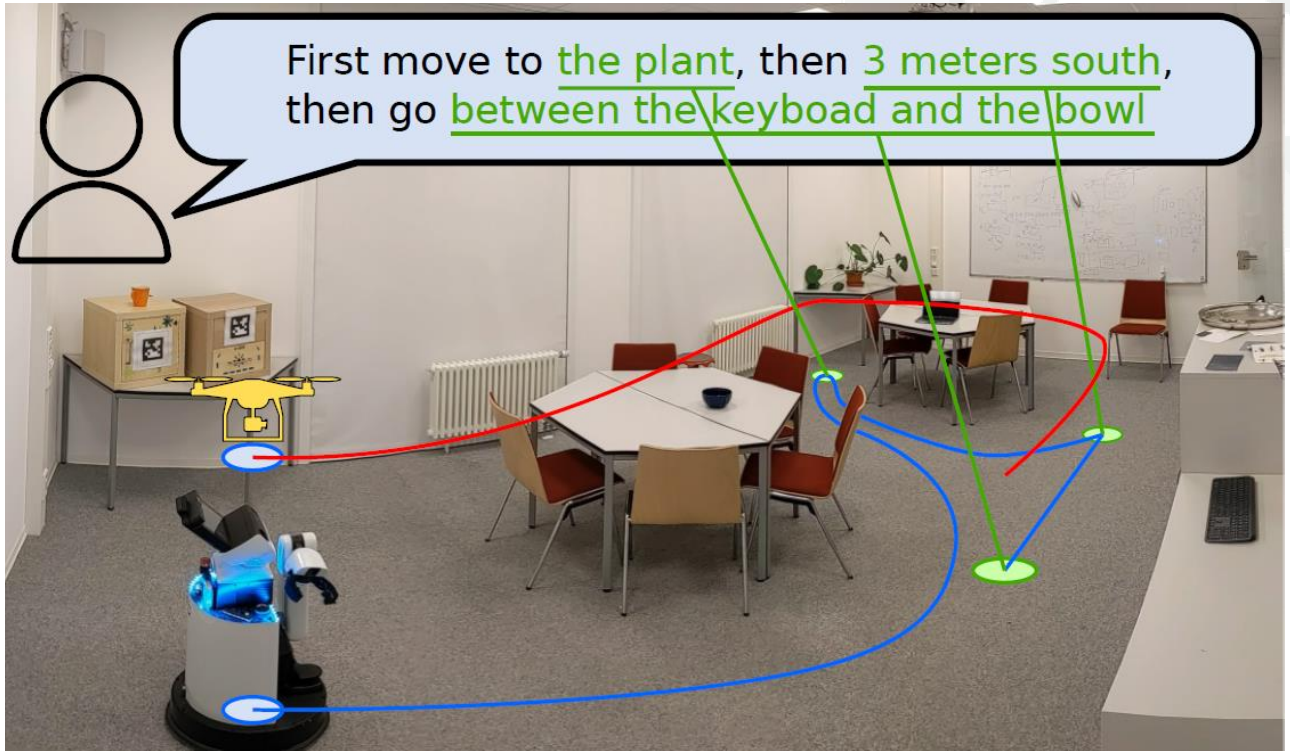
Introduction

Background:

- The need for effective autonomous robot navigation is growing, requiring accurate environmental perception and dynamic mapping.
- Current approaches often struggle with integrating real-time video data, spatial features, and obstacle detection for efficient navigation in complex environments.

Motivation:

- Enhancing autonomous robots' ability to perceive and navigate 3D environments is critical for applications in industries like manufacturing, logistics, and exploration.
- Accurate and real-time mapping and obstacle detection enable safer and more efficient robot decision-making in dynamic settings.



Introduction

Objectives:

- To develop a comprehensive framework for 3D scene reconstruction using video data for dynamic environmental mapping.
- To integrate advanced object detection algorithms, depth mapping, and path planning for real-time navigation and obstacle avoidance.

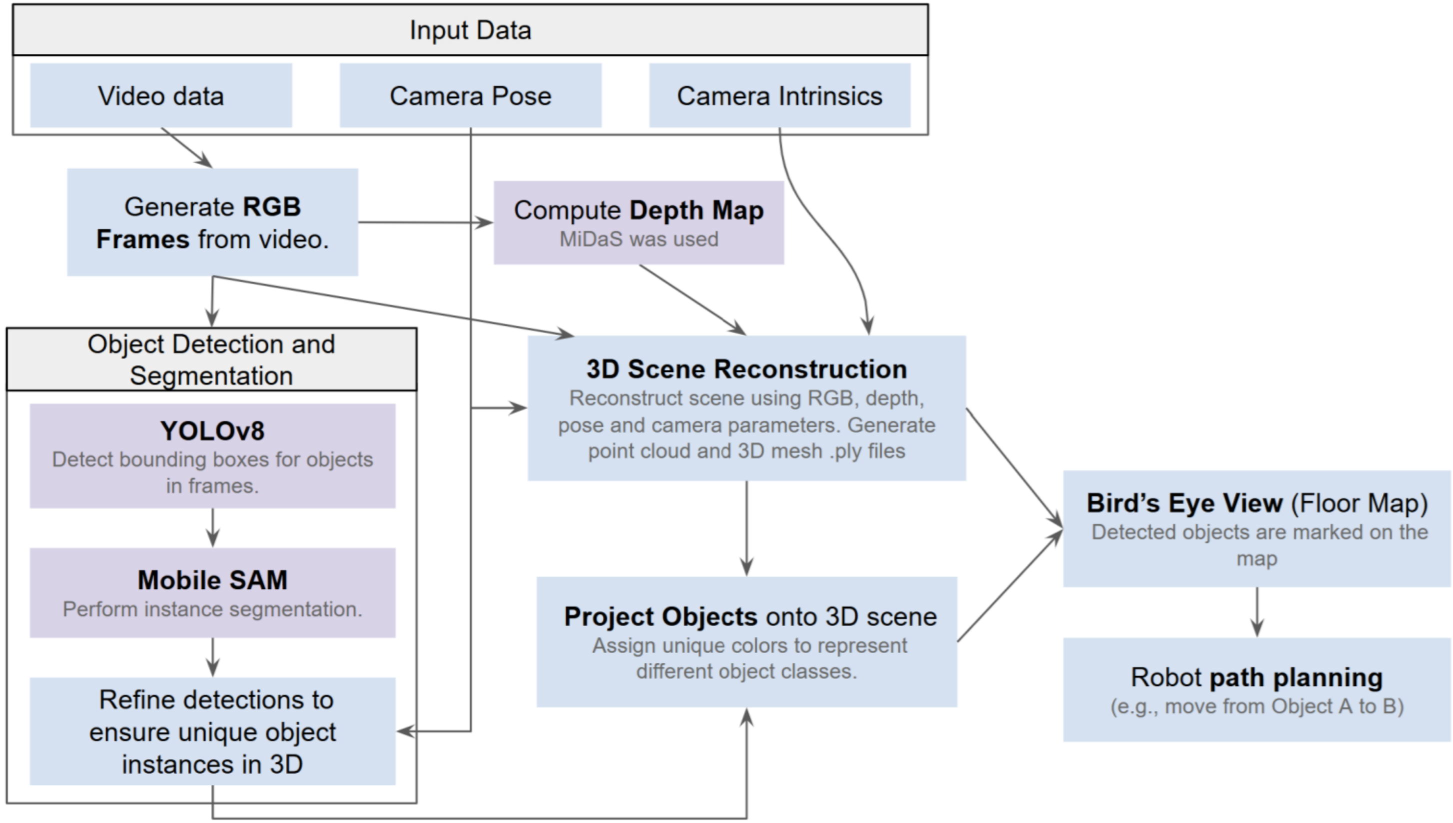
Dataset Used:

- We utilized the ScanNet sensor dataset, specifically scene0000, which contains RGB images, depth maps, and camera pose information. The dataset includes a total of 5,578 frames.

Approach

To achieve our objectives, we divided the problem into multiple sub-tasks:

- Depth Estimation
- 3D Scene Reconstruction
- Object Detection
- Instance Segmentation
- 3D Object Mapping
- Bird's-Eye View Generation
- Optimal Path Planning



Monocular 3D Scene Reconstruction and Object Mapping Pipeline

Depth Estimation

Definition: Depth estimation is the process of predicting the distance of objects in a scene from a single 2D RGB image, enabling 3D scene understanding.

Application in Project:

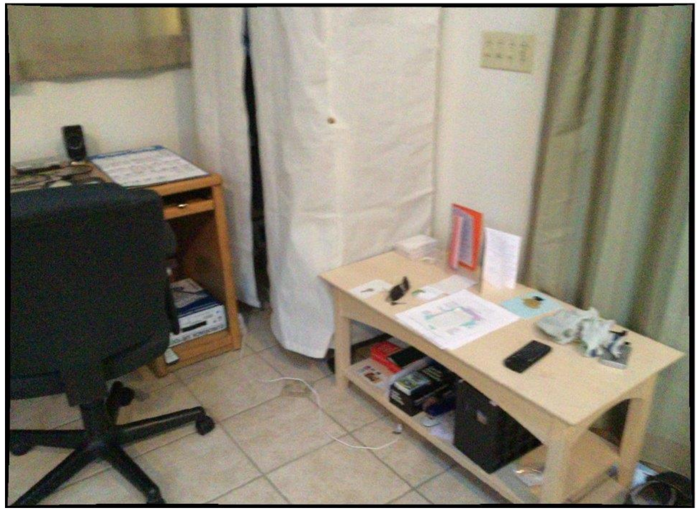
- Provides spatial context by generating depth maps for RGB images.
- Facilitates the placement of objects in 3D reconstruction tasks.

Key Techniques:

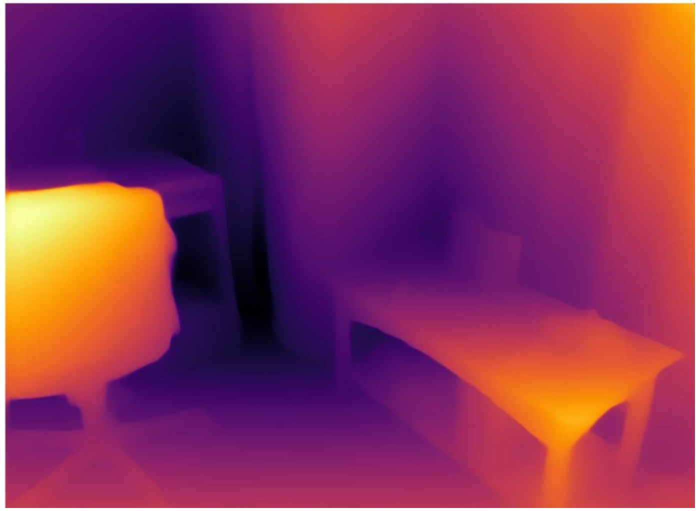
- *MiDAS Model:* A deep learning approach that utilizes monocular images to estimate depth accurately.
- *Depth Anything*: Developed to improve monocular depth estimation by leveraging large-scale unlabeled data.

Depth Estimation Performance & Results

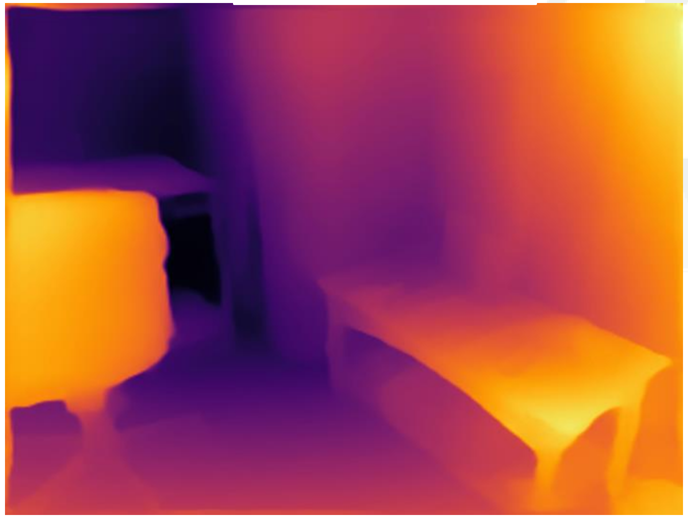
Image



Depth Anything

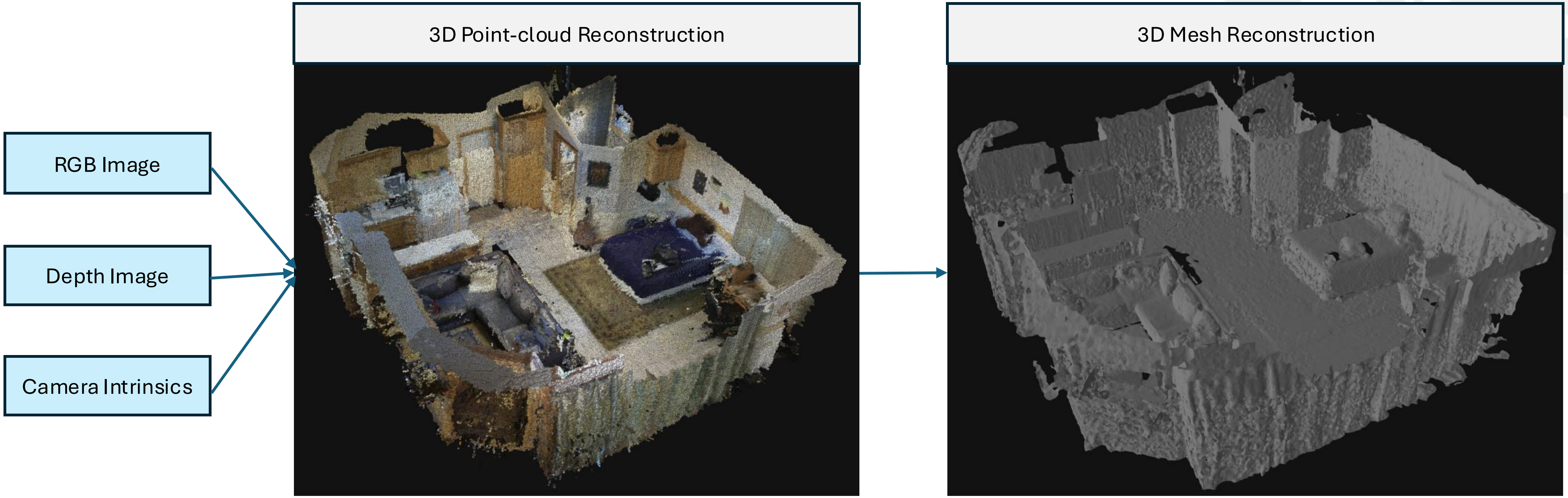


MiDAS



Model	Abs Rel	Log10	Delta
Depth Anything	0.6319	0.5768	0.3692
Midas	0.6435	0.5787	0.3294

3D Reconstruction

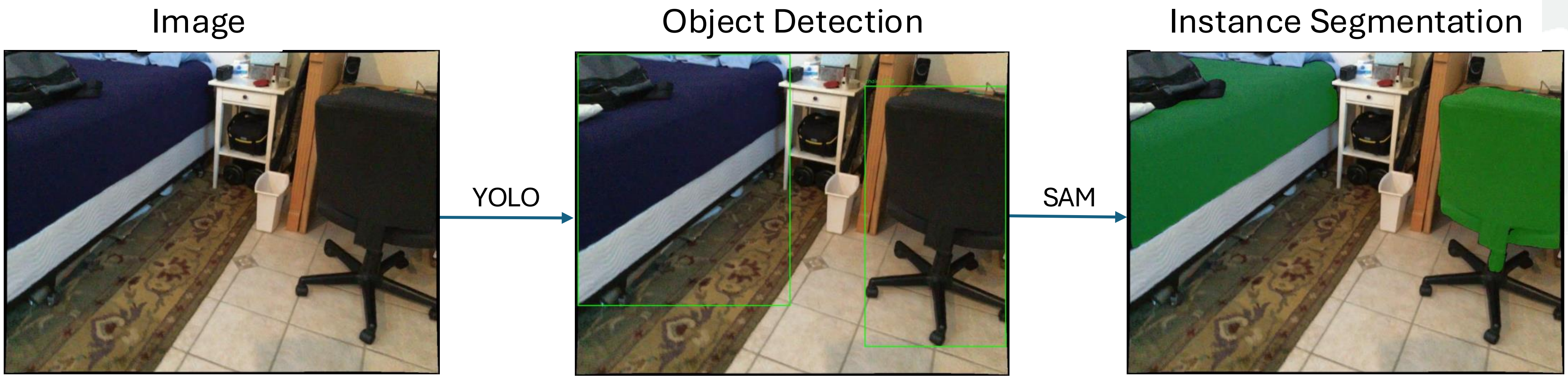


Objective: Build a 3D representation using RGB-D data and camera poses.

Methodology:

- Depth images are backprojected to 3D points in camera coordinates, then transformed to world coordinates using pose matrices. Corresponding colors from RGB images are mapped to the points, creating a colored point cloud.
- A Poisson surface reconstruction creates a mesh, which is cropped based on density, simplified, and saved for visualization.

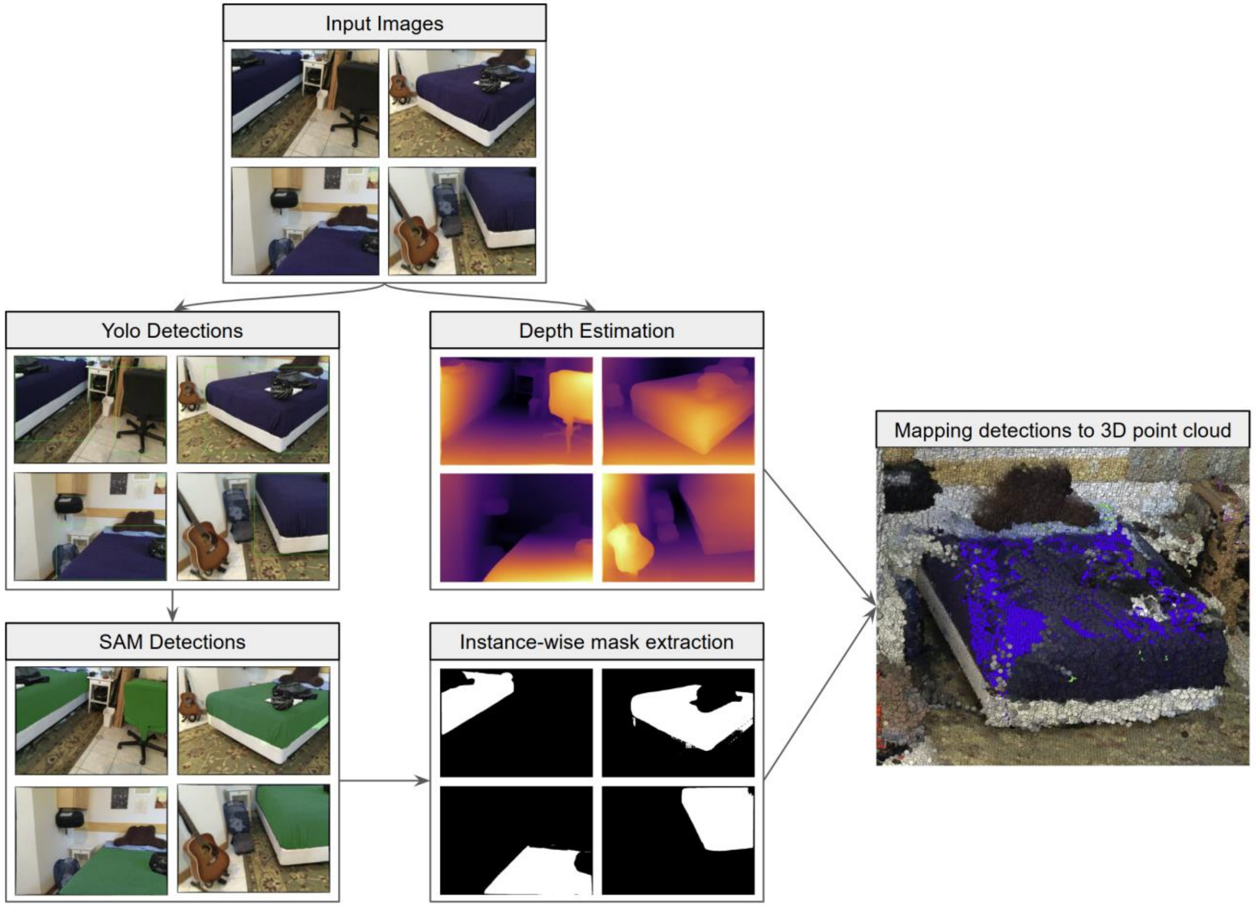
Object Detection & Instance Segmentation



Objective: To accurately identify and segment objects for 3D scene reconstruction
Methodology:

- YOLO large is used to detect objects and provide bounding boxes around them (middle image).
- Mobile SAM (Segment Anything Model) refines the detections by generating segmentation masks for each object (final image), highlighting objects with precise boundaries.

3D Object Mapping



Monocular 3D Scene Reconstruction and Object Mapping Pipeline

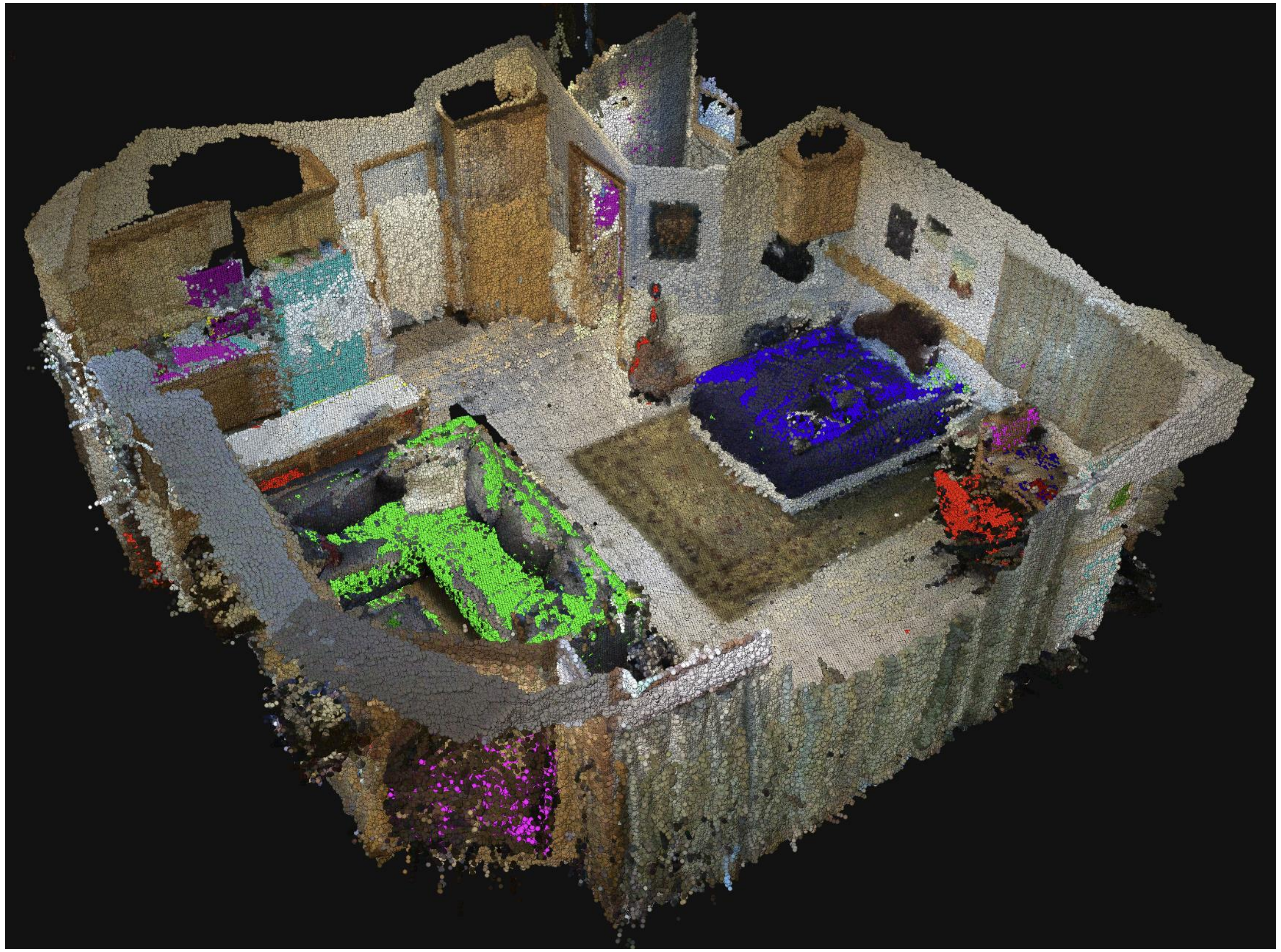
Objective:

Integrate segmented objects into a 3D scene.

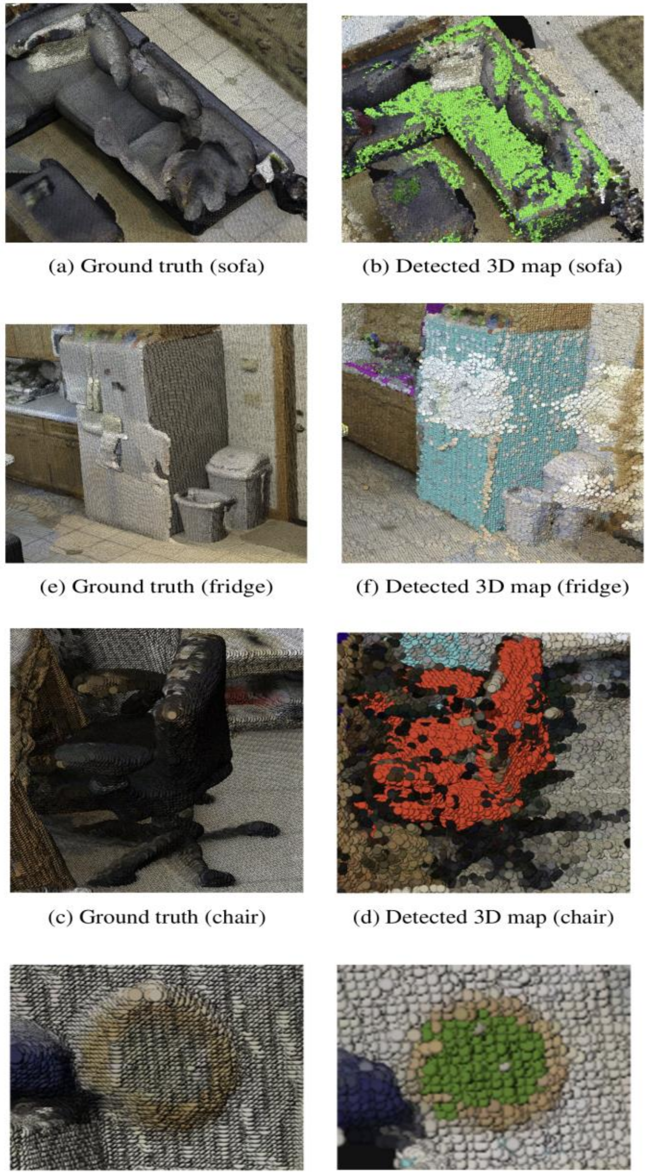
Methodology:

- **Map Segmented Pixels to 3D Points**: Convert 2D segmentation masks into 3D points using depth information and camera parameters.
- **Merge Objects**: Cluster 3D points into objects by class, based on spatial proximity or bounding box overlap, ensuring meaningful groupings.
- **Noise Removal**: Apply statistical outlier removal to clean noisy points from each object.
- **Assign Colors**: Use a predefined color map to assign unique colors to object classes.

3D Object Mapping (Results)



3D object mapping performed on scene 0000. Different colors denote different classes detected. Ex: Sofa(green), Chair(red)



(a) Ground truth (sofa) (b) Detected 3D map (sofa)

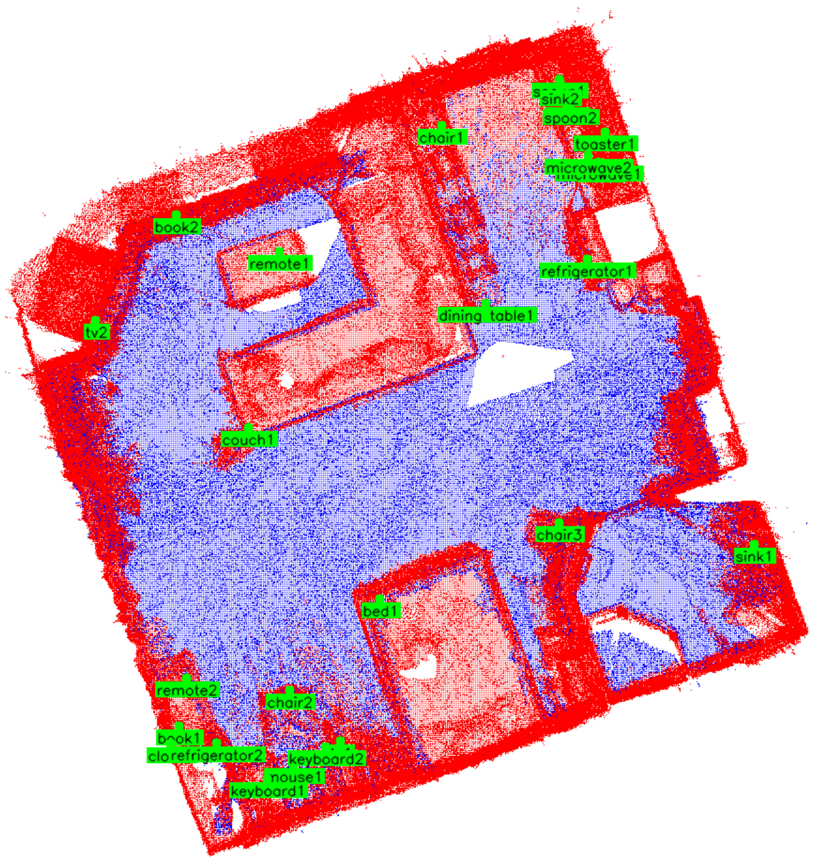
(e) Ground truth (fridge) (f) Detected 3D map (fridge)

(c) Ground truth (chair) (d) Detected 3D map (chair)

(g) Ground truth (clock) (h) Detected 3D map (clock)

Bird's-Eye View

Ground truth



Birdseye view

Objective:

Create a 2D occupancy grid from 3D point cloud data for simplified navigation.

Methodology:

- **Classify Floor and Obstacles**: Based on the Z-coordinate of points, classify them as floor or obstacles, and fill the corresponding pixels in the bird’s-eye view images.
- **Annotate Objects**: Use density centers from precomputed 3D annotations to draw labeled circles (green) on the bird’s-eye view visual map, indicating object class, id and locations.

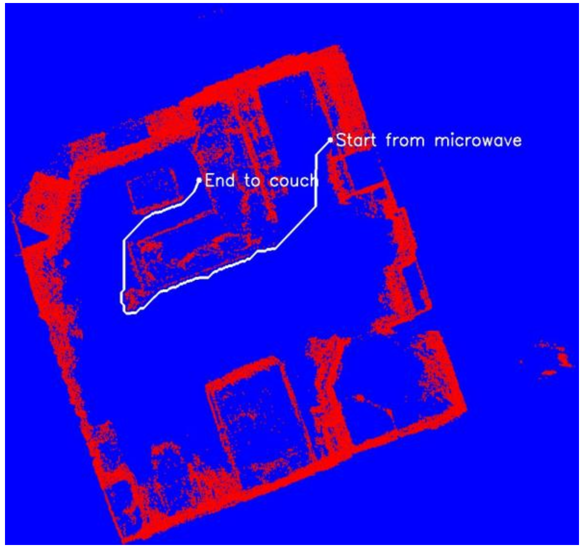
Optimal Path Planning

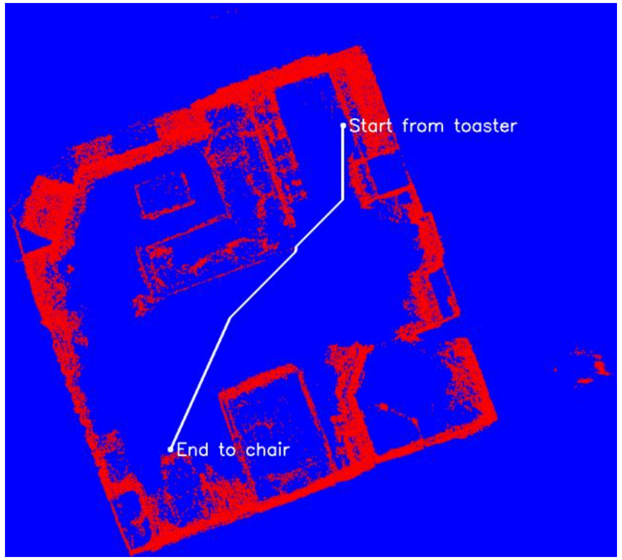
Objective:
Determine the shortest route from a start point to a destination while avoiding obstacles and maintaining a safe distance from them.

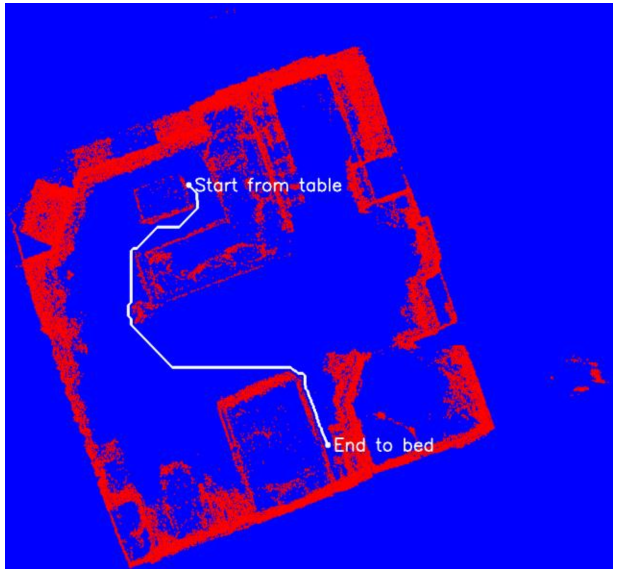
Methodology:
The grid map is converted into a binary map, where obstacles are marked as 1s and free space as 0s. A distance transform precomputes the nearest obstacle distances. The A* algorithm is employed to evaluate neighboring cells based on a combined cost: the actual path cost and the Euclidean distance to the goal. This heuristic guides the search efficiently toward the destination.

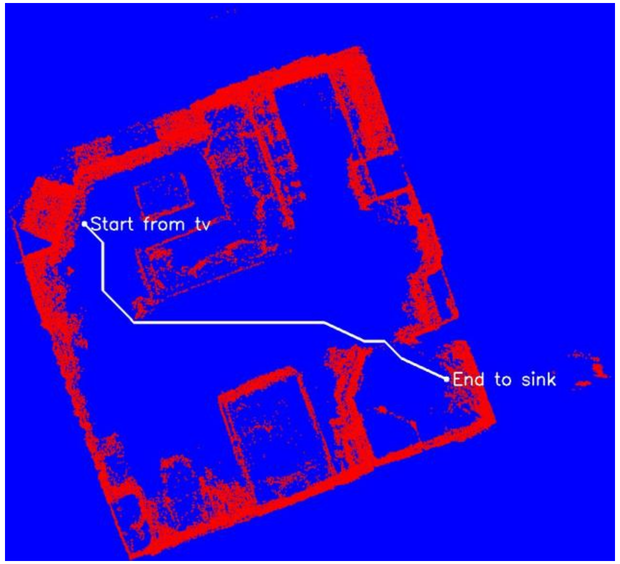
Results:
The optimal path is reconstructed and visualized, ensuring adherence to constraints. Obstacles are successfully avoided, and the route maintains the required minimum distance from them.

Optimal Path Planning









Discussion

- 3D scene reconstruction combines RGB-D data, camera poses, and point cloud processing to create accurate spatial models for downstream tasks.
- Object detection and segmentation with YOLOv8 and Mobile SAM effectively identify and map diverse objects, including smaller ones like remotes and clocks.
- Bird's-eye view generation and path planning convert spatial data into actionable navigation insights, with the A* algorithm ensuring obstacle-free paths.
- The system demonstrates high accuracy and sensitivity, successfully handling a wide range of object types and scenarios.
- The modular and scalable design ensures flexibility, seamless integration, and adaptability to both simple and complex environments.

Future Work

- Improve depth estimation accuracy and robustness across diverse scenarios.
- Extend navigation capabilities to handle three-dimensional spaces and vertical complexity.
- Optimize real-time processing and enhance scalability for larger environments and complex datasets.

Thank You!