

## 1 A Presentation of the Real-world VLN

2 Figure 1 shows the lab environment for the real-world navigation. The left part presents the arrange-  
3 ment of various types of rooms, and the right part presents the visualization of a house layout.



Figure 1: The arrangement and layout of the real-world lab environment.

4 Figure 2 shows two examples of the Interbotix LoCoBot navigating in the real-world environment.  
5 The robot can identify landmarks mentioned in the language instructions and successfully avoid  
6 obstacles during navigation.

**Instruction:** Go past the green plant into the bathroom, and stop in front of the toilet.



**Instruction:** Pass by the table and stop in front of the green plant next to the basketball.



Figure 2: Examples of the vision-and-language navigation in the real-world environment.

## 7 B Visualization of the Semantic Traversable Map

8 Figure 3 and Figure 4 illustrate the construction and updating of the semantic map and traversable  
 9 map during navigation. Both maps are agent-centered to better predict candidate waypoints. In the  
 10 traversable map, warmer color indicates higher traversability probabilities, while the square black  
 11 spots in the semantic map represent predicted navigable waypoints. In step 1 of navigation, the  
 12 VLN model executes a 360-degree rotation to construct a more comprehensive semantic map and  
 13 3D feature fields. In the subsequent steps, the agent can only observe the forward-facing view. These  
 14 examples illustrate that our approach can predict high-quality semantic maps of the environment, the  
 15 traversable map can identify candidate waypoints well and achieve obstacle avoidance.

**Instruction:** Turn around, walk out the door on the right, and walk across the hall into the bedroom on the left.

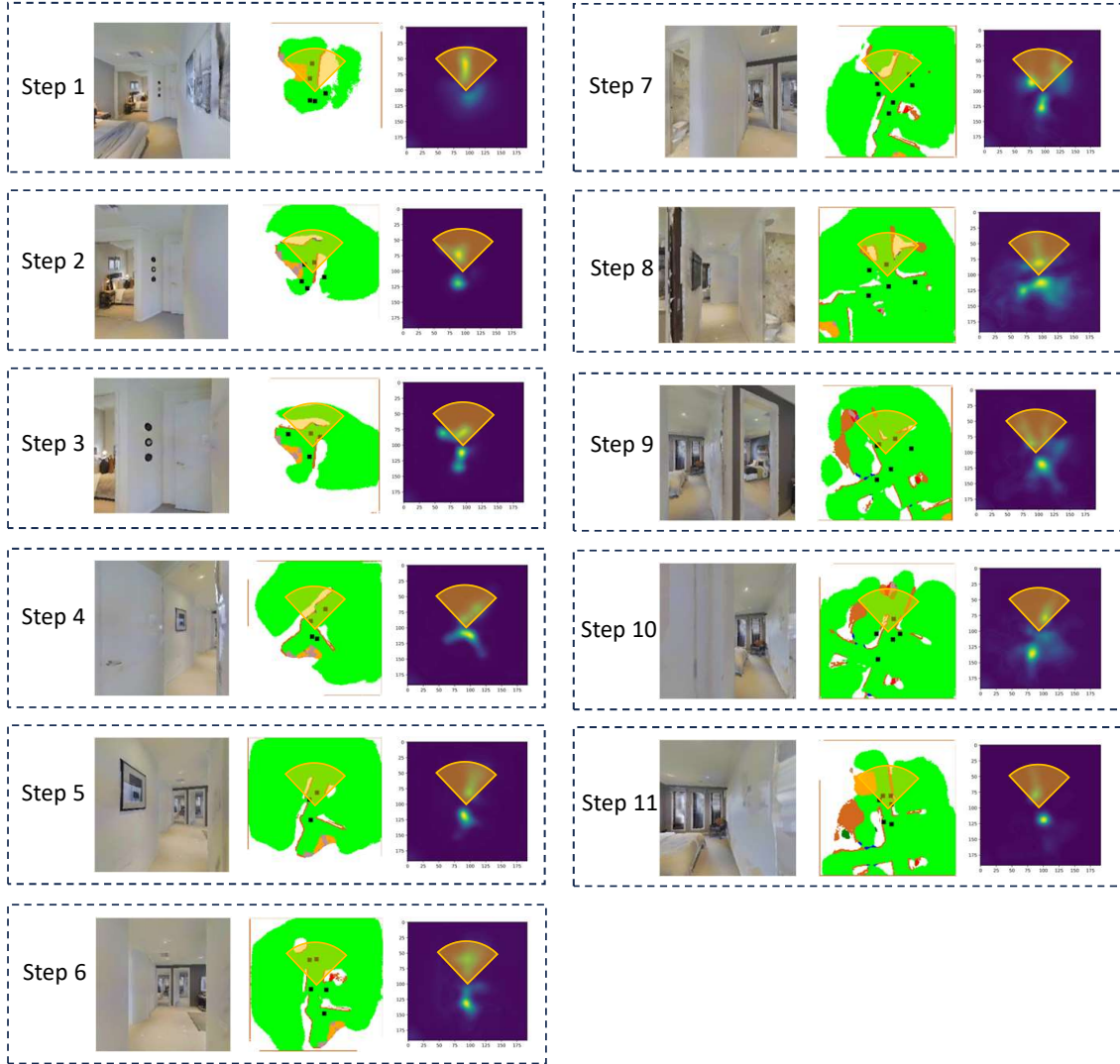


Figure 3: (1/2) The visualization of the RGB observation, semantic map with candidate waypoints, and traversable map during navigation.

**Instruction:** Exit the room. Turn left and walk straight toward the pool. Wait near the pool.

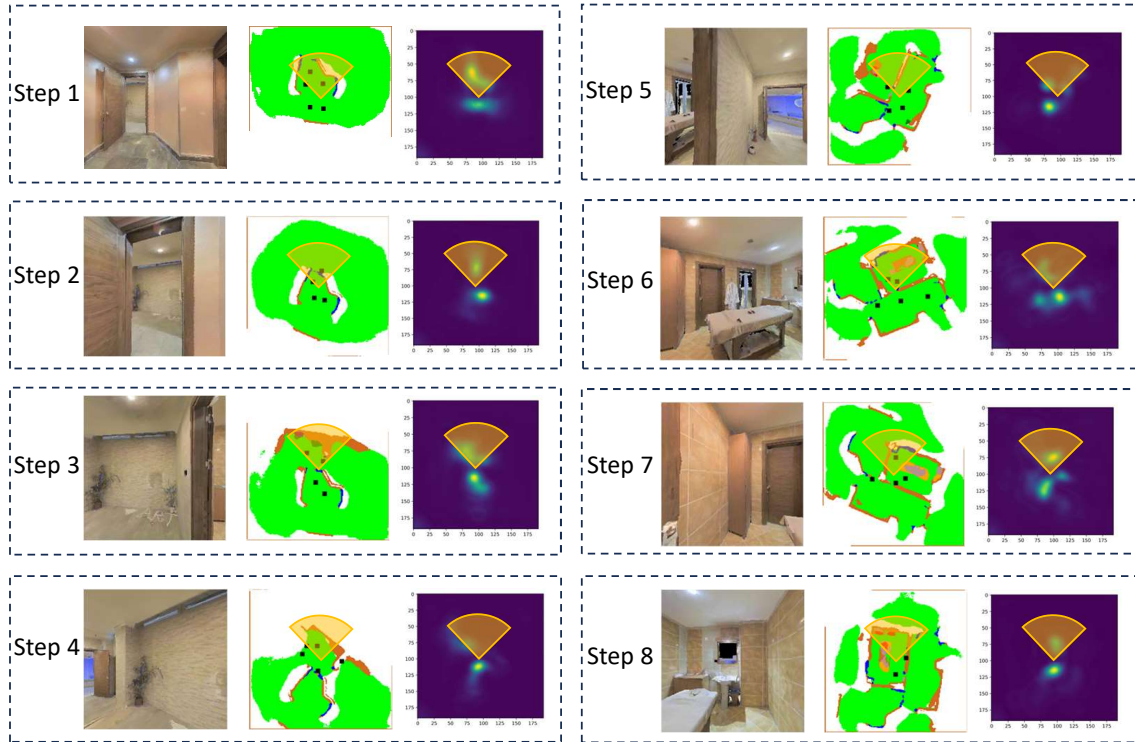


Figure 4: (2/2) The visualization of the RGB observation, semantic map with candidate waypoints, and traversable map during navigation.