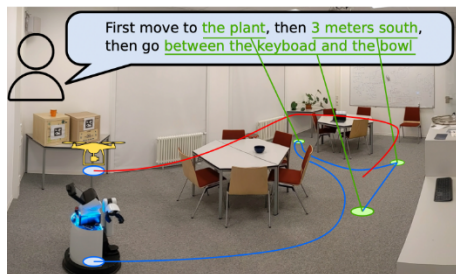


## Idea & Contributions

### Question

VLM - Image is OK.  
Environment - No!

### Solution



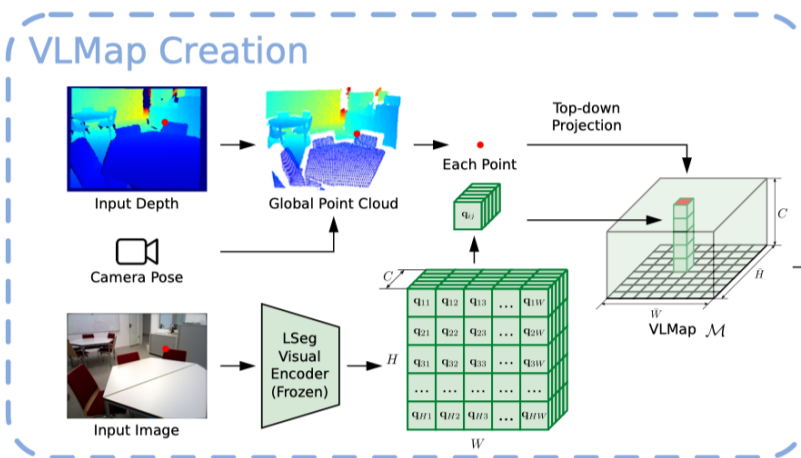
VL features +  
3D reconstruction  
= VLMs

### Contribution

- Combined with LLMs
1. language - localization
  2. generate obstacles maps

## Approaches

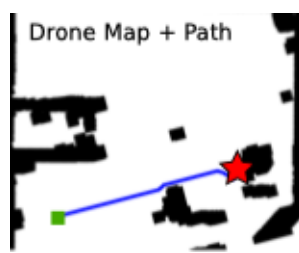
### S1 Build a VLMap



The resulting features contain the averaged embeddings from multiple views of the same object.

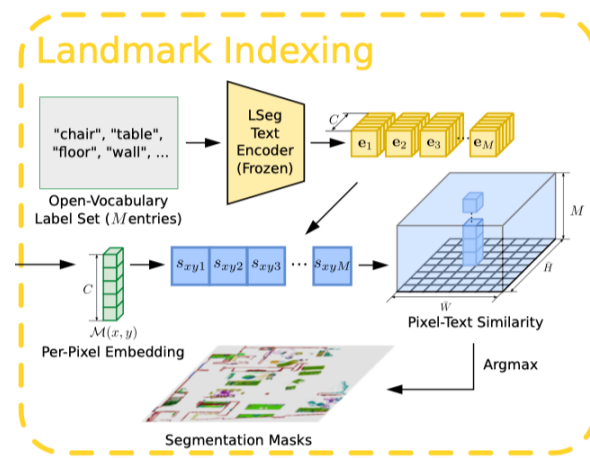
### S3 Obstacle Maps

$$\mathcal{O}_{ij} = \begin{cases} 1, & t_1 \leq P_W^y \leq t_2 \text{ and } p_{map}^x = i \text{ and } p_{map}^y = j \\ 0, & \text{otherwise} \end{cases}$$



Avoid collision and get the shortest path.  
Provide different lists of obstacle categories.

### S2 Localize Landmarks



In grid map, every pixel maps the most related language-based category.

### S4 Navigation with LLM

primitives	functions
move_to(pos)	move to a position on the map.
move_to_left(object_name)	move to the left side of the nearest front object.
move_to_right(object_name)	move to the right side of the nearest front object.
get_pos(object_name)	get the map position of the nearest front object.
get_contour(object_name)	get the contour turning points of the nearest front object on the map.
with_object_on_left(object_name)	turn until the nearest object is on the robot's left side.
with_object_on_right(object_name)	turn until the nearest object is on the robot's right side.
move_in_between(object_a, object_b)	move in between two objects.

Allow referencing precise spatial goals.  
subgoals - generate code - unseen CMD

## Experiments

### Multi-Object Navigation

Tasks	No. Subgoals in a Row				Independent Subgoals
	1	2	3	4	
LM-Nav [13]	26	4	1	1	26
CoW [12]	42	15	7	3	36
CLIP Map	33	8	2	0	30
VLMs (ours)	<b>59</b>	<b>34</b>	<b>22</b>	<b>15</b>	<b>59</b>
GT Map	91	78	71	67	85

### Zero-Shot Goal Navigation

Tasks	No. Subgoals in a Row			
	1	2	3	4
LM-Nav [13]	5	5	0	0
CoW [12]	33	5	0	0
CLIP Map	19	0	0	0
VLMs (ours)	<b>62</b>	<b>33</b>	<b>14</b>	<b>10</b>
GT Map	76	48	33	29

### Cross-Embodiment Navigation

Tasks	No. Subgoals in a Row								Independent Subgoals
	1	2	3	4	1	2	3	4	
	SR SPL		SR SPL		SR SPL		SR SPL		SR
LoCoBot (ground map)	53	<b>49.0</b>	28	<b>17.8</b>	14	6.7	6	<b>2.5</b>	52.3
Drone (ground map)	53	41.8	28	15.5	14	5.3	6	2.0	53.3
Drone (drone map)	<b>56</b>	45.4	<b>30</b>	16.3	<b>17</b>	<b>7.0</b>	<b>7</b>	<b>2.5</b>	<b>55.0</b>

### Real Robot Experiments

#### Limitations

1. 3D reconstruction Noise
2. odometry drift
3. could not deal with similar objects