Domestic Robot Navigation Using Vision-Language Models

Topic Introduction

Robot Navigation in Domestic Contexts

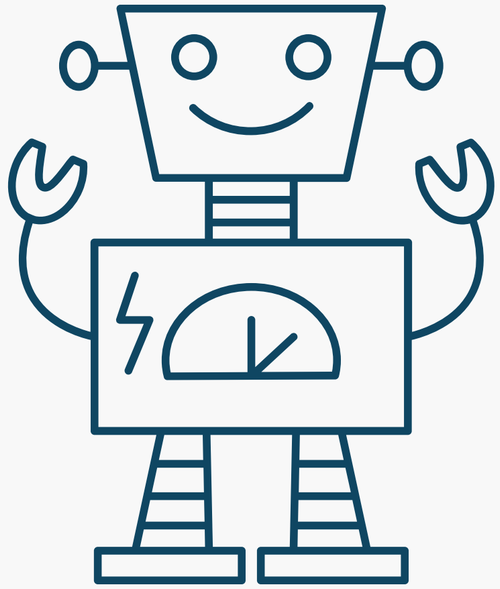
[Video Intro]

[Video Intro 2]

Motivation

Robustness

Adaptability



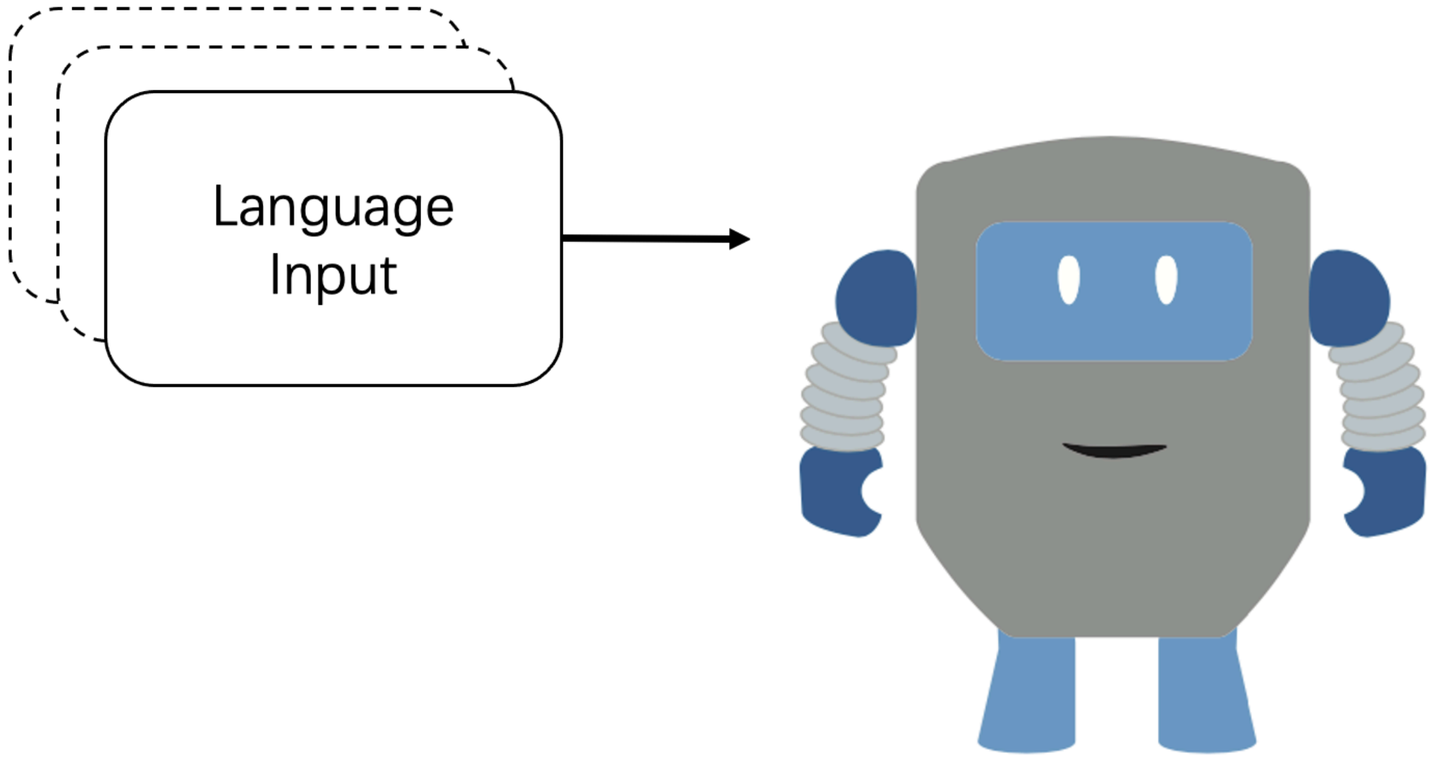
Generalizability

Solution: Using Vision-Language Models!

Suggested Approaches

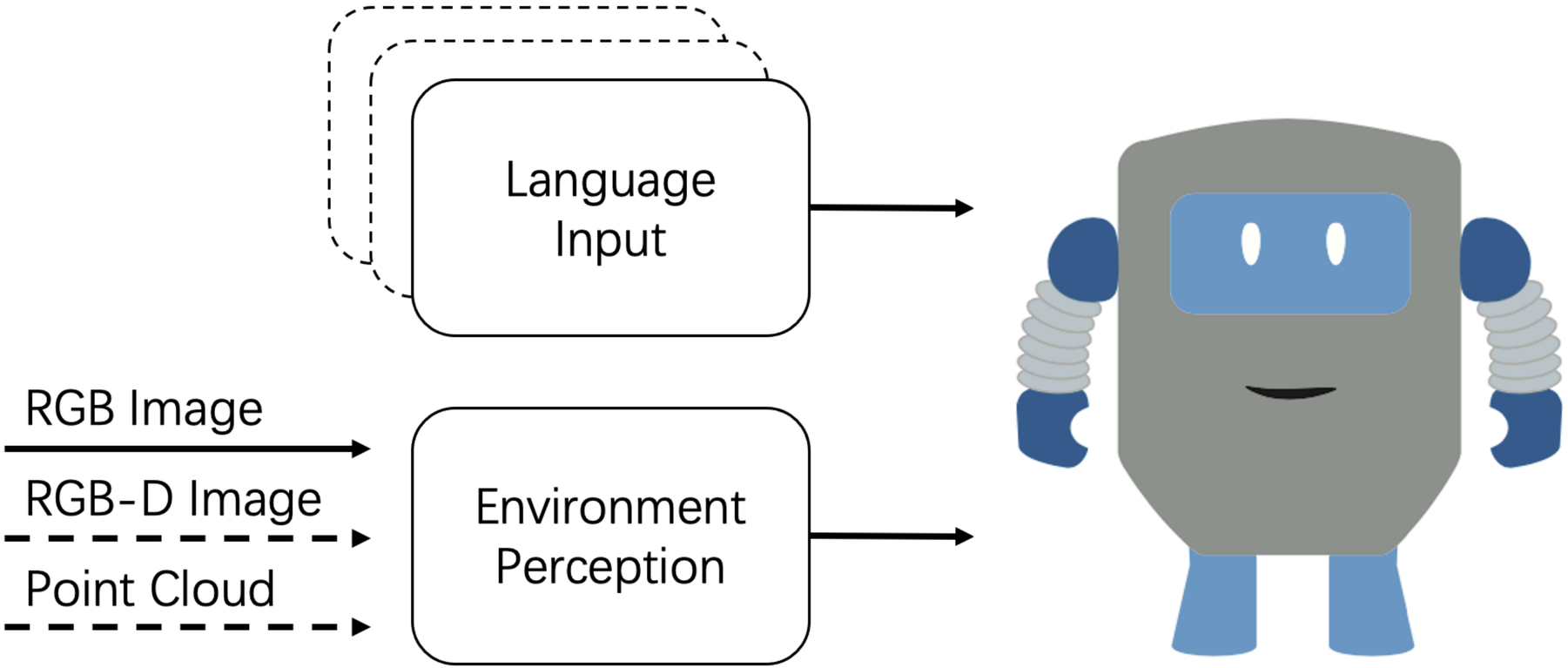
- Large Model Navigation (LM-Nav)
- Visual Language Maps (VLMaps)

Vision-Language Models (VLMs)?



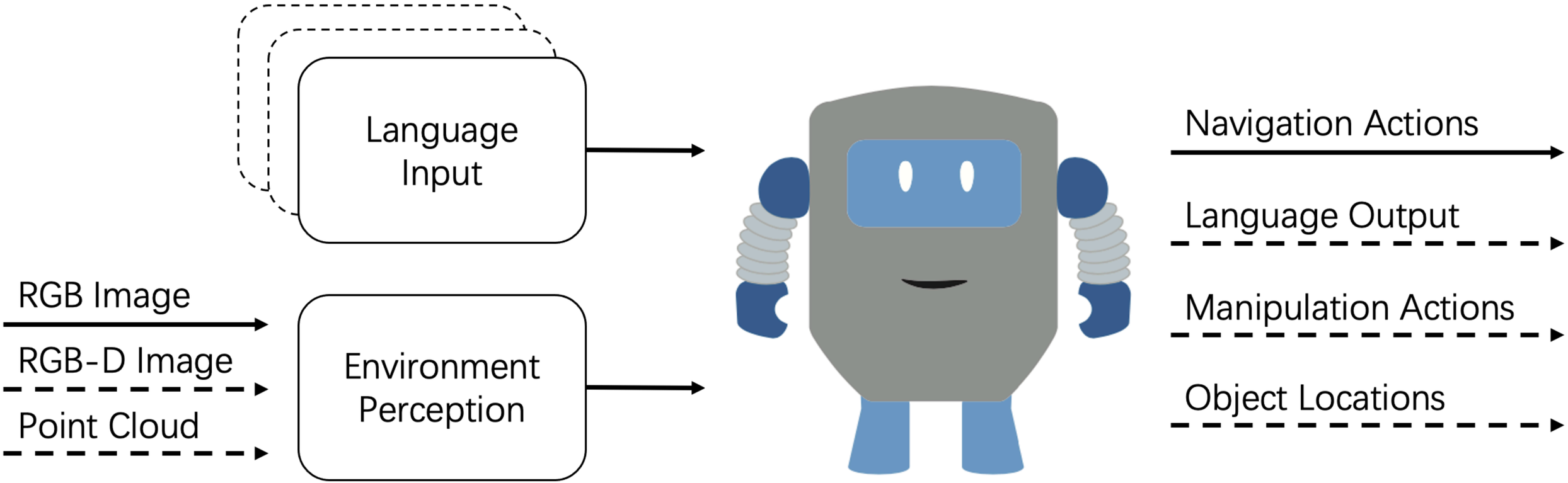
Ref: [1]

Vision-Language Models (VLMs)?



Ref: [1]

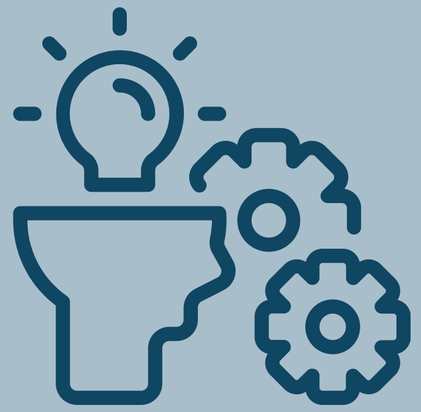
Vision-Language Models (VLMs)?



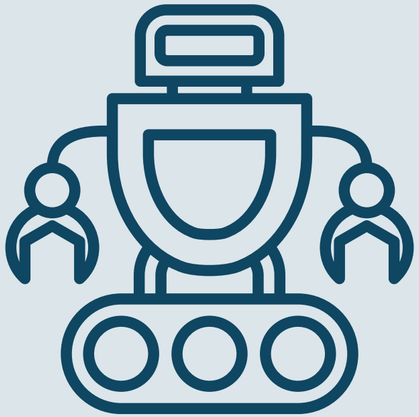
Ref: [1]

Advantages of VLMs in robotics

Generalization and Adaptability



Multimodal Command Interpretation

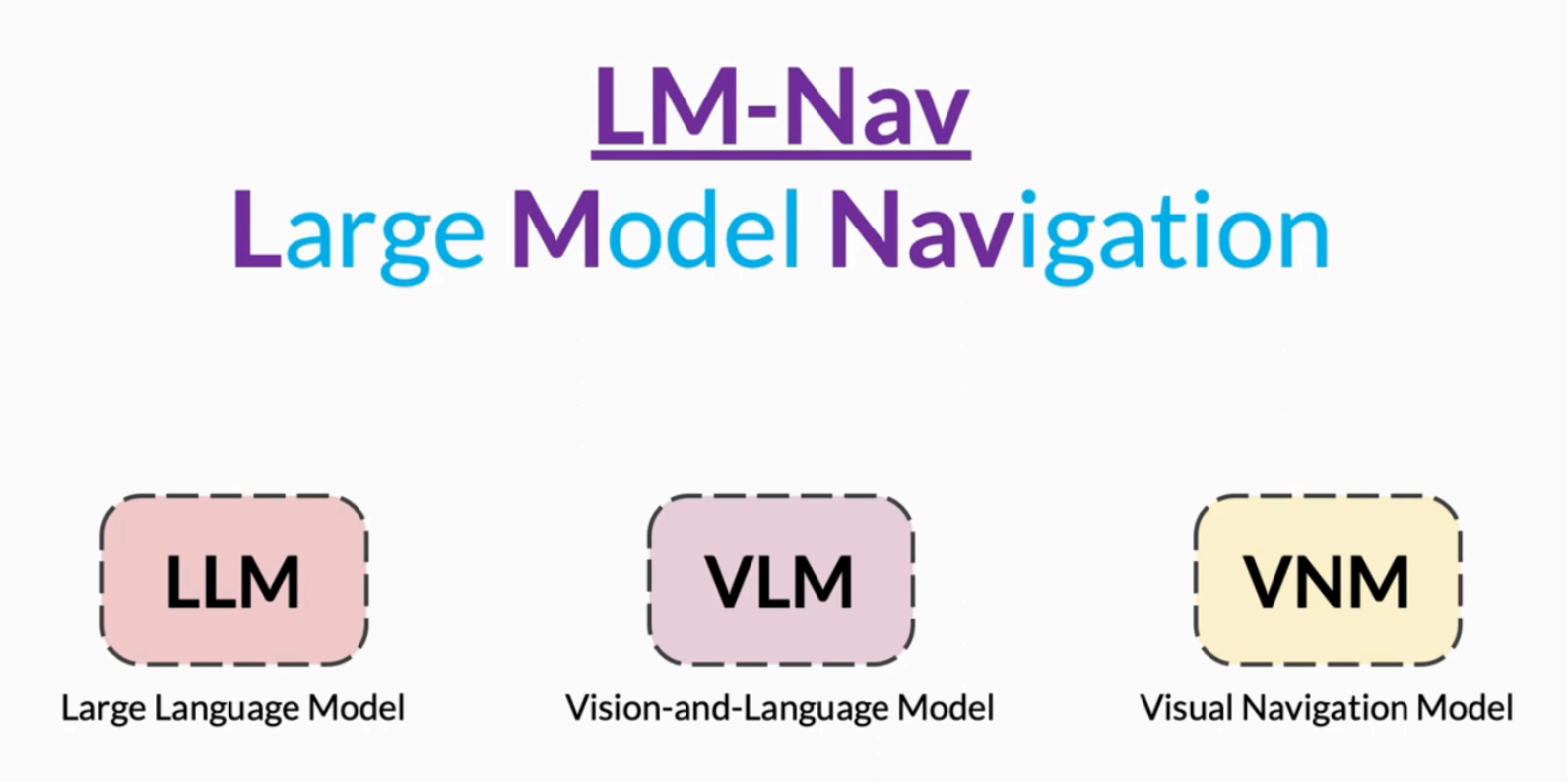


Zero-Shot Task Execution

Suggested Approaches

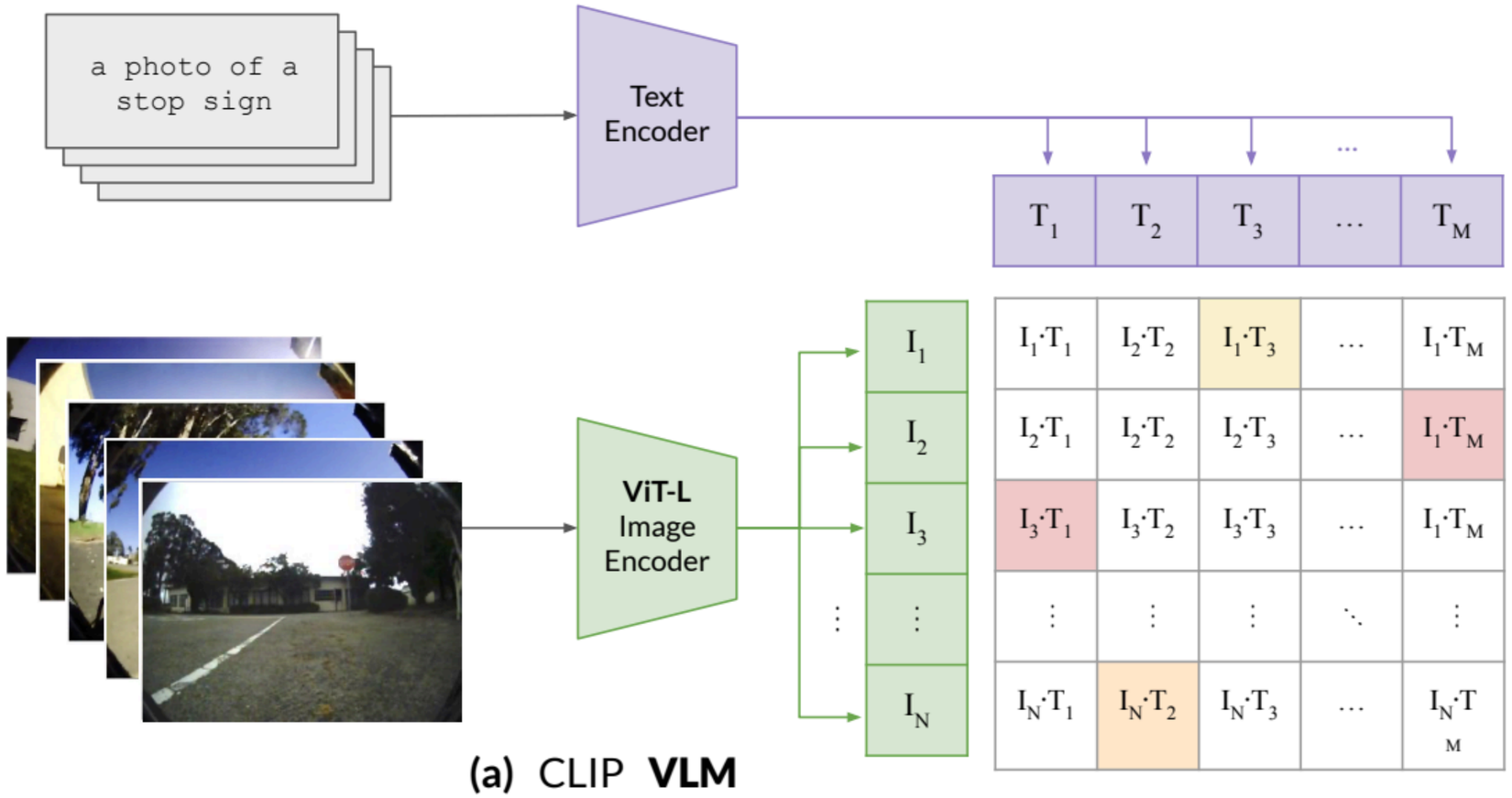
- Large Model Navigation (LM-Nav)
- Visual Language Maps (VLMaps)

LM-Nav: Methodology



Ref: [2]

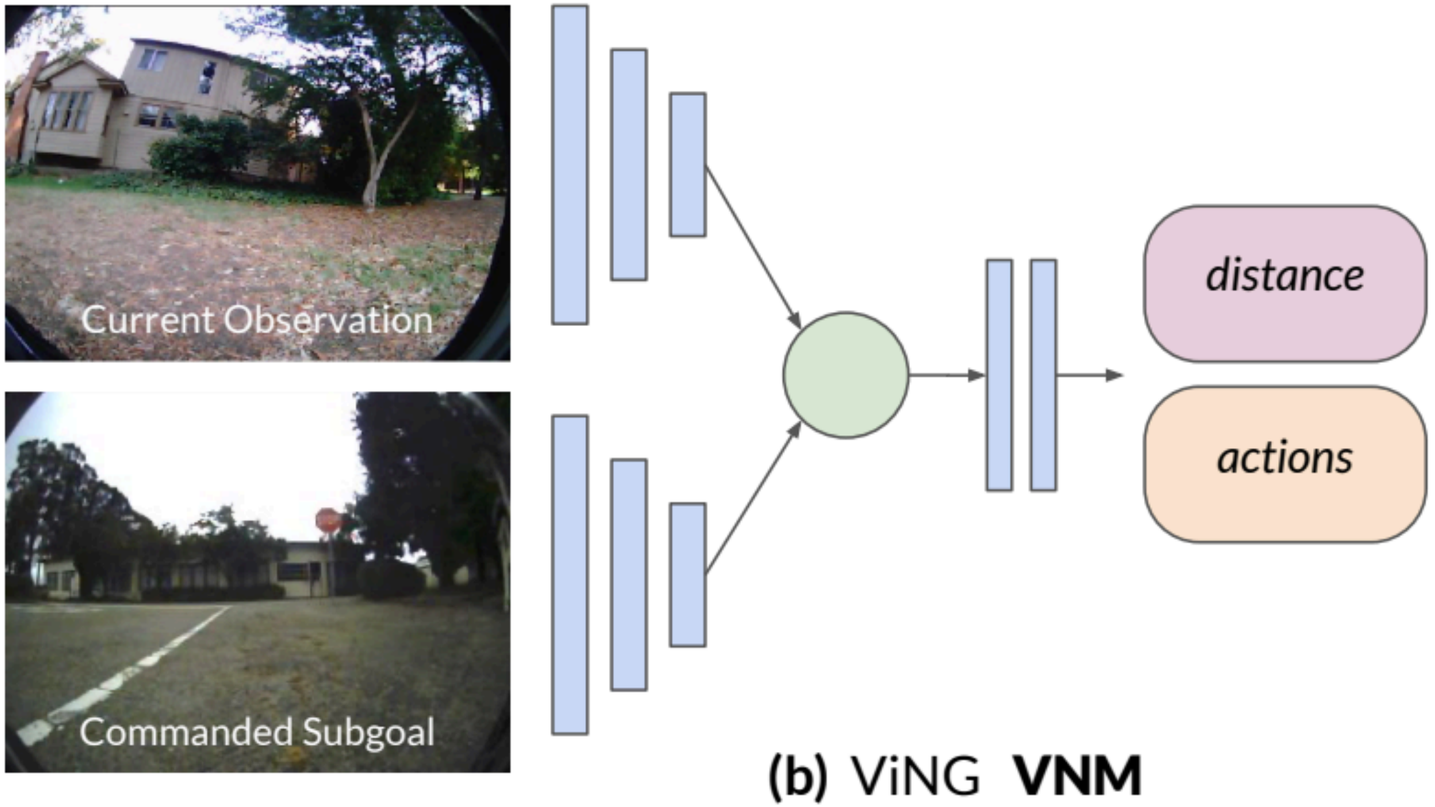
LM-Nav: Methodology



(a) CLIP **VLM**

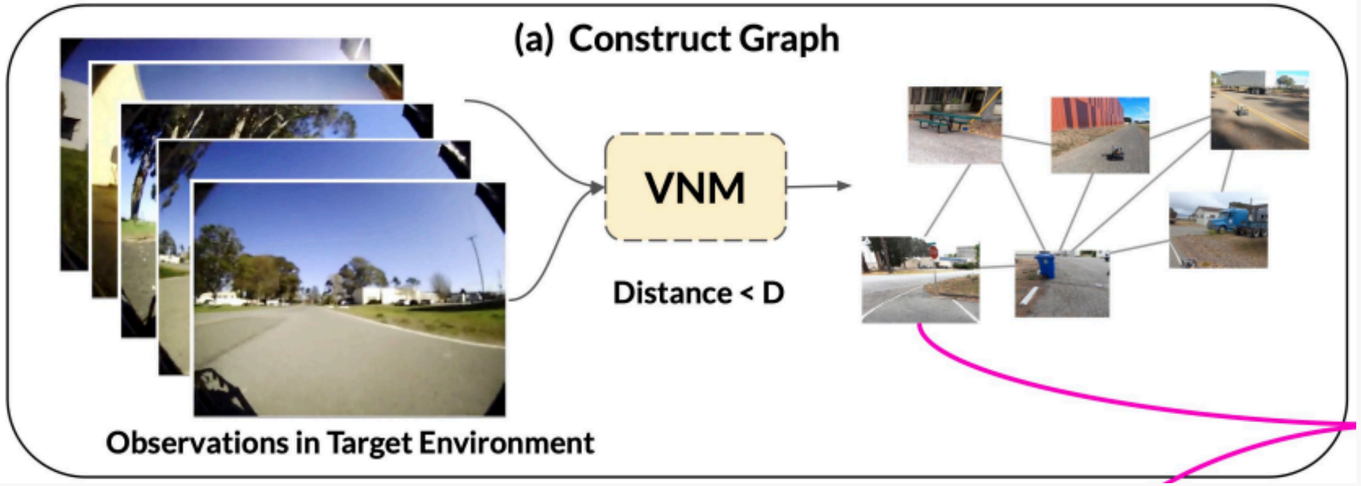
Ref: [2]

LM-Nav: Methodology



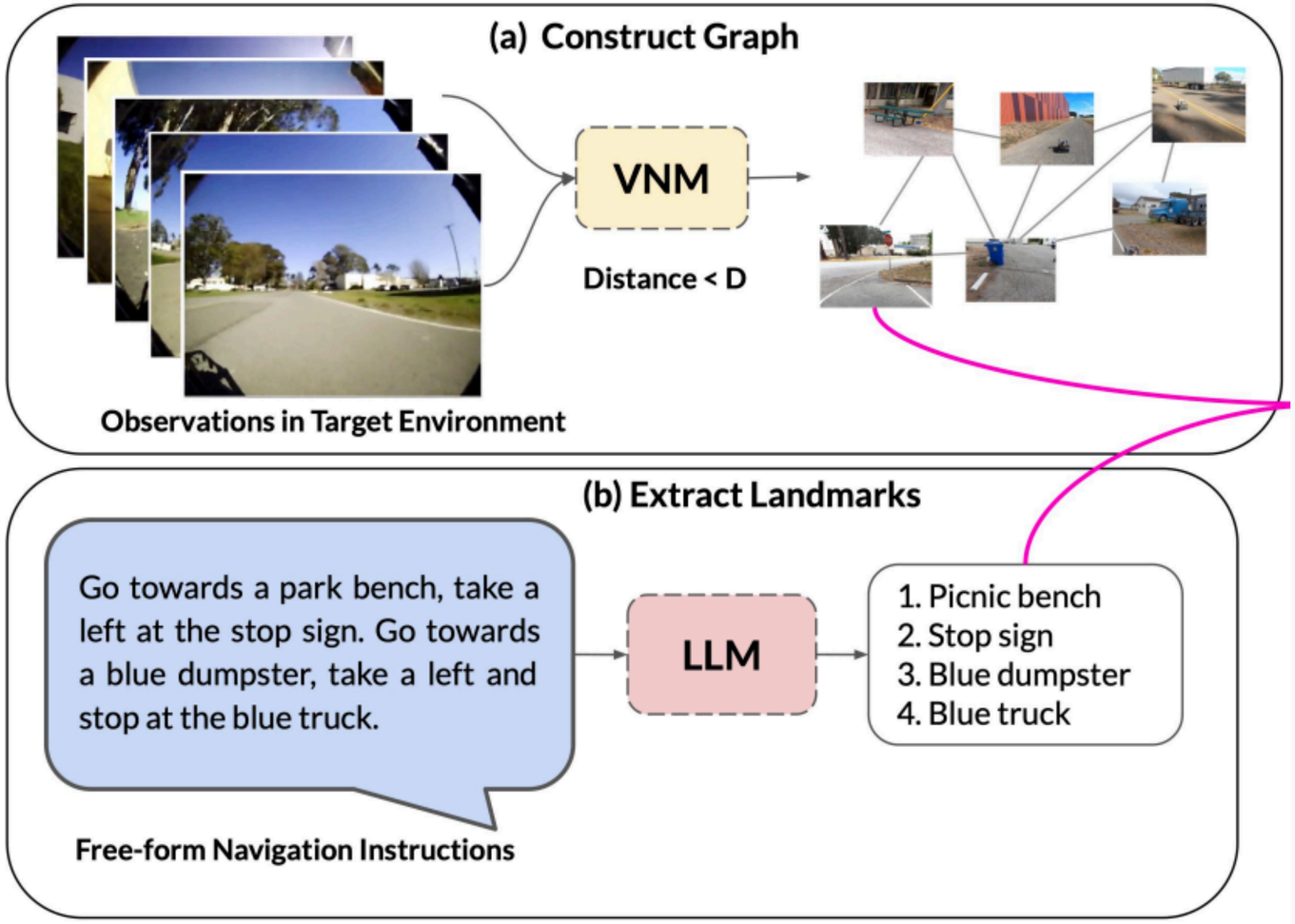
Ref: [2]

LM-Nav: Methodology



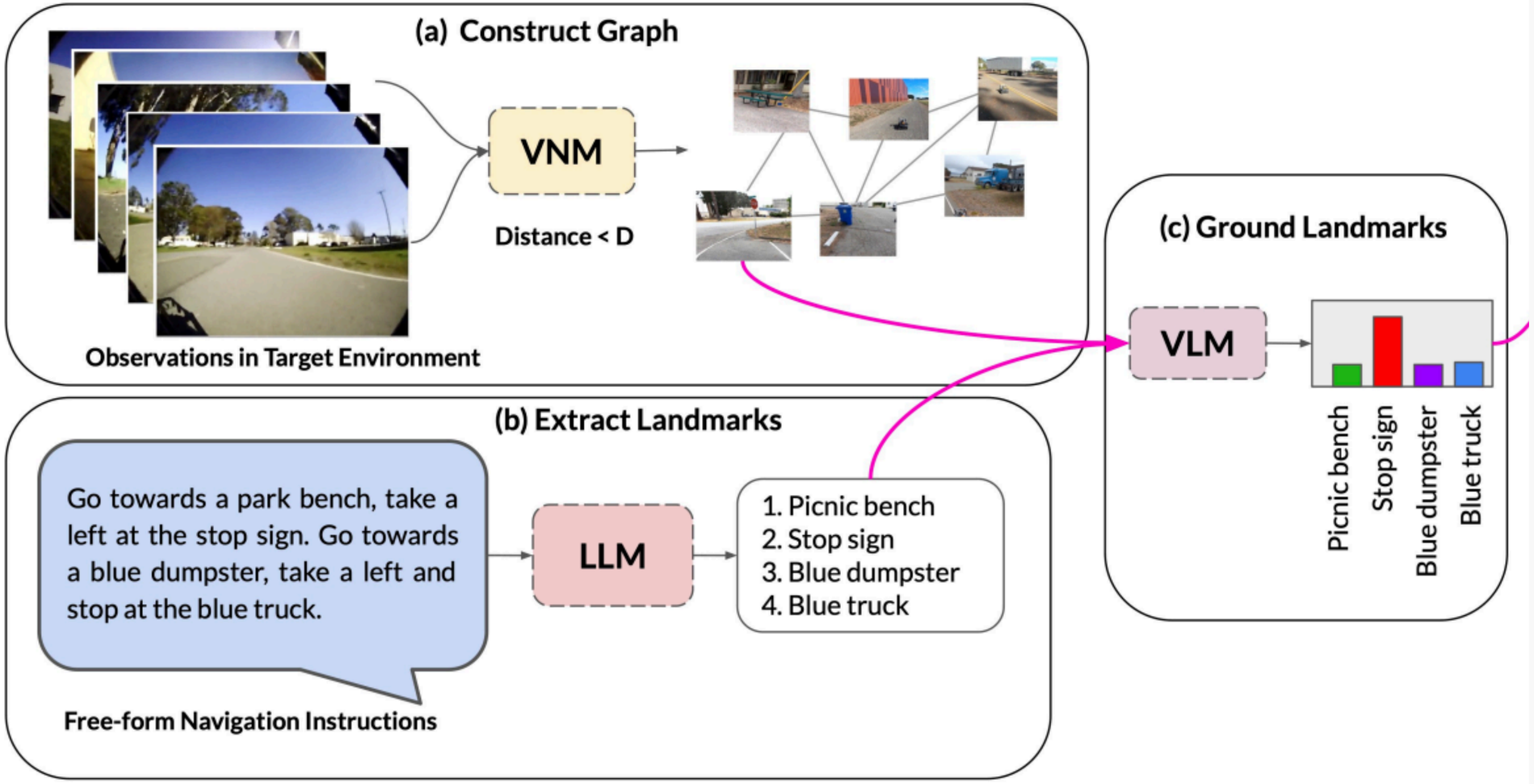
Ref: [2]

LM-Nav: Methodology



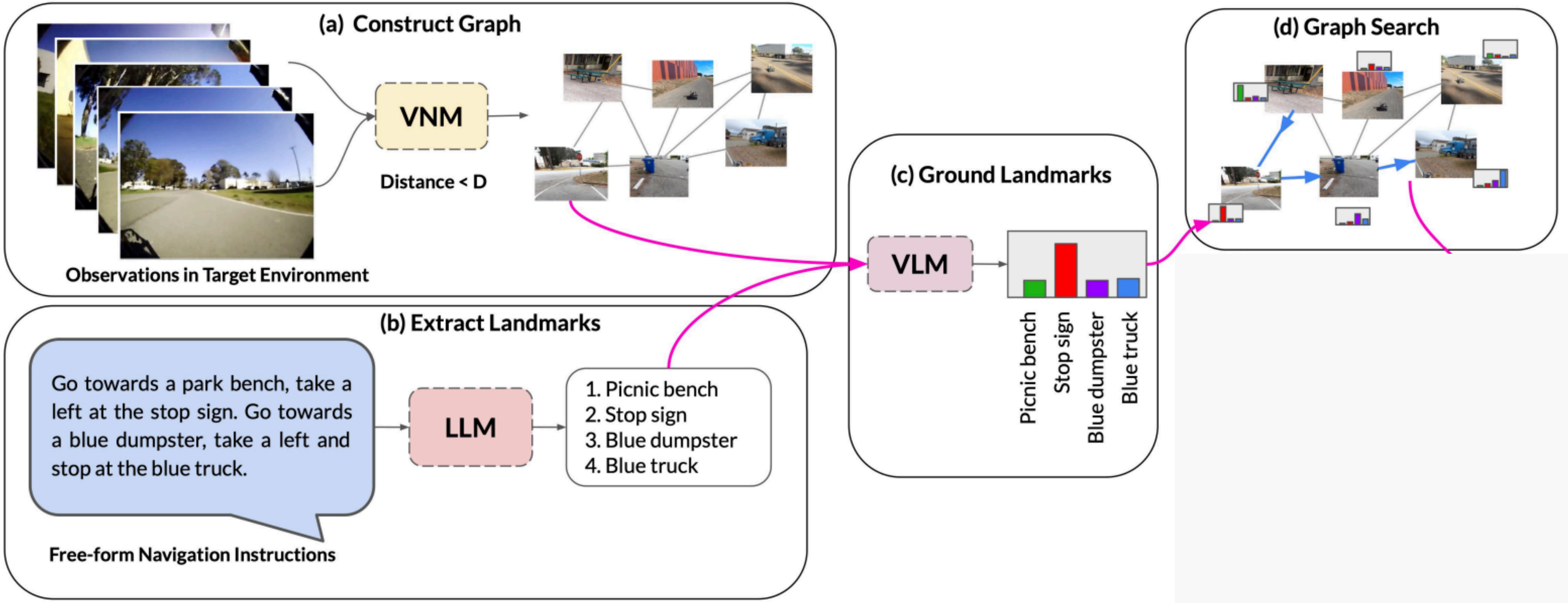
Ref: [2]

LM-Nav: Methodology



Ref: [2]

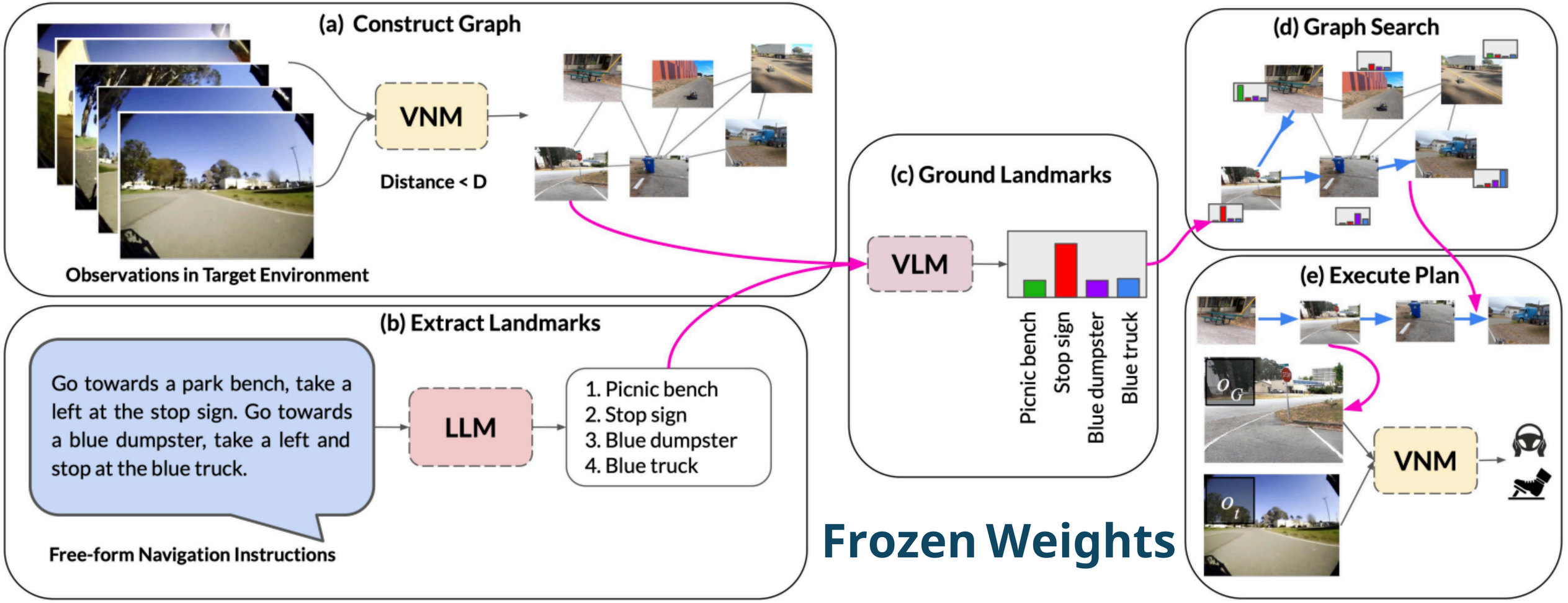
LM-Nav: Methodology



Ref: [2]

LM-Nav: Methodology

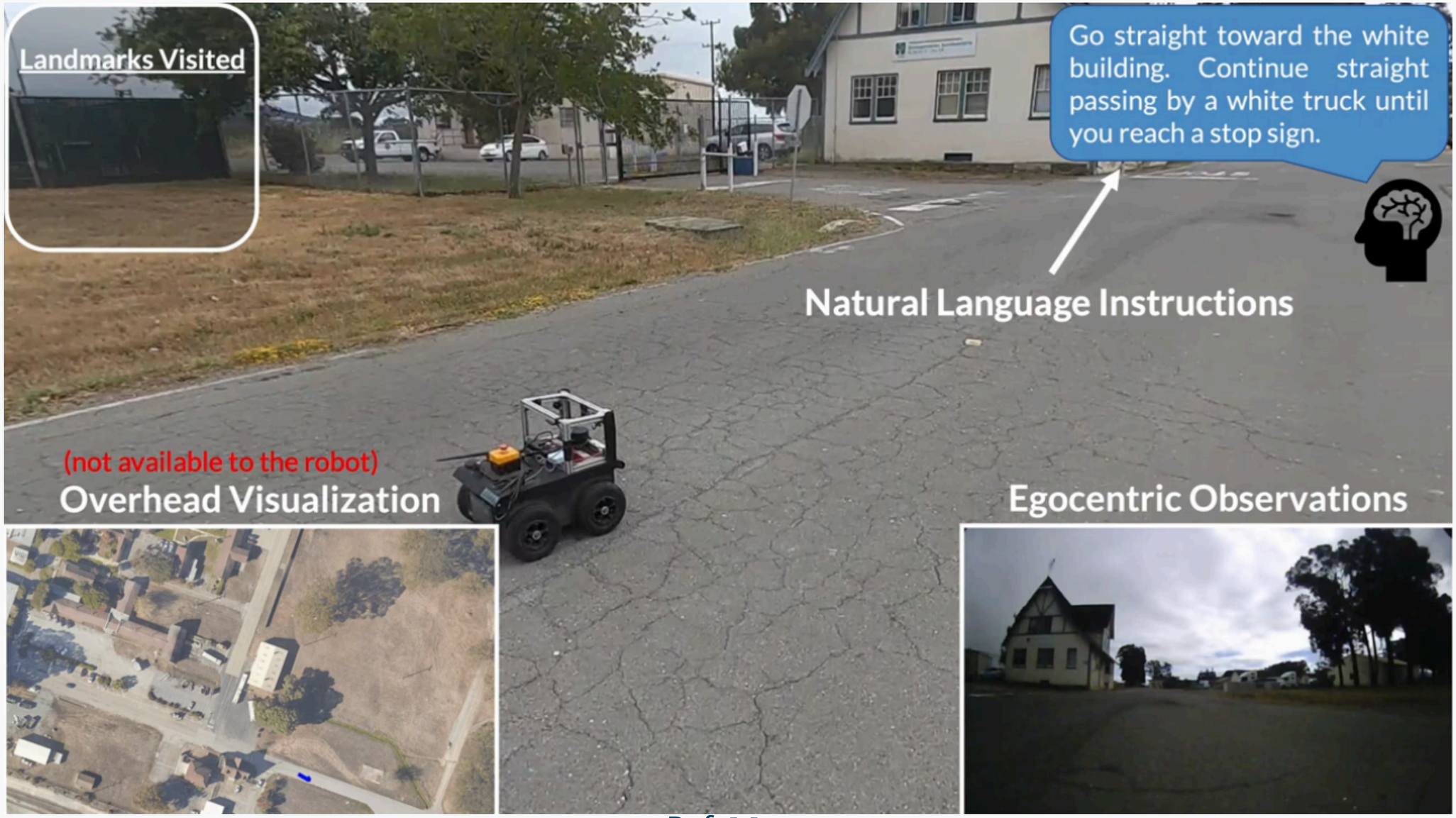
No Global Information



Frozen Weights

Ref: [2]

LM-Nav: Experiments



Ref: [2]

LM-Nav: Experiments

Ref: [2]

LM-Nav: Advantages and Disadvantages

The Advantages:

- Simple and attractive prototype
- Efficient landmark-based navigation

The Disadvantages:

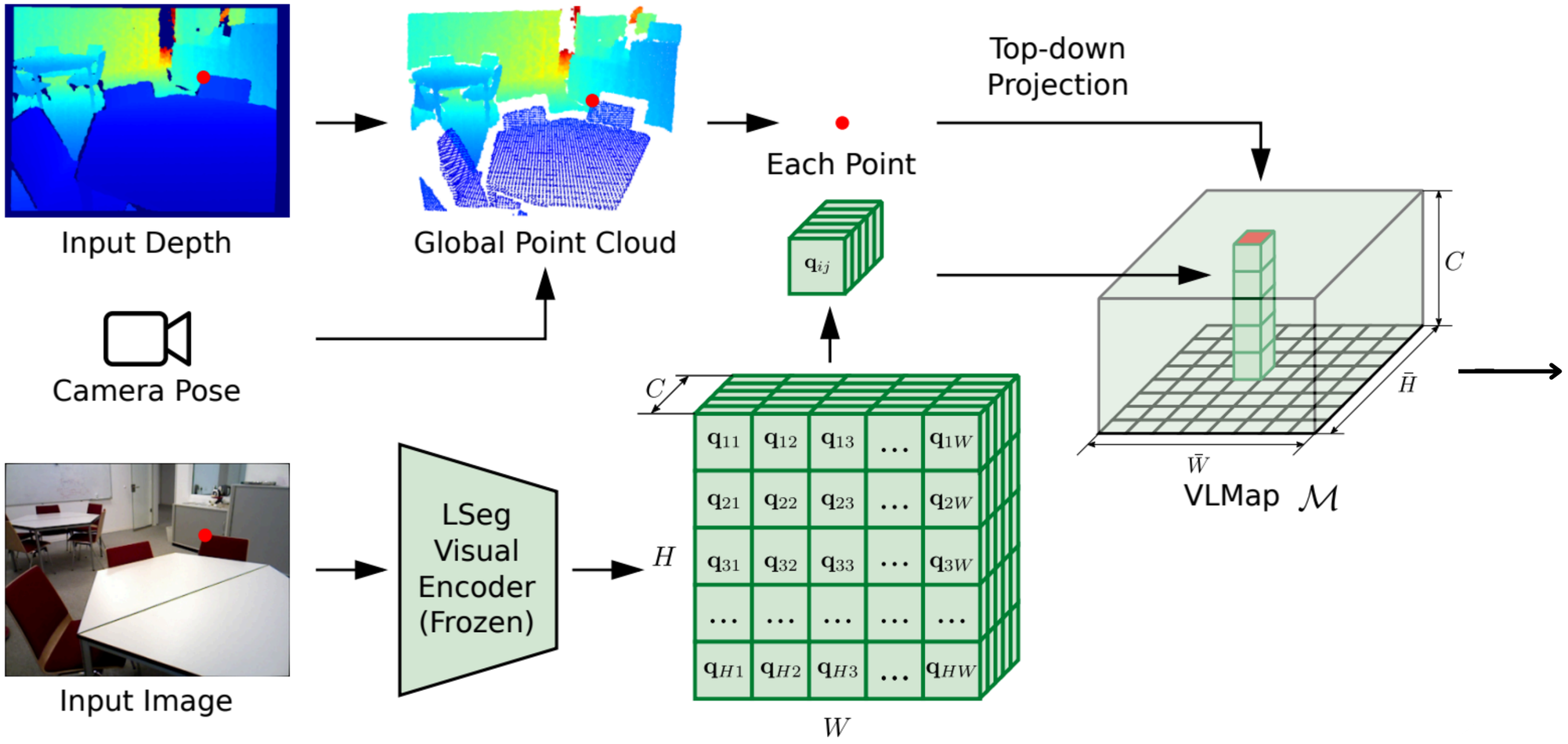
- Reliance on landmarks and pre-explored environments
- Limited spatial precision
- Limitation on wider adoption (Jackal robot)

Suggested Approaches

- Large Model Navigation (LM-Nav)
- Visual Language Maps (VLMaps)

VLMaps: Methodology

VLMap Creation



Ref: [3]

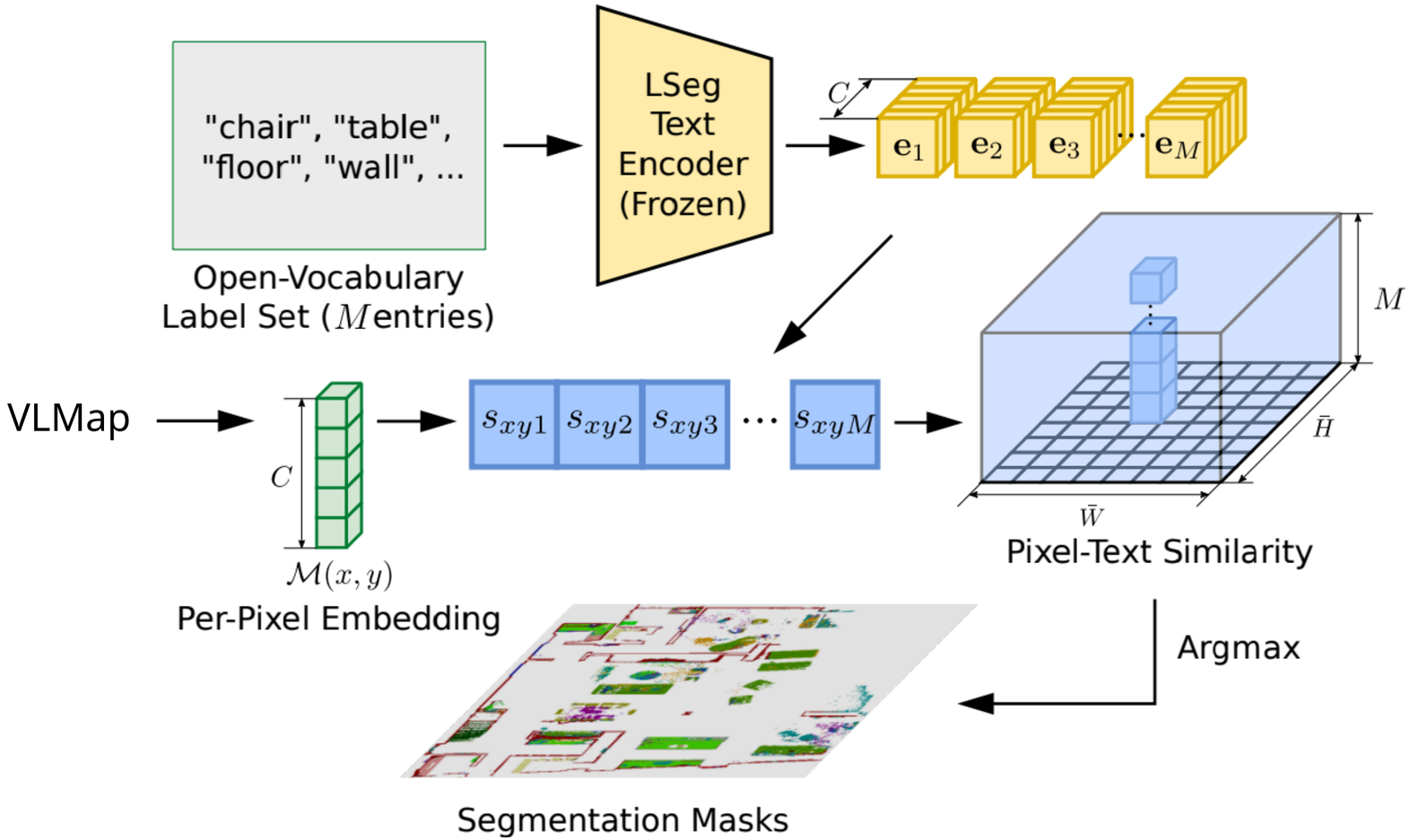
VLMaps: Methodology

Visual Language Map Creation

Ref: [3]

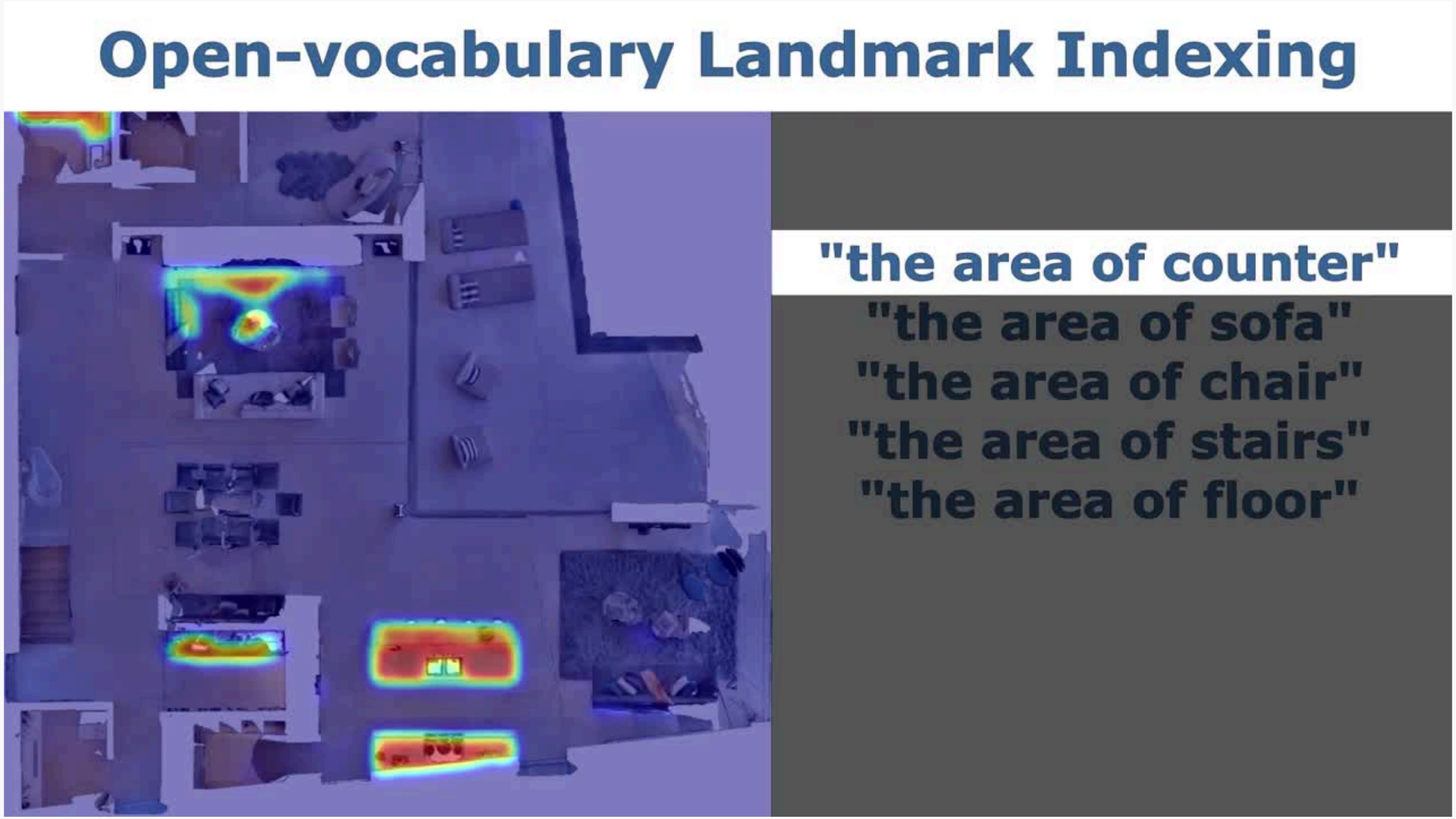
VLMaps: Methodology

Open-Vocabulary Landmark Indexing



Ref: [3]

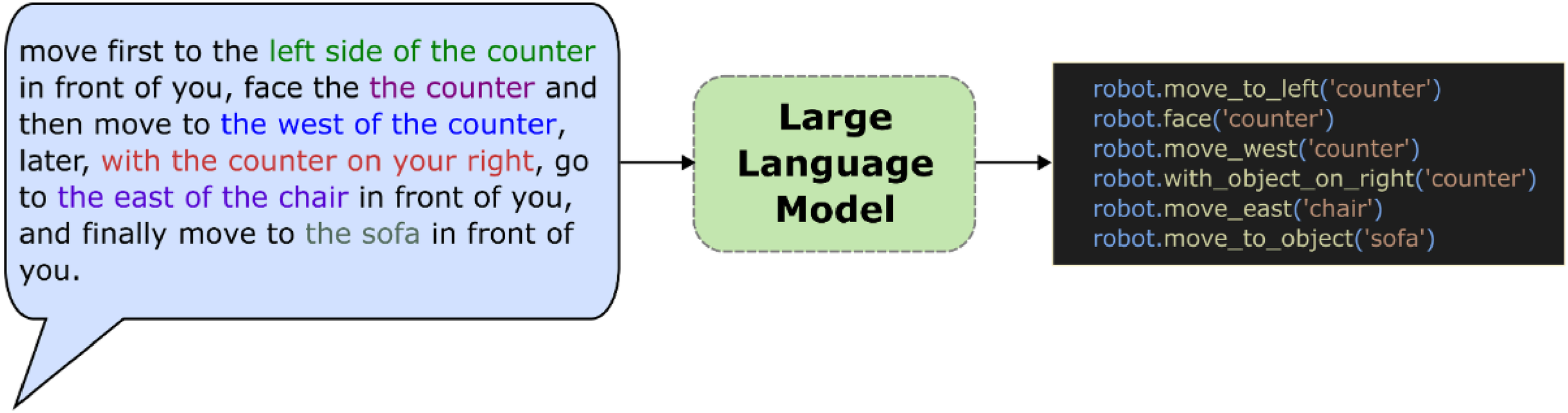
VLMaps: Methodology



Ref: [3]

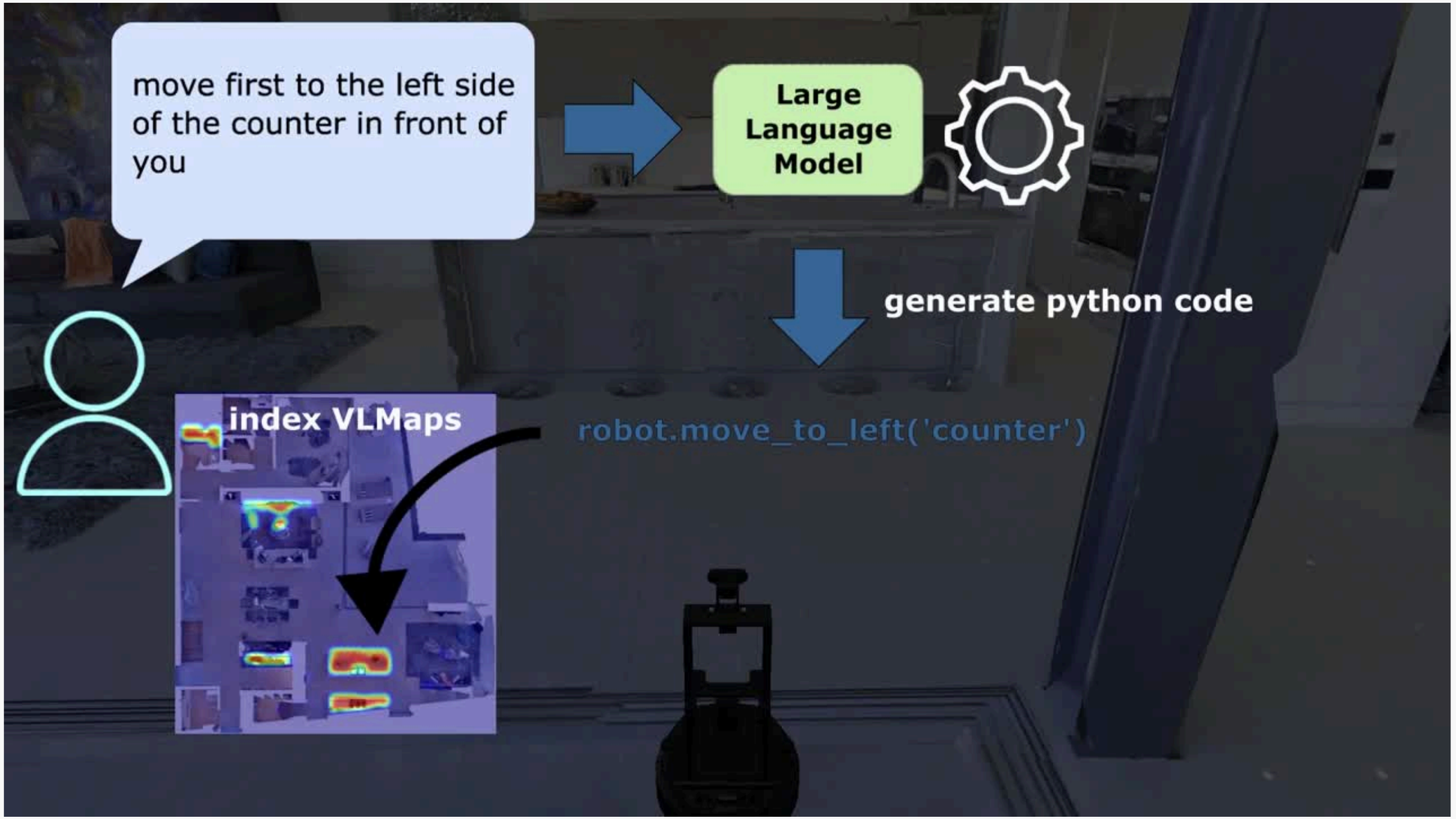
VLMaps: Methodology

Navigation Policies Generation



Ref: [3]

VLMaps: Methodology



Ref: [3]

VLMaps: Experiment

Long-Horizon Spatial Goal Navigation from Language

Ref: [3]

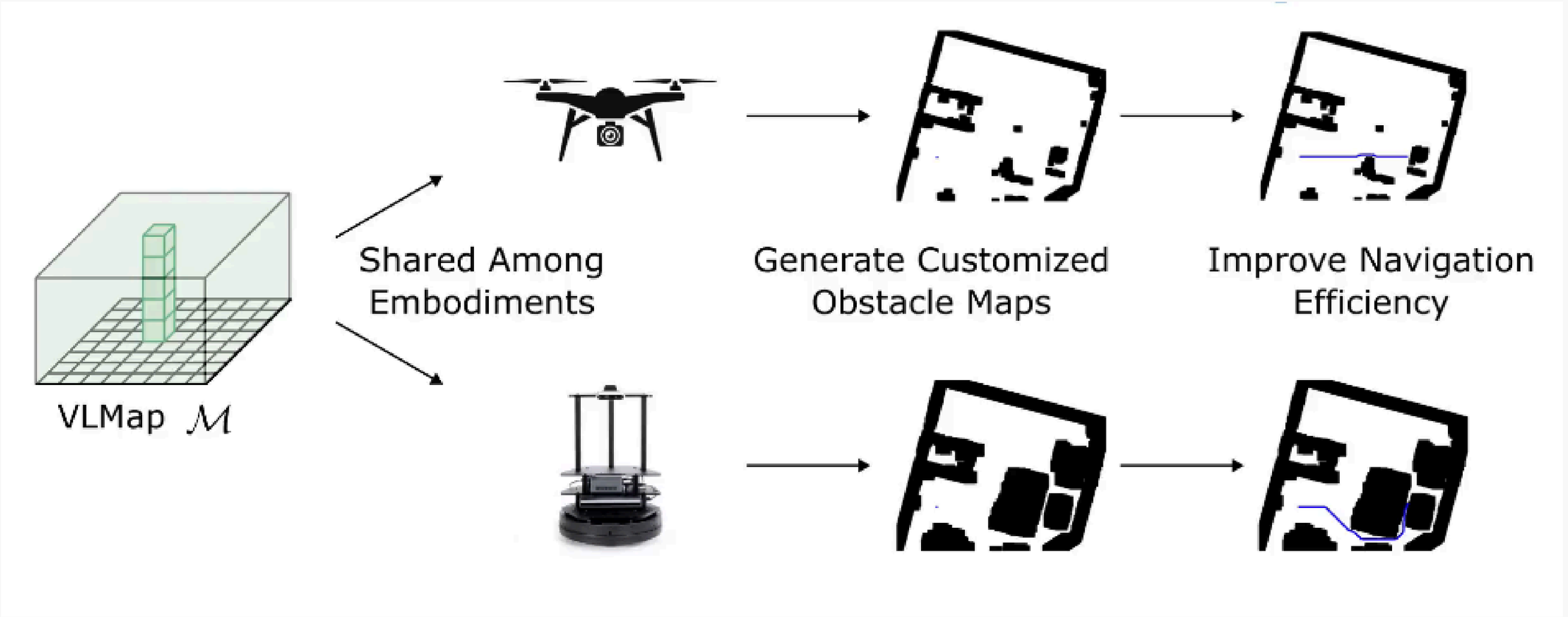
VLMaps: Experiment

Long-Horizon Spatial Goal Navigation from Language

Ref: [3]

VLMaps: Experiment

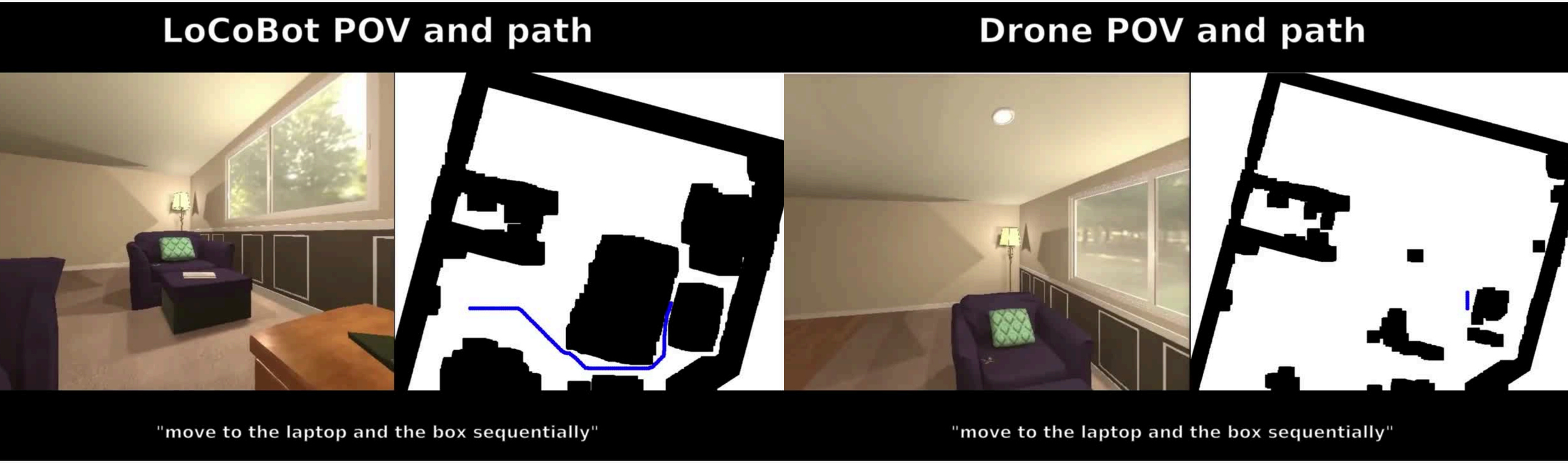
Multi-Embodiment Navigation



Ref: [3]

VLMaps: Experiment

Multi-Embodiment Navigation



Ref: [3]

VLMaps: Advantages and Disadvantages

The Advantages:

- Long-Horizon Navigation
- Cross-Robot Map Sharing
- High Spatial Precision
- Flexible Goal Setting

The Disadvantages:

- Sensitive to 3D reconstruction noise and odometry drift
- Cannot resolve object ambiguities

Comparison of LM-Nav and VLMaps

Shared Strengths & Core Capabilities

- Utilize VLMs to bridge visual and natural language processing
- Designed for adaptable, zero-shot navigation capabilities

Key Differences in Methodology

- LM-Nav: graph-based approach to break down and execute complex tasks
- VLMaps: visual-language maps for open-ended spatial queries

Comparison of LM-Nav and VLMaps

Complementary Functions

LM-Nav

- Long-range planning
- High-level commands converting

VLMaps

- Precise spatial localization
- Spatial understanding
- Flexible goal setting

Future Directions for Robot Navigation Using VLMs

- Employ more advanced and robust visual language models
- Extend to dynamic environments
- Enhance methods for incorporating external knowledge into robot decision-making

NEXT

Thank you

References

1. Wu, Wansen, et al. "Vision-language navigation: a survey and taxonomy." Neural Computing and Applications 36.7 (2024): 3291-3316.
2. Shah, Dhruv, Błażej Osiński, and Sergey Levine. "Lm-nav: Robotic navigation with large pre-trained models of language, vision, and action." Conference on robot learning. PMLR, 2023.
3. Huang, Chenguang, et al. "Visual language maps for robot navigation." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.