

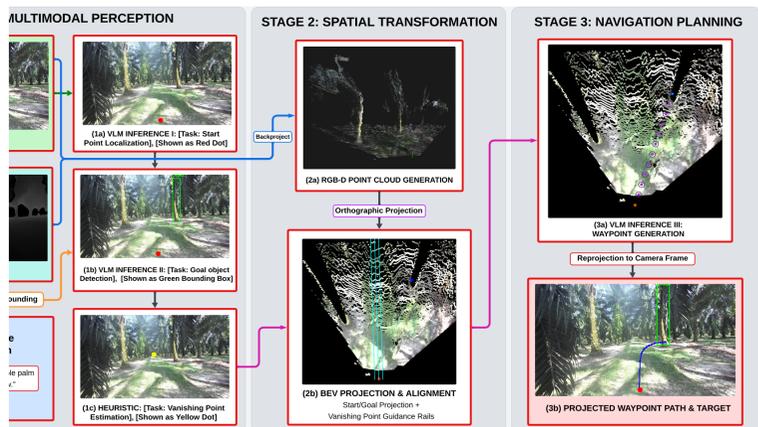


Motivation

requires extensive site-specific tuning. Traditional SLAM and perception degrade in unstructured vegetation and lack interfaces for conditioned path planner operating directly in 2D BEV space, using VLM Reasoning to choose MPPI Trajectories that respect NL preferences.

NL commands via open-vocabulary grounding to synthesize constraint-aware waypoints. A novel semi-supervised labeling pipeline farm environments.

Method: NL Command Generation



ocabulary objects; SAM3 produces instance masks and centroids; DAM generates mask-level NL descriptions. (2) Using a small human-labeled seed, Qwen3-MoE generates a VLM judge. (3) For each command, Qwen3 selects a target ROI, MPPI proposes 7 feasible BEV trajectories, and Qwen3 selects the best one to export waypoint supervision.

mmands and 2D BEV waypoints.

TS.
 ds mapping to objects (generated via QWEN-30B-A3B).
 nd NL Templates for CSL SC.
 ummaries, RGBD data, and BEV Image overlaid with start/goal points and camera viewpoint.

arning Setup:
 the "Egocentric Human View" based on NL templates for in-context learning.
 ported for datasets: Citrus, Palm, Solar, Corn, Soybean, and TreeScope.

Method: Semi-supervised dataset generation

Step 3: Semi-Supervised Dataset Generation Pipeline:

- Inference 1 (Target Grounding):** QWEN-MoE-30B-A3B selects a *single Target ROI* based on the input NL command and available data.
- Inference 2 (Trajectory Selection):** QWEN-MoE-30B-A3B dynamically selects the best path from *7 candidate MPPI trajectories*.
- Export:** Waypoints generated to navigate to target ROI.

Related Work

Baselines

- Zero shot VLMs:** Zero-shot VLM waypoint generation directly from raw images.
- GeNIE [4]:** Generalizable VLM navigation. **Limitation:** Struggles to strictly adhere to fine-grained geometric and structural constraints. **WaypointGen** is better as it enforces constraints by explicitly verifying and selecting feasible MPPI trajectories in 2D BEV space.
- 3D Reconstruction:** Dense point cloud fusion with RRT* planning.

Datasets

- Palm plantation:** 500 m rows with 30+ trees and water channels.
- Solar farm:** Panel arrays, maintenance paths, and vegetation.
- Greenhouse (High tunnels):** Variable row spacing, Dynamic Lighting, Dynamically moving obstacles with Rigid Non Traversable leaves, uneven terrain.
- Corn and Soybean fields:** SLAM benchmark with stereo images, IMU, odometry, and GPS across growth stages [2].
- CitrusFarm:** Multimodal dataset featuring 9 sensing modalities across 7.5 km of trajectories in citrus tree farms [3].
- TreeScope:** LiDAR mapping dataset from forests and orchards for evaluating 3D mapping in natural vegetation [1].

Conclusion and Future Work

- Conclusion:** WaypointGen is a language-conditioned path planner that generates constraint-aware waypoints in 2D BEV space.
- Future Work:** Support Multiple Target ROIs, MPPI Trajectory selection for subpaths in long path navigation, and predict temporal dynamics.

References

- [1] Derek Cheng et al. *TreeScope: An Agricultural Robotics Dataset for LiDAR-Based Mapping of Trees in Forests and Orchards*. 2023. arXiv: 2310.02162 [cs/ML].
- [2] Jose Cuaran et al. "Under-canopy dataset for advancing simultaneous localization and mapping in agricultural robotics". In: *The International Journal of Robotics Research* 10.1177/02783649231215372.
- [3] Hanzhe Teng et al. "Multimodal Dataset for Localization, Mapping and Crop Monitoring in Citrus Tree Farms". In: *International Symposium on Visual Computing* 2023.
- [4] Jiaming Wang et al. "GeNIE: A Generalizable Navigation System for In-the-Wild Environments". In: *IEEE Robotics and Automation Letters* 10.12 (2025), 1-8.

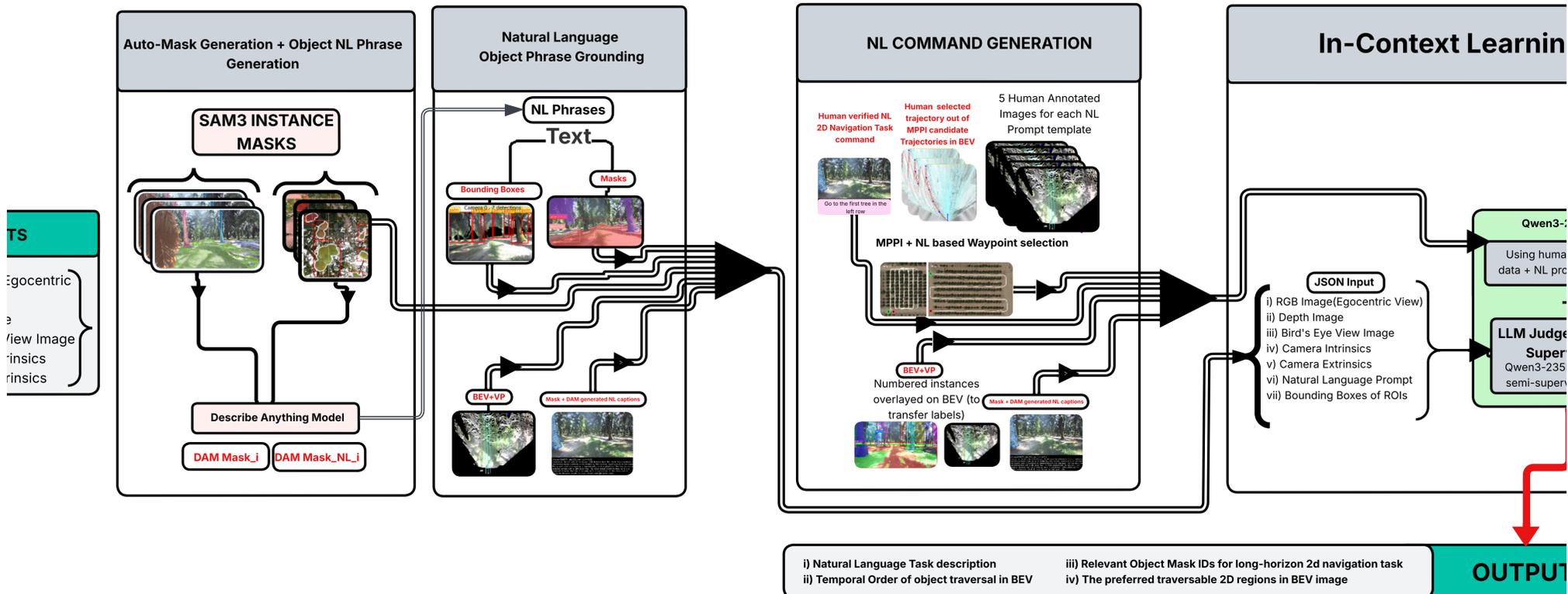


Fig. 2: Semi-supervised dataset generation pipeline overview. (i) SAM3 and DAM generate Masks and NL descriptions. (ii) QWEN3 is used to generate relevant 2D NL Navigation task descriptions. (iii) Human annotations are used for in-context learning for weak supervision using QWEN3: 235B VLM for image space Target ROI selection and trajectory selection.