LAMP: Implicit Language Map for Robot Navigation
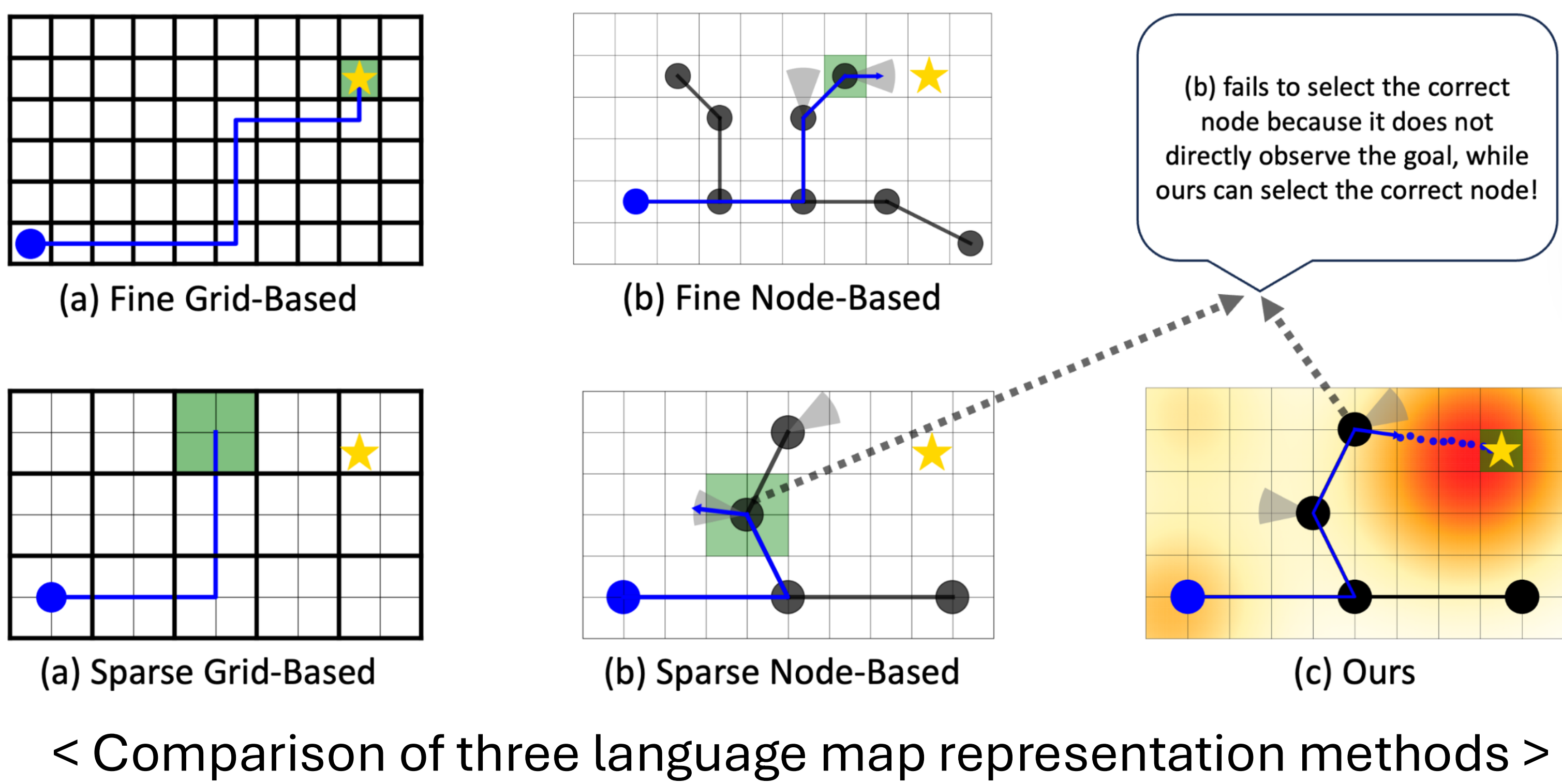
| Sibaek Lee[1,2] | Hyeonwoo Yu[1] | | Giseop Kim[3] | Sunwook Choi[2] |

Sungkyunkwan University[1], NAVER LABS[2], DIGIST[3]

1. Introduction & Problem Definition

(a) Fine Grid-Based
(b) Fine Node-Based
(b) fails to select the correct node because it does not directly observe the goal, while ours can select the correct node!
(a) Sparse Grid-Based
(b) Sparse Node-Based
(c) Ours

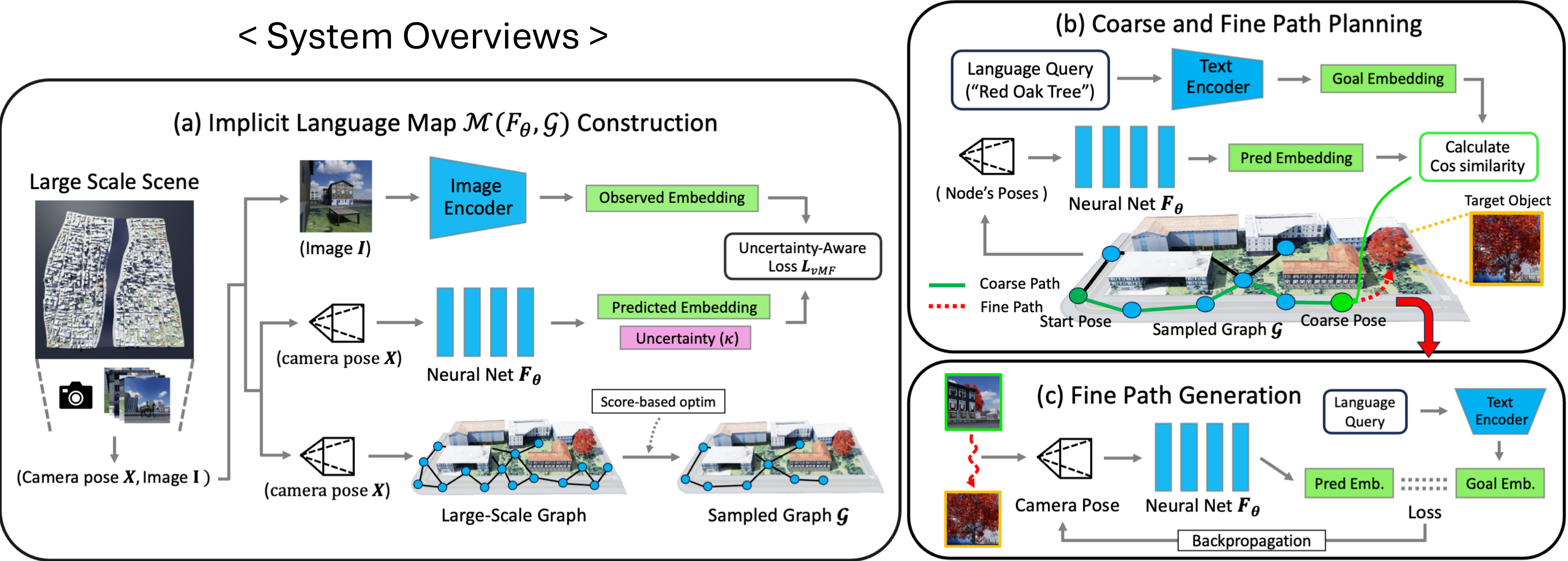< Comparison of three language map representation methods >

Current methods rely on explicit structures such as grids or topological nodes.

• **Fine-level (Top):** Captures the goal, but explicitly storing language information is infeasible due to memory constraints at large scales.

• **Coarse-level (Bottom):** Reduces memory usage but leads to navigation failure.

So, we propose LAMP, an Implicit Language Map that breaks this trade-off, allowing for fine-grained navigation with efficient memory usage.

2. Method

< System Overviews >

(a) Implicit Language Map $\mathcal{M}(F_\theta, \mathcal{G})$ Construction

Large Scale Scene

(Camera pose X, Image I)

(Image I) → Image Encoder → Observed Embedding

(camera pose X) → Neural Net F_θ → Predicted Embedding / Uncertainty (κ)

Uncertainty-Aware Loss L_{vMF}

(camera pose X) → Large-Scale Graph → Score-based optim → Sampled Graph $\mathcal{G}$

(b) Coarse and Fine Path Planning

Language Query ("Red Oak Tree") → Text Encoder → Goal Embedding

(Node's Poses) → Neural Net F_θ → Pred Embedding → Calculate Cos similarity

Target Object

Coarse Path — Fine Path ⋯⋯ Start Pose   Sampled Graph $\mathcal{G}$ Coarse Pose

(c) Fine Path Generation

Camera Pose → Neural Net F_θ → Pred Emb. ⋯ Goal Emb.

Language Query → Text Encoder → Goal Emb.

Loss / Backpropagation

(a): Implicit Map Construction
- Learns the implicit function $z = F_\theta(x)$ mapping a pose x to a language embedding z.
- Constructs a sampled graph $\mathcal{G}$ via score-based optimization for memory efficiency.

(b): Coarse Path Planning
- Encodes the user's language query into a semantic goal embedding.
- Performs A* search on the sampled graph $\mathcal{G}$ to find the closest semantic node.

(c): Fine Path Generation
- Optimizes the continuous pose using $z = F_\theta(x)$ to maximize cosine similarity.
- Locally refines the path to observe the target object precisely.

3. Experiments Results

Method	Memory (GB)	SR (Easy)	SPL (Easy)	Gdist (Easy)	SR (Hard)	SPL (Hard)	Gdist (Hard)
Grid-Based	0.005	0.08	0.07	15.50	0.0	0.0	-
Node-Based	0.057	0.41	0.36	12.69	0.21	0.17	4.14
Ours	0.057	0.67	0.62	6.36	0.42	0.38	1.74

Method	Memory (GB)	SR (Easy)	SPL (Easy)	Gdist (Easy)	SR (Hard)	SPL (Hard)	Gdist (Hard)
Grid-Based	56.34	0.83	0.80	2.44	0.36	0.33	1.27
Node-Based	3.962	0.67	0.61	7.51	0.47	0.41	2.38
Ours	0.057	0.67	0.67	6.36	0.42	0.38	1.74

Goal : Drinks

Goal : Sink

Papaer

Website