

A SURVEY

Semantic Mapping in Indoor Embodied AI

Advances, Challenges, and Future Directions



Sonia Raychaudhuri



Angel X. Chang

Motivation



"Go to the red and blue striped zebra in the nursery"

01 ·
LANGUAGE

abc

*Understand the
request*

02 ·
PERCEPTION 

See & recognize

03 ·
MEMORY 

*Where have I
seen it before?*

04 ·
NAVIGATION 

Plan & act



"Go to the red and blue striped zebra in the nursery"

01 ·

How to structure
the map?

*Understand the
request*

02 ·

What to store in
the map?

See & recognize

03 ·

MEMORY 

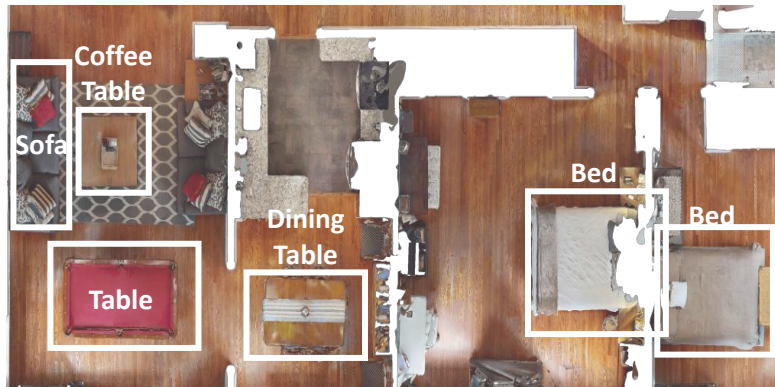
*Where have I
seen it before?*

04 ·

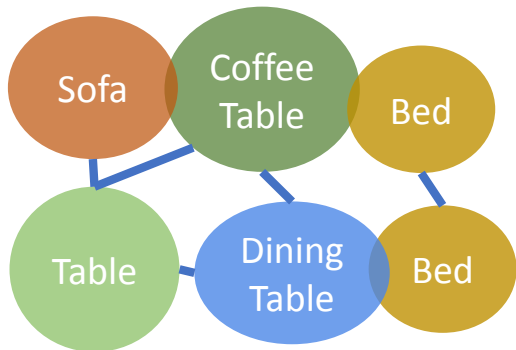
NAVIGATION 

Plan & act

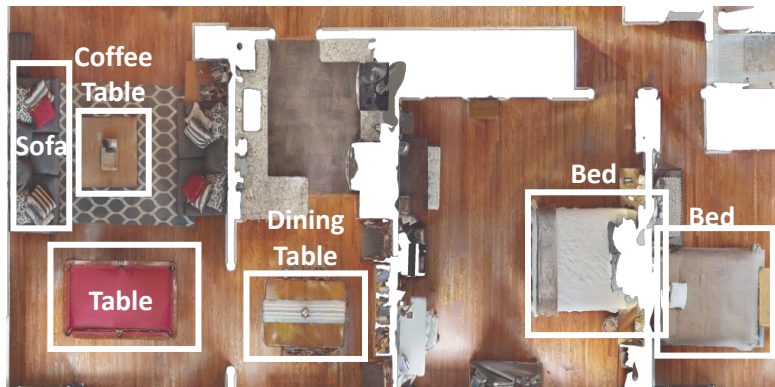
How to structure the map?



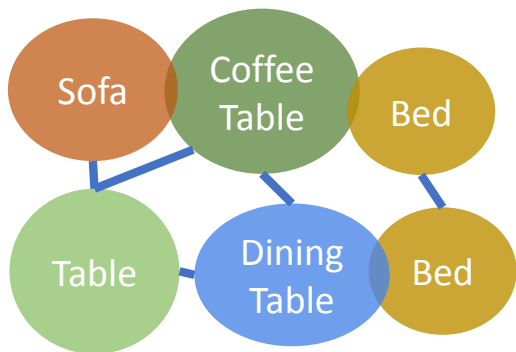
TOPOLOGICAL MAP



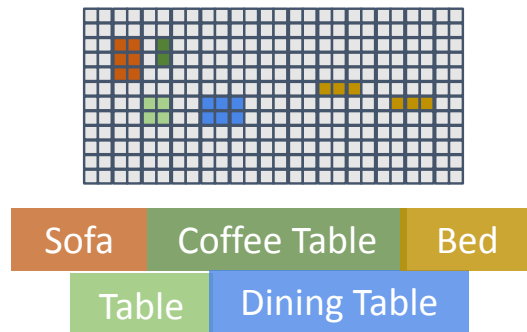
How to structure the map?



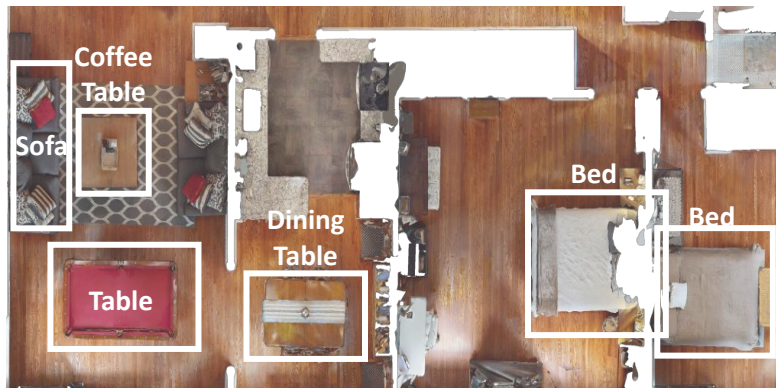
TOPOLOGICAL MAP



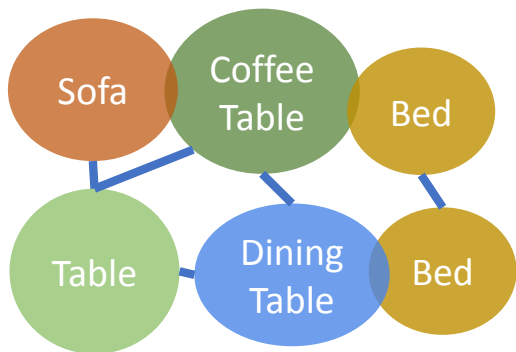
GRID MAP



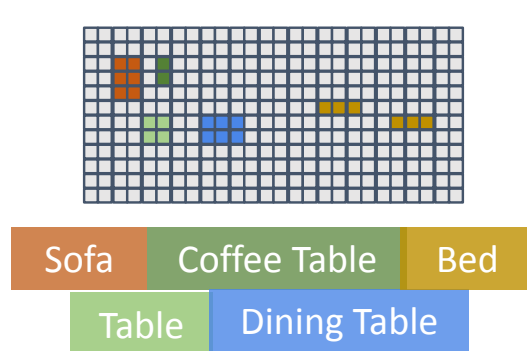
How to structure the map?



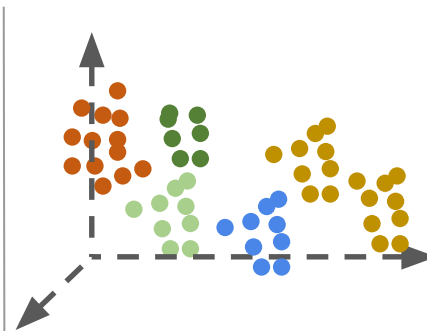
TOPOLOGICAL MAP



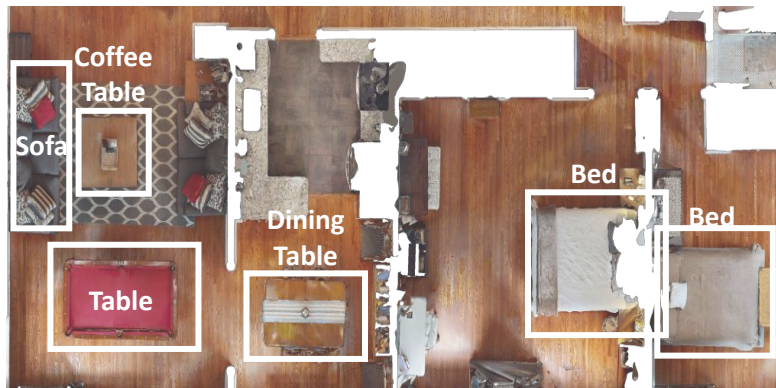
GRID MAP



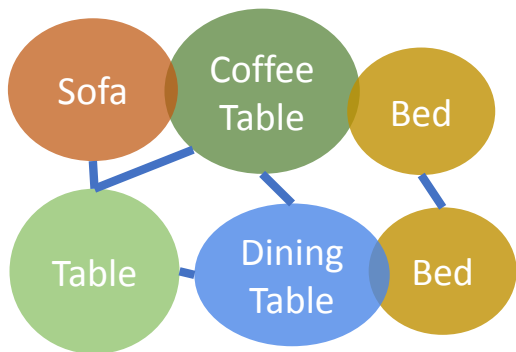
GEOMETRIC MAP



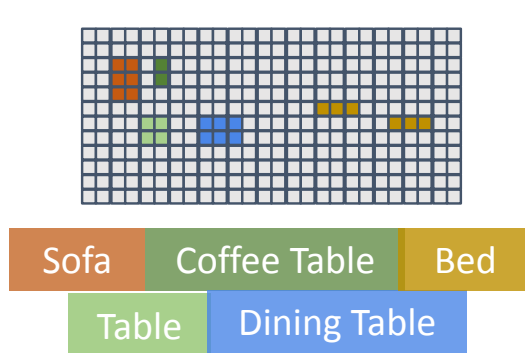
How to structure the map?



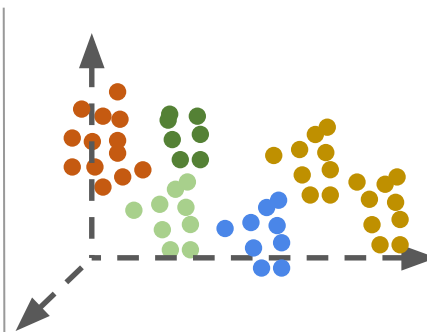
TOPOLOGICAL MAP



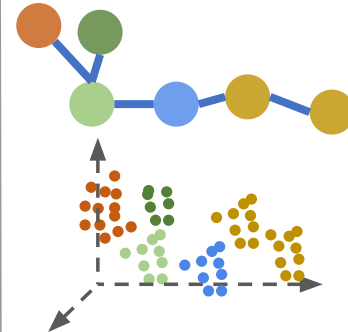
GRID MAP



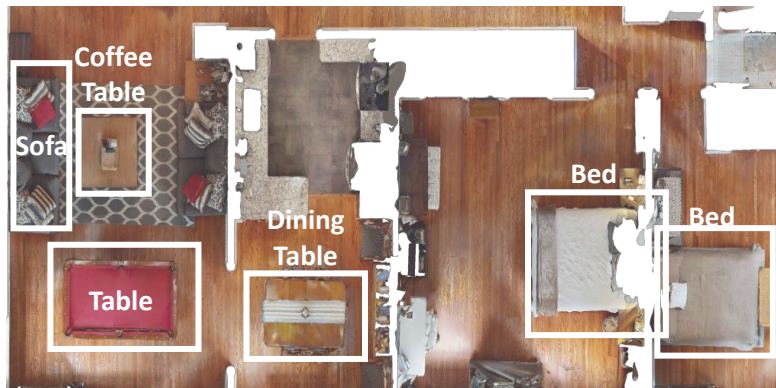
GEOMETRIC MAP



HYBRID



How to structure the map?

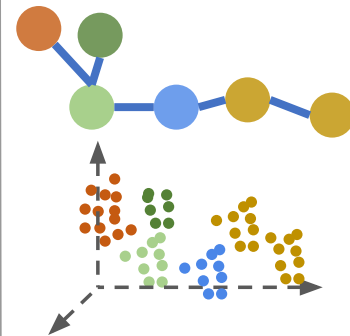
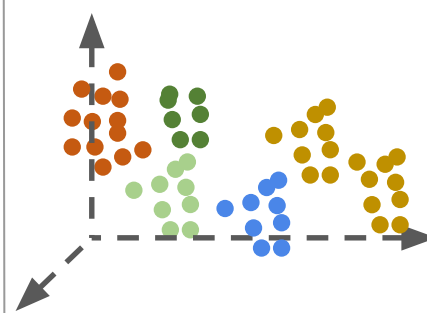
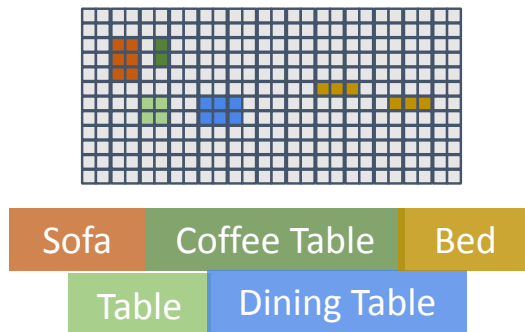
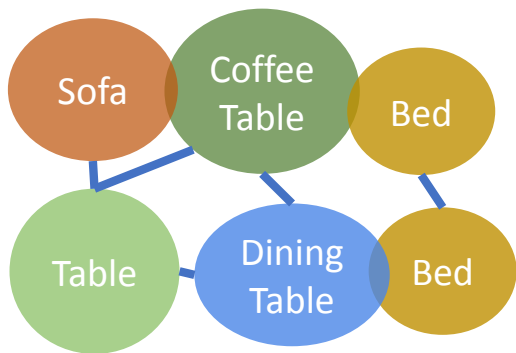


TOPOLOGICAL MAP

GRID MAP

GEOMETRIC MAP

HYBRID

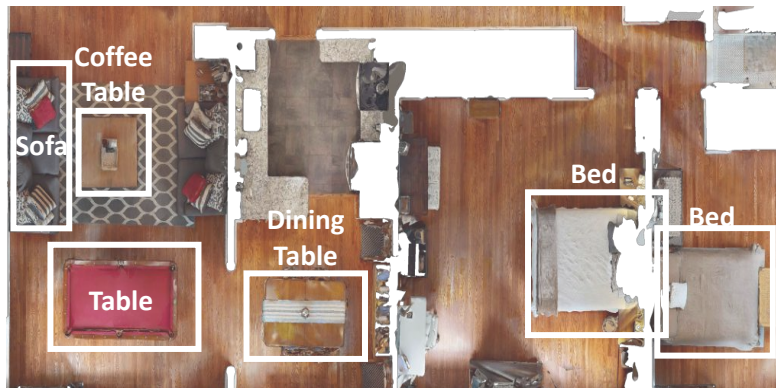


Sparse

Dense



What to store in the map?



EXPLICIT ENCODING

Occupancy Binary [0,1]

Explored Binary [0,1]

Image RGB pixel values



Objects RGB pixel values



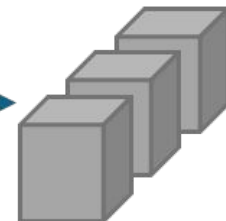
IMPLICIT ENCODING



Open-vocab
 Detection



LVLm

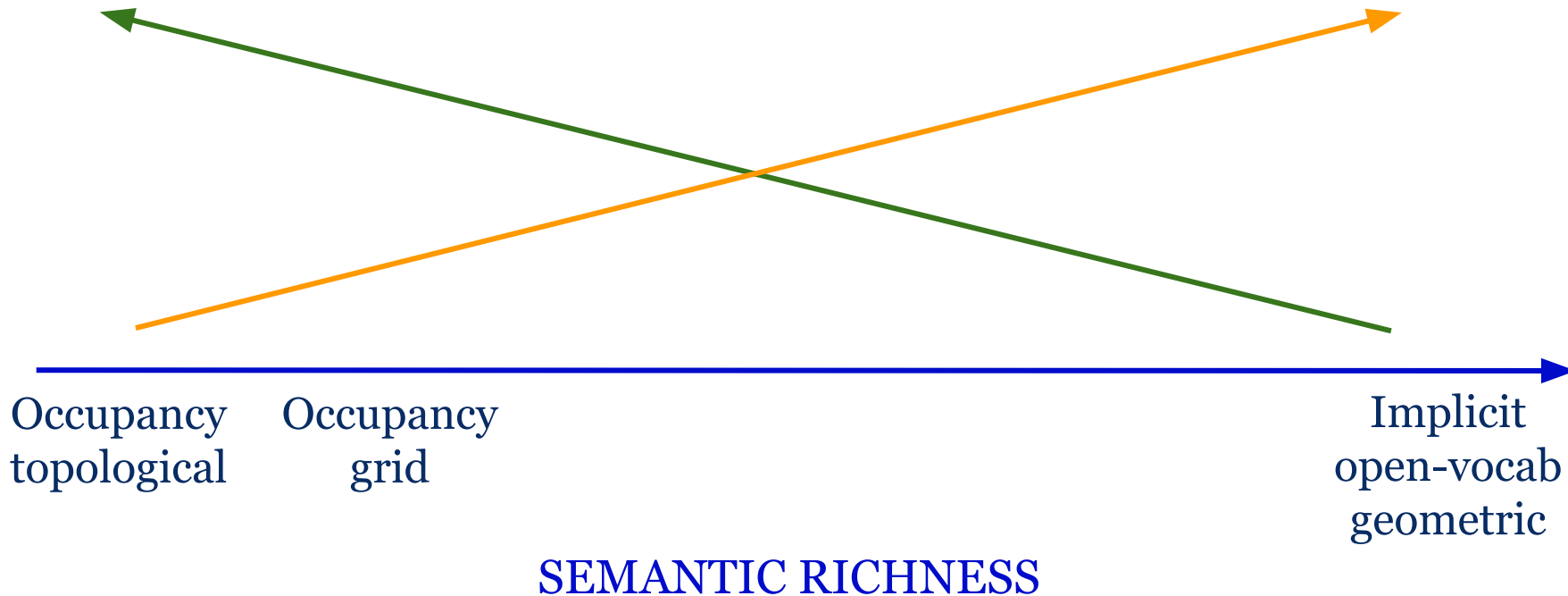


Open problems

The Richness–Efficiency Trade-off

PROCESSING SPEED

MEMORY COST

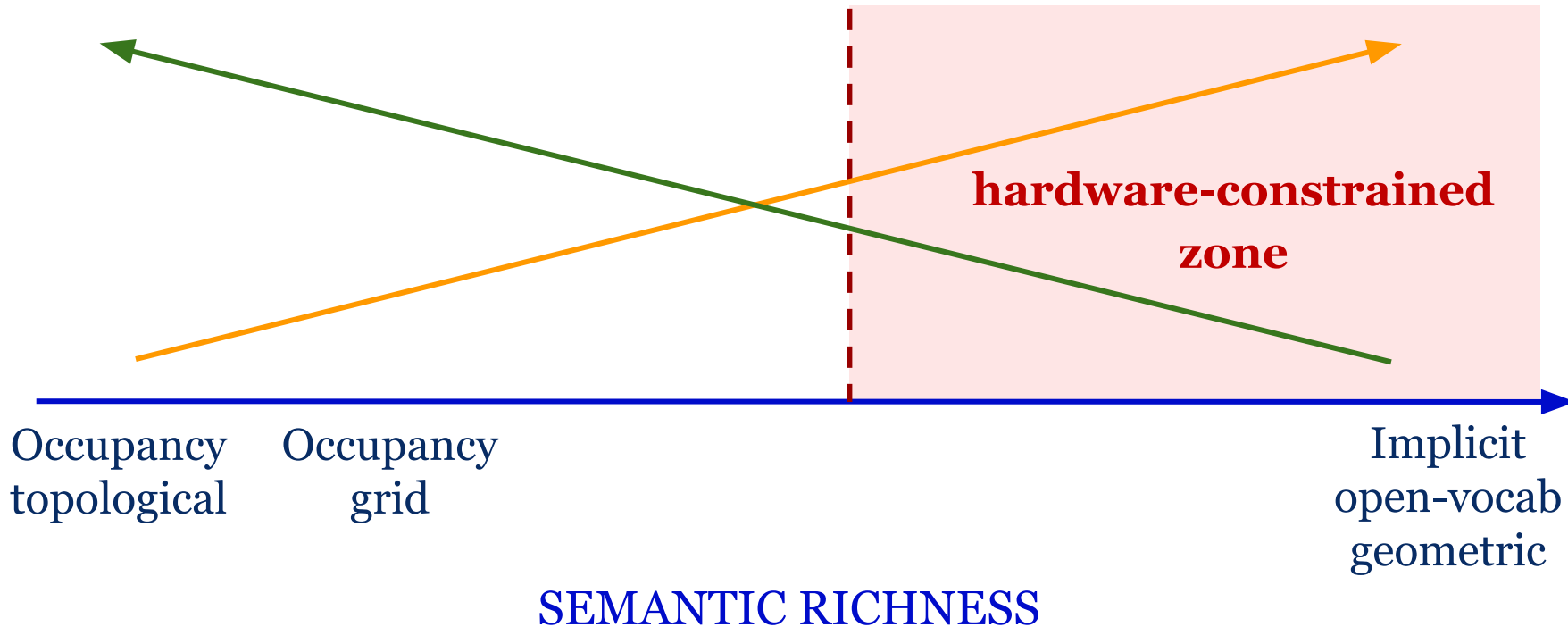


Open problems

The Richness–Efficiency Trade-off

PROCESSING SPEED

MEMORY COST

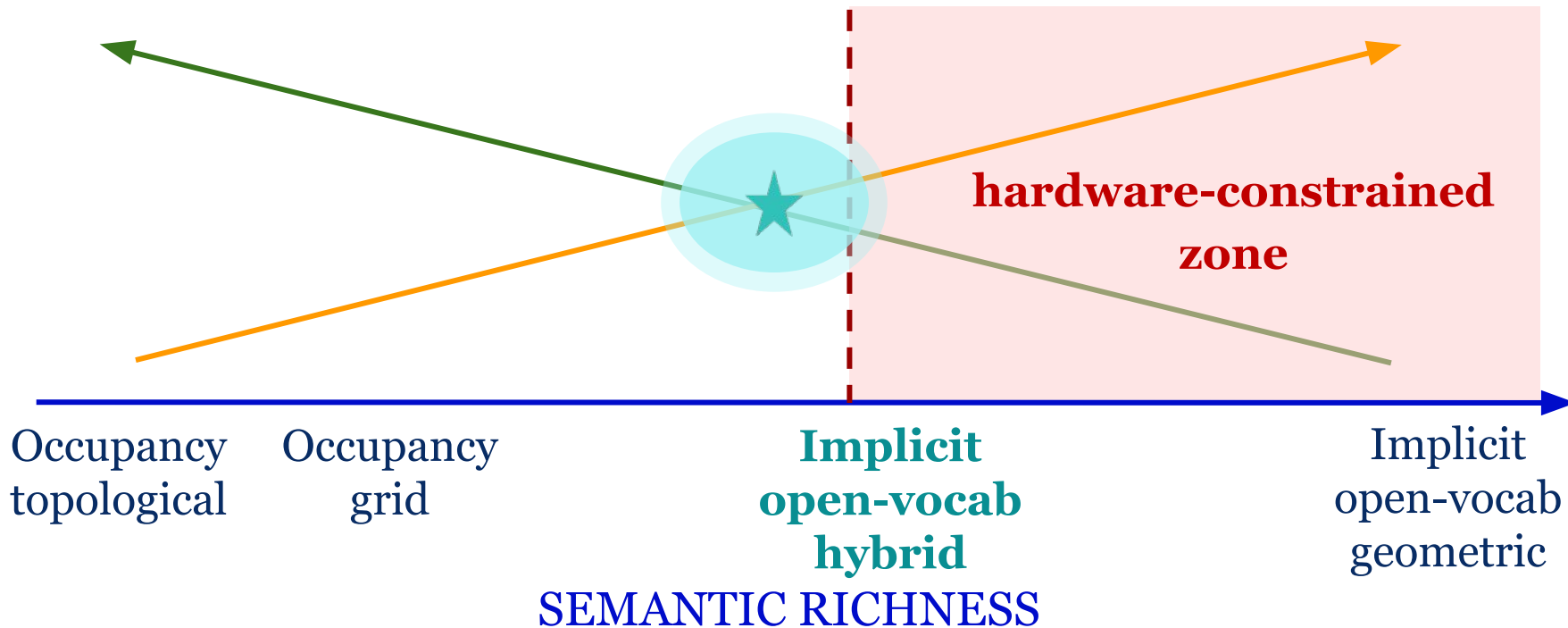


Open problems

The Richness–Efficiency Trade-off

PROCESSING SPEED

MEMORY COST



The Robustness–Generalization Gap

NOISE AND UNCERTAINTY

errors compound → semantic drift

uncertainty not modelled explicitly

LIFELONG LEARNING

adapt to real-world changes

MULTI-MODAL FUSION

vision, depth, audio, language, ...

STANDARDIZED EVALUATION

map quality evaluation lacking

