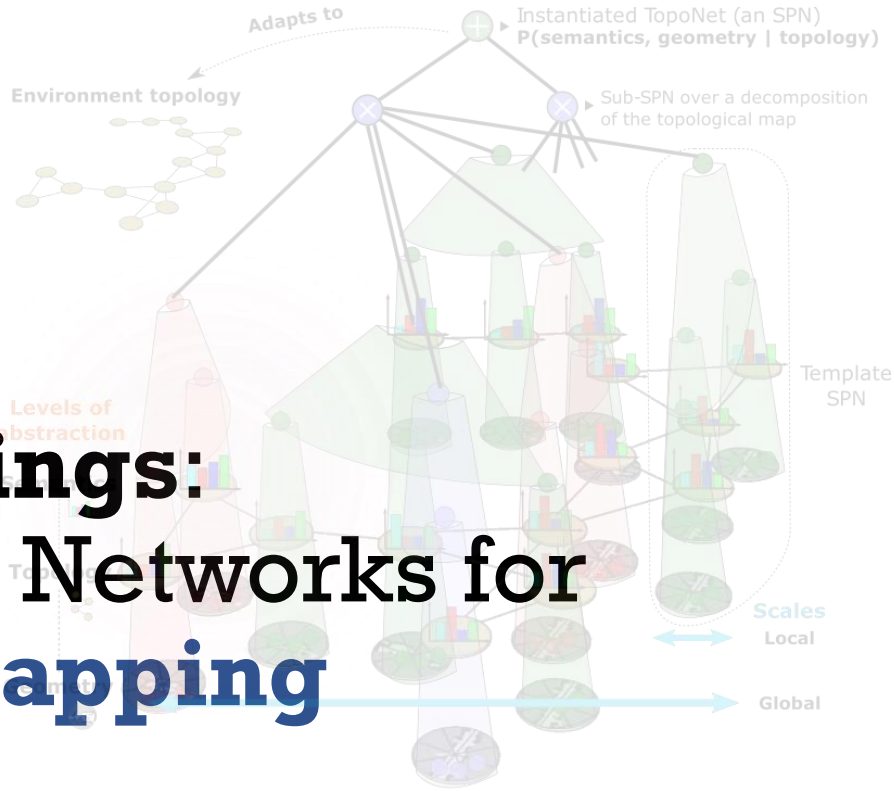


From Pixels to Buildings: End-to-end Probabilistic Deep Networks for Large-scale Semantic Mapping

Kaiyu Zheng[1*], Andrzej Pronobis[2,3]

[1]Brown University

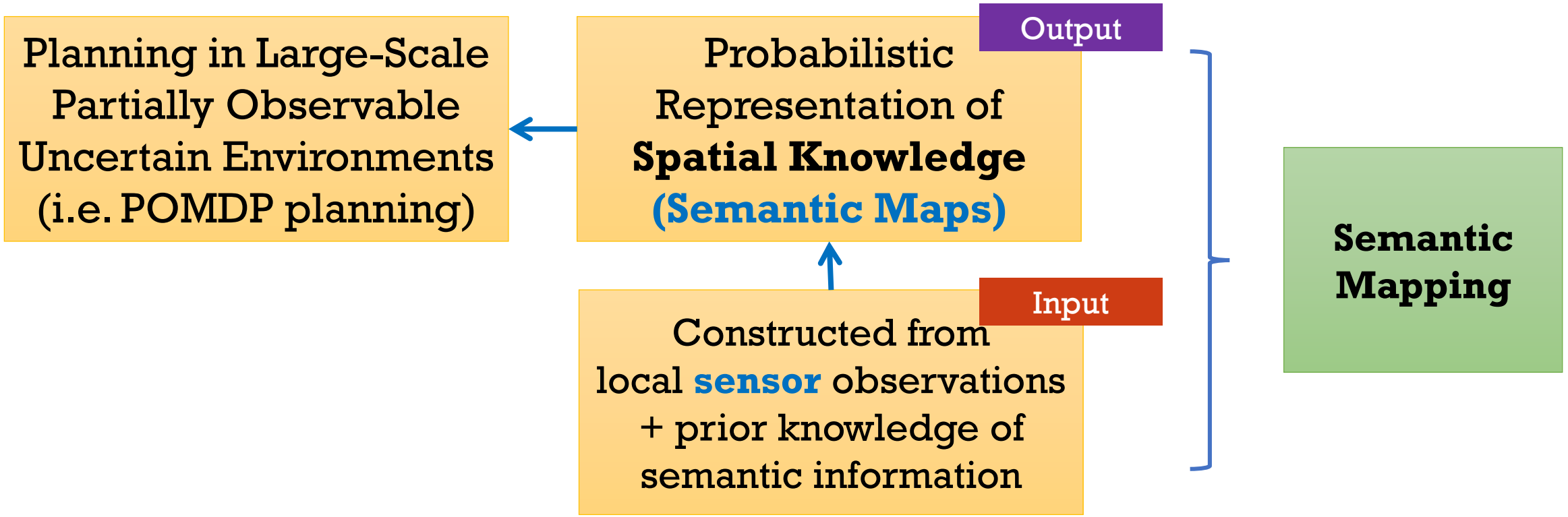
[2]University of Washington

[3]KTH Royal Institute of Technology

*work done while studying at [2]UW

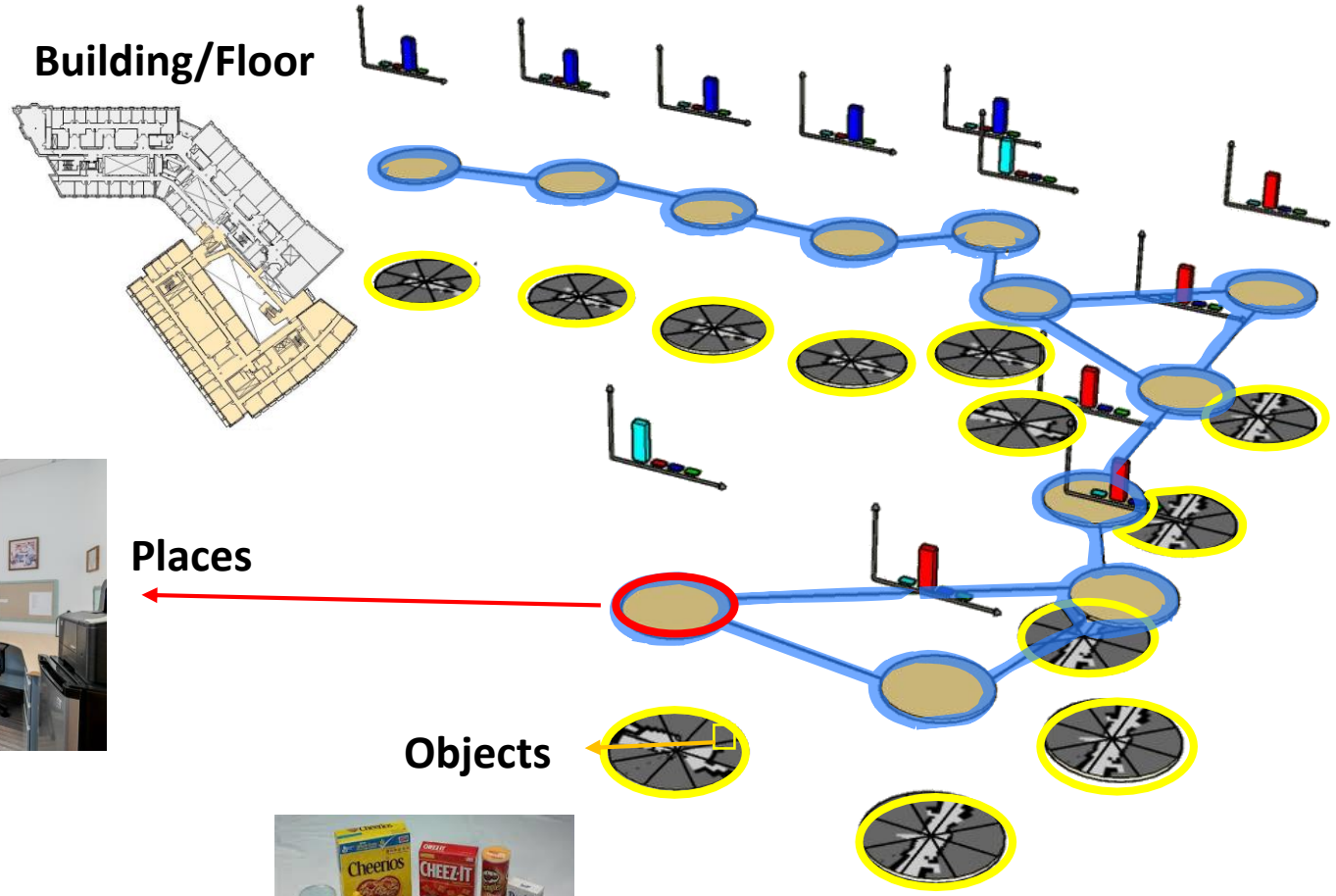
IROS 2019

Motivation: Semantic Mapping



Semantic Mapping: Challenges

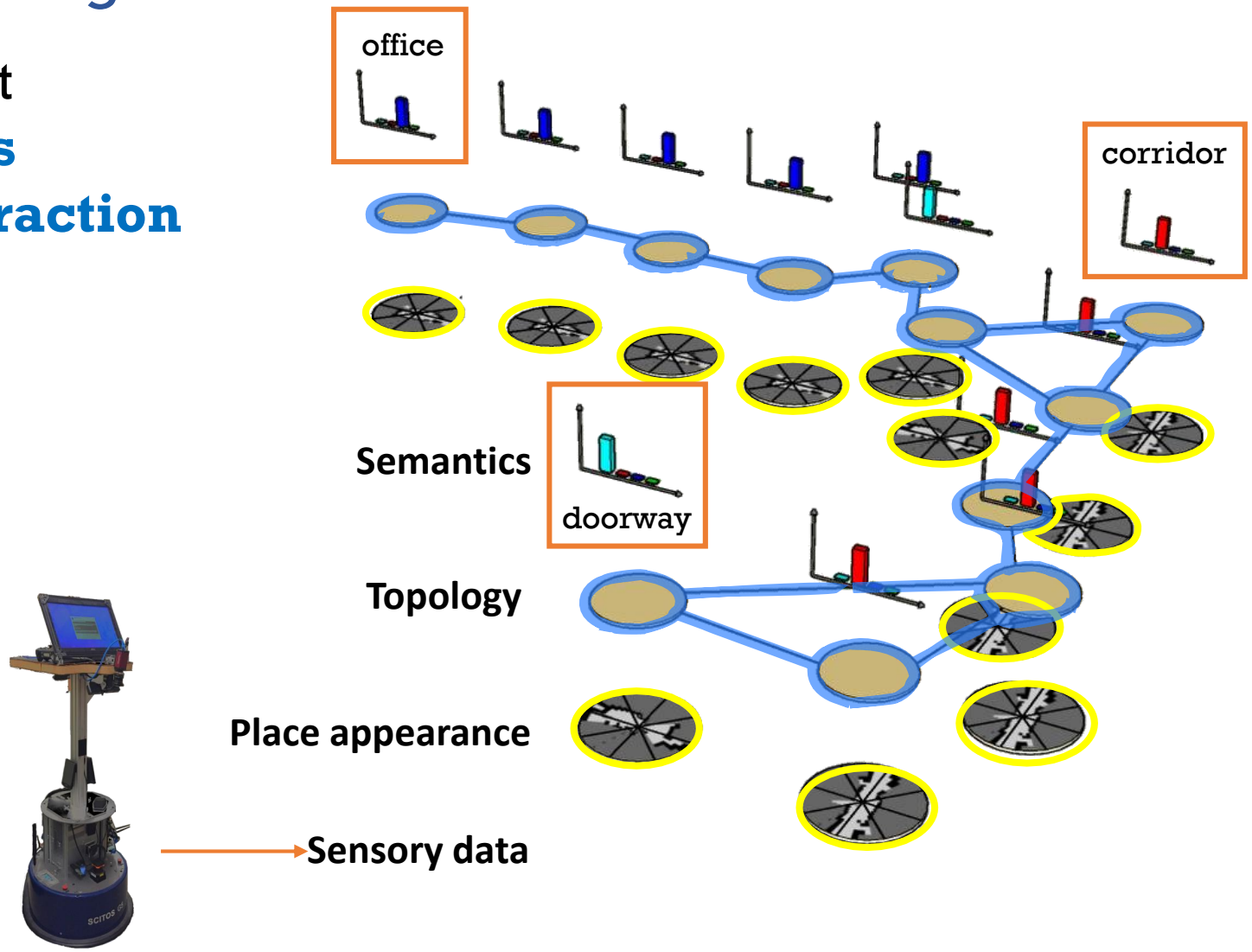
- Spatial knowledge exists at
  - Different **spatial scales**



YCB dataset [Calli et al, 2015]

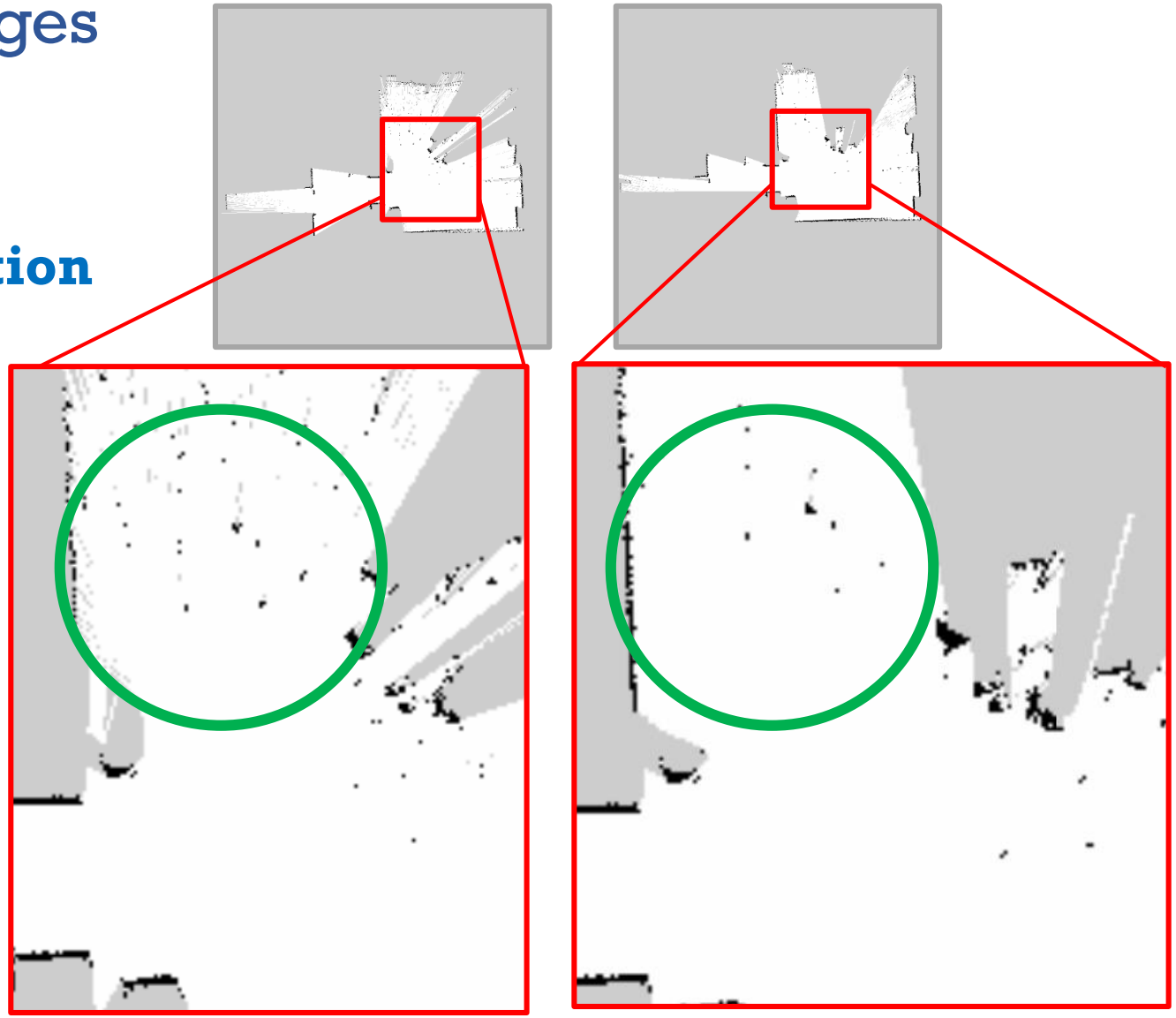
Semantic Mapping: Challenges

- Spatial knowledge exists at
  - Different **spatial scales**
  - Multiple **levels of abstraction**



Semantic Mapping: Challenges

- Spatial knowledge exists at
  - Different **spatial scales**
  - Multiple **levels of abstraction**
- Sensory observations are
  - **Local, Partial, Noisy**



Local, **Partial** laser-range observations with **Noisy** occupancy

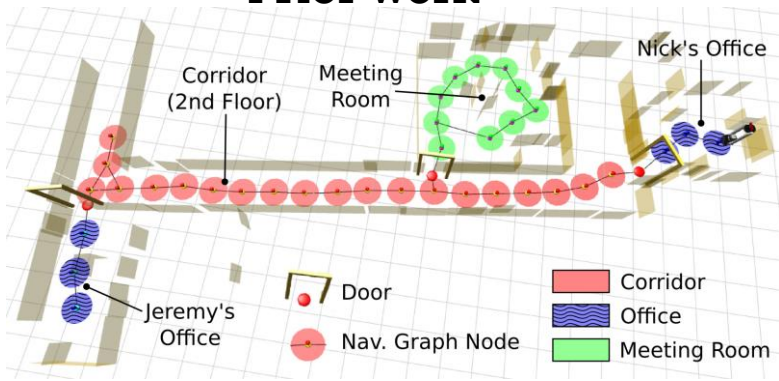
Credit of Images: Kousuke Ariga

Semantic Mapping: Challenges

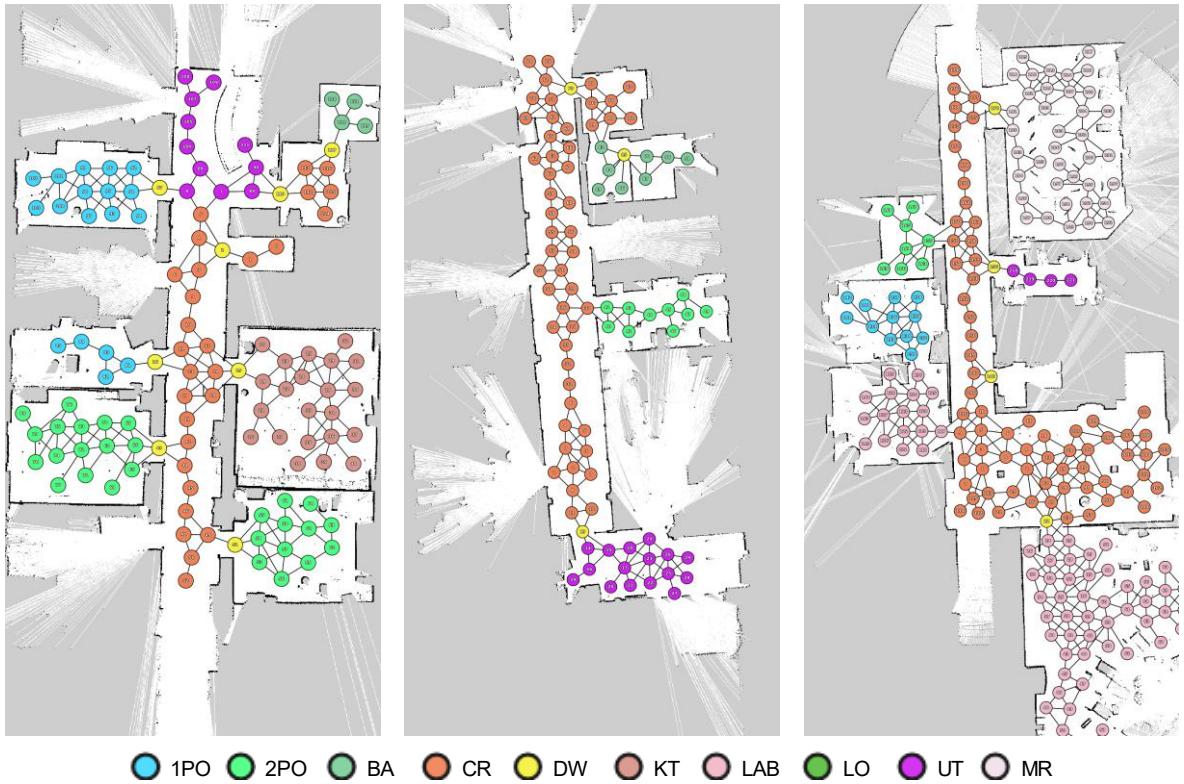
- Spatial knowledge exists at
  - Different **spatial scales**
  - Multiple **levels of abstraction**
- Sensory observations are
  - **Local, Partial, Noisy**
- Relationships in human world are
  - **Complex, Noisy**

Complex: Large number of connections

Prior work



Ours

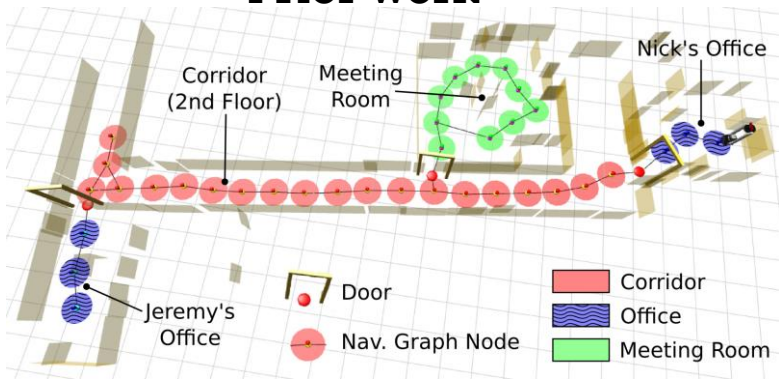


Semantic Mapping: Challenges

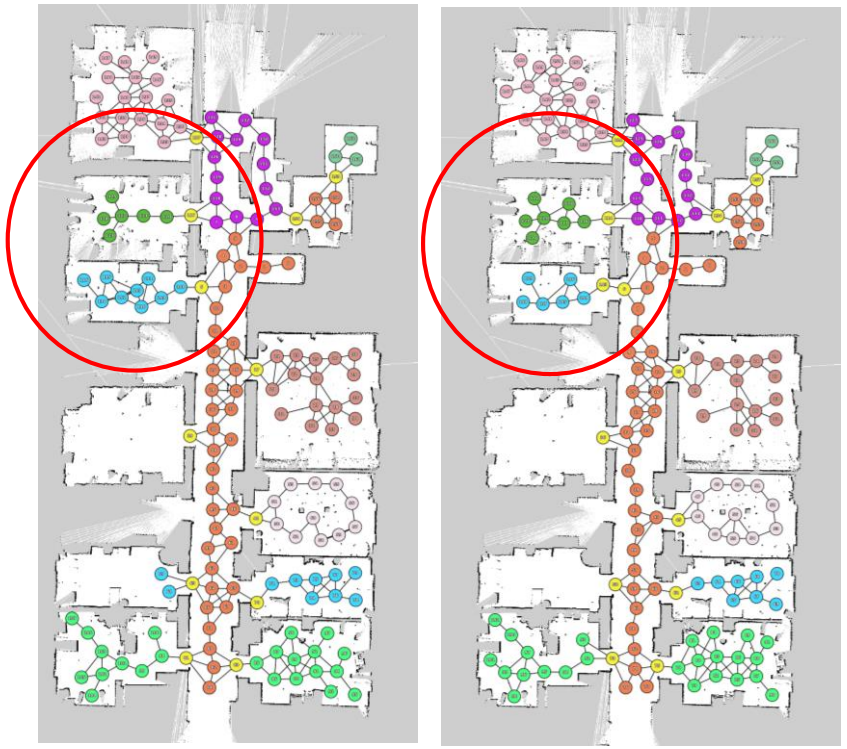
- Spatial knowledge exists at
  - Different **spatial scales**
  - Multiple **levels of abstraction**
- Sensory observations are
  - **Local, Partial, Noisy**
- Relationships in human world are
  - **Complex, Noisy**

Complex: Large number of connections
Noisy: Variability across floors/runs

Prior work



Ours



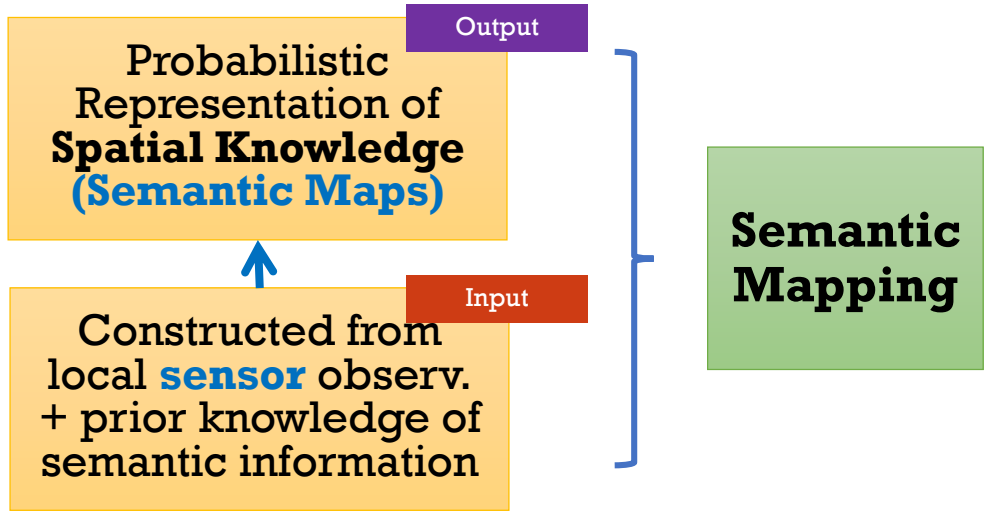
Topological graph constructed on the same floor in two runs.

Semantic Mapping: Challenges

- Spatial knowledge exists at
  - Different **spatial scales**
  - Multiple **levels of abstraction**
- Sensory observations are
  - **Local, Partial, Noisy**
- Relationships in human world are
  - **Complex, Noisy**
- Agent operates in **new** environments
  - Vary in scale and structure
  - Reason about unexplored places

Semantic Mapping: Desired Properties

A. Captures spatial **scales** and **abstractions**
B. Is **probabilistic**, captures uncertainty
C. Allows **real-time**, **efficient** inference
D. Leverages **relationships** between spatial concepts to
   - Improve robustness
   - resolve ambiguities
   - predict latent information (e.g. about unexplored places)



Existing Work: Robotics

Structured prediction in semantic mapping

- Assembly of independent components

  (e.g. Conditional Random Field + CNN)
  - Bottleneck in communication between components
  - Cannot be learned end-to-end
- Approximate inference for graphical models
  - Convergence issues
- Unable to reason about unexplored space

Our method doesn't require segmentation, or room/door detection

[Mozos et al. 2007] [Friedman et al. 2007]
[Pronobis et al. 2012] [Sünderhauf et al 2015]
[Brucker et al. 2018]

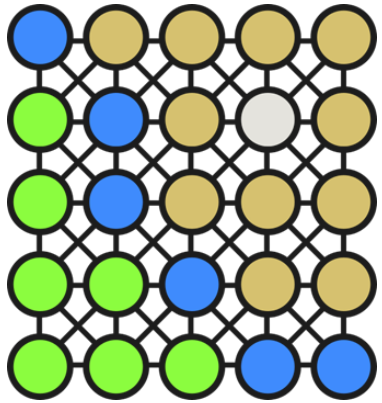
Existing Work: Computer Vision

Deep structured prediction approaches

(e.g. image generation, semantic segmentation)

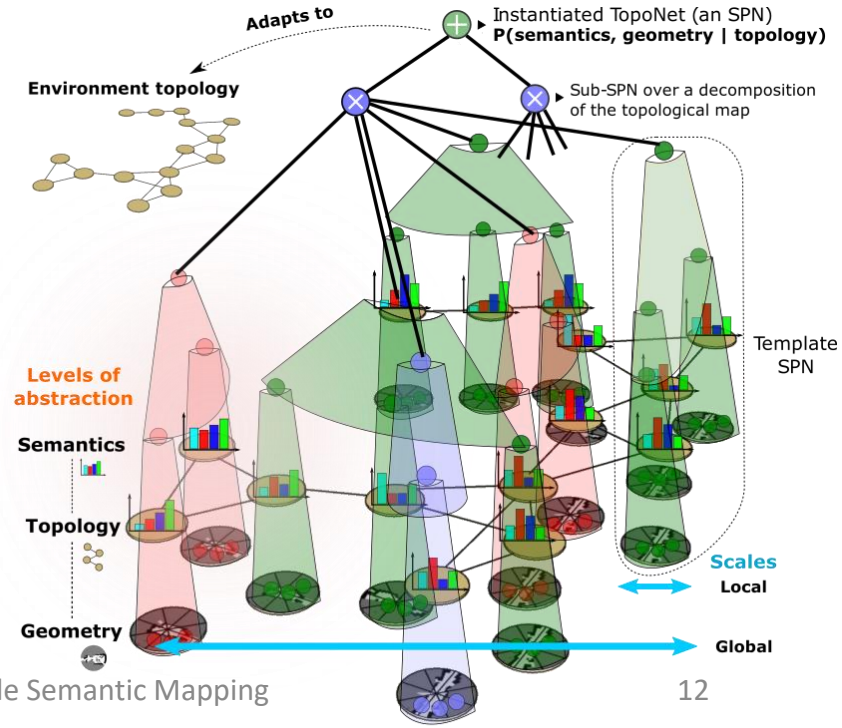
- Fixed number of variables
- Static global structure
- Some not probabilistic

[Wu et al.'16][Mahmood et al.'19]
[Chen et al.'18][Schwing & Urtasun,'15]
[Belanger & McCallum,'16]
[Shelhamer et. al.'16]



TopoNets: Overview

- Take-away I : End-to-end Unified Deep Probabilistic Spatial Model
- Take-away II: Tractable Exact Inference (real time)
- Take-away III: Template-based method
  - **Learn** template networks during **training**
  - **Instantiate** complete network while to **infer** semantics for any test environment
    - *Pr(semantics (**Y**), geometry (**X**) | topology)*

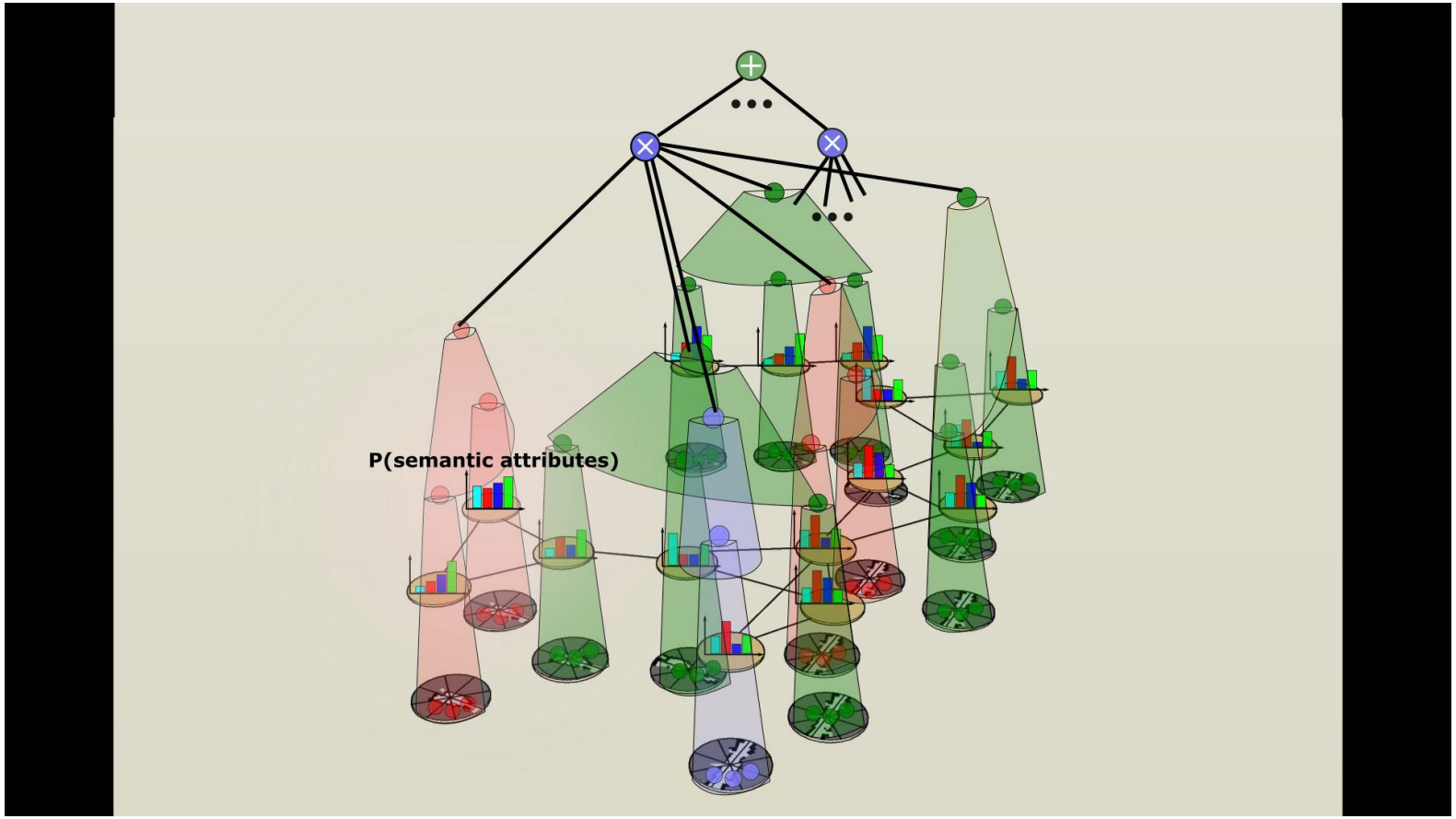


TopoNets

Take-away I : End-to-end Unified Deep Probabilistic Spatial Model

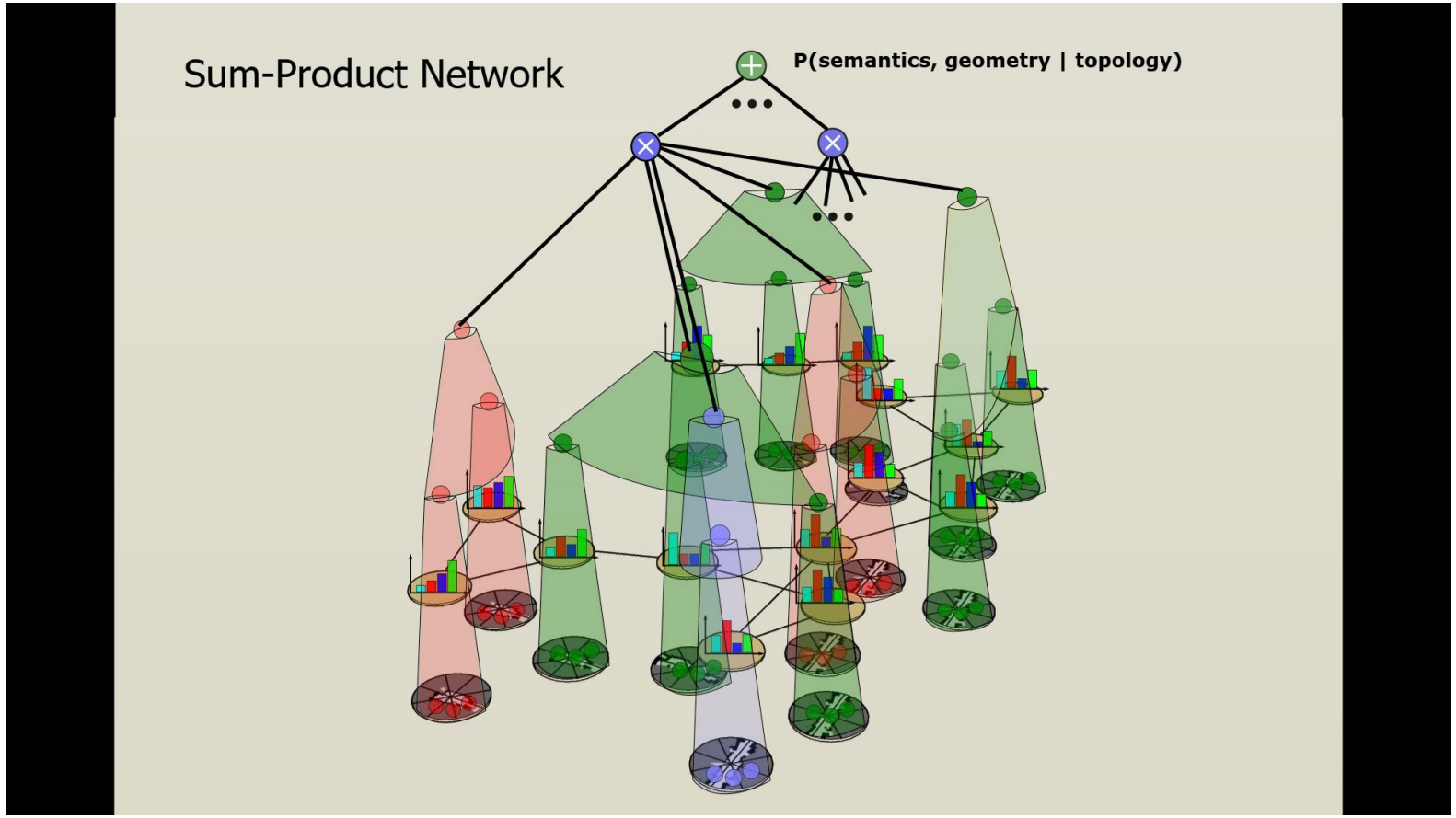
TopoNets

Take-away I : End-to-end Unified Deep Probabilistic Spatial Model



TopoNets

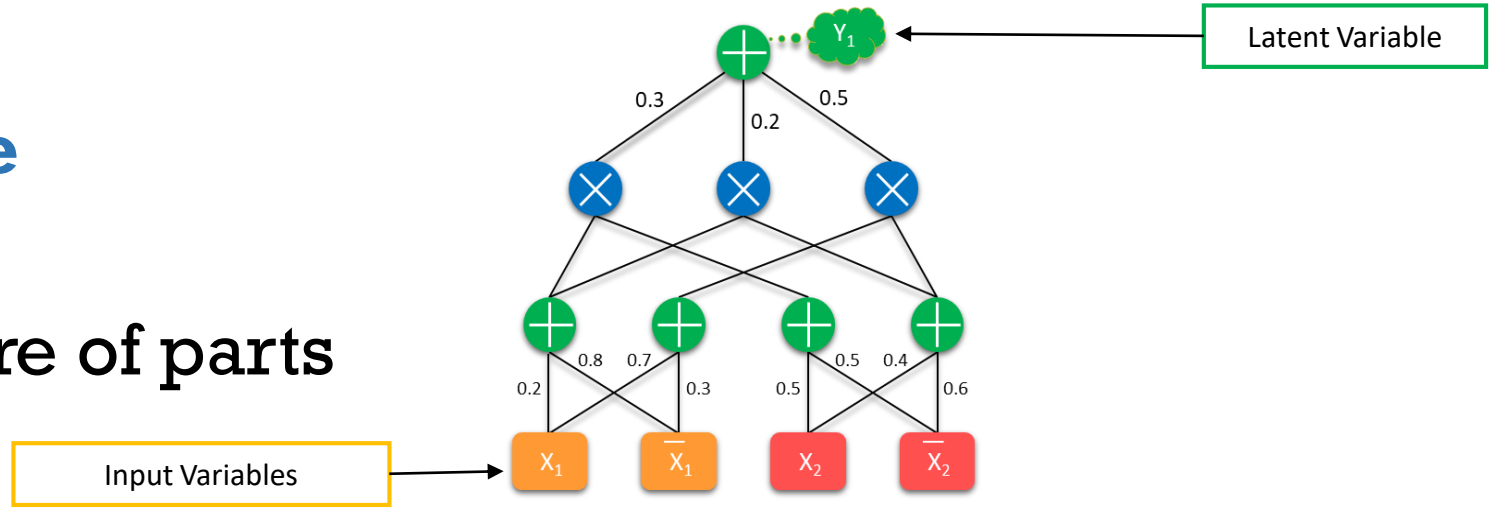
Take-away II: Tractable Exact Inference



TopoNets: Sum Product Networks

Sum-Product Networks, a recent deep architecture

- Solid theoretical foundations [Poon&Domingos'11] [Gens&Domingos'12] [Peharz et al.'17]
  - Learn conditional or joint distributions
  - **Tractable** partition function, exact inference
- Applied in a variety of problems (vision, NLP, robotics etc.)
- Viewed in 2 ways:
  - **Graphical model**
  - **Deep architecture**
- Structure semantics:
  - Hierarchical mixture of parts

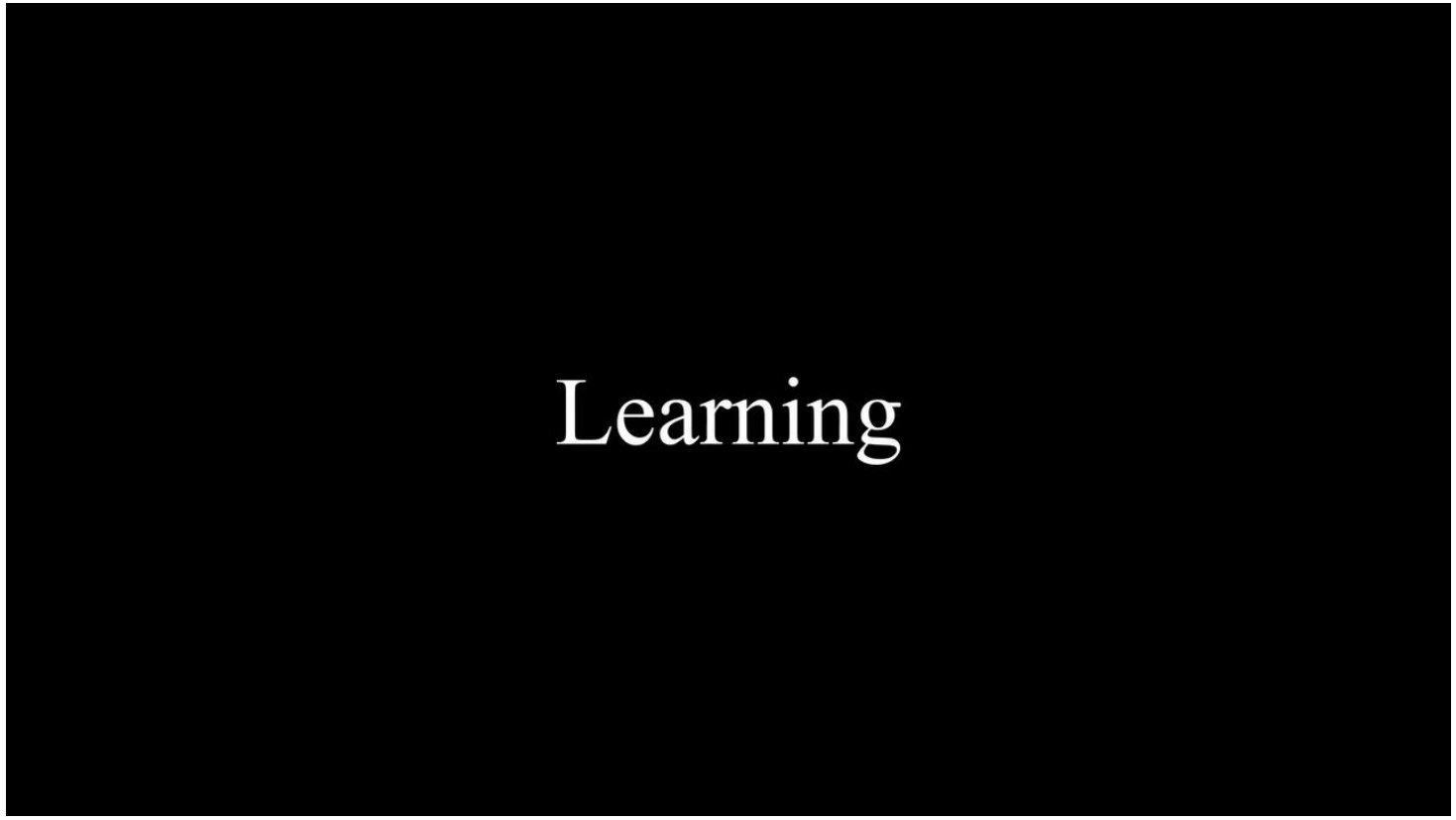


TopoNets

Refer to [van de Wolfshaar and Pronobis 2019] for convolutional representations of visual/spatial data.
url: https://arxiv.org/pdf/1902.06155.pdf

Take-away III: Template-based method

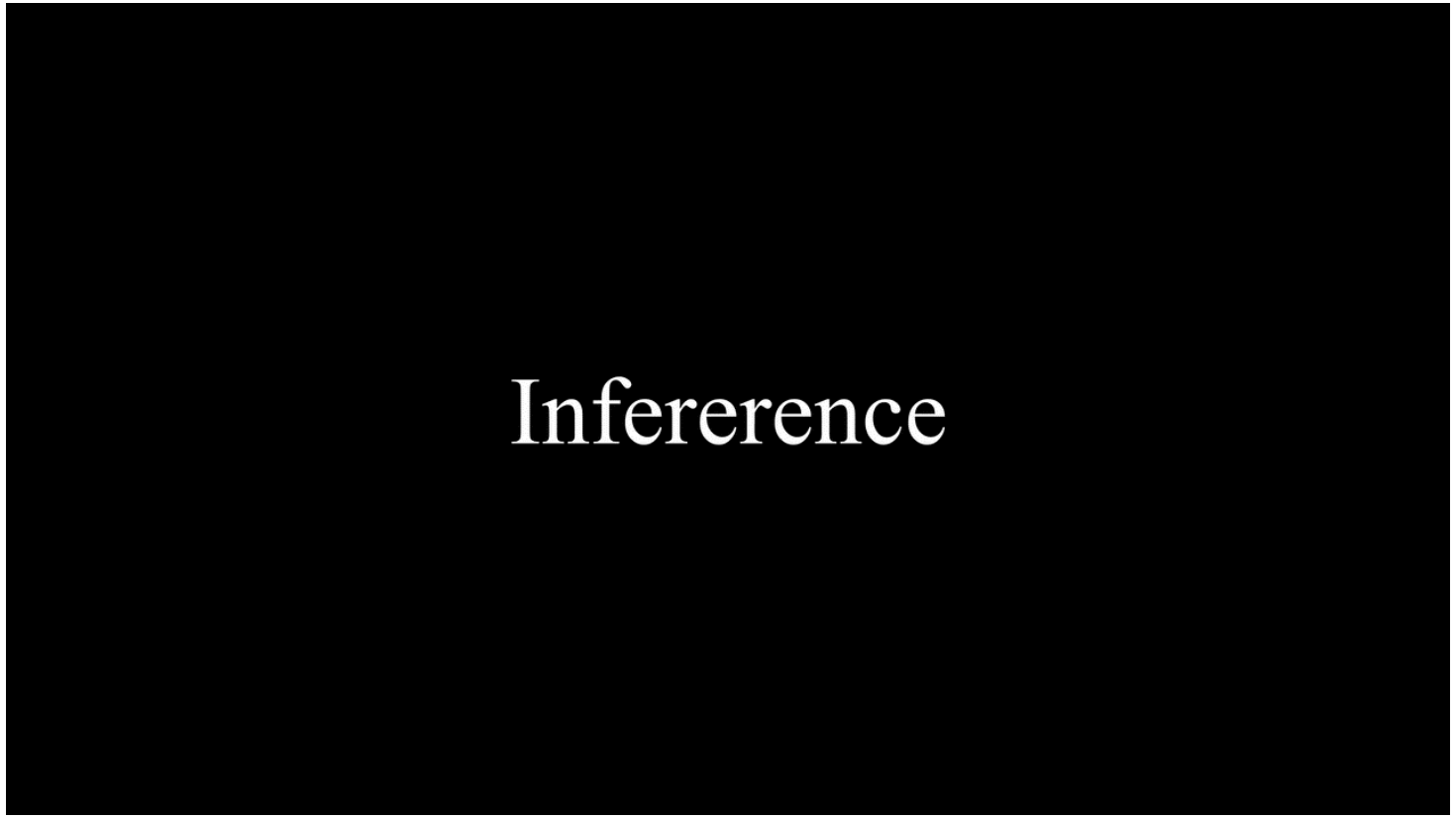
- **Learn** template networks during **training**



TopoNets

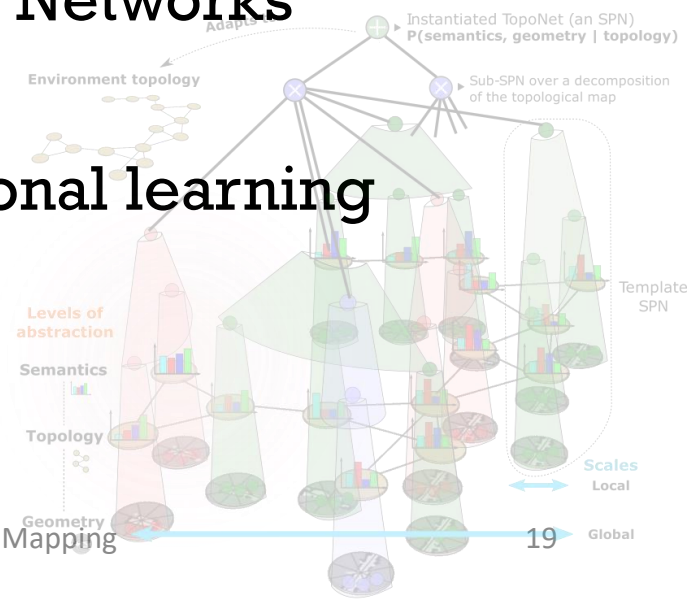
Take-away III: Template-based method

- **Instantiate** complete network to **infer** semantics of any test environment



TopoNets: Recap of Merits

- Builds a **unified deep** model (an SPN) instead of an assembly of independent models
  - Can be **learned end-to-end** from robot sensor input
- **Template-based** method
  - **Adapts** to different environments
- **Tractable, exact inference** (real-time)
  - **Theoretically guaranteed** thanks to Sum-Product Networks
- Fully **probabilistic** and **generative**
  - Can **detect novel** semantic maps to trigger additional learning



Experiments

Experiments: Inference Tasks

Task 1: Semantic place classification (accuracy)

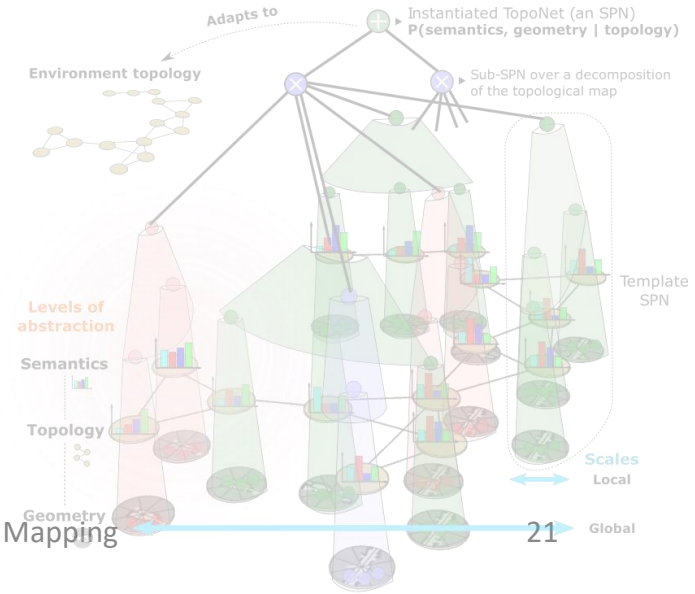
$$\hat{\boldsymbol{y}}_{explored} = \text{argmax}_{\boldsymbol{y}_{explored}} P(\boldsymbol{y}_{explored}|\boldsymbol{x})$$

Task 2: Inferring placeholders (unexplored) (accuracy of placeholders)

$$\hat{\boldsymbol{y}}_{explored}, \hat{\boldsymbol{y}}_{unexplored} = \text{argmax}_{\substack{\boldsymbol{y}_{unexplored} \\ \boldsymbol{y}_{explored}}} P(\boldsymbol{y}_{explored}, \boldsymbol{y}_{unexplored}|\boldsymbol{x})$$

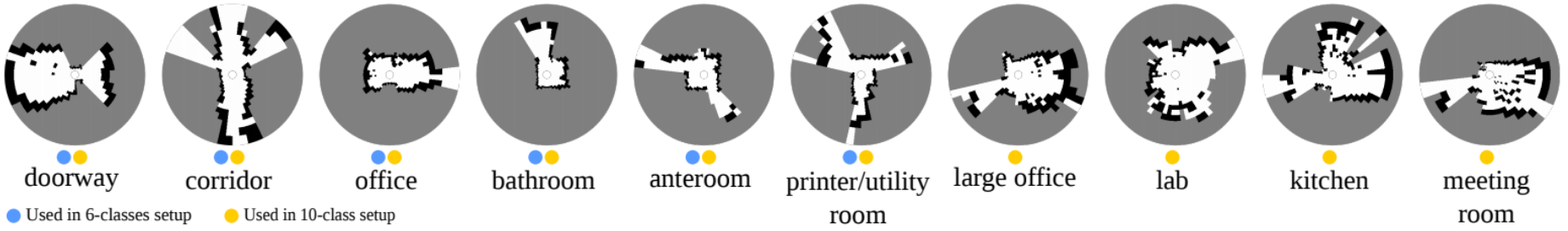
Task 3: Novelty detection (ROC curve)

$$\sum_{\boldsymbol{y}_{explored}} P(\boldsymbol{y}_{explored}, \boldsymbol{x}) > threshold$$



Experiments: Dataset

- Collected by a mobile robot
  - **32** semantic maps on 4 floors
  - Built from laser-range and odometry data
- Two experimental setups (6 or 10 semantic clases)



- Cross-validation:
  - **Trained** on data from **3 floors**
  - **Tested** on data from **remaining floor**

Experiments: Baseline

An assembled approach consisting of

- SPN-based Local Place Classifier
- Markov Random Field (MRF)
- Similar to [Pronobis et al. 2012]
  - Markov Random Field + door detector + SVM

Experiments: Semantic Place Classification

Task 1: Semantic place classification

$$\hat{\boldsymbol{y}}_{explored} = \text{argmax}_{\boldsymbol{y}_{explored}} P(\boldsymbol{y}_{explored} | \boldsymbol{x})$$

Our approach **consistently** improves classification accuracy and disambiguates semantic information.

Semantic Place Classification

	#classes	Local		Local + MRF		TopoNet	
		avg.	std.	avg.	std.	avg.	std.
Overall	6	95.96%	2.83%	95.47%	3.00%	**96.91%**	2.04%
	10	79.06%	6.45%	75.15%	9.34%	**80.14%**	6.35%
456-7	6	95.22%	1.70%	96.35%	2.68%	**97.50%**	1.11%
	10	73.48%	1.92%	68.03%	4.55%	**74.69%**	2.73%
457-6	6	96.75%	1.98%	93.87%	2.24%	**97.39%**	1.25%
	10	81.49%	1.93%	81.63%	6.90%	**82.55%**	1.37%
467-5	6	92.70%	1.52%	**95.66%**	3.14%	94.46%	1.62%
	10	73.41%	2.06%	66.63%	5.38%	**74.58%**	2.48%
567-4	6	**99.16%**	0.94%	96.00%	3.21%	98.30%	1.64%
	10	87.88%	2.83%	84.31%	1.56%	**88.73%**	2.35%

Experiments: Inferring placeholders (unexplored)

Task 2: $\hat{\boldsymbol{y}}_{explored}, \hat{\boldsymbol{y}}_{unexplored}$

$$= \underset{\substack{\boldsymbol{y}_{unexplored} \\ \boldsymbol{y}_{explored}}}{\text{argmax}} P\left(\boldsymbol{y}_{explored}, \boldsymbol{y}_{unexplored} \middle| \boldsymbol{x}\right)$$

Our approach **significantly** outperforms the baseline on this task.

Placeholder Inference

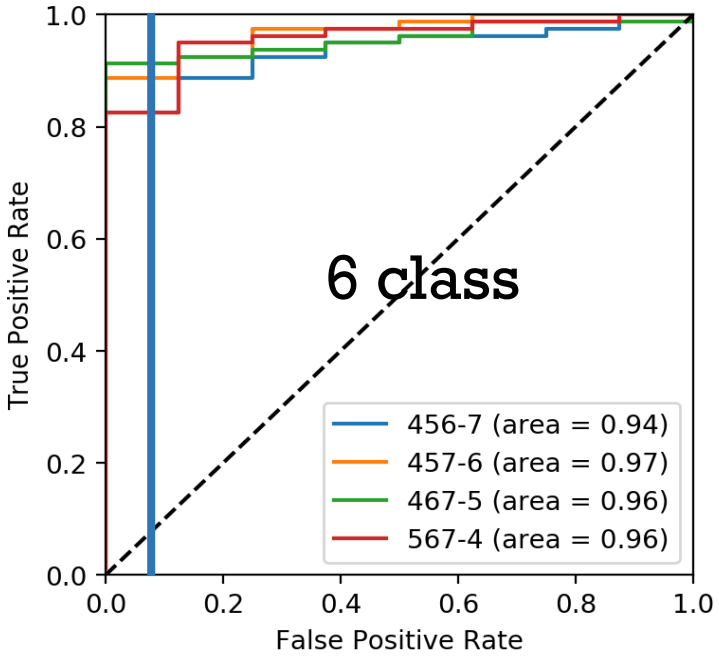
	#classes	Local + MRF avg.	Local + MRF std.	TopoNet avg.	TopoNet std.
Overall	6	91.16%	8.27%	**94.46%**	7.35%
	10	60.45%	9.84%	**64.49%**	10.11%
456-7	6	94.07%	5.65%	**99.46%**	1.44%
	10	57.94%	3.33%	**70.24%**	10.18%
457-6	6	79.77%	6.49%	**83.29%**	4.26%
	10	**61.62%**	9.48%	61.30%	4.83%
467-5	6	94.72%	3.48%	**96.35%**	3.62%
	10	50.50%	5.48%	**54.48%**	3.38%
567-4	6	96.09%	3.50%	**98.75%**	3.31%
	10	71.72%	4.78%	**71.94%**	8.45%

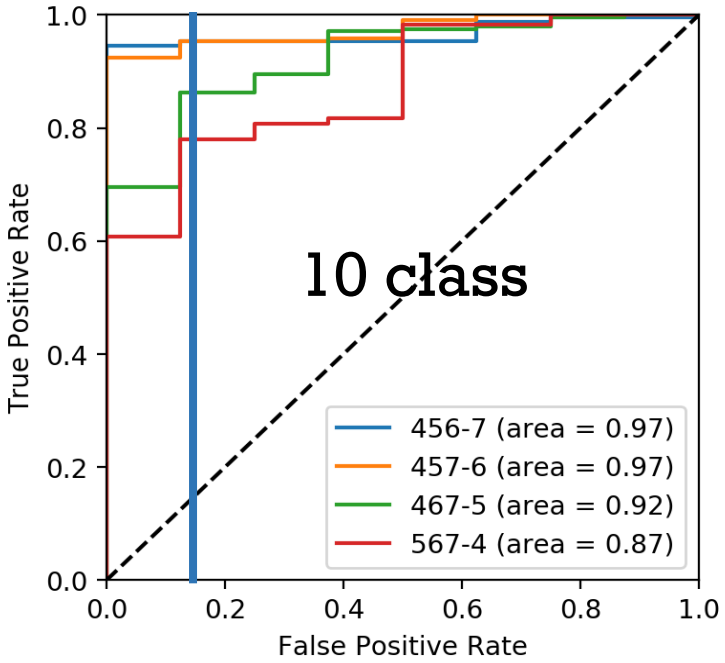
Experiments: Novelty Detection

Task 3: Novelty detection

$$\sum_{\boldsymbol{y}_{explored}} P(\boldsymbol{y}_{explored}, \boldsymbol{x}) > threshold$$

85-90% True Positive, 10-15% False Negative.

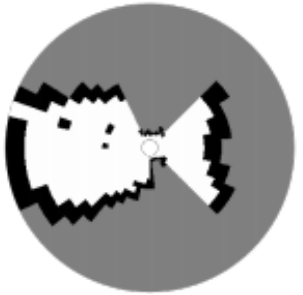




True positive:
Semantic map is novel, classified as novel

False positive:
Semantic map is NOT novel, classified as novel

Experiments: Performance



- Each local laser range observation:
  - **1176 pixels**, each 3 possible values
    - >3500 indicator variables
- Topological graph size: **~100-150 nodes**
- NVidia GeForce 1080Ti, LibSPN library [Pronobis et al.'17]

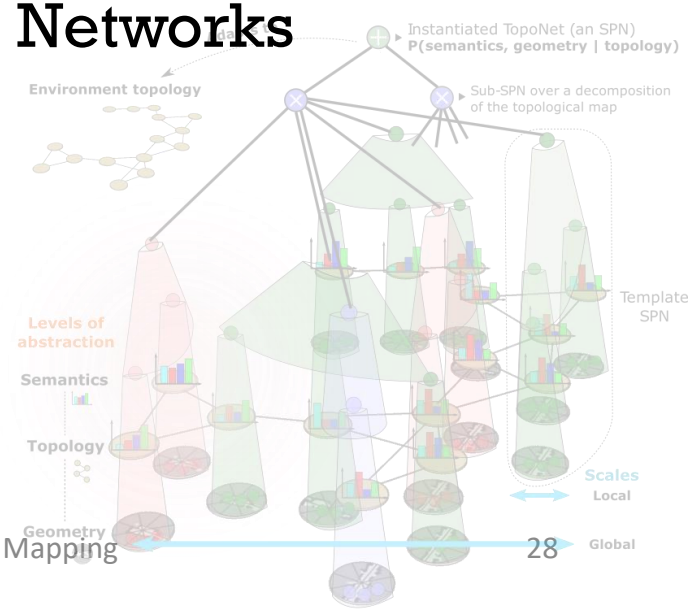
size	TopoNets	Base line
105	0.36s	> 45s
155	0.49s	

Worst case run time (empirical), 10 class setup
(Evaluate P(**X**,**Y**), for 30 random different **Y** settings)

TopoNets **infers in real-time**, while MRF suffers from **convergence issues**

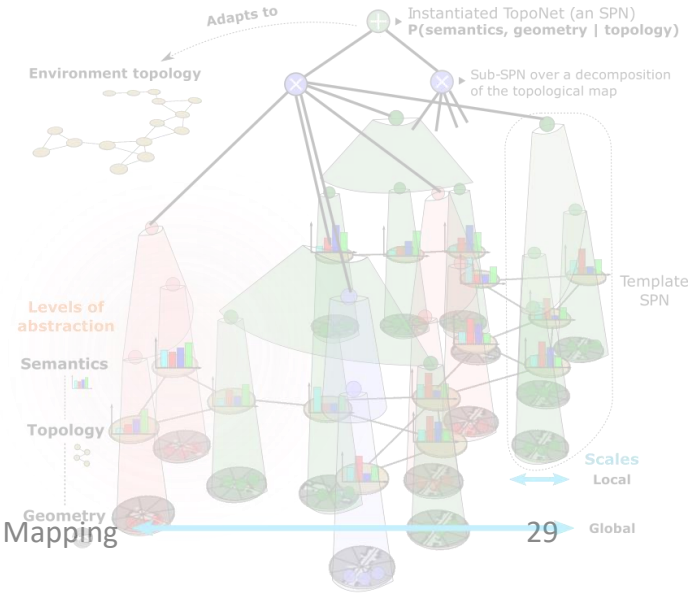
Summary

- Take-away I : **End-to-end** Unified Deep Probabilistic Spatial Model
  - Builds a **unified deep** model (a SPN) that can be **learned end-to-end**
  - Fully **probabilistic** and **generative**
    - Capable to **detect novel** semantic maps
- Take-away II: **Tractable, exact inference** (real-time)
  - **Theoretically guaranteed** thanks to Sum-Product Networks
- Take-away III: **Template-based** method
  - **Adapts** to different environments



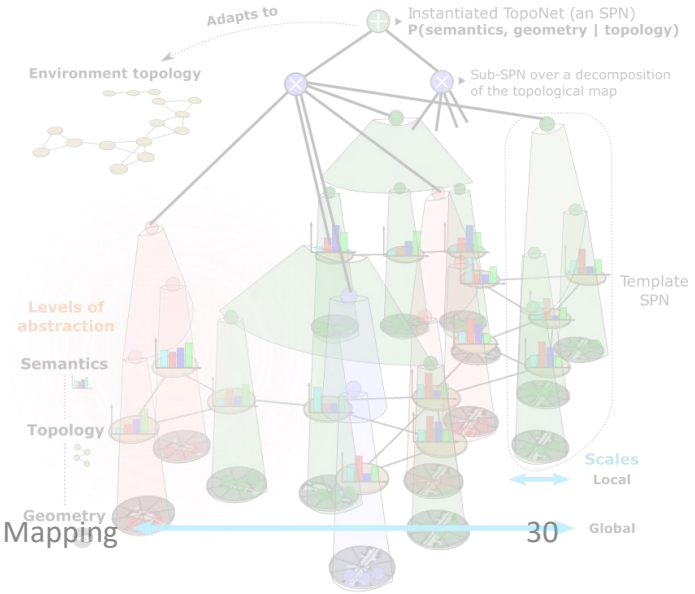
Summary

- TopoNets introduce novel, probabilistic deep learning techniques to robotics
- Ideal model for partially-observable planning in large, unknown environment



Video link

https://www.youtube.com/watch?v=luv2XpaHeTU



References

- [Calli et al. 2015] Benchmarking in Manipulation Research: The YCB Object and Model Set and Benchmarking Protocols
- [Pronobis 2011] Semantic mapping with mobile robots
- [Mozos et al. 2007] Supervised semantic labeling of places using information extracted from sensor data
- [Friedman et al. 2007] Voronoi random fields : Extracting the topological structure of indoor environments via place labeling
- [Brucker et al 2018] Semantic labeling of indoor environments from 3d rgb maps
- [Wu et al. 2016] Deep markov random field for image modeling
- [Mahmood 2019] Structured prediction using cgans with fusion discriminator
- [Chen et al 2018] Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRF
- [Schwing et al 2015] Fully connected deep structured network
- [Belanger and McCallum 2016] Structured prediction energy networks
- [Pronobis et al ICAPS Workshop'17] Deep spatial affordance hierarchy: Spatial knowledge representation for planning in large-scale environments
- [Peharz et al 2017] On the latent variable interpretation in Sum-Product network
- [Poon and Domingos 2011] Sum-product networks: A new deep architecture
- [Gens and Domingos 2012] Discriminative learning of sum-product networks
- [Pronobis et al 2010] Semantic Modeling of Space
- [Zheng et al. 2018] Learning Graph-Structured Sum-Product Networks for Probabilistic Semantic Maps
- [van de Wolfshaar and Pronobis 2019] Deep Generalized Convolutional Sum-Product Networks for Probabilistic Image Representations
- [Sünderhauf et al 2015] Place Categorization and Semantic Mapping on a Mobile Robot