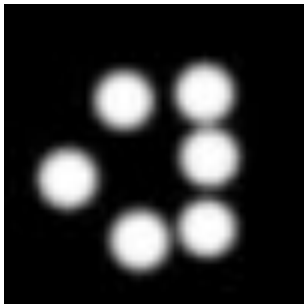
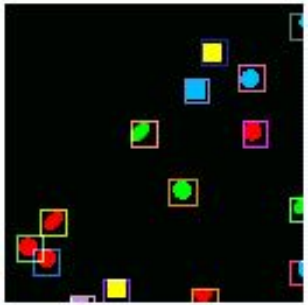
Improving Generative Imagination in Object-Centric World Models

Zhixuan Lin, Yi-Fu Wu, Skand, Bofeng Fu, Jindong Jiang, Sungjin Ahn

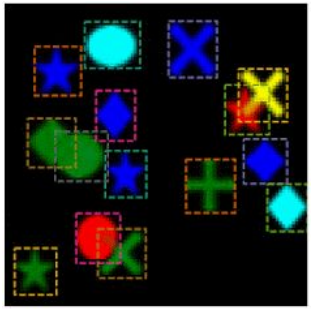
Object-Centric Temporal Generative Models



STOVE
(Kossen et al., 2019)



SCALOR
(Jiang et al., 2019)

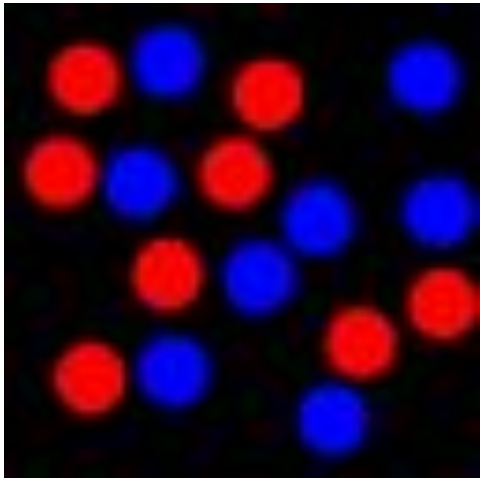


SILOT
(Crawford & Pineau, 2020)

OP3
(Veerapaneni et al., 2019)

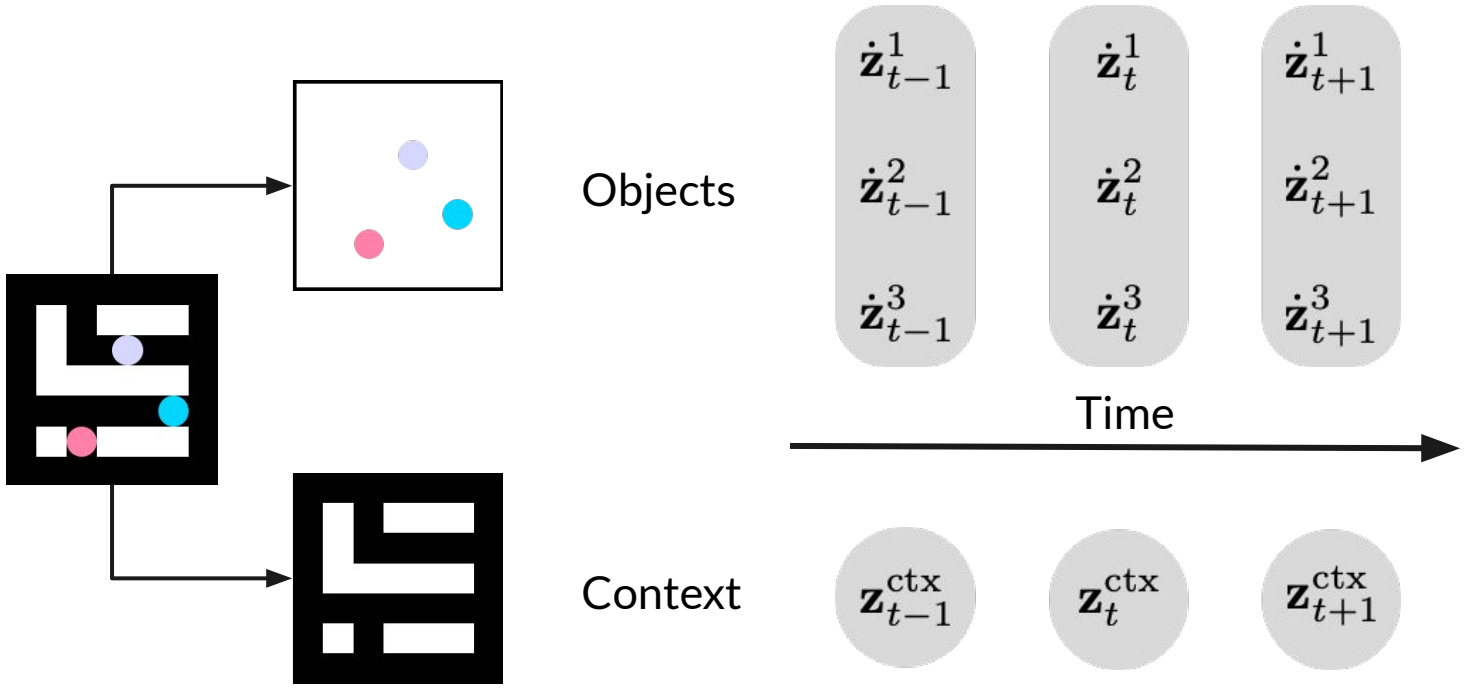
What's Missing

- Interaction
- Occlusion
- Scalability

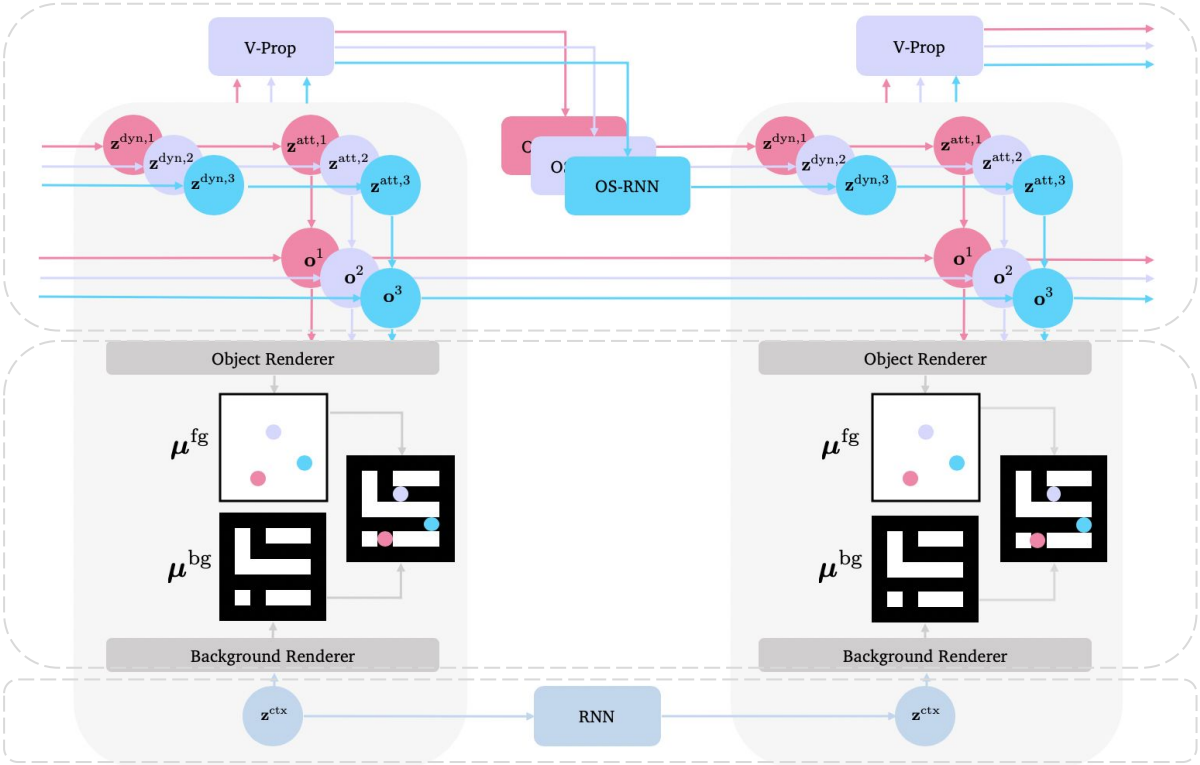


- Multimodal Uncertainty
- Situation Awareness

G-SWM: Generative Structured World Models



G-SWM: Generative Process



Object Dynamics
(Versatile Propagation)

Rendering

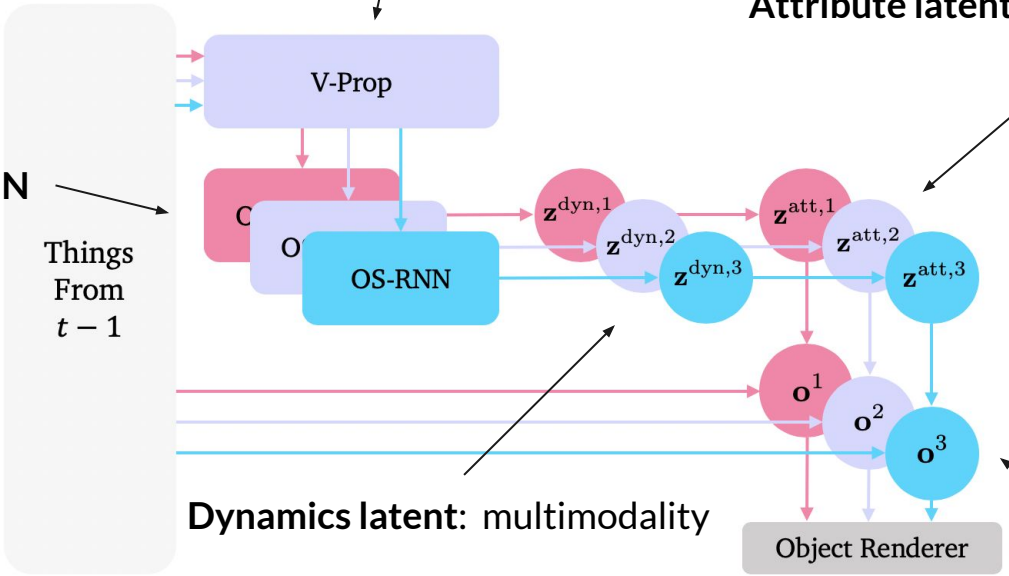
Context Dynamics

Versatile Propagation

Versatile propagation: the core

$$\mathbf{z}^{\text{att}} = [\mathbf{z}^{\text{pres}}, \mathbf{z}^{\text{depth}}, \mathbf{z}^{\text{where}}, \mathbf{z}^{\text{what}}]$$

Attribute latents: position, size, appearance

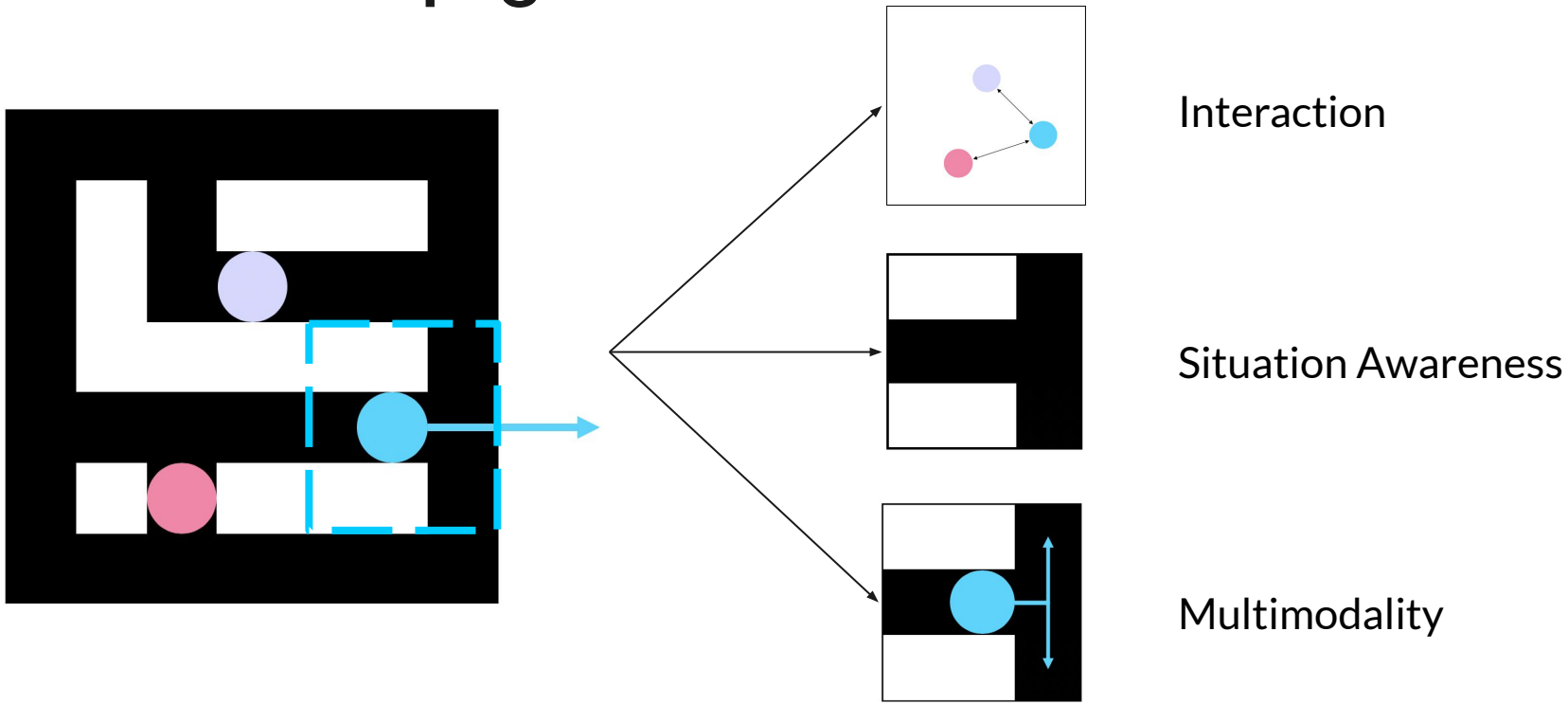


Dynamics latent: multimodality

Object attributes: position, size, appearance, ...

$$\mathbf{o} = [\mathbf{o}^{\text{pres}}, \mathbf{o}^{\text{depth}}, \mathbf{o}^{\text{where}}, \mathbf{o}^{\text{what}}]$$

Versatile Propagation

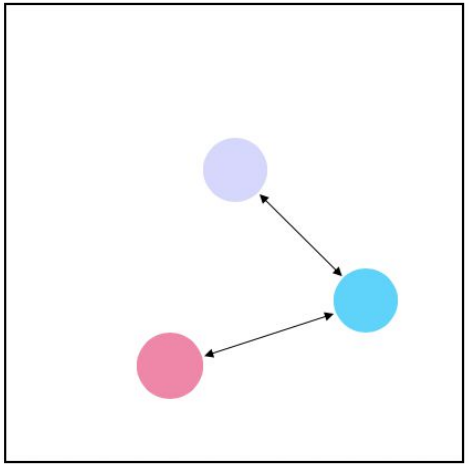


Interaction

Situation Awareness

Multimodality

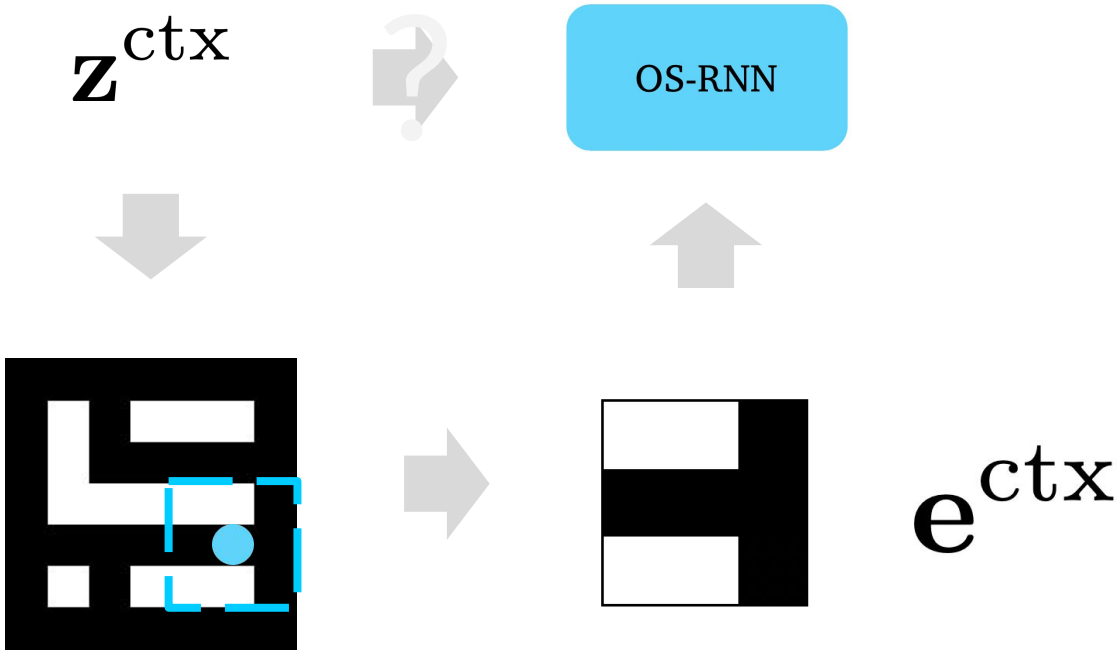
Object Interaction: Graph Neural Network



GNN: $\mathbf{e}^{\mathrm{rel},k} = \mathbf{e}^{k,k} + \sum_{j \neq k} w^{k,j} \mathbf{e}^{k,j}$

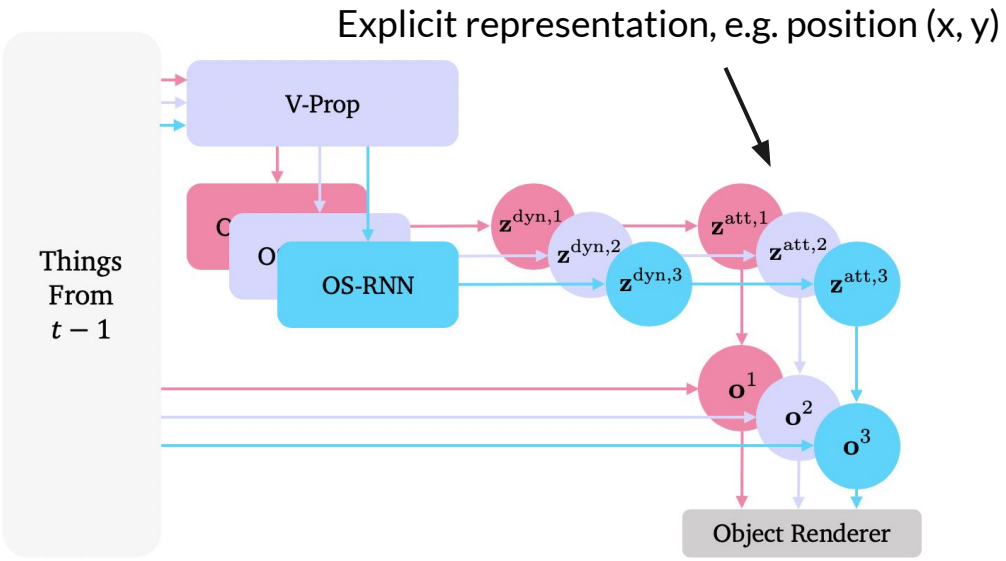
$\mathbf{z}^{\mathrm{att}} = [\mathbf{z}^{\mathrm{pres}}, \mathbf{z}^{\mathrm{depth}}, \mathbf{z}^{\mathrm{where}}, \mathbf{z}^{\mathrm{what}}]$

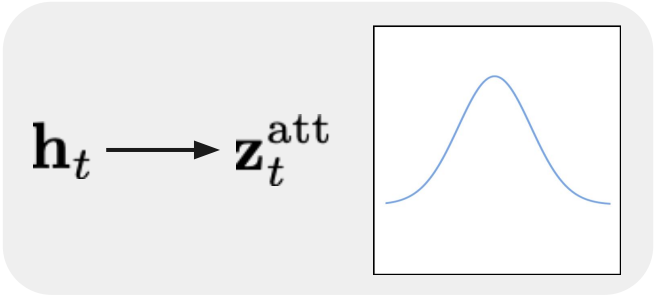
Situation Awareness: Attention on Environment

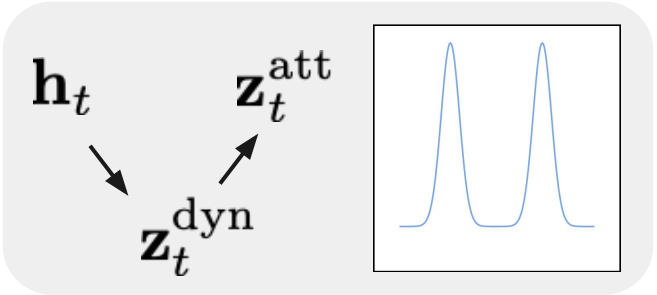


AOE: Attention on Environment

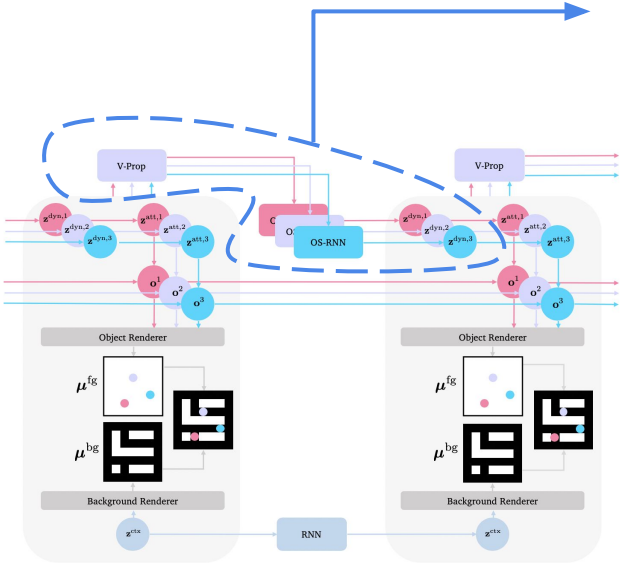
Multimodal Uncertainty: Hierarchical Dynamics

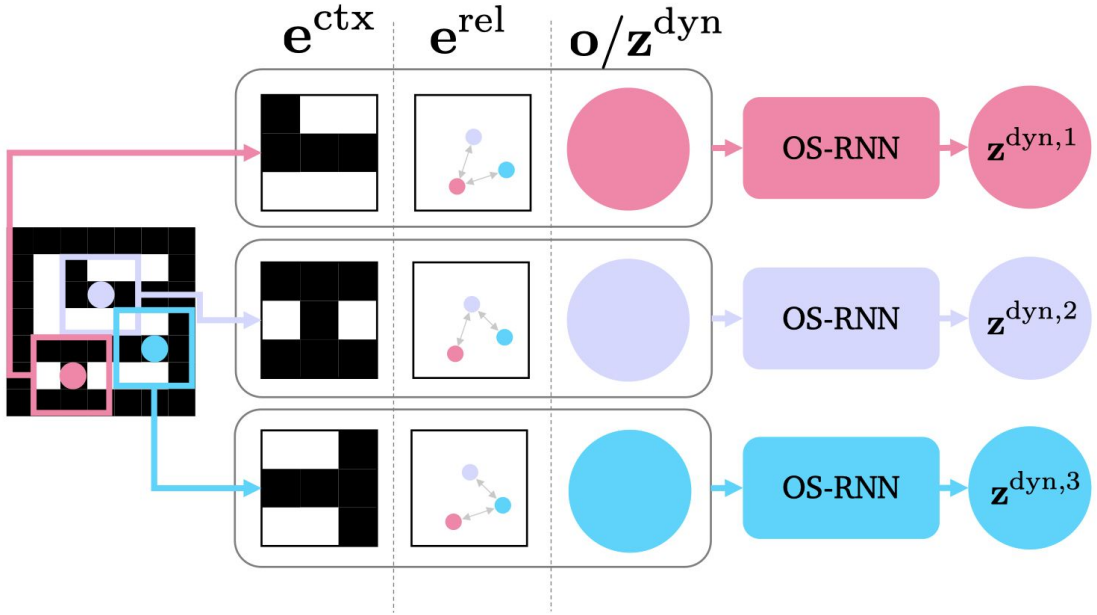




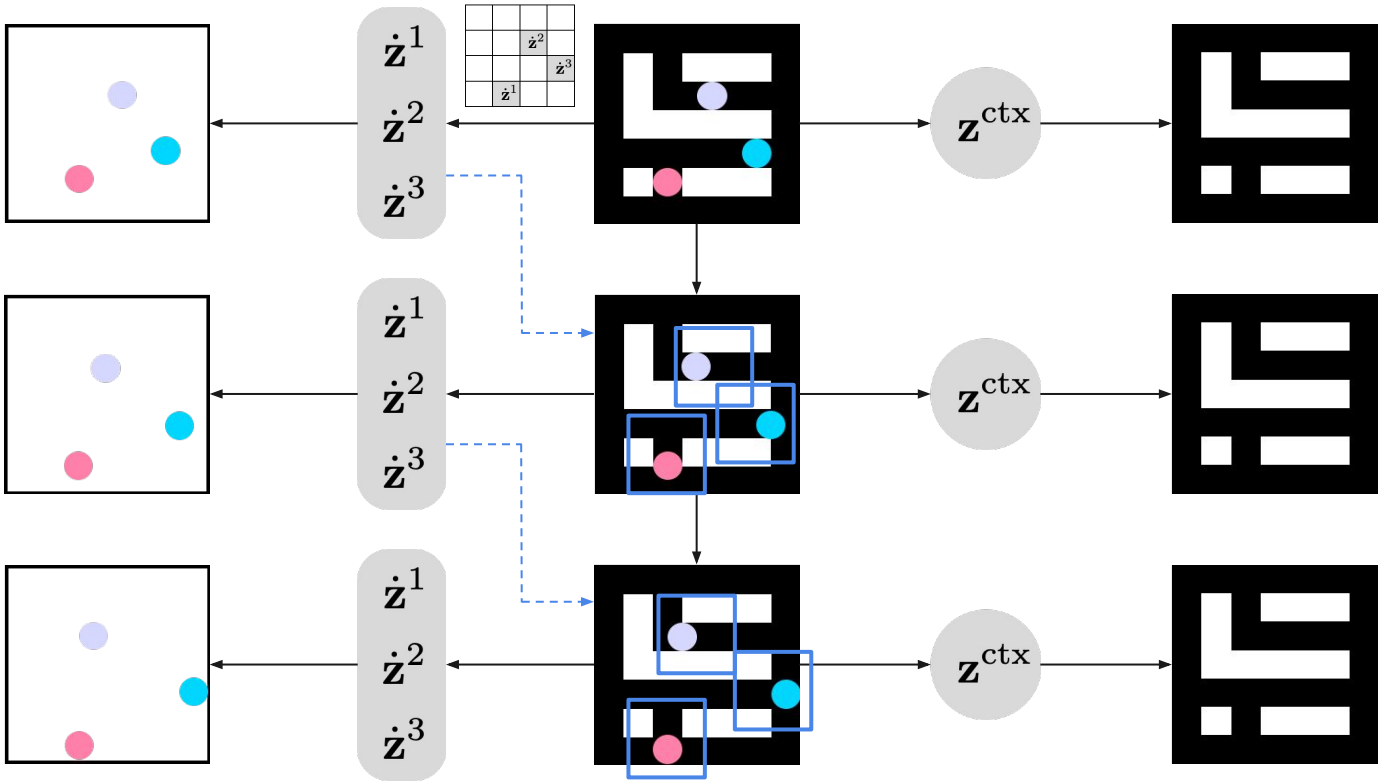


Summary of Versatile Propagation

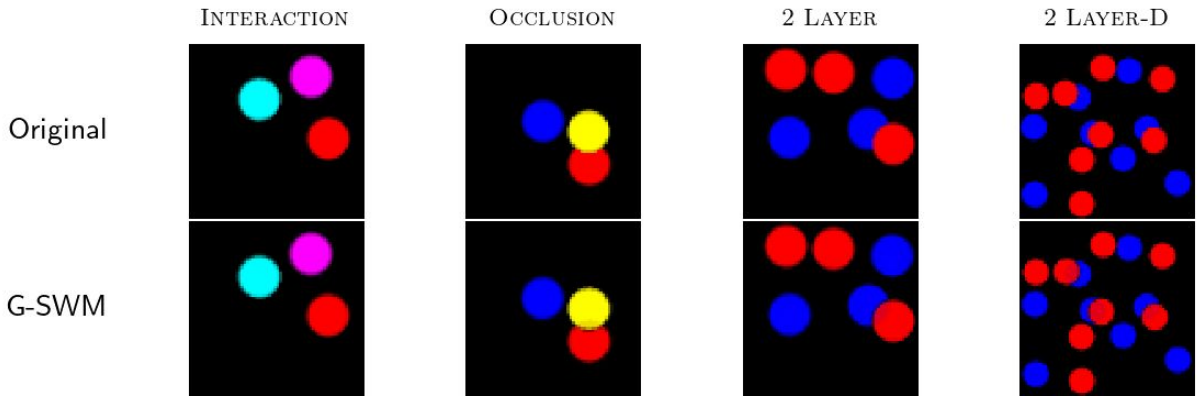


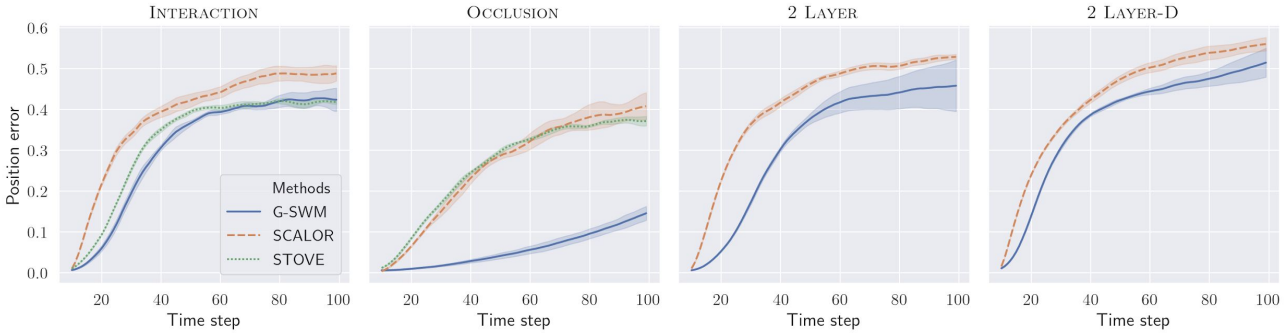


Inference: Scalable Discovery and Propagation

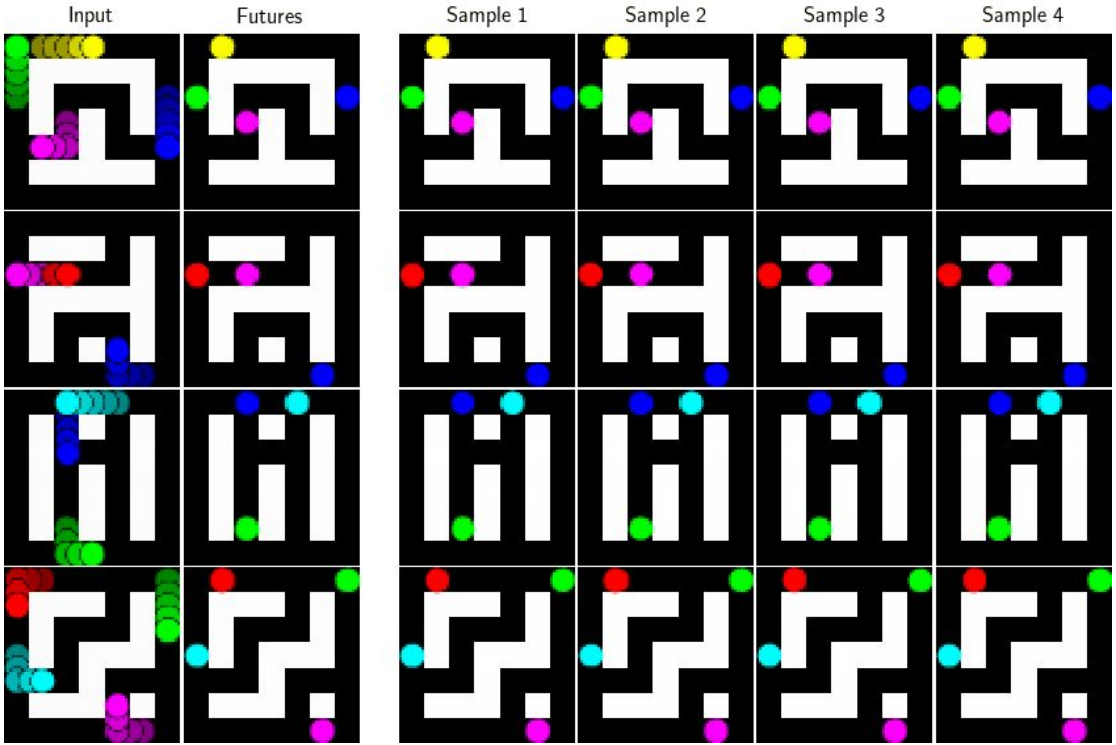


Scalability, Occlusion and Interaction

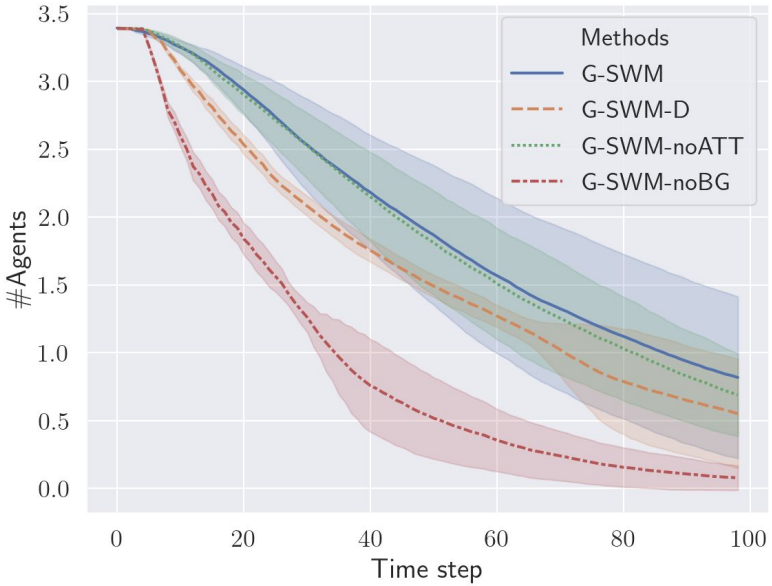


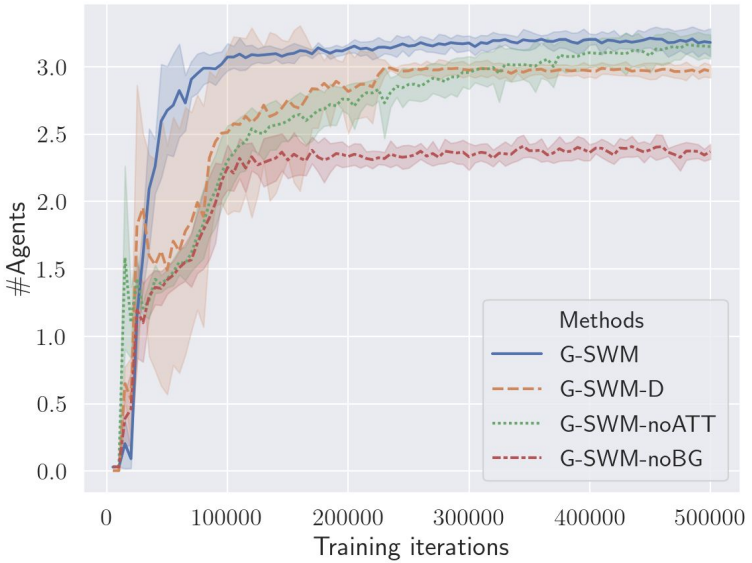


Situation Awareness and Multimodal Uncertainty

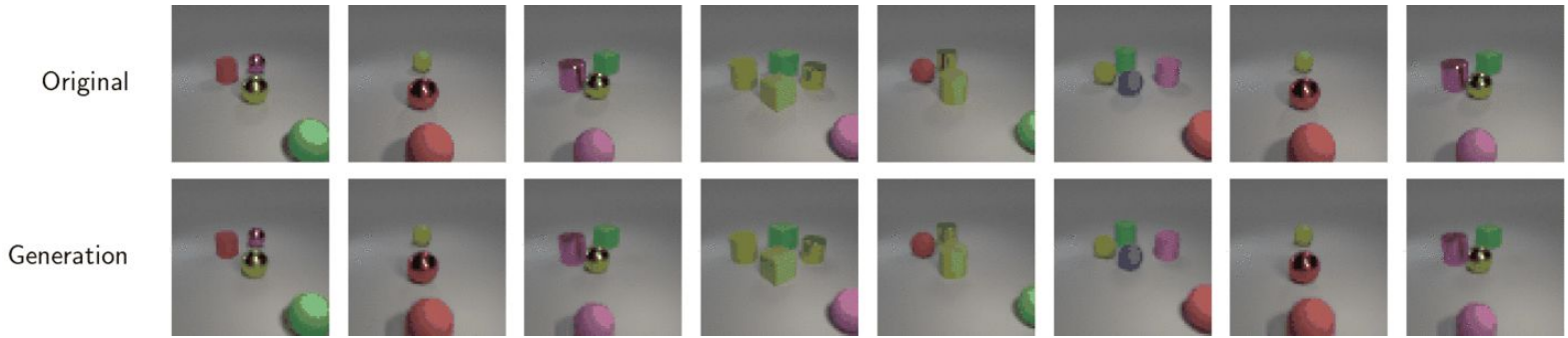


Ablation Study



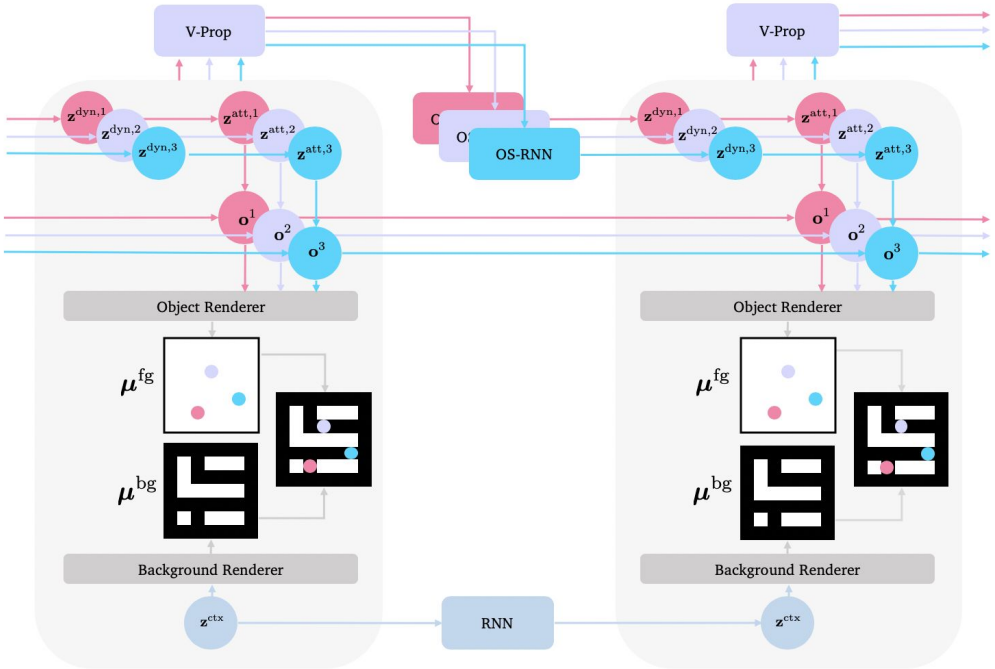


3D Interactions



Summary

- G-SWM:
  - Object-centric
  - Interaction
  - Multimodal uncertainty
  - Situation awareness
  - Occlusion Handling
  - Scalability
- Limitations
  - Quite complex
  - Many assumptions and priors



Thank you!