

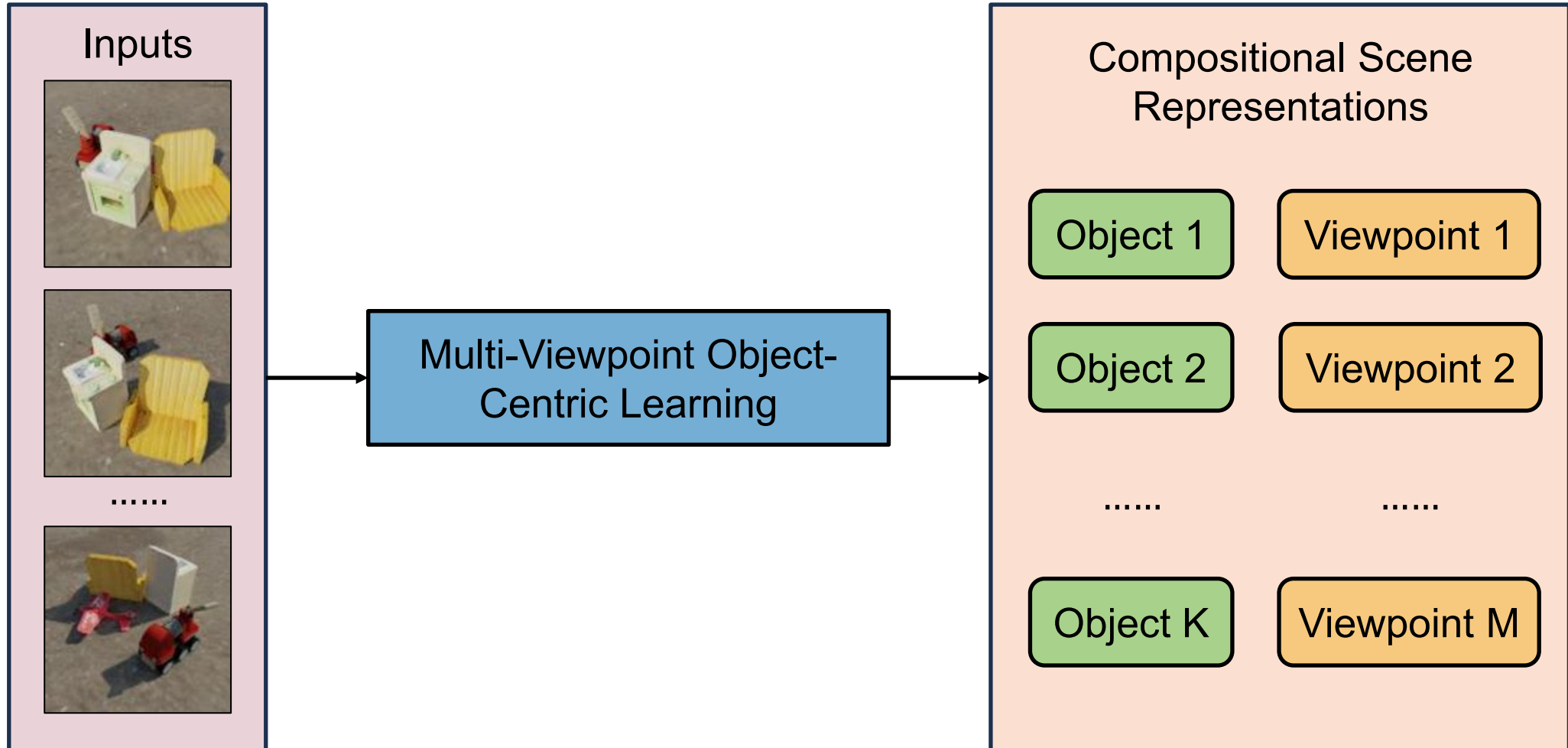
# Improving Viewpoint-Independent Object-Centric Representations through Active Viewpoint Selection

Yinxuan Huang, Chengmin Gao, Bin Li, Xiangyang Xue

School of Computer Science, Fudan University

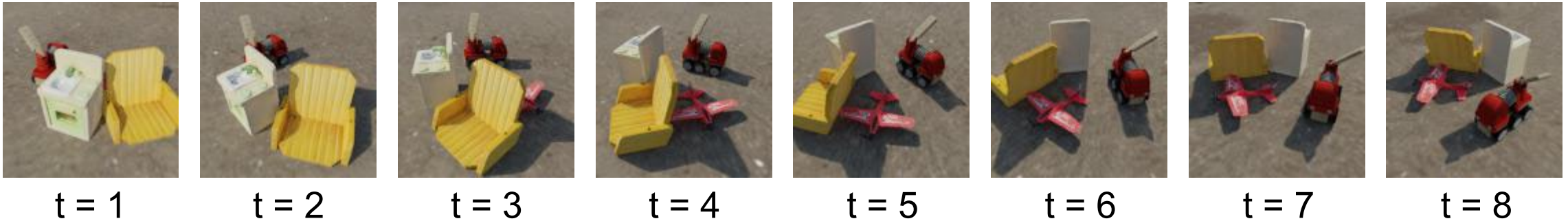
# Object-Centric Learning in Multi-Viewpoint Scenes

**Goal:** Represent a multi-viewpoint scene as a set of **objects** and a set of **viewpoints**



# Viewpoint Selection Strategy

## Multi-Viewpoint Images



### ▶ Random Selection

- ▶ Observation Set:  $t = 1, 4, 5, 8$
- ▶ Example: OCLOC [1]

### ▶ Sequential Selection

- ▶ Observation Set:  $t = 3, 4, 5, 6$
- ▶ Example: SIMONe [2]

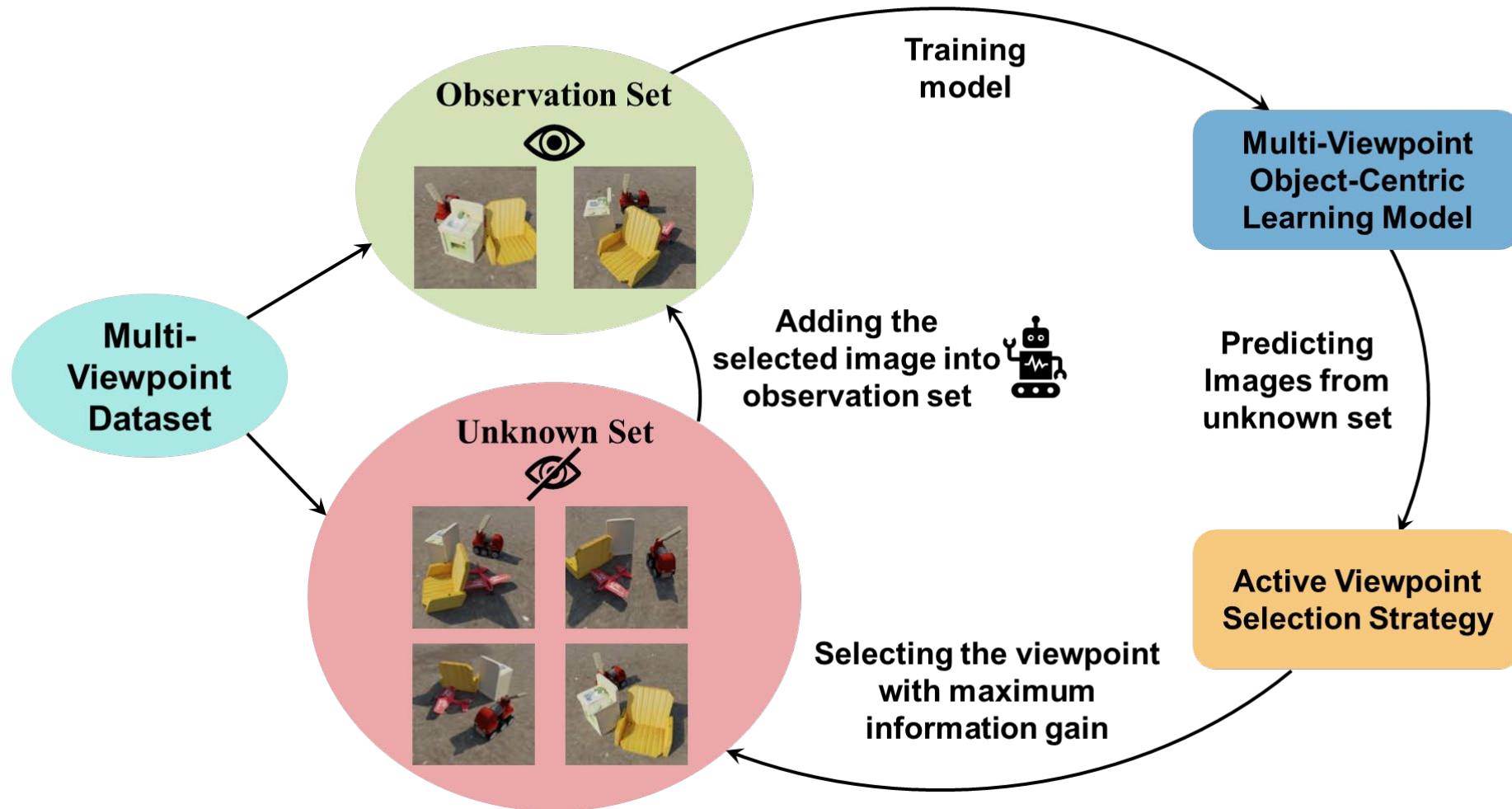
### ▶ **Active Selection:** Select the **next observation viewpoint** for each scene based on **the information of the observed images**

[1] Yuan et al. Unsupervised learning of compositional scene representations from multiple unspecified viewpoints. AAAI 2022.

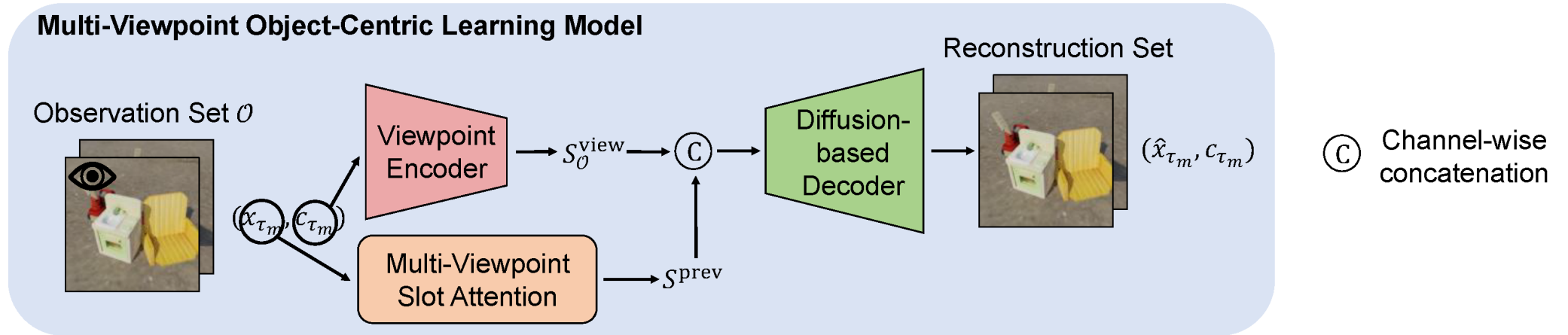
[2] Kabra et al. Simone: View-invariant, temporally-abstracted object representations via unsupervised video decomposition. NeurIPS 2021.

# Active Viewpoint Selection

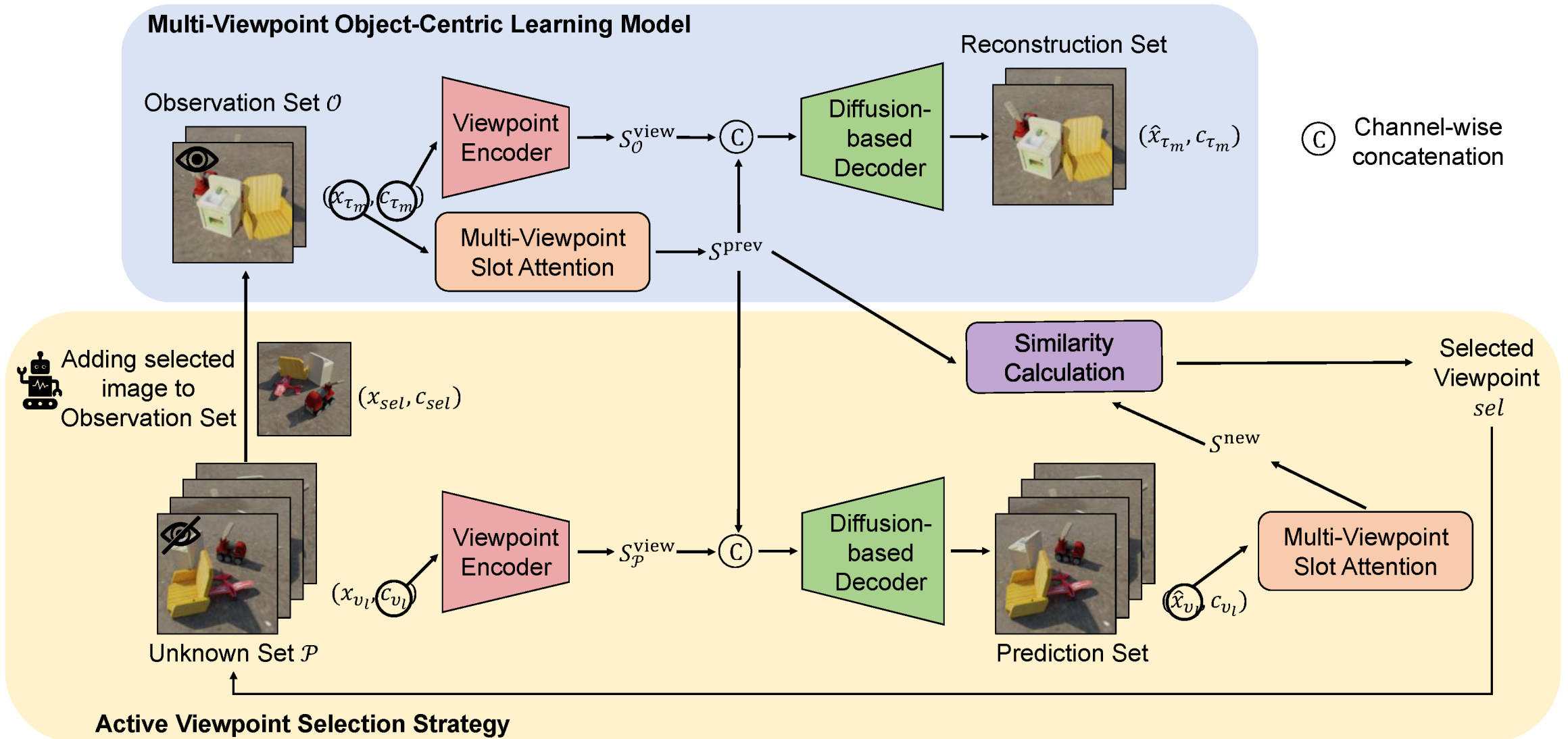
**Goal:** *Iteratively* select viewpoints from the unknown set to form a *small yet informative* observation set, enabling effective training with fewer images



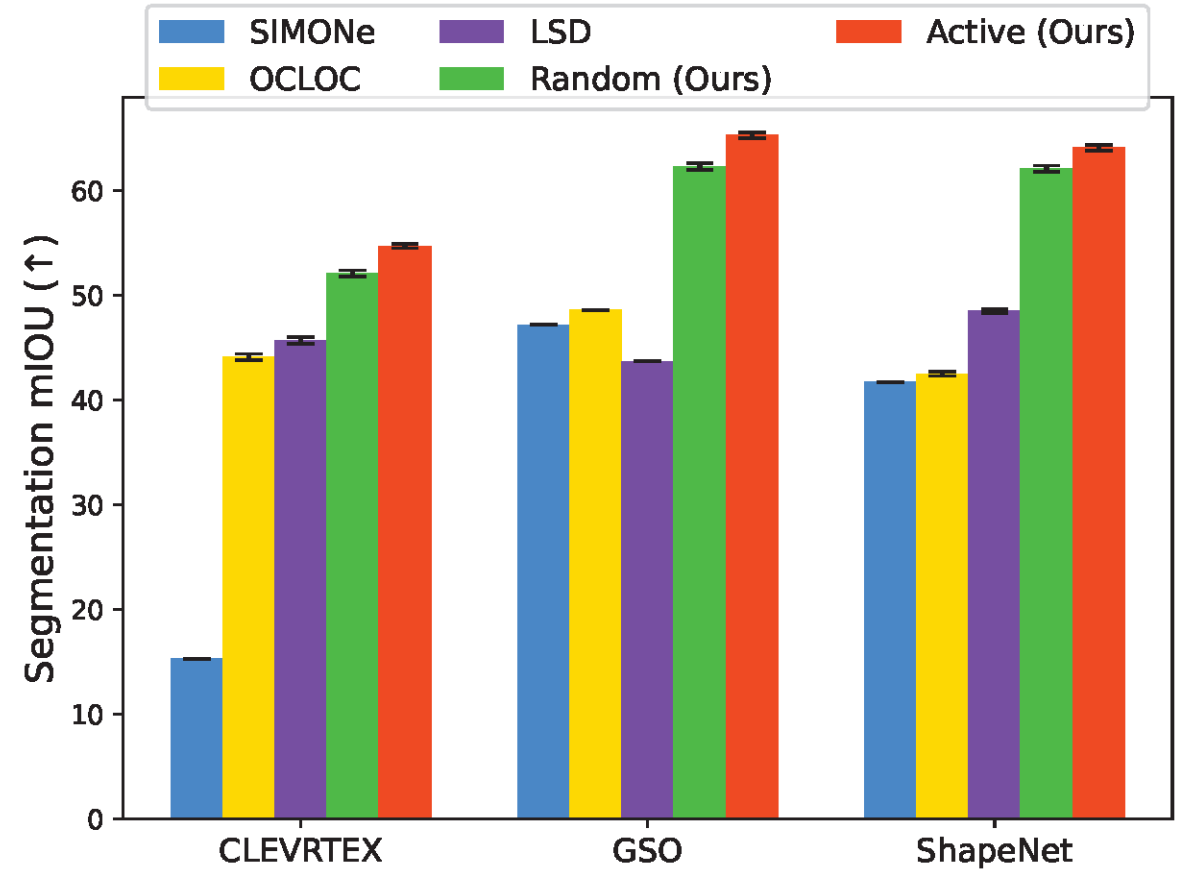
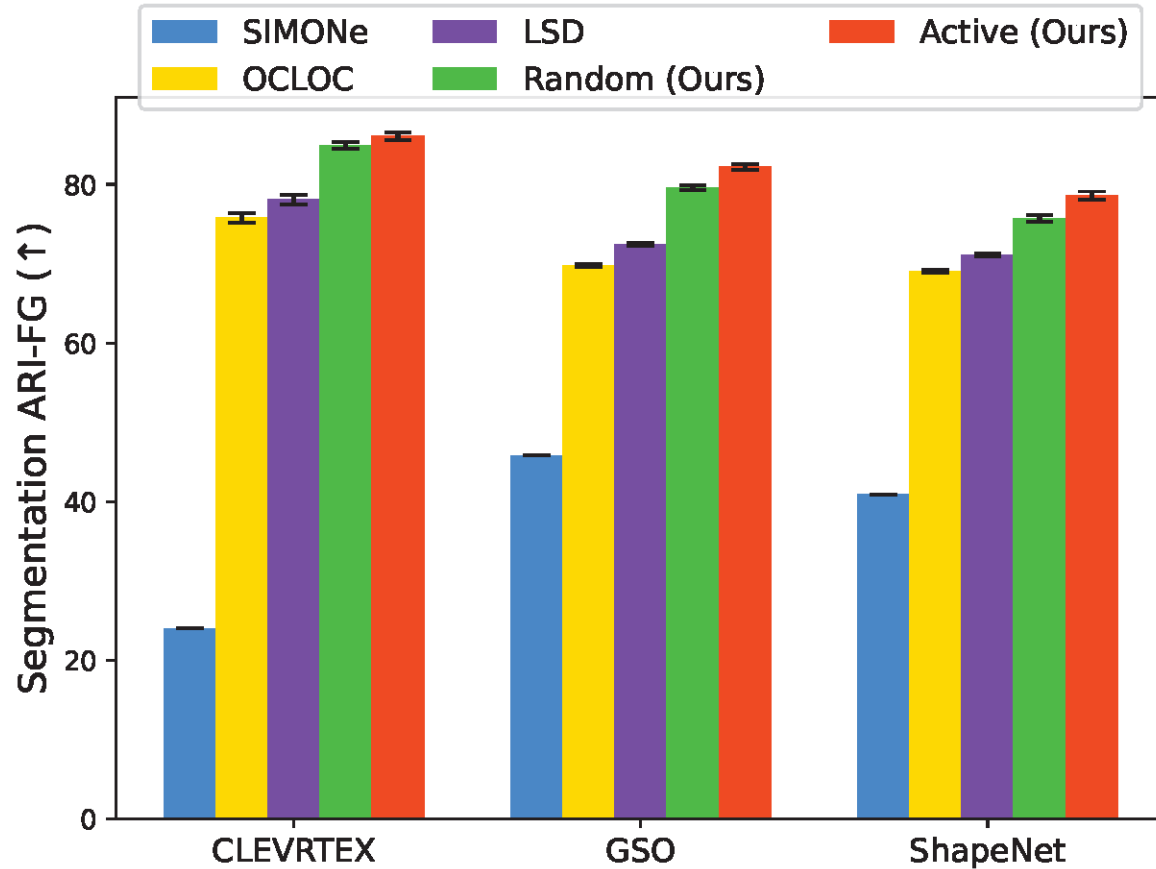
# Active Viewpoint Selection



# Active Viewpoint Selection

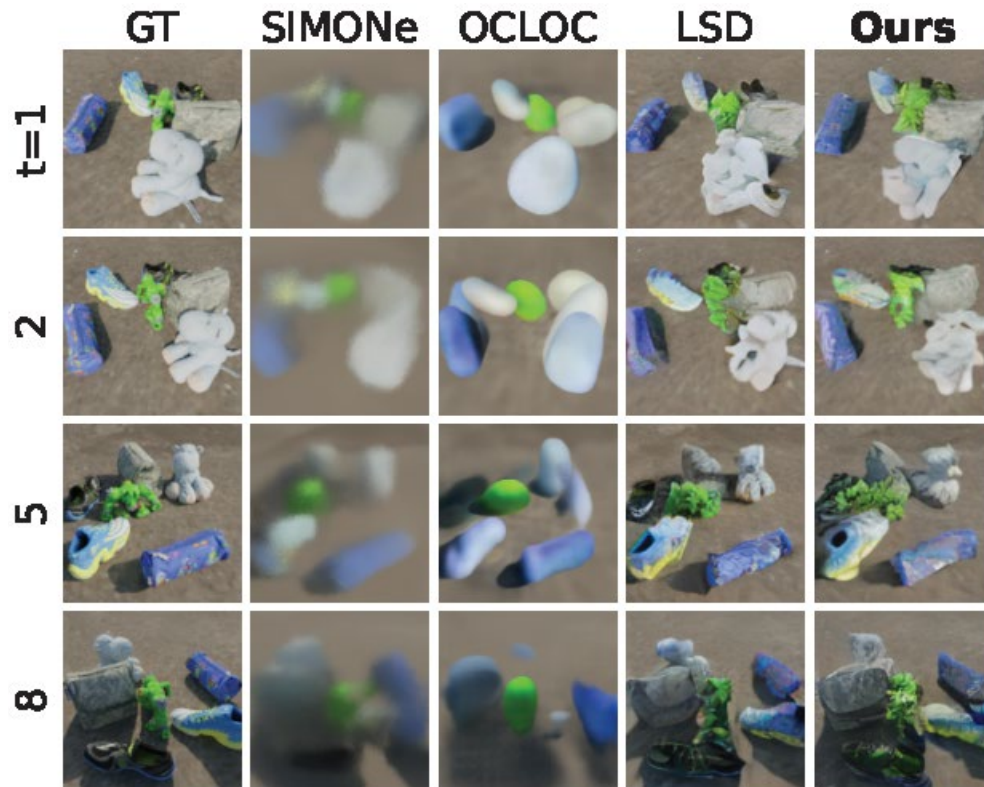


# Experiments: Unsupervised Object Segmentation

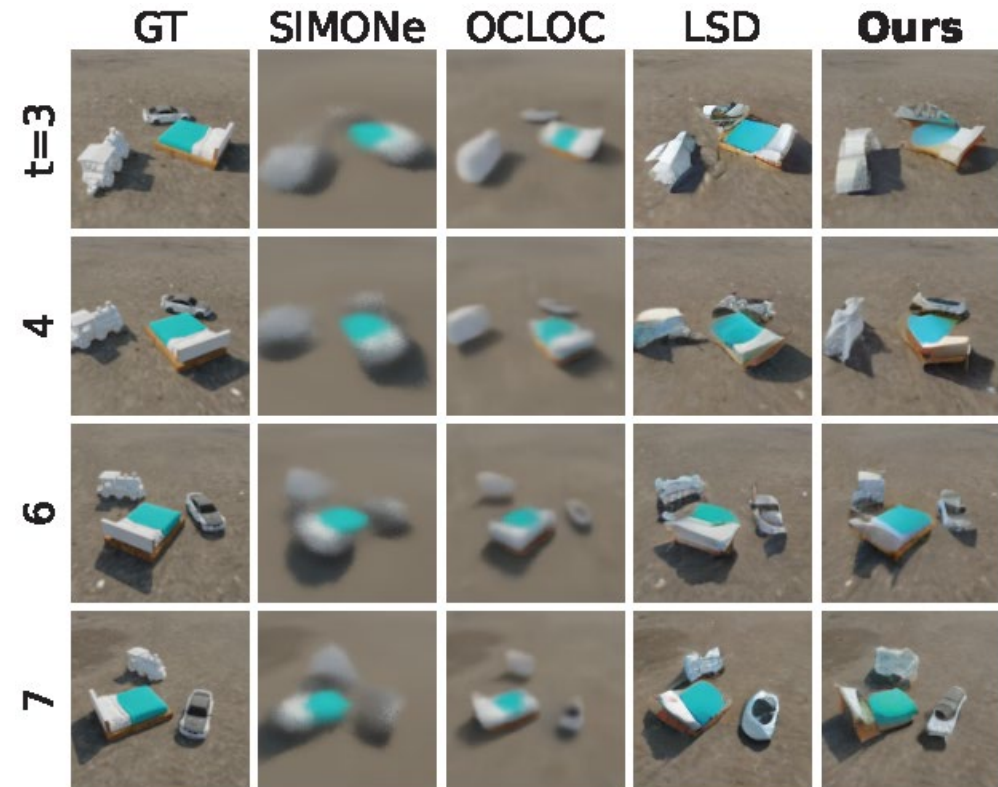




# Experiments: Scene Reconstruction



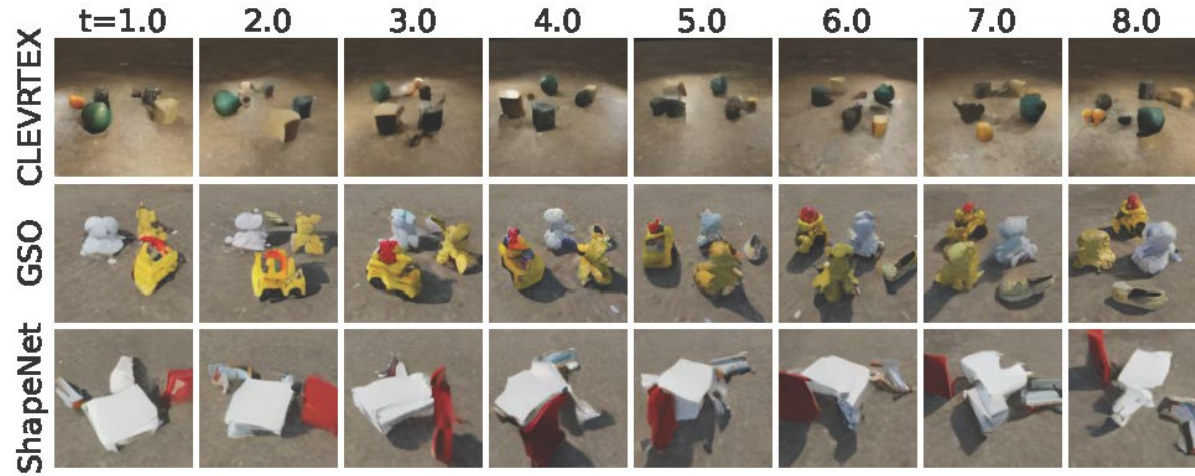
GSO



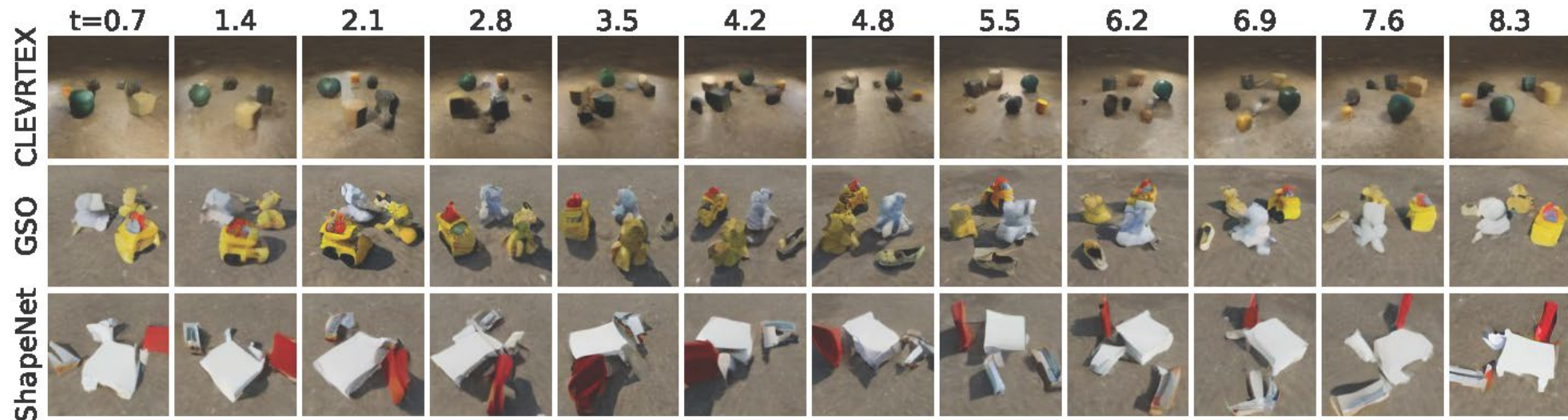
ShapeNet



# Experiments: Compositional Generation

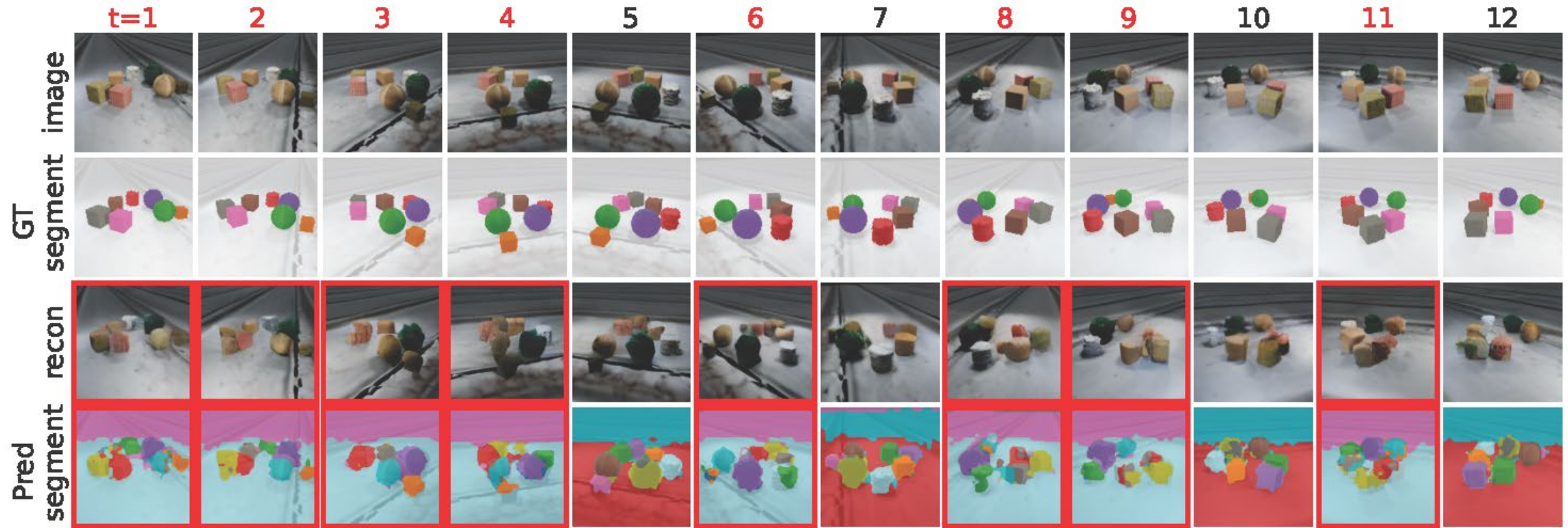


(a) Multi-Viewpoint Generation



(b) Multi-Viewpoint Interpolation

# Experiments: Novel Viewpoint Synthesis



# Conclusion

---

- ▶ We propose a multi-viewpoint object-centric learning model with an active viewpoint selection strategy.
- ▶ Our model achieves outstanding performance in unsupervised object segmentation and image generation.
- ▶ Compared to random viewpoint selection, our active selection strategy significantly enhances viewpoint-independent object-centric representations, improving scene understanding.
- ▶ The model can also predict images from unknown viewpoints and generate images from novel viewpoints.