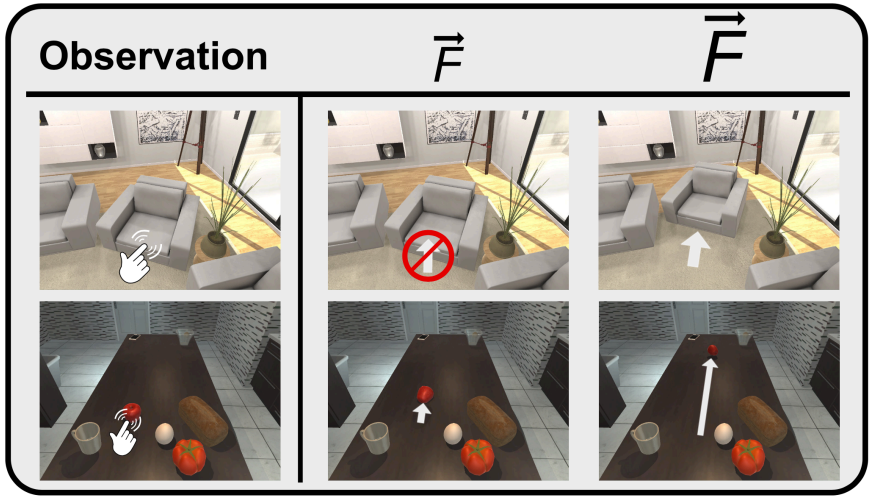
Learning About Objects by Learning to Interact with Them

Martin Lohmann, Jordi Salvador, Aniruddha Kembhavi, Roozbeh Mottaghi

Embodied self supervised learning

How can embodied agents learn about objects in their environment without any supervision?

- Use environment interaction
- Changes to environment = rich learning signal
- Challenges: Noise and sparsity in learning targets

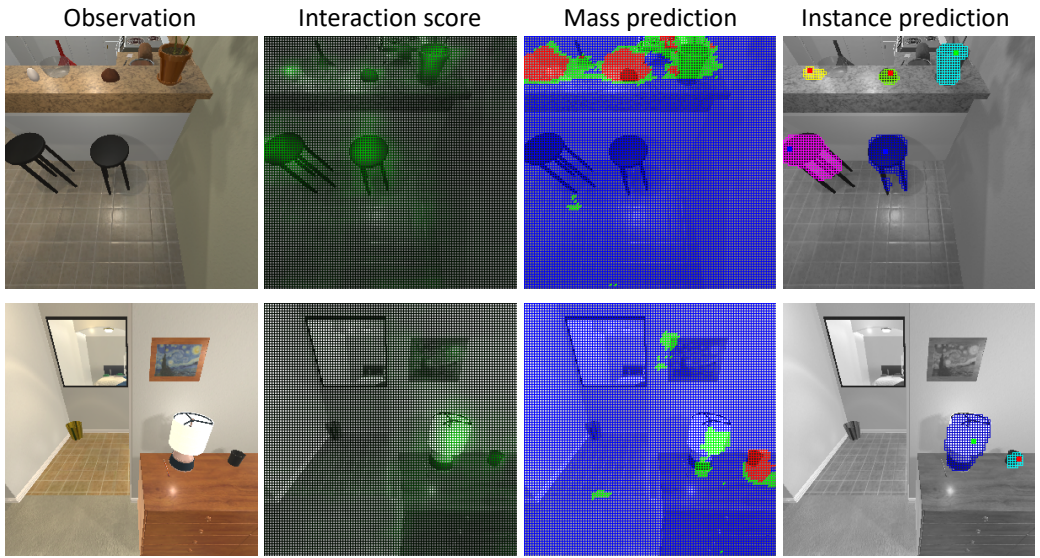


➢ We discover objects and learn about their properties by interacting with them

AI2

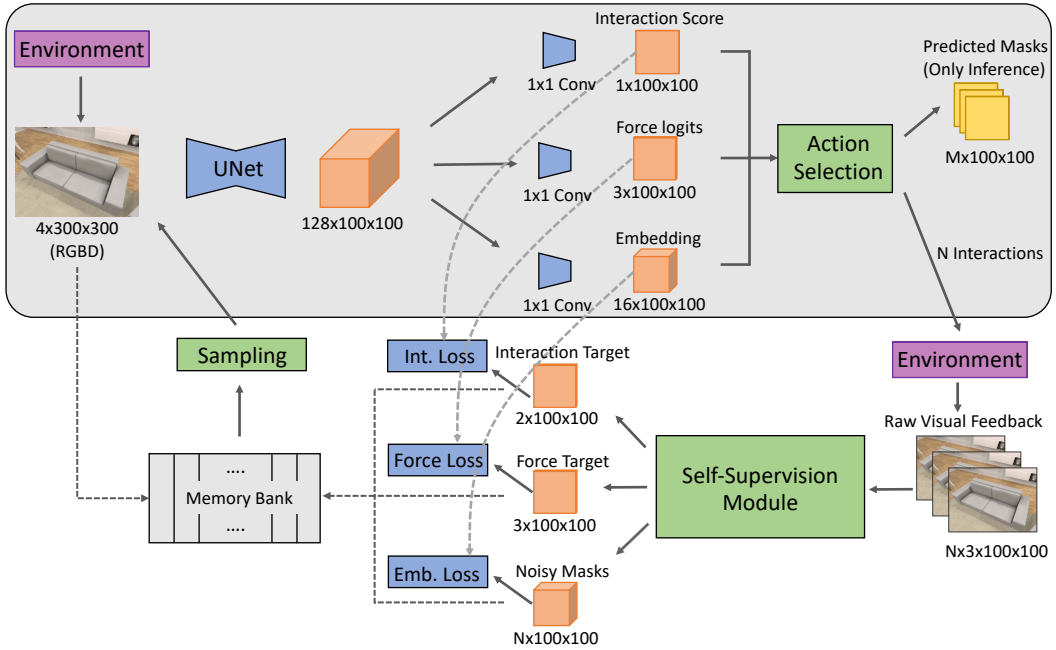
The task

- **Input:** Single RGB+D observation from random agent pose in AI2-THOR household scene
- **Output:**
  - Instance segmentation masks
  - Pixel-wise probability of successful interaction
  - Pixel-wise relative mass estimate (light/medium/heavy)
- **Feedback:** RGB observation after interaction by agent. No labels!
- **Embodiment:** Apply chosen force magnitude to pixel (light/medium/strong)



Our approach

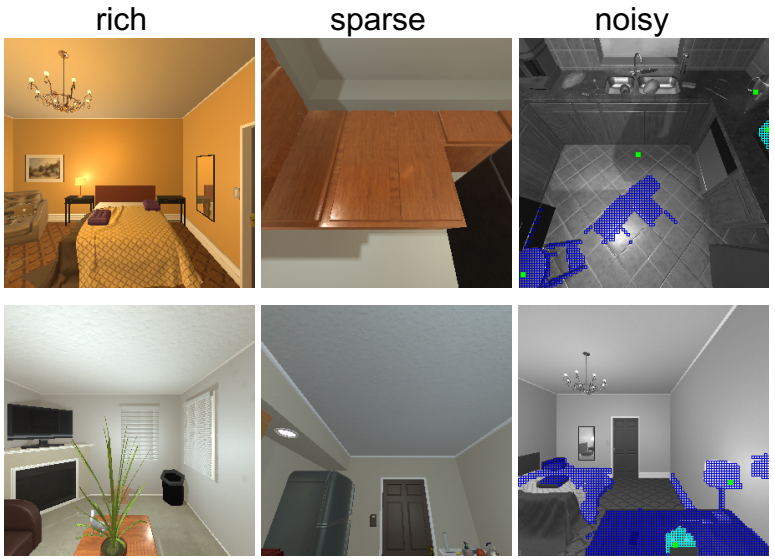
1. **Self-supervision module:** execute model's sampled actions, extract noisy learning targets from before/after RGB images
2. **Clustering-based segmentation model:** learning targets suitable for instance segmentation via clustering from learned pixel embeddings, using specialized loss functions
3. **Memory bank:** for efficient offline learning



Challenges

Self-supervised training from scratch

- Interactive object discovery and learning of properties must happen simultaneously
- Sparsity of interactable objects
- Noisy self-supervised learning targets
- Rich variety of scenes
- Strong generalization requirement (new object/scene types)



Results

- Our model **generalizes** to new scenes and objects types
- Our design choices **outperform** baselines
- Different supervision scenarios and ablations **illustrate** challenges of the task