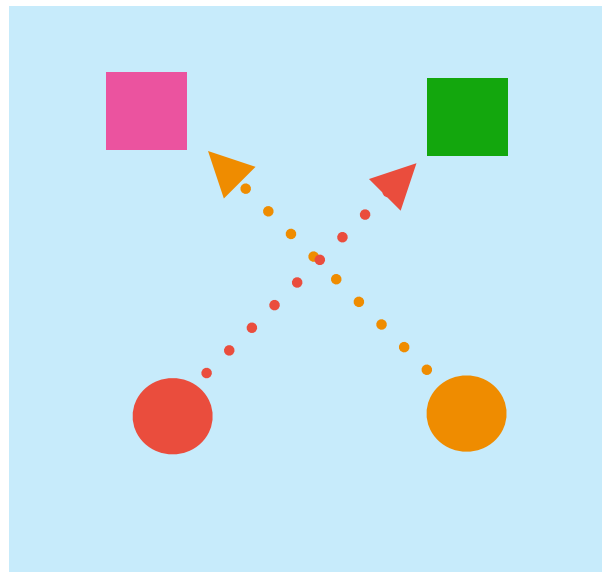
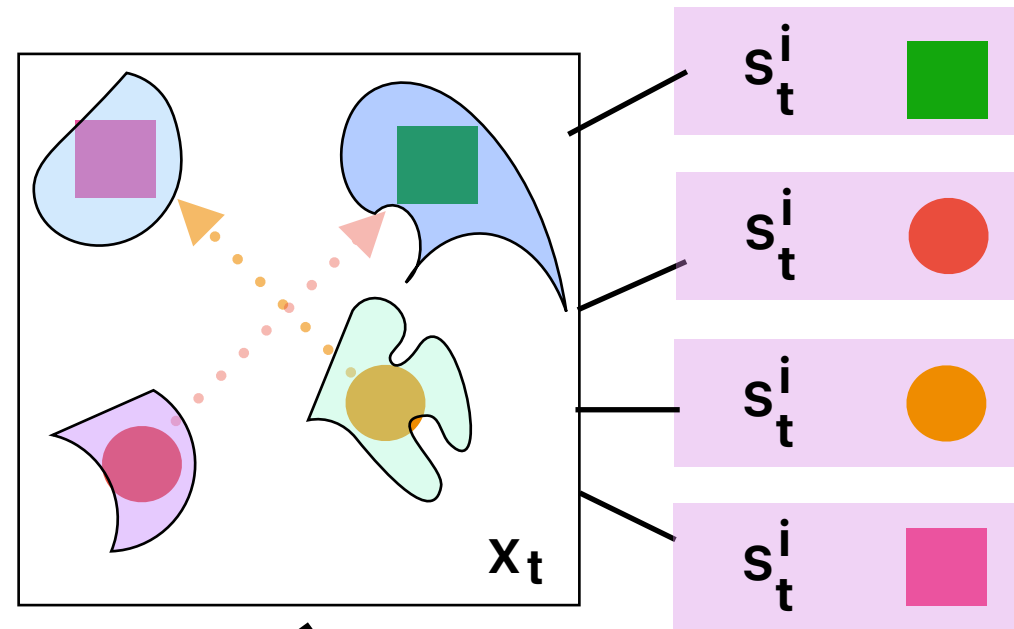


## Multi-Agent Observation



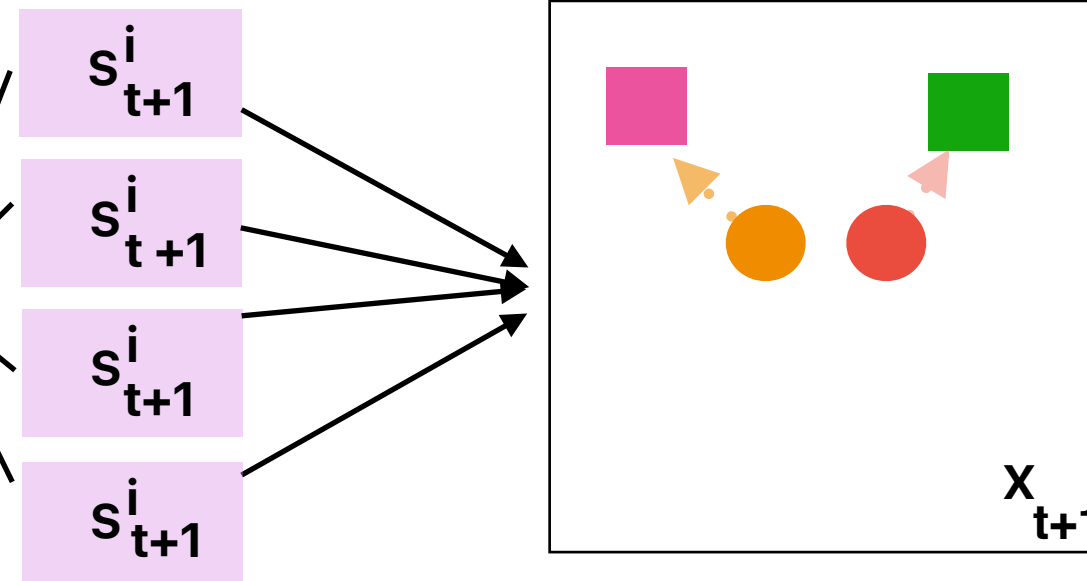
   
Agents

   
Goals



**Visually Segmented  
Attention-based  
Latent representations**

**State Dynamics  
Transformer**



**Next time step  
image decoded  
prediction**