Graph-Structured Visual Imitation

Maximilian Sieb*, Zhou Xian*, Audrey Huang, Oliver Kroemer, Katerina Fragkiadaki

Visual Imitation Learning

Learn policy from visual input so that the robot's actions create the same effect on the environment as the human

Problem Statement

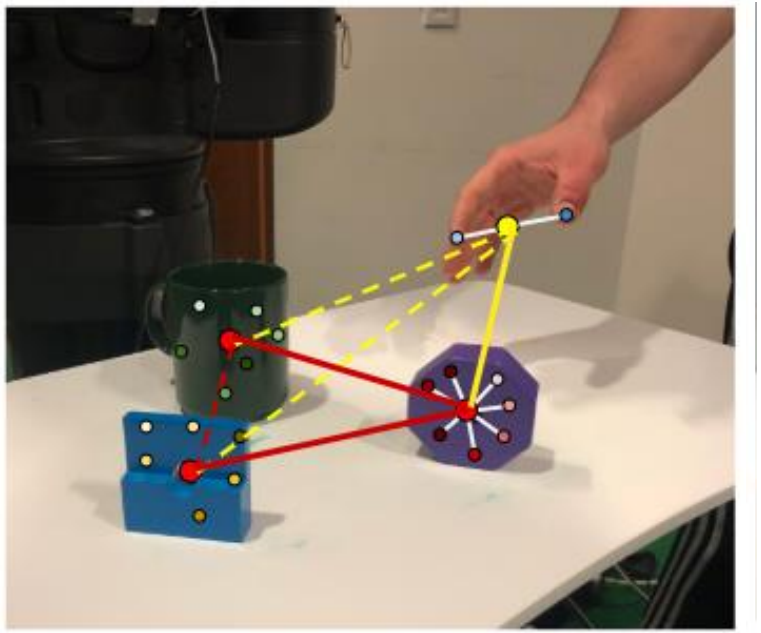
Focus of this work:

Find interpretable state representation that allows

for sample-efficient visual imitation learning from

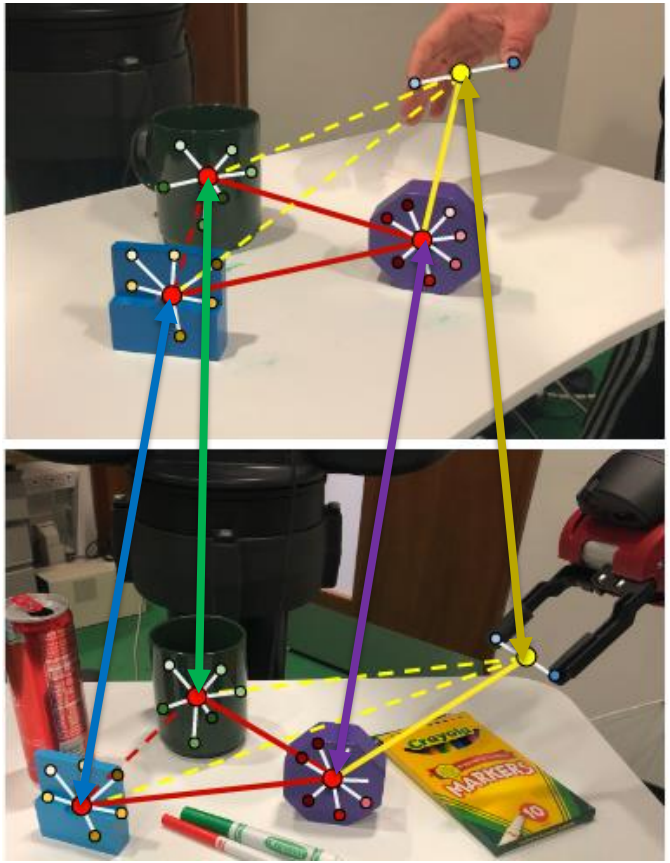
single demonstration

Visual Entity Graphs for Visual Imitation



1. Detect **object** and **hand entities** in image
2. Detect object-level **pixel entities**
3. Establish **geometric relations** between the entities
4. Place **attention** on “important” edges

Correspondence of Visual Entity Graphs



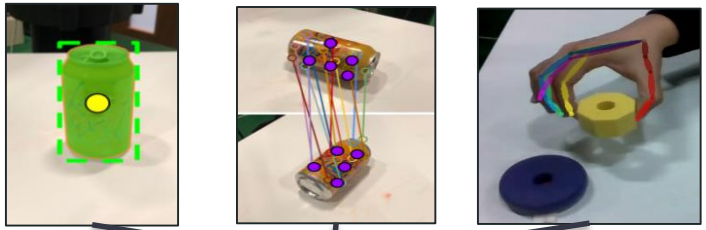
Establish correspondence between demonstration and imitation

From Visual Entity Graphs to Policy Learning

Object entities

Pixel entities

Hand entities



Visual Entity Graph Encoding $\mathcal{G}$

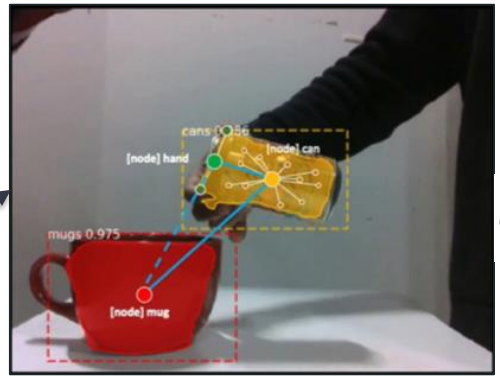
Demonstration

Encode

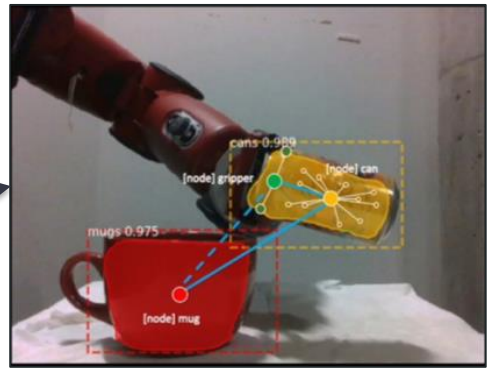
Imitation

Encode

$$\mathcal{C}(\mathcal{G}_D^t, \mathcal{G}_I^t) = \sum_{i,j,i<j} w(\mathcal{E}^{t,(i,j)}) \cdot \mathrm{att}(\mathcal{E}^{t,(i,j)}) \cdot \left\| \left(\mathbf{x}_D^{t,i} - \mathbf{x}_D^{t,j}\right) - \left(\mathbf{x}_I^{t,i} - \mathbf{x}_I^{t,j}\right) \right\|$$

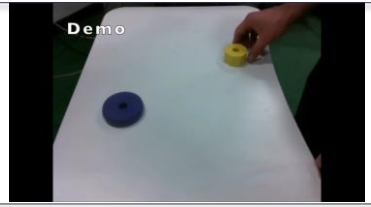
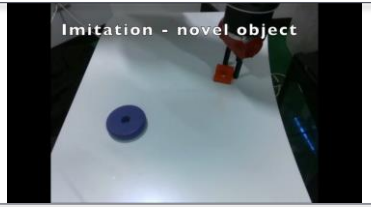
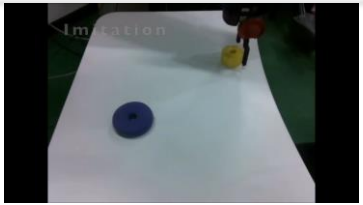
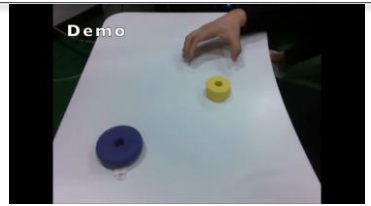


$\mathcal{G}_D^t$



$\mathcal{G}_I^t$

$\mathrm{Cost}\left(\mathcal{G}_D^t, \mathcal{G}_I^t\right)$ → **Policy Learning**

	Demonstration	Imitation	Imitation different object instance
Pushing			
Pushing – Direction Change			
Stacking			
Pouring