



input video

Attribute $\{\mathbf{o_i}, \alpha_s^t\}$

1. {microwave_1: closed}

2. {bottle_1: closed}

Relationship $\{\mathbf{o_i}, \gamma_{\mathbf{n}}^{\mathbf{t}}, \mathbf{o_j}\}$

1. {hand_1, not_holding, cup_1}

2. {hand_2, holding, apple_1}

3. {head_1, contacting, apple_1}

4. {head_1, apart, cup_1}

video graph

Attribute-based $\{A_k^{\alpha}, \mathbf{o_i}, \mathbf{t_s}\}$ 1. {close, microwave_1, #40}

Relationship-based $\{A_k^{\gamma}, \mathbf{o_i}, \mathbf{o_j}, \mathbf{t_s}\}$

2. {pick, hand_1, cup_1, #126}

3. {drink, head_1, cup_1, #150}

4. {place, hand_1, cup_1, #194}

5. {pick, hand_2, apple_1, #225}

6. {eat, head_1, apple_1, #310}

. .

output actions

Training

Testing