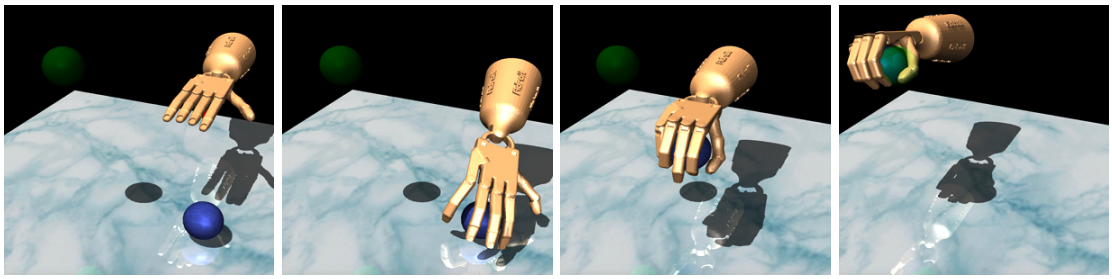
Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations: Supplementary notes

Author Names Omitted for Anonymous Review. Paper-ID 166

1 Task details

1.1 Object relocation



Task: Pick up the blue ball and move it to the green target location.

Randomization: The initial positions of the ball is randomized through the entire workspace of the table. The position of the target is randomized through the entire workspace of the table. Additionally, the target height is also randomized.

State: $s_{relo} = [hand_{joints}; palm_{pos}, object_{pos}; object^{goal}_{pos}]$

Success measure: $\psi_{relo} = \mathbf{I}(||object_{pos} - object^{goal}_{pos}||_{l2} < 0.05)$, where $\mathbf{I}$ is the indicator function

Reward (sparse):

$$r^{sparse}_{relo} = 10\mathbf{I}(||object_{pos} - object^{goal}_{pos}||_{l2} < 0.1) + 20\mathbf{I}(||object_{pos} - object^{goal}_{pos}||_{l2} < 0.05)$$

Note that this reward is quite sparse, and could be applied to real world scenarios with minimal environment augmentation.

Reward (shaped):

$$r^{shaped}_{relo} = r^{sparse}_{relo} - 0.1||palm_{pos} - obj_{pos}||_{l2} + \mathbf{I}(obj_z > 0.04)(1.0 - 0.5||palm_{pos} - obj^{goal}_{pos}||_{l2} - 0.01||obj_{pos} - obj^{goal}_{pos}||_{l2})$$

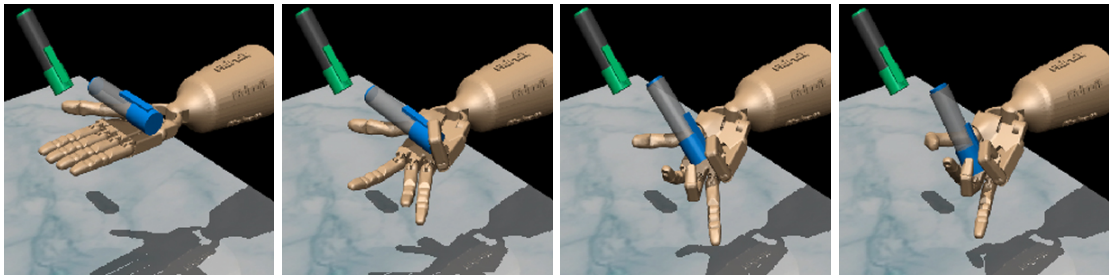
1.2 In-hand Manipulation – Repositioning a pen

Task: Reposition the blue pen to a desired target orientation, visualized by the green pen. The base of the hand is fixed. The pen is highly underactuated and requires careful application of forces by the hand to reposition it.

Randomization: All possible target orientation of the pen.

State: $s_{pen} = [hand_{joints}; pen_{pos,rot}; pen^{goal}_{pos,rot}]$

Success measure: $\psi_{pen} = \mathbf{I}(||pen_{rot} - pen^{goal}_{rot}||_{cosine} > 0.95)$



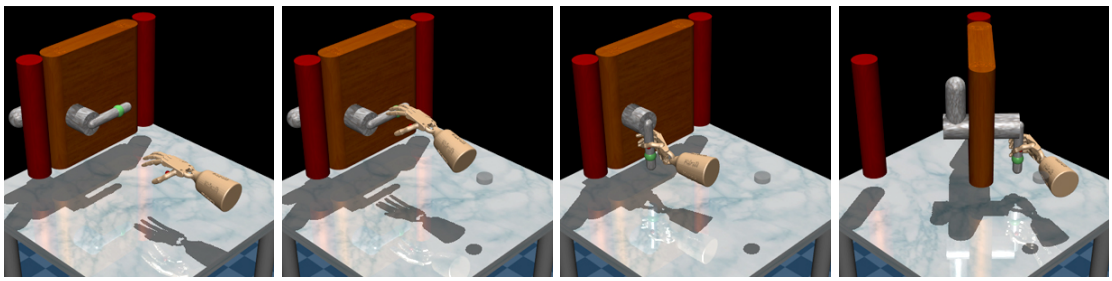
Reward (sparse):

$$r_{pen}^{sparse} = 50(\mathbf{I}(||pen_{pos}^{goal} - pen_{pos}||_{l2} < 0.075) \otimes \mathbf{I}(||pen_{rot} - pen_{rot}^{goal}||_{cosine} > 0.95))$$

Reward (shaped):

$$r_{pen}^{shaped} = r_{pen}^{sparse} - ||pen_{pos}^{goal} - pen_{pos}||_{l2} + ||pen_{rot} - pen_{rot}^{goal}||_{cosine} + 10\mathbf{I}(||pen_{rot} - pen_{rot}^{goal}||_{cosine} > 0.9) - 5\mathbf{I}(pen_z < 0.15)$$

1.3 Manipulating Environmental Props: Door Opening



Task: Swing the doors open. The hand has to undo the latch before the door can be opened. The latch has a significant dry friction and a bias torque that forces the door to be closed.

Randomization: The x, y and z position of the door is randomized.

State: $s_{door} = [hand_{joints}; palm_{pos}; door_{handle\ pos,latch,hinge}]$

Success measure: $\psi_{door} = \mathbf{I}(door_{joint} > 1.4)$

Reward(sparse):

$$r_{door}^{sparse} = 10\mathbf{I}(door_{pos} > 1.35) + 8\mathbf{I}(door_{pos} > 1.0) + 2\mathbf{I}(door_{pos} > 1.2) - 0.1||door_{pos} - 1.57||_{l2}$$

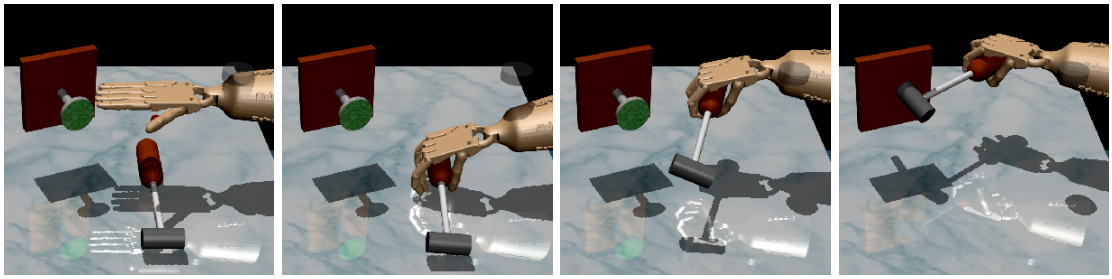
Reward (shaped):

$$r_{door}^{shaped} = r_{door}^{sparse} - ||palm_{pos} - handle_{pos}||_{l2}$$

Note that the reward does not explicitly encode the information that the latch needs to be undone before the door can be opened. The agent needs to develop this understanding through multiple interactions with the environment.

1.4 Tool Use–Hammer

Task: Hammer to drive in a nail into the board. The hand needs to pick up the hammer from the ground, move it over to the nail and hammer in with a significant force to get the nail to move



into the board. The nail has dry friction capable of absorbing up of 15N of force. There are more than one steps needed to perform this task, which require accurate grasping and positioning.

Randomization: The vertical position of the nail is randomized.

State: $s_{nail} = [hand_{joints,velocity}; palm_{pos}; hammer_{pos,rot}; nail^{goal}_{pos}; nail_{impactforce}]$

Success measure: $\psi_{nail} = \mathbf{I}(||nail_{pos} - nail^{goal}_{pos}||_{l2} < 0.01)$

Reward(sparse):

$$r^{sparse}_{nail} = 75 * \mathbf{I}(||nail^{goal}_{pos} - nail_{pos}||_{l2} < 0.10) + 25 * \mathbf{I}(||nail^{goal}_{pos} nail_{pos}||_{l2} < 0.02) - 10||nail^{goal}_{pos} - nail_{pos}||_{l2}$$

Reward(shaped):

$$r^{shaped}_{nail} = r^{sparse}_{nail} - 0.1||palm_{pos} - nail_{pos}||_{l2} - 0.1||hand_{jointvelocity}||_{l2} + 2\mathbf{I}(hammer_z > 0.04)$$

Note that the reward function here depends only on the nail position relative to the final position in the board, and doesn't involve the position of the hammer or the hand.