Robust Policy Design Using a Value Function-Based Disturbance Observer

Edgar Lee(PhD), Tae Min Kim(Undergraduate)

RL + Control Theory(Disturbance Observer)

- Research on enhancing the robustness of reinforcement learning policies using Disturbance Observer, a method from control theory

Controller + DOB (Disturbance Observer)	RL + DOB (Proposed Idea)
Nominal Plant (P_n)	Train Environment
Real Plant (P)	Deploy Environment
Control Input (u)	Policy action (a_t)
Plant Output (y)	Observation, reward, action

Fig. 1 Comparison of conventional DOB and proposed value function based DOB

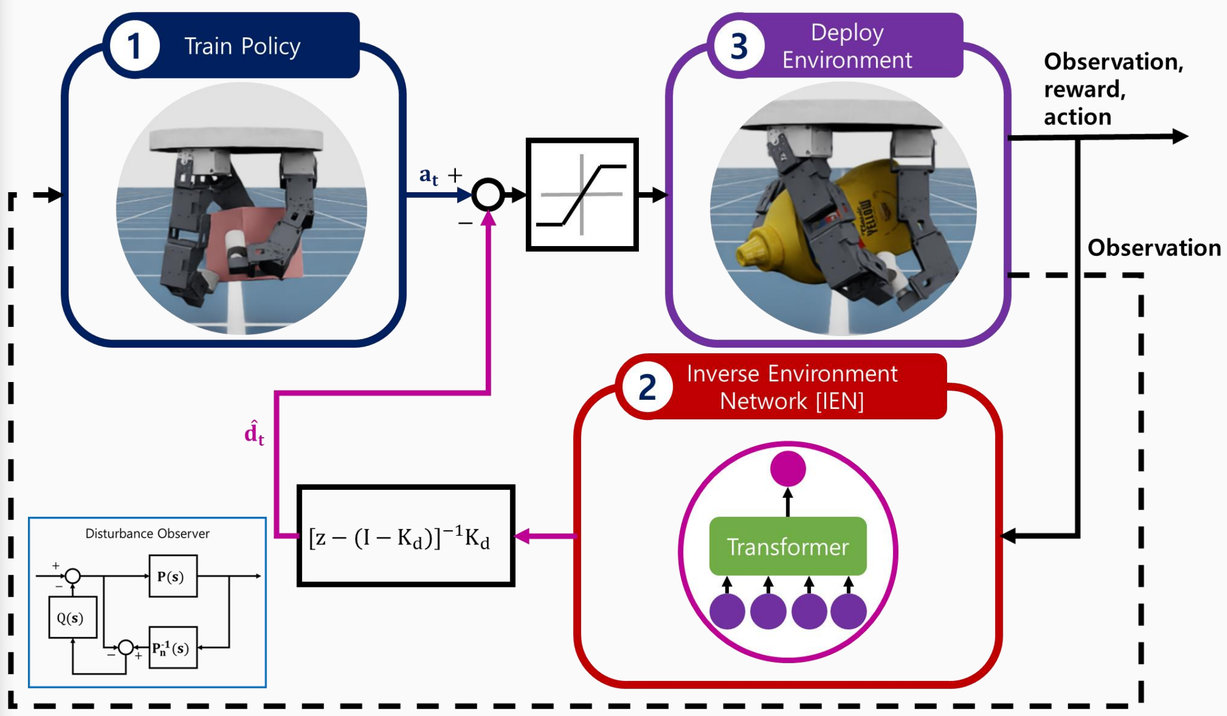


Fig. 2 Diagram of proposed value function-based DOB

Training IEN for Robust Policy

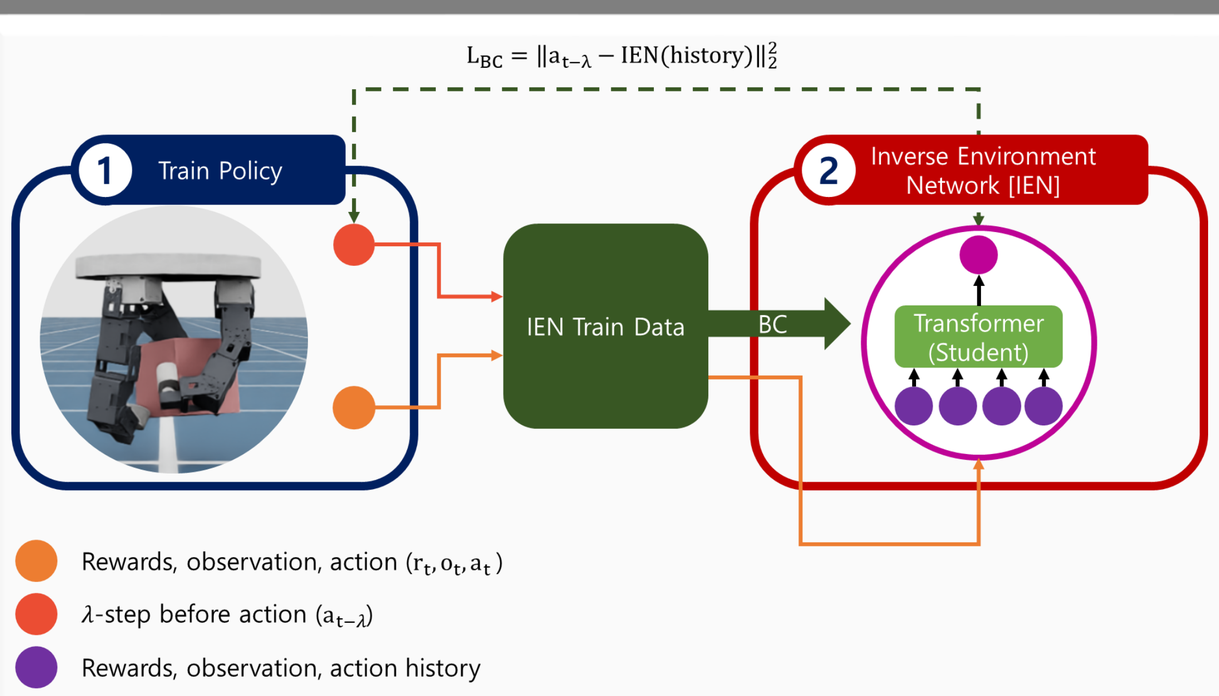


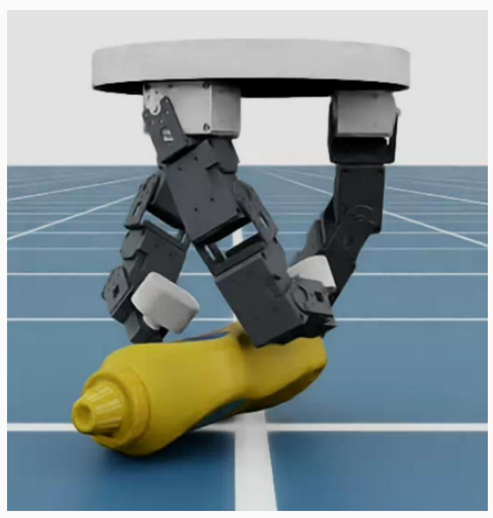
Fig. 3 Diagram of training Inverse Environment Network(IEN)

- The proposed DOB is trained via a neural network
- The neural network uses a transformer decoder
- It takes in rewards, observations, and actions, and predicts nominal actions.

Simulation Test

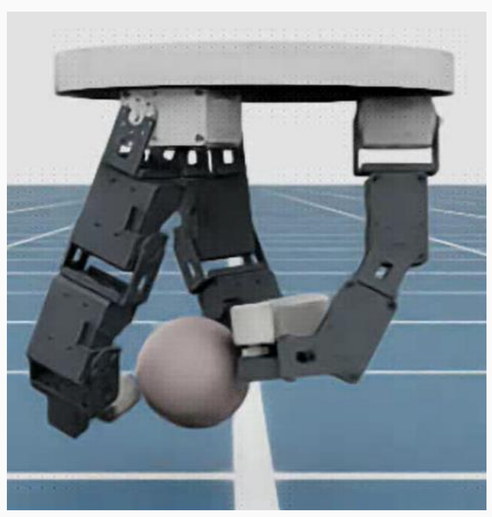
- YCB Object – Mustard Bottle

With DOB: success

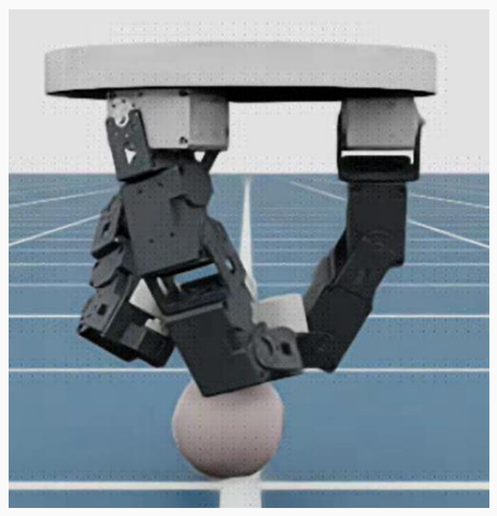


Without DOB: fail

- Other Shapes - Sphere



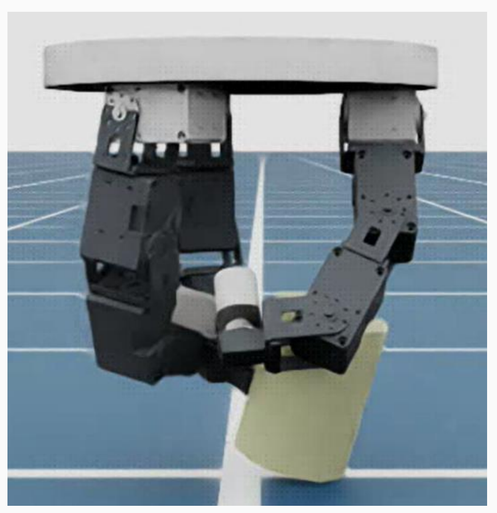
With DOB: success



Without DOB: fail

- Other Shapes - Cylinder

With DOB: success



Without DOB: fail

Conclusion

- By training on a single cuboid, a variety of objects were successfully grasped
- Although extreme disturbances(such as extremely large on small sizes or very heavy masses) were not fully compensated, disturbances within a certain range were effectively mitigated
- This approach is expected to be extended to a broader range of tasks in the future