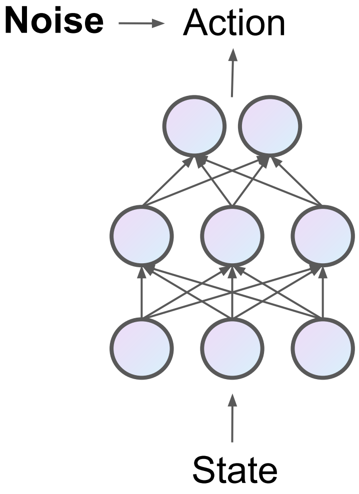
Deep Coherent Exploration for Continuous Control

Yijie Zhang[1] & Herke van Hoof[2]

[1]University of Copenhagen [2]University of Amsterdam

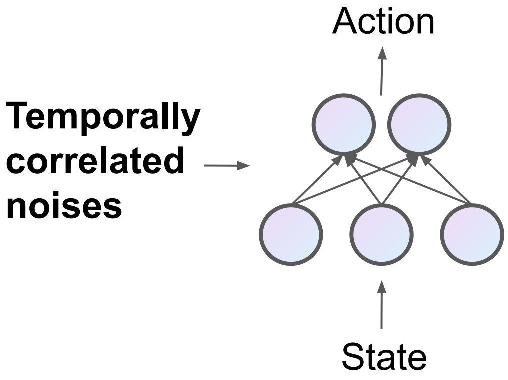
Undirected Exploration for (Deep) RL

Action space exploration
Step-based
(Sutton, 1995; Williams, 1992)



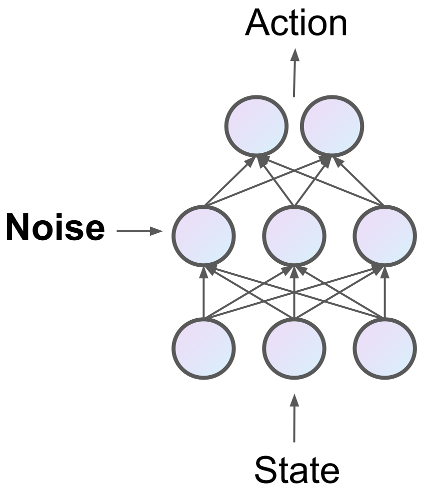
- Straightforward and easy to understand
- High-frequency perturbations can be unstable

Generalized exploration (GE)
Intermediate
(van Hoof et al., 2017)



- Better balance between stability and stochasticity
- Unscalable for complex models and long trajectories

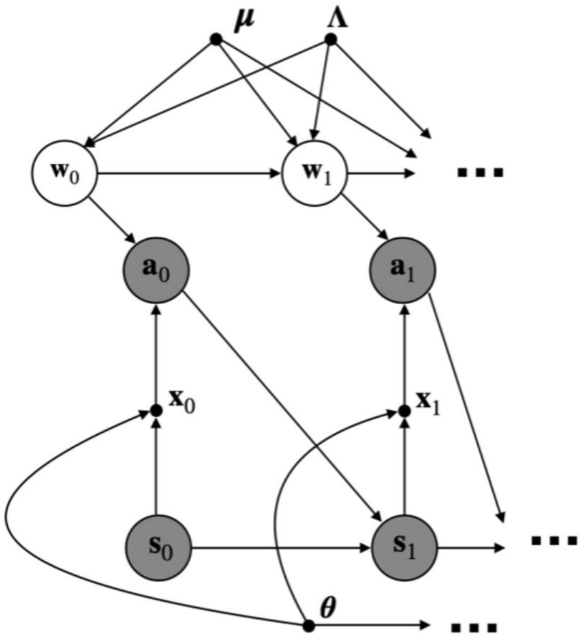
Parameter space exploration
Trajectory-based
(Ruckstieß et al., 2010; Sehnke et al., 2010; Fortunato et al., 2018; Plappert et al., 2018)



- Consistent, structure, and global exploration behaviors
- Inefficient evaluation and insufficient stochasticity

Can we make GE more scalable for deep RL?

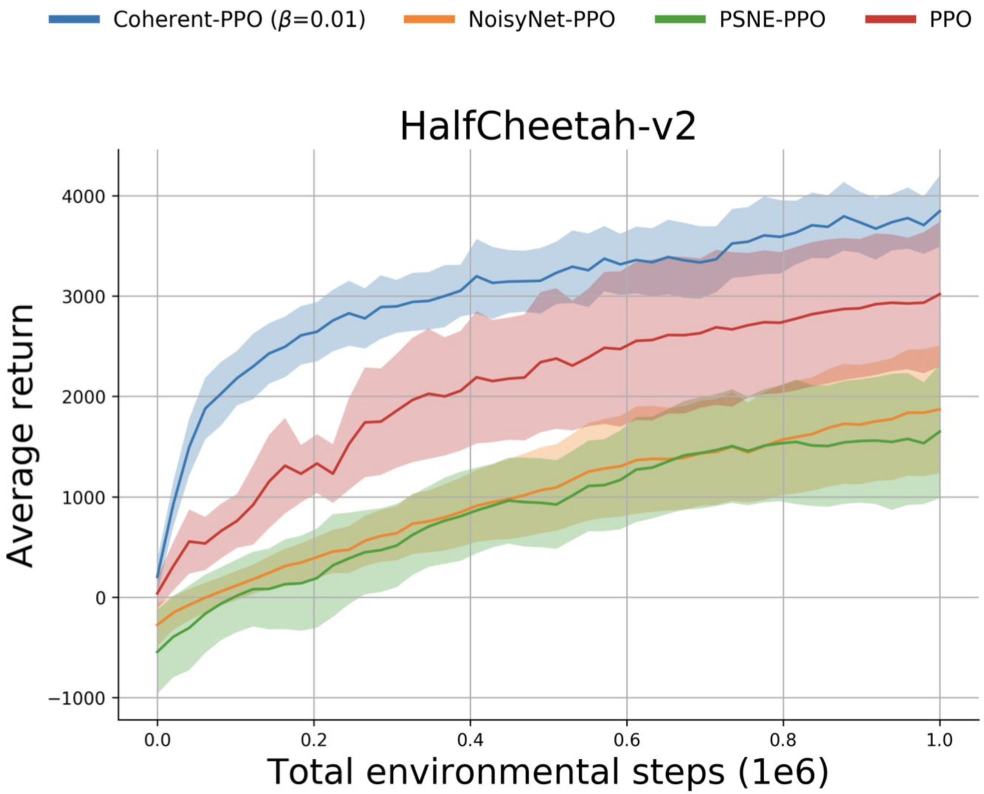
Deep Coherent Exploration



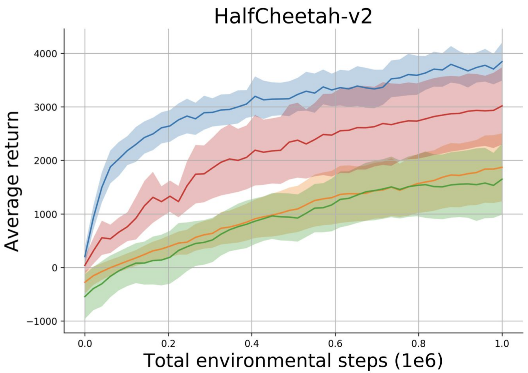
Characteristics of deep coherent exploration

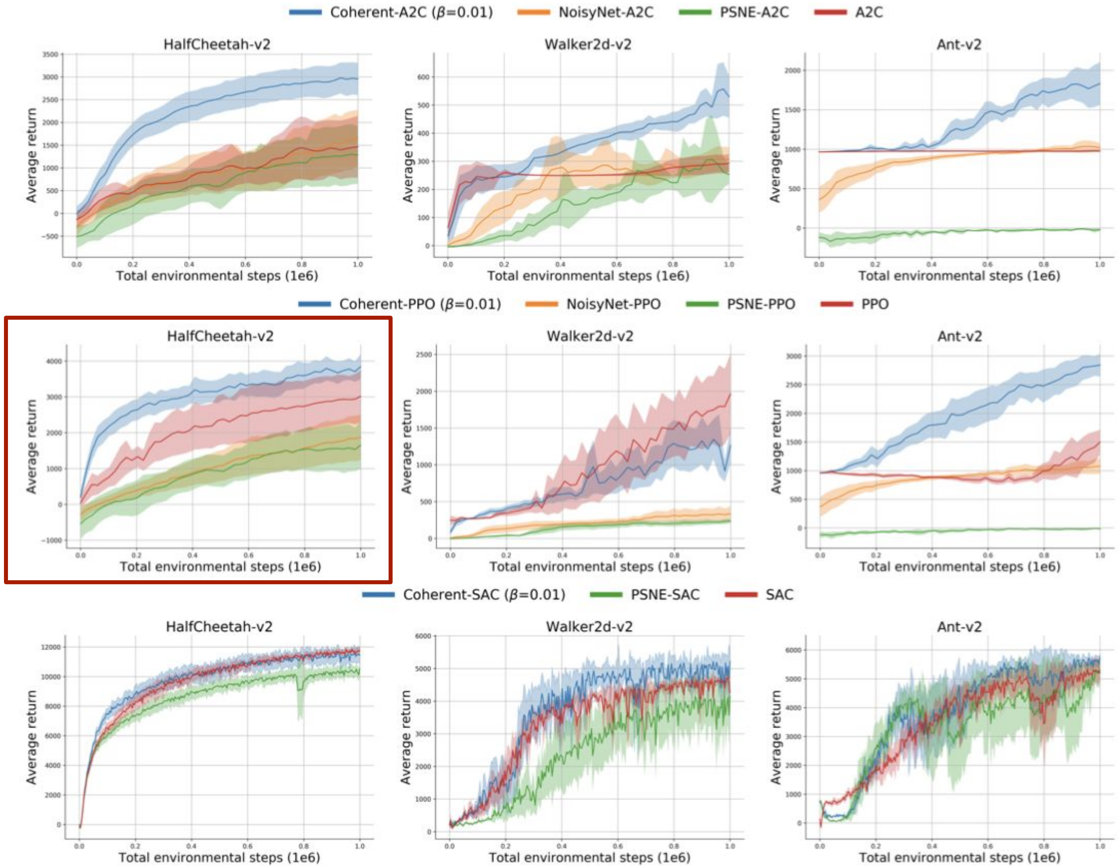
1. Generalizing step-based and trajectory-based exploration (following GE)
    - Balanced trade-off between stability and stochasticity
2. Recursive exact integration of latent exploring policies
    - Scalable policy updates compared to GE
    - Lower-variance gradient estimates compared to reparameterization trick
3. Perturbing only last layers of policy networks
    - Controllable noise injection

Experiments: Comparative Evaluation

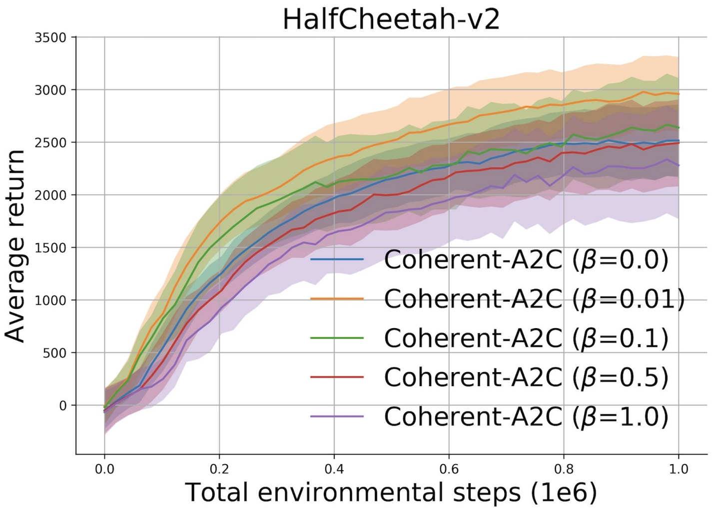


Experiments: Comparative Evaluation



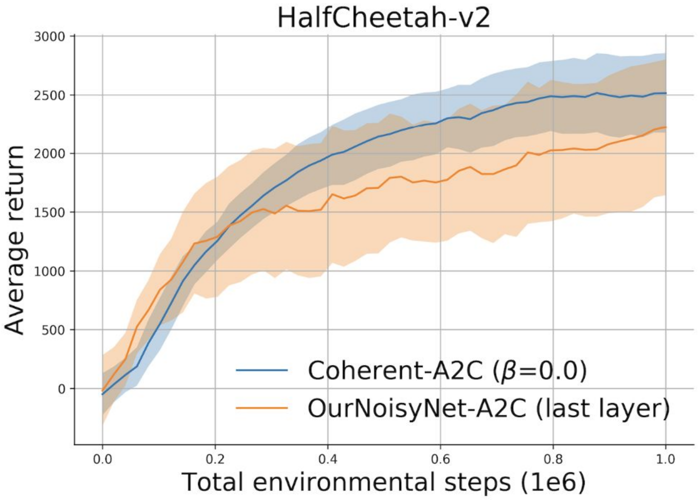


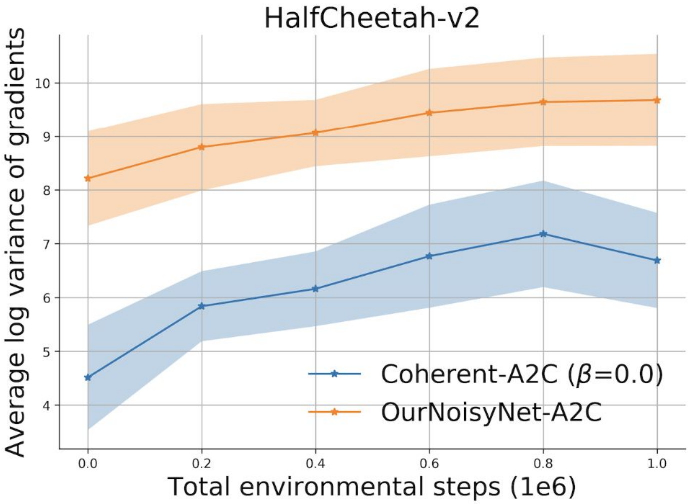
Ablation Study #1: Generalizing Step-Based and Trajectory-Based Exploration



A more balanced trade-off between stability and stochasticity brings more efficient exploration and learning.

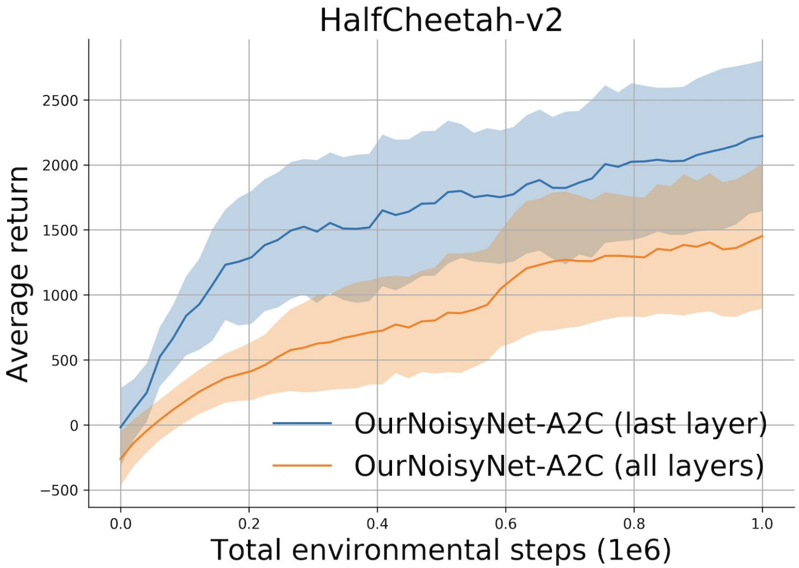
Ablation Study #2: Recursive Analytical Integration of Latent Exploring Policies





Recursive analytical integration leads to more scalable and stable policy updates.

Ablation Study #3: Perturbing Only Last Layers of Policy Networks



Controllable noise injection is beneficial for faster learning.

Conclusion

- General and easy to implement for both on- and off-policy deep RL algorithms
- Generalizing step-based and trajectory-based exploration allows more delicate trade-off between stability and stochasticity
- Recursive analytical integration enables more scalable, stable updates and faster learning
- Perturbing only the last layers of policy networks brings controllable noise injection

Thank you!