Continuous Neural Algorithmic Planners

Yu He, Petar Veličković, Pietro Lió, Andreea Deac

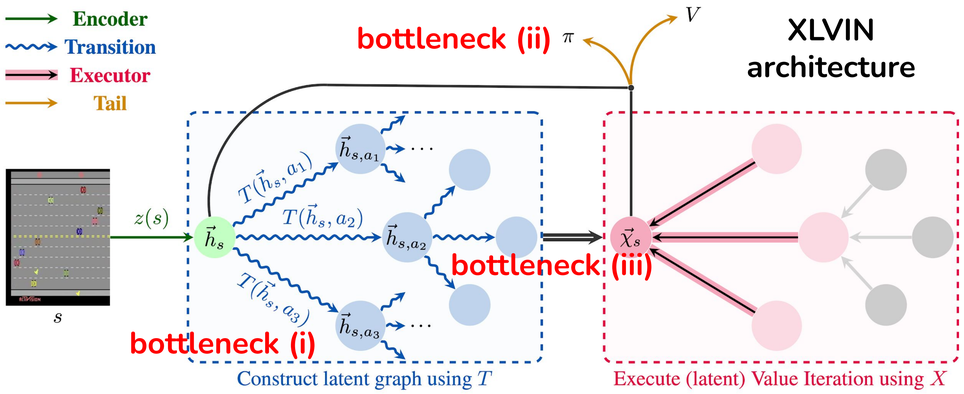
Summary: We propose **CNAP**, an implicit planner to perform in ***high-dimensional continuous control problems*** by simulating value iteration with **neural algorithmic reasoning**. CNAP inherits **XLVIN**'s low data merits while addressing its limitations.

Neural algorithmic Planners

eXecuted Latent Value Iteration Nets (XLVINs):

- **neural algorithmic reasoning**: pre-train a GNN Executor to simulate value iteration behaviour in the latent space
- **implicit planning**: no explicit modelling of the MDPs required
- **low data regime**: overcome algorithmic bottleneck suffered by other VI-based implicit planners (e.g. ATreeC)

The CNAP architecture



(i) Deal with continuous action space:

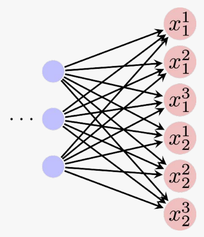
- **Discretisation by binning:** evenly spaced discrete action bins
- However, discretisation results in **a combinatorial increase** in action space

(ii) Deal with multidimensional action space:

- **Factorised joint policy:** linear increase with dimensionality $\pi(\vec{a}|s) := \prod_{i=1}^{D} (\pi_{a_i}|s)$

(iii) Deal with complex action space:

- **Neighbour sampling policy:** selectively expand actions
  - **Manual-Gaussian**: Gaussian(mean=bins/2, std=bins/4)
  - **Learned-Gaussian**: Gaussian with learnt parameters
  - **Reuse-Policy**: reuse policy distribution
  - **Learned-Sampling**: learnt sampling distribution



Results in Classic Control

(i) Best performance:

Model	MountainCarContinuous-v0
PPO Baseline	-4.96 ± 1.24
CNAP-B	55.73 ± 45.10
CNAP-R	**63.41** ± 37.89

(ii) Varying GNN depth:

Model	GNN Steps	MountainCar-Continuous
CNAP-B	1	55.73 ± 45.10
	2	46.93 ± 44.13
	3	40.58 ± 48.20
CNAP-R	1	**63.41** ± 37.89
	2	34.49 ± 47.77
	3	43.61 ± 46.16

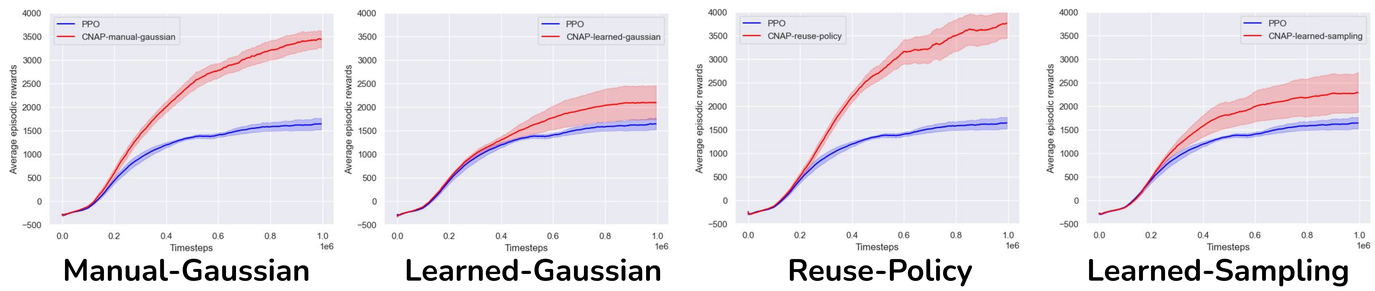
(iii) Varying GNN width:

Model	Action Bins	MountainCar-Continuous
PPO	5	-2.16 ± 1.25
	10	-4.96 ± 1.24
	15	-3.95 ± 0.77
CNAP-B	5	29.46 ± 57.57
	10	55.73 ± 45.10
	15	22.79 ± 41.24
CNAP-R	5	20.32 ± 53.13
	10	**63.41** ± 37.89
	15	26.21 ± 46.44

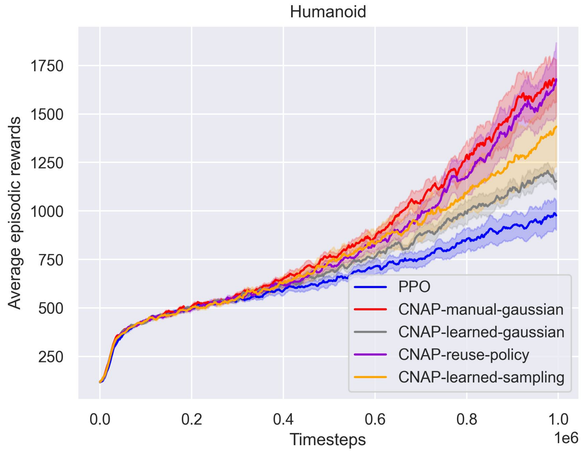
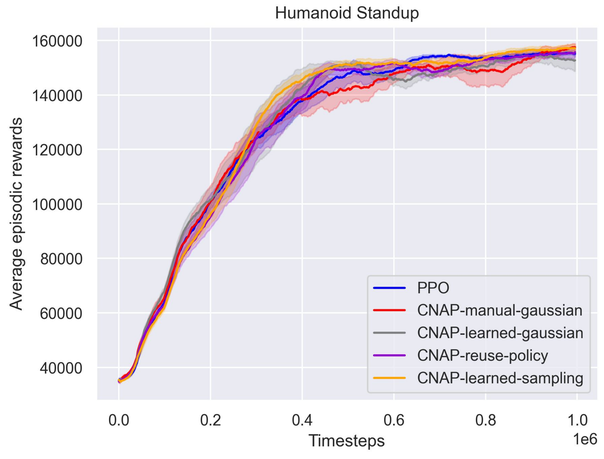
PPO Baseline: Encoder + Tail
CNAP-B: pre-trained GNN with binary graphs
CNAP-R: pre-trained GNN with Erdős–Rényi graphs

Results in MuJoCo

(i) Halfcheetah:



(ii) Humanoid:



(ii) Humanoid Standup: