

# Contrastive Inductive Bias Controlling Networks for Reinforcement Learning

**Dongxu Li**  
**Shaochen Wang**  
**Kang Chen**  
**Bin Li**

*University of Science and Technology of China*

LIDONGXU6C@MAIL.USTC.EDU.CN

SAMWANG@MAIL.USTC.EDU.CN

CK6@MAIL.USTC.EDU.CN

BINLI@MAIL.USTC.CN

## Appendix A. Hyperparameters

The full list of hyper-parameters is presented in Table 1.

Parameters	Settings
Replay buffer capacity	$10^6$
Action repeat	2
Initial steps	1000
Stacked frames	3
Encoder learning rate	$1e^{-4}$
Actor learning rate	$1e^{-3}$
Critic learning rate	$1e^{-3}$
Batch size	128
Discount $\gamma$	0.99
Optimizer	Adam
Transfer update interval	100
Encoder update intervals	100

Table 1: Default set of hyper-parameters used in our experiments

## Appendix B. PlaNet benchmark

Scores achieved by Our methods (mean standard deviation for 10 seeds) and baselines on DMControl500k and 1DMControl100k. Our method achieves state-of-the-art performance on the majority (8 out of 12) environments benchmarked on DMControl500k. The baselines are DrQ, CURL, PlaNet, Dreamer, SAC+AE, SLAC, and state-based SAC.

500k step scores	DrQ	CURL	PlaNet	SAC-AE	SLAC	Ours	SAC State
Finger Spin	<b>938±103</b>	874±151	718±40	914±107	771±203	638±42	927±43
Cartpole Swingup	868±10	861±30	787±46	730±152	-	<b>868±5</b>	870±7
Reacher Easy	942±71	904±94	588±471	601±135	-	<b>965±30</b>	975±5
Cheetah Run	<b>660±96</b>	500±91	568±21	544±50	629±74	535±98	772±60
Walker Walk	<b>921±45</b>	906±56	478±164	858±82	865±97	853±48	964±8
Ball In Cup Catch	963±9	958±13	939±43	810±121	959±4	<b>970±7</b>	979±6
100k step scores							
Finger Spin	<b>901±104</b>	779±108	560±77	747±130	680±130	582±102	672±76
Cartpole Swingup	759±92	592±170	563±73	276±38	-	<b>838±53</b>	812±45
Reacher Easy	601±213	517±113	82±174	225±164	-	<b>625±230</b>	919±123
Cheetah Run	344±67	307±48	165±123	252±173	<b>391±47</b>	170±98	228±95
Walker Walk	402±164	344±132	221±43	395±58	428±74	<b>605±121</b>	604±317
Ball In Cup Catch	780±53	772±241	710±217	338±196	607±173	<b>888±58</b>	957±26

Figure 1: The PlaNet benchmark