Supplementary Material: Self-Imitation Learning

Junhyuk Oh [*1] **Yijie Guo** [*1] **Satinder Singh** [1] **Honglak Lee** [2 1]

A. Hyperparameters

Table 1. A2C+SIL hyperparameters on Atari games.

Hyperparameters	Value
Architecture	Conv(32-8x8-4) -Conv(64-4x4-2) -Conv(64-3x3-1) -FC(512)
Learning rate	0.0007
Number of environments	16
Number of steps per iteration	5
Entropy regularization (α)	0.01
SIL update per iteration (M)	4
SIL batch size	512
SIL loss weight	1
SIL value loss weight ($\beta^s il$)	0.01
Replay buffer size	10^5
Exponent for prioritization	0.6
Bias correction for prioritized replay	0.1 for hard exploration experiment (Section 5.3) 0.4 for overall evaluation (Section 5.4)

Table 2. PPO+SIL hyperparameters on MuJoCo.

Hyperparameters	Value
Architecture	FC(64)-FC(64)
Learning rate	Best chosen from $\{$0.0003, 0.0001, 0.00005, 0.00003$\}$
Horizon	2048
Number of epochs	10
Minibatch size	64
Discount factor (γ)	0.99
GAE parameter (λ)	0.95
Entropy regularization (α)	0
SIL update per batch	10
SIL batch size	512
SIL loss weight	0.1
SIL value loss weight (β)	Best chosen from $\{$0.01, 0.05$\}$
Replay buffer size	50000
Exponent for prioritization	Best chosen from $\{$0.6, 1.0$\}$
Bias correction for prioritized replay	0.1

B. Performance on Atari Games

Table 3. Performances on 49 Atari games with 30 random no-op after 50M steps of training (200M frames).

	A2C	ACPER	A2C+SIL
Alien	1859.2	390.2	**2242.2**
Amidar	739.9	424.8	**1362.0**
Assault	**1981.4**	818.2	1812.0
Asterix	16083.3	3533.1	**17984.2**
Asteroids	2056.0	1780.1	**2259.4**
Atlantis	3032444.2	58012.5	**3084781.7**
BankHeist	**1333.7**	1203.2	1137.8
BattleZone	10683.3	15025.0	**25075.0**
BeamRider	**3931.7**	2602.4	2366.2
Bowling	31.2	**59.3**	31.1
Boxing	99.7	**100.0**	99.6
Breakout	**501.6**	118.5	452.0
Centipede	3857.8	**7790.1**	7559.5
ChopperCommand	3464.2	1307.5	**6710.0**
CrazyClimber	129715.8	19918.8	**130185.8**
DemonAttack	**18331.4**	4777.5	10140.5
DoubleDunk	-0.5	-9.8	**21.5**
Enduro	0.0	**3113.3**	1205.1
FishingDerby	39.1	**59.8**	55.8
Freeway	0.0	31.4	**32.2**
Frostbite	339.5	2342.5	**6289.8**
Gopher	9358.5	3919.5	**23304.2**
Gravitar	329.2	627.5	**1874.2**
Hero	28008.1	13299.1	**33156.7**
IceHockey	-4.3	**0.0**	-2.4
Jamesbond	399.2	**598.1**	310.8
Kangaroo	1563.3	**5875.0**	2888.3
Krull	8883.9	**11323.2**	10614.6
KungFuMaster	32507.5	20485.0	**34449.2**
MontezumaRevenge	5.8	0.0	**1100.0**
MsPacman	2843.4	1016.0	**4025.1**
NameThisGame	11174.2	2888.0	**14958.2**
Pong	20.8	20.9	**20.9**
PrivateEye	210.8	100.0	**661.2**
Qbert	17605.2	657.2	**104975.6**
Riverraid	13036.0	2224.5	**14306.1**
RoadRunner	39874.2	8925.0	**57071.7**
Robotank	3.2	7.7	**10.5**
Seaquest	1795.2	804.5	**2456.5**
SpaceInvaders	2466.1	729.5	**2951.7**
StarGunner	**57371.7**	1107.5	31309.2
Tennis	**-10.3**	-17.0	-17.3
TimePilot	5346.7	3952.5	**10811.7**
Tutankham	305.6	270.7	**340.5**
UpNDown	48131.8	9562.5	**53314.6**
Venture	**0.0**	**0.0**	**0.0**
VideoPinball	391241.6	21797.7	**461522.4**
WizardOfWor	4196.7	1550.0	**7088.3**
Zaxxon	124.2	4278.8	**9164.2**

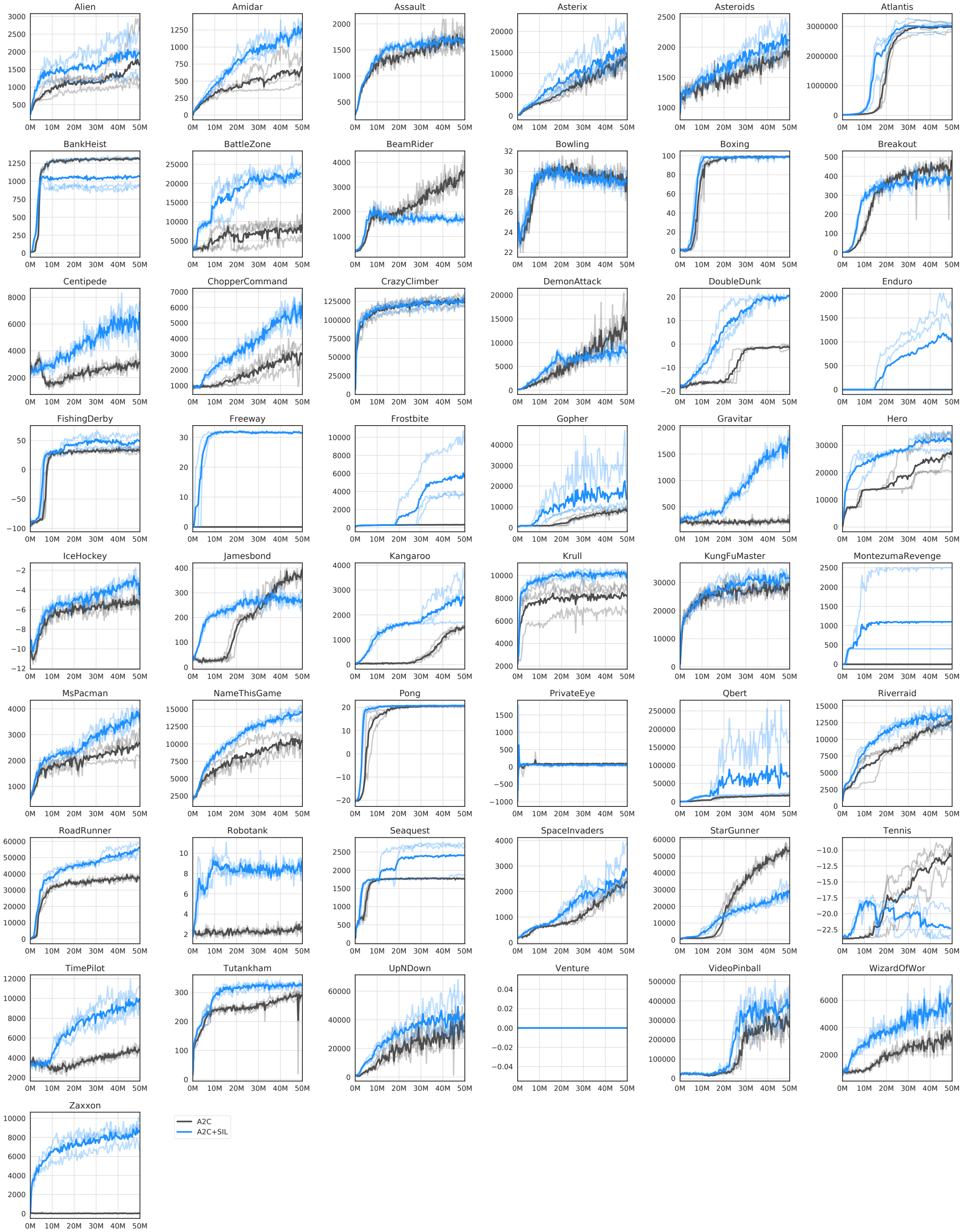

Figure 1. Learning curves on 49 Atari games.