Supplementary Material

A. Experimental Setup

A.1. A3C

For our Breakout experiments we use the standard high-performance architecture implemented in (Kostrikov, 2018a).

Table 3. A3C hyperparameters

Hyperparameter	Value
architecture	LSTM-A3C
state size	$1 \times 80 \times 80$
# actor learners	32
discount rate	0.99
Adam learning rate	0.0001
step-returns	20
entropy regularization weight	0.01

A.2. A2C

We use the implementation in (Kostrikov, 2018b) for comparison and as a skeleton for our method implementation.

Table 4. A2C hyperparameters

Hyperparameter	Value
architecture	FF-A2C
state size	$4 \times 84 \times 84$
# actor learners	84
discount rate	0.99
RMSprop learning rate	0.0007
step-returns	20
entropy regularization weight	0.01

A.3. A2C with Imitation Learning

Table 5. A2C with Imitation Learning algorithm hyperparameters

Hyperparameter	Value
trajectories	5
β_1	0.75
β_2	0.6
Supervised_Iterations	500
SGD learning rate	0.0007
SGD momentum	0.9
b	4
op_interval	100

B. Fine-tuning Settings

We consider the following settings for our Fine-tuning experiments on Breakout:

- From-Scratch: The game is being trained from scratch on the target game.
- Full-FT: All of the layers are initialized with the weights of the source task and are fine-tuned on the target task.
- Random-Output: The convolutional layers and the LSTM layer are initialized with the weights of the source task and are fine-tuned on the target task. The output layers are initialized randomly.
- Partial-FT: All of the layers are initialized with the weights of the source task. The three first convolutional layers are kept frozen, and the rest are fine-tuned on the target task.
- Partial-Random-FT: The three first convolutional layers are initialized with the weights of the source task and are kept frozen, and the rest are initialized randomly.

C. GAN Comparison Evaluation

Table 6. The scores accumulated by an Actor-Critic RL agent using UNIT and Cycle-GAN. We examine both methods by running the RL agent with each every 1000 GAN training iterations and considering the maximum score after $500k$ iterations.

Method	**UNIT**		**CycleGAN**	
	Frames	Score	Frames	Score
A Constant Rectangle	333K	399	358K	26
A Moving Square	384K	300	338K	360
Green Lines	378K	314	172K	273
Diagonals	380K	338	239K	253
Road Fighter - Level 2	274K	5750	51K	6000
Road Fighter - Level 3	450K	5350	20K	3200
Road Fighter - Level 4	176K	2300	102K	2700