A Appendix

A.1 Learning Curves

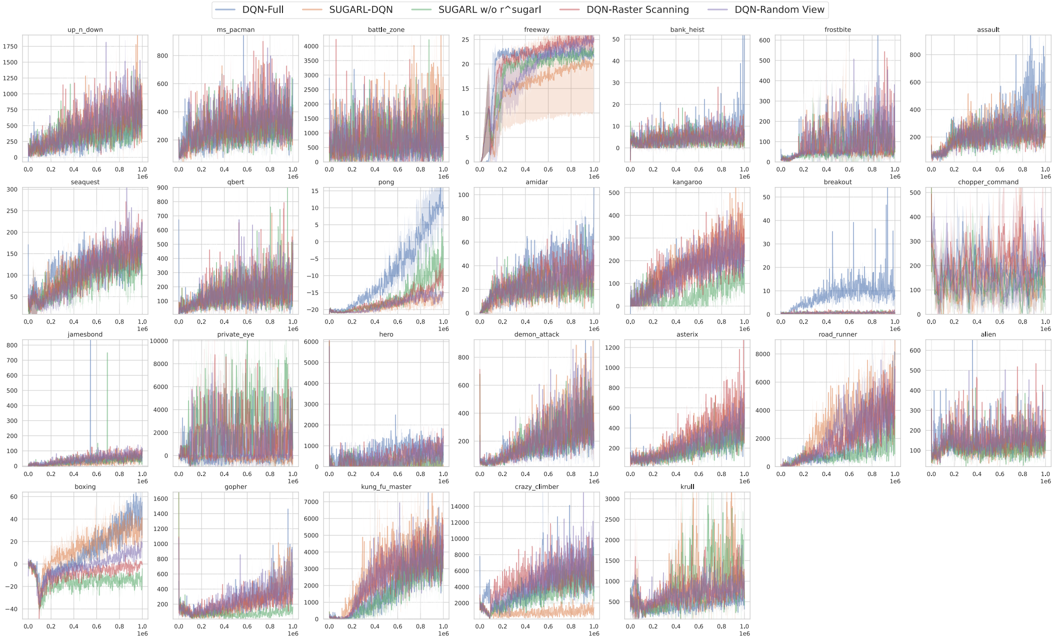

Figure 8: Learning curves of 26 Atari games, under the setting of 50x50 foveal observation size and 20x20 peripheral observation.

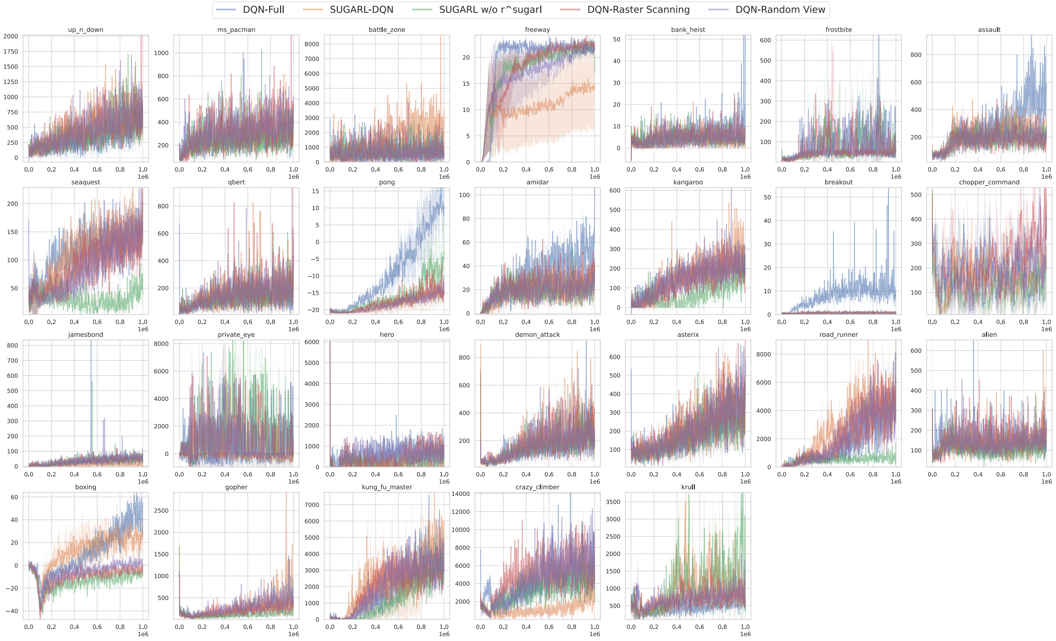

Figure 9: Learning curves of 26 Atari games, under the setting of 30x30 foveal observation size and 20x20 peripheral observation.

A.2 Hyper-parameter Settings

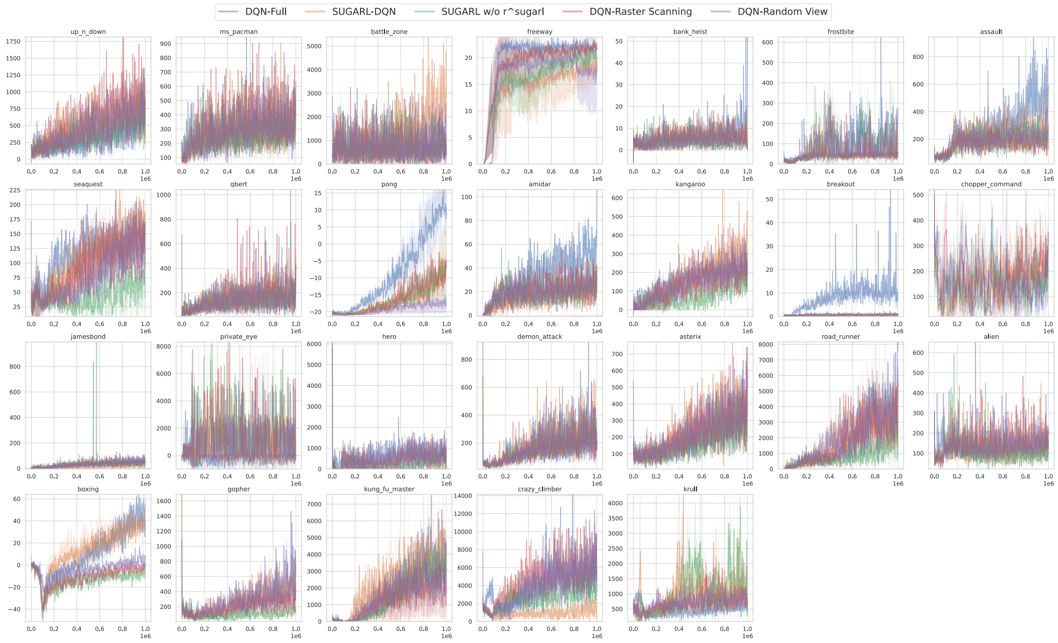

Figure 10: Learning curves of 26 Atari games, under the setting of 20x20 foveal observation size and 20x20 peripheral observation.

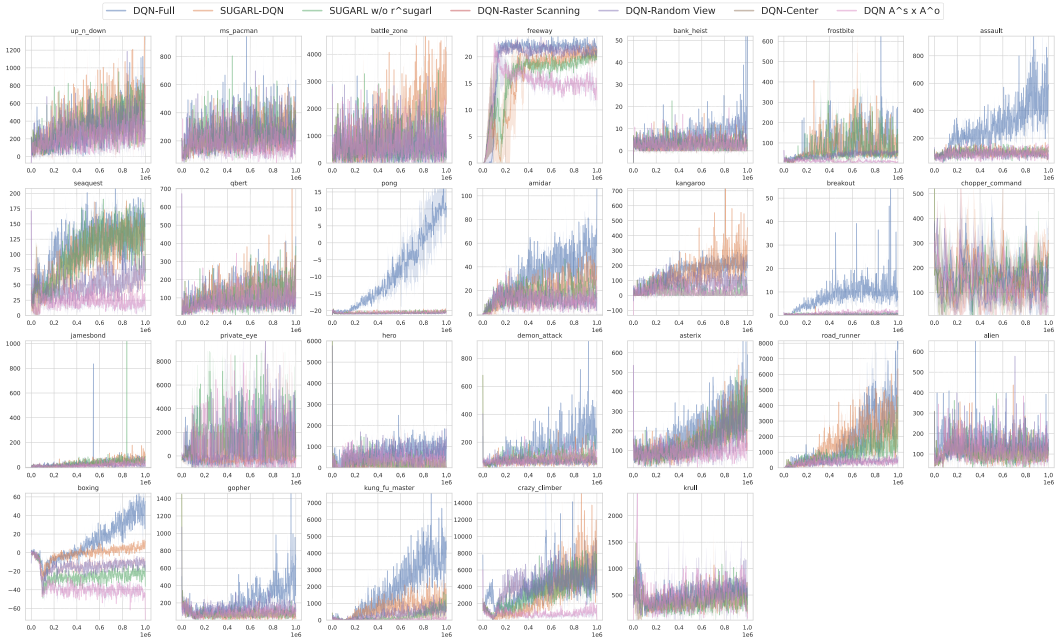

Figure 11: Learning curves of 26 Atari games, under the setting of 50x50 foveal observation size and w/o peripheral observation.

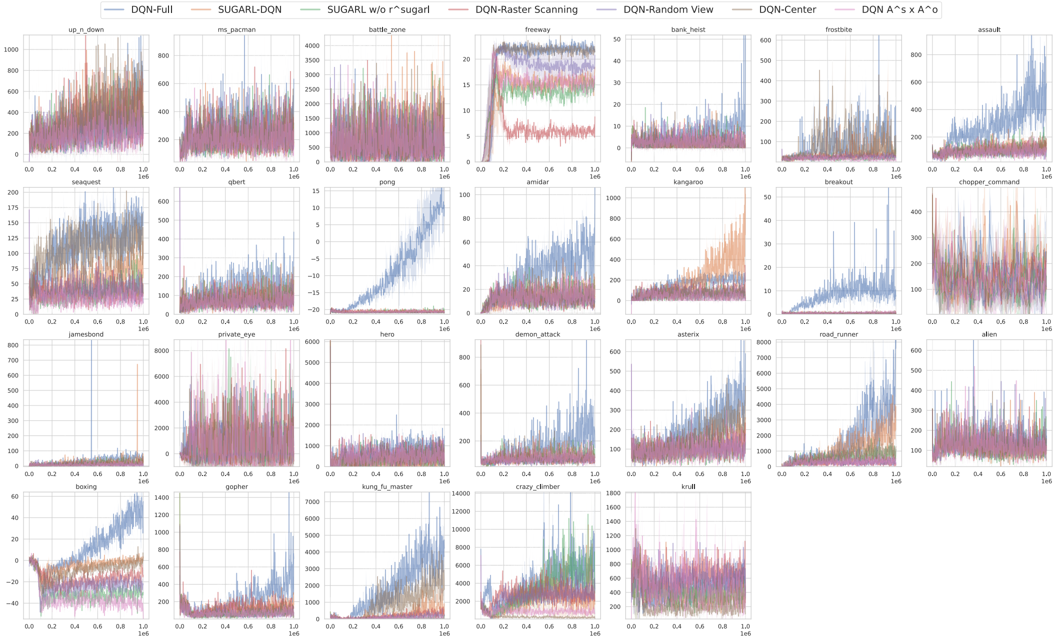

Figure 12: Learning curves of 26 Atari games, under the setting of 30x30 foveal observation size and w/o peripheral observation.

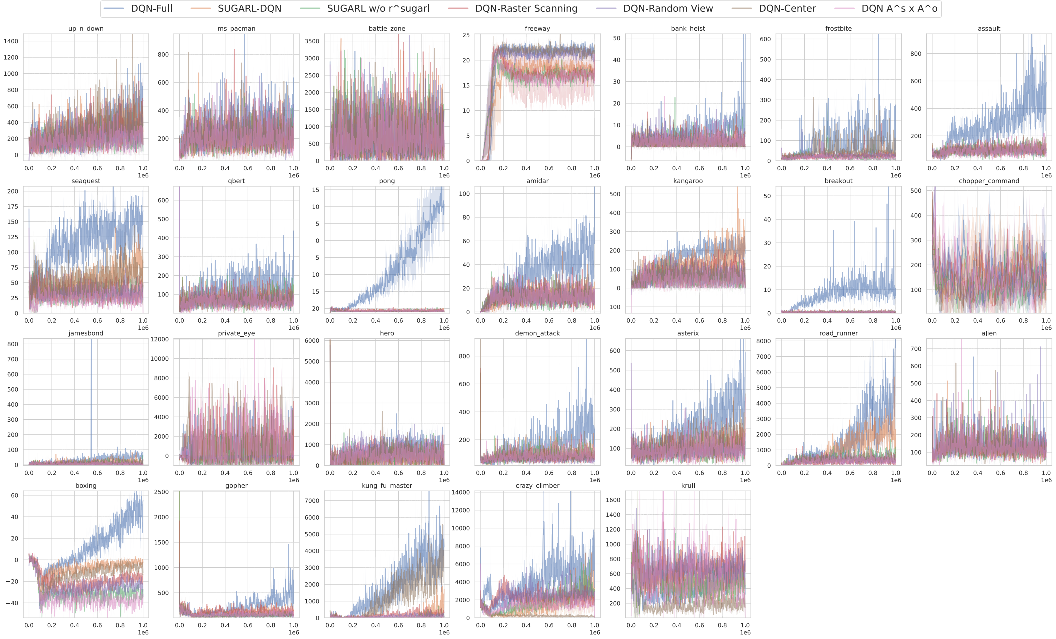

Figure 13: Learning curves of 26 Atari games, under the setting of 20x20 foveal observation size and w/o peripheral observation.

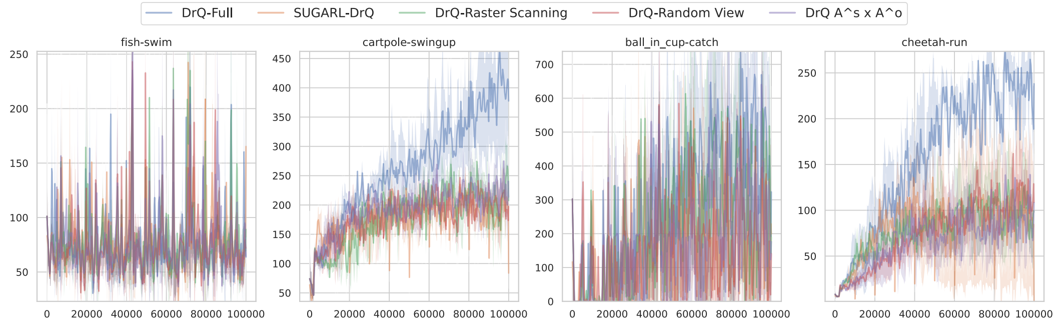

Figure 14: Learning curves of 4 DMC environments, under the setting of 50x50 foveal observation size and w/o peripheral observation.

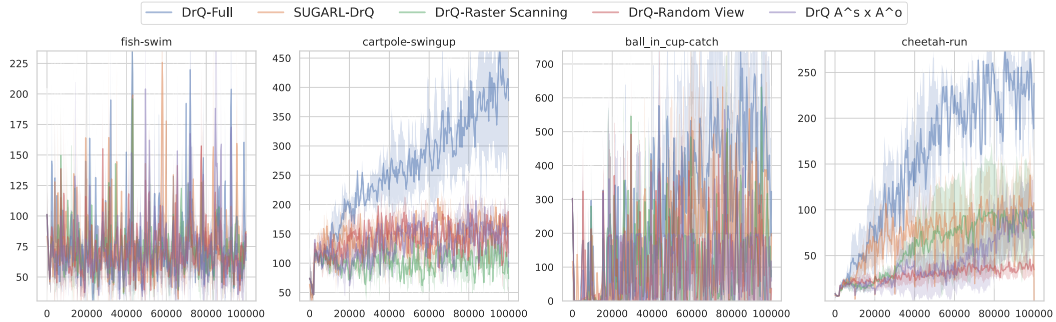

Figure 15: Learning curves of 4 DMC environments, under the setting of 30x30 foveal observation size and w/o peripheral observation.

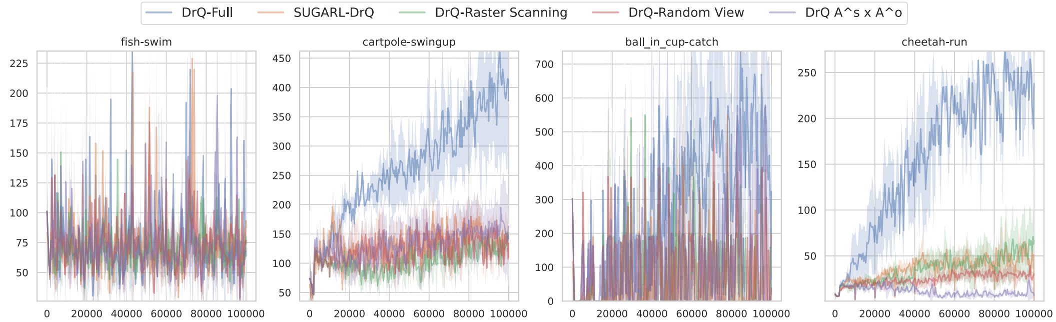

Figure 16: Learning curves of 4 DMC environments, under the setting of 20x20 foveal observation size and w/o peripheral observation.

Table 5: Hyper-parameters for DQN / SUGARL-DQN (on Atari)

Total steps	1,000,000 or 5,000,000
Replay buffer size	100,000
ϵ start	1.0
ϵ end	0.01
$\min \epsilon$ step	100,000
γ	0.99
Learning start	80,000
Q network train frequency	4
Target network update frequency	1,000
Learning rate	10^{-4}
Batch size	32
Self-understanding module train frequency	4
Self-understanding module learning rate	10^{-4}

Table 6: Hyper-parameters for SAC (on Atari)

Total steps	1,000,000
Replay buffer size	100,000
γ	0.99
Learning start	80,000
Actor train frequency	4
Critic train frequency	4
Target network update frequency	8,000
Actor Learning rate	3×10^{-4}
Critic Learning rate	3×10^{-4}
Batch size	64
Self-understanding module train frequency	4
Self-understanding module learning rate	3×10^{-4}
Visual policy alpha	0.2
Physical policy alpha	autotune
Physical policy target entropy scale	0.2

Table 7: Hyper-parameters for DrQv2 (on DMC)

Total steps	100,000
Replay buffer size	100,000
γ	0.99
Standard deviation start	1.0
Standard deviation end	0.1
Standard deviation end step	50,000
Standard deviation clip	0.3
Learning start	2,000
Actor train frequency	2
Critic train frequency	2
Target network update frequency	2
Target network exponential moving average weight	0.01
Actor Learning rate	10^{-4}
Critic Learning rate	10^{-4}
Batch size	256
Self-understanding module train frequency	2
Self-understanding module learning rate	10^{-4}
Multiple-step reward	3

Table 8: Environment Settings

Atari	
Gray-scale	True
Full observation size	84x84
Frame stacking	4
Action repeat (frame skipping)	4
Observable area initial location	$(0, 0)$
Visual action options	4×4 grid
Visual action space size	16 (abs) or 5 (rel)
PVM number of steps	3
DMC	
Gray-scale	True
Full observation size	84x84
Frame stacking	3
Action repeat (frame skipping)	2
Observable area initial location	$(0, 0)$
Visual action options	4×4 grid
Visual action space size	5 (rel)
PVM number of steps	3