TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KHOA HỌC MÁY TÍNH

BÀI TẬP MÔN
TRÍ TUỆ NHÂN TẠO NÂNG CAO

Deep Q-Network (DQN) và Double DQN (DDQN)

Giảng viên hướng dẫn: Lương Ngọc Hoàng

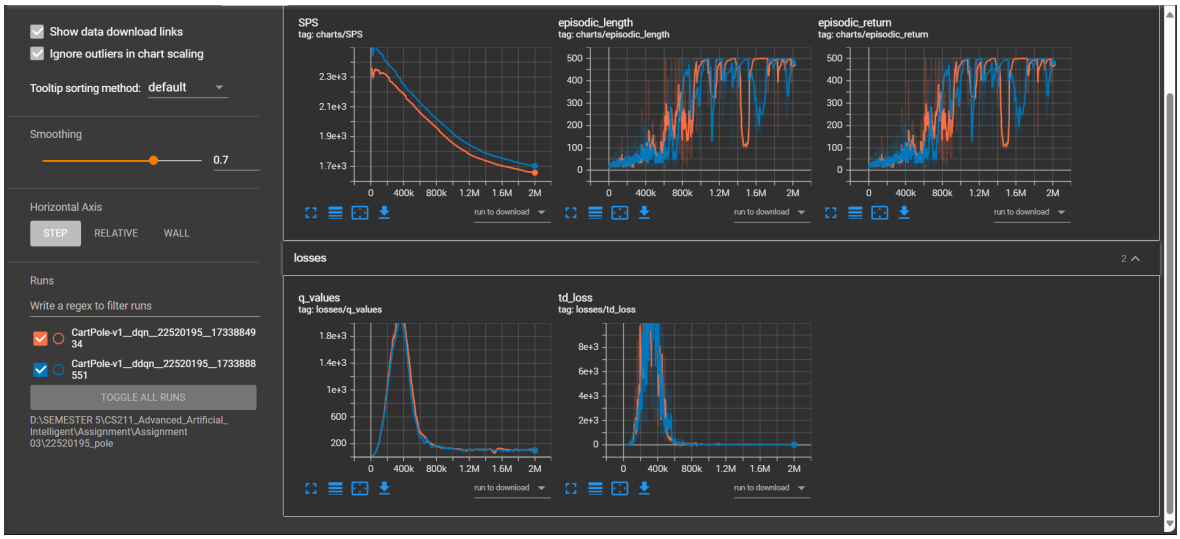
Họ và tên	MSSV	Mã lớp
Trần Đình Khánh Đăng	22520195	CS211.P11

TP. Hồ Chí Minh, ngày 12 tháng 12 năm 2024

Kết quả thực nghiệm:

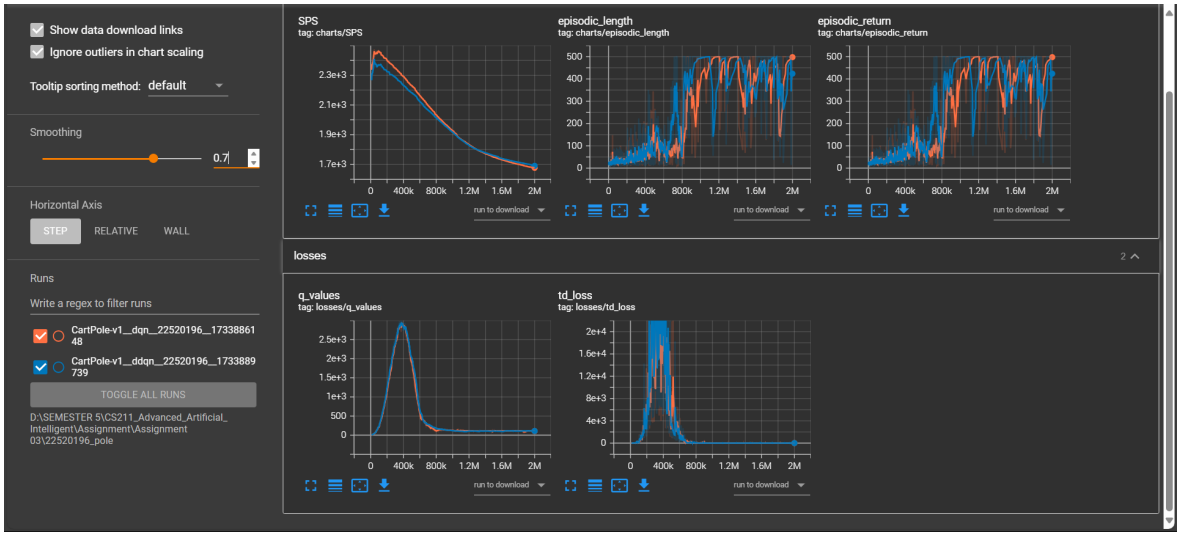
CartPole-v0

Seed = 22520195



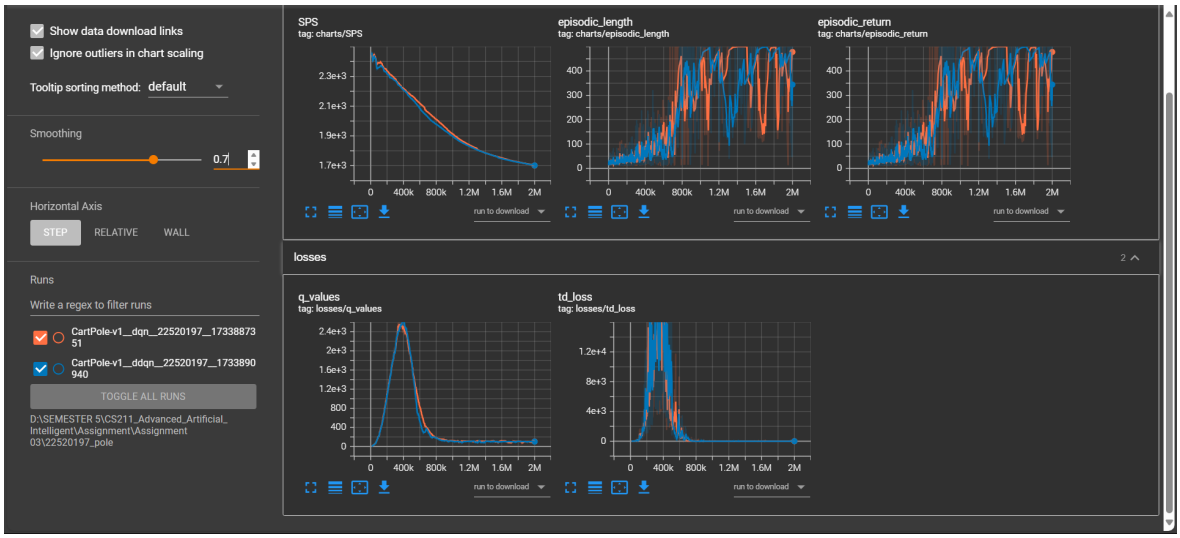
Hình 1: cartpole_seed22520195_timestamp2,000,000

Seed = 22520196



Hình 2: cartpole_seed22520196_timestamp2,000,000

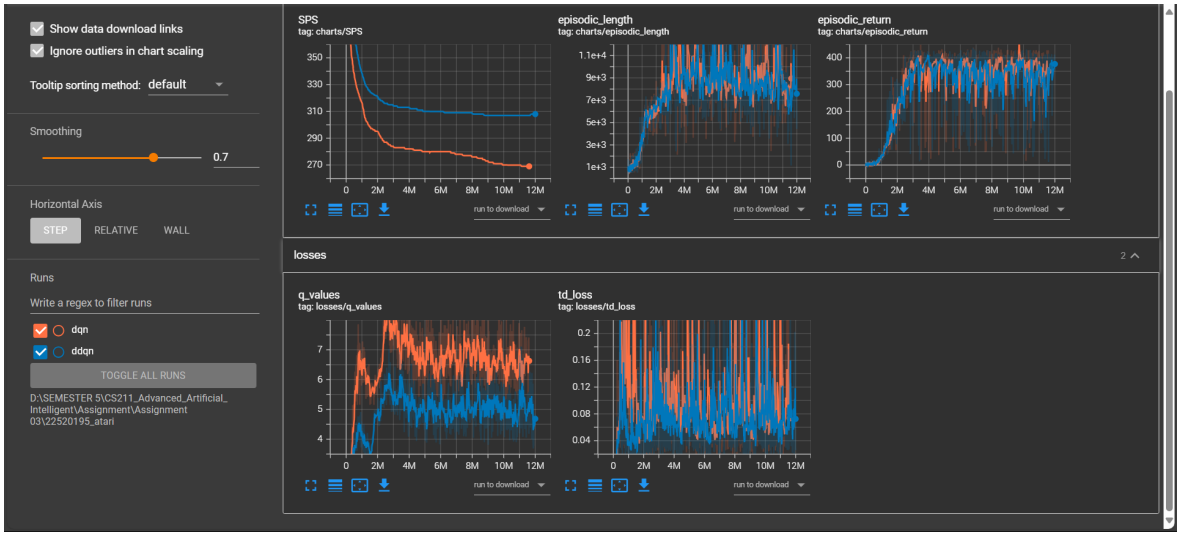
Seed = 22520197



Hình 3: cartpole_seed22520197_timestamp2,000,000

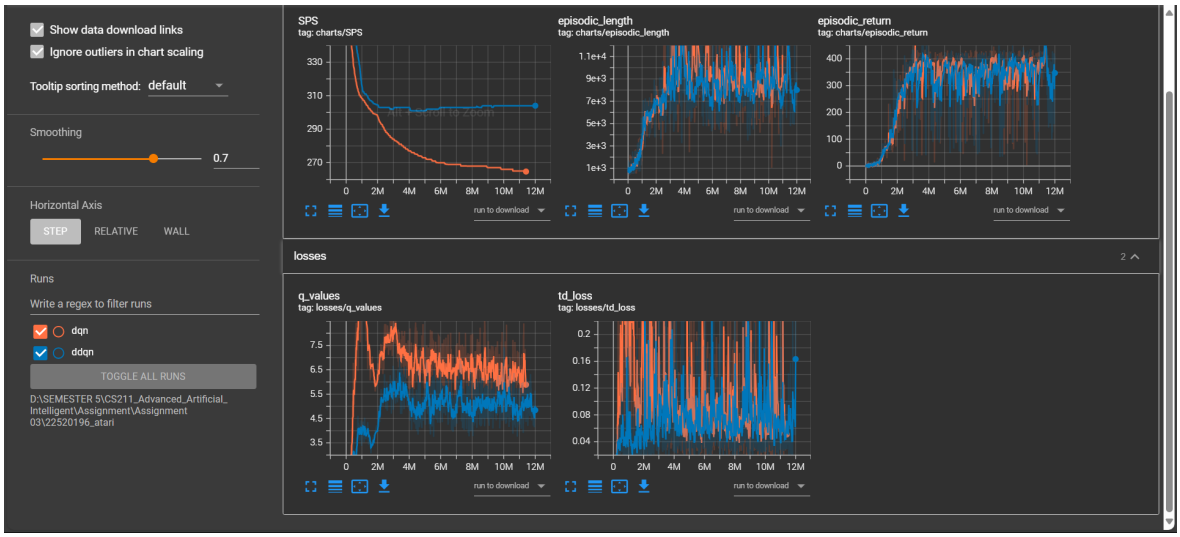
BreakoutNoFrameskip-v4

Seed = 22520195



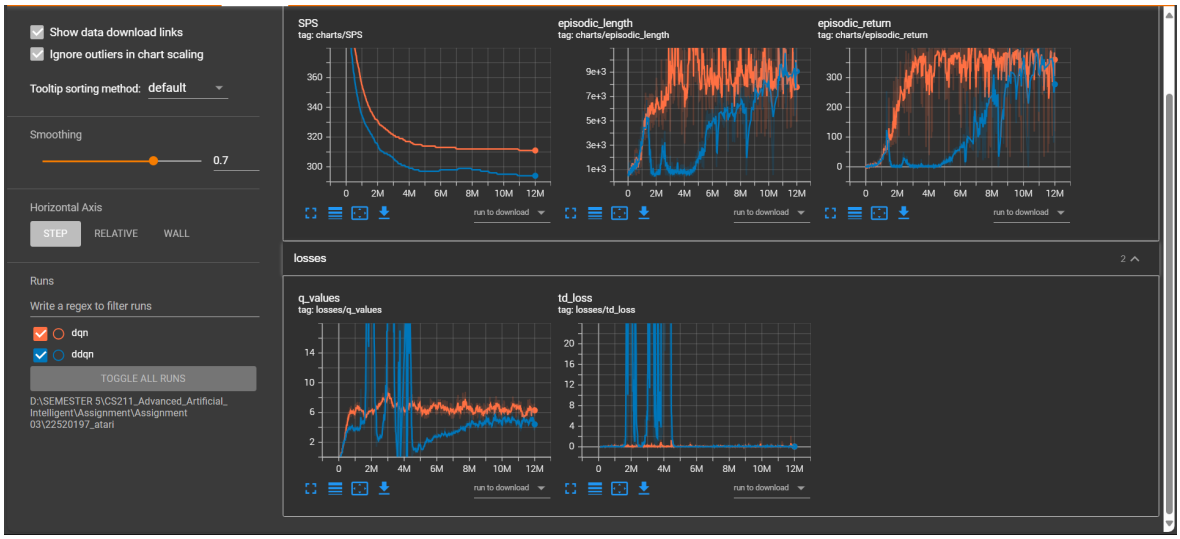
Hình 4: atari_seed22520195_timestamp12,000,000

Seed = 22520196



Hình 5: atari_seed22520196_timestamp12,000,000

Seed = 22520197



Hình 6: atari_seed22520197_timestamp12,000,000

Nhận xét

CartPole-v0

- **SPS**: Cả DDQN và DQN bắt đầu với SPS khá cao nhưng dần giảm đi khi quá trình huấn luyện tiến triển. Điều này là bình thường vì khi mô hình học được nhiều hơn, các tính toán trở nên phức tạp hơn. Dù vậy, DDQN có vẻ giữ được hiệu suất tốt hơn một chút so với DQN trong khoảng cuối của quá trình huấn luyện.
- **Episodic Length, Episodic Return**: DDQN có thể sẽ cao hơn so với DQN trong các tập dài, thể hiện sự ổn định và chính xác hơn trong việc chọn hành động.
- **Q Values**: Q Value của DDQN thấp hơn một cách hợp lí và ổn định hơn so với DQN, vì DDQN giảm thiểu vấn đề overestimation bias.
- **TD Loss**: TD loss của DDQN cao hơn DQN, cho thấy DDQN gặp khó khăn trong hội tụ.

Tổng kết

DDQN mang lại hiệu suất ổn định hơn và ít biến động hơn so với DQN. Nhờ việc giảm thiểu vấn đề overestimation bias, DDQN cho phép học tập nhanh hơn và đạt được kết quả tốt hơn trong việc tối ưu hóa chính sách và hành động. Trong khi đó, DQN, dù có thể đạt được kết quả tốt, nhưng lại không ổn định bằng DDQN, đặc biệt khi gặp phải các tình huống phức tạp hoặc cần phải ước lượng chính xác giá trị của các hành động.

BreakoutNoFrameskip-v4

- **SPS**: Ở 2 seed 22520195 và 22520196, DDQN cao hơn DQN, điều này phản ánh việc DDQN yêu cầu tính toán nhiều hơn. Riêng ở seed 22520197, có thể do tính ngẫu nhiên của môi trường mà dẫn đến việc DDQN thấp hơn DQN.
- **Episodic Length, Episodic Return**: Ở 2 seed 22520195 và 22520196, cả 2 thuật toán hoạt động gần như giống nhau, nhưng DDQN thấp hơn một chút, cho thấy sự tối ưu hóa hành động hiệu quả hơn của DDQN. .
- **Q Values**: Ở 2 seed 22520195 và 22520196, DDQN đều thấp hơn DQN, riêng ở seed 22520197 thì thuật toán hành xử không ổn định khi số bước dưới 4 triệu, sau đó thuật toán ổn định hơn và vẫn thấp hơn DQN, điều này chứng tỏ DDQN ước lượng Q-values cẩn trọng hơn để tránh overestimation.
- **TD Loss**: TD loss của DDQN thấp hơn và ổn định hơn DQN, cho thấy khả năng hội tụ tốt hơn và cập nhật chính xác hơn.

Tổng kết

Trong môi trường phức tạp như BreakoutNoFrameskip-v4, DDQN thể hiện sự vượt trội trong TD Loss và khả năng giảm overestimation, đổi lại bằng tốc độ tính toán lâu hơn. Trong khi đó, DQN tính toán nhanh nhưng lại không ổn định trong việc ước lượng.

So sánh cơ chế hoạt động của DQN và Double DQN.

Tiêu chí	**DQN**	**Double DQN**
Thiên vị ước lượng	Có thể đánh giá quá cao giá trị của các hành động	Giảm thiểu thiên vị ước lượng
Cách cập nhật hàm Q	Sử dụng cùng một mạng để chọn hành động và đánh giá hành động	Sử dụng hai mạng khác nhau (online network và target network)
Tốc độ hội tụ	Hội tụ chậm hơn và có thể không ổn định	Hội tụ nhanh hơn và ổn định hơn
Ứng dụng	Phù hợp cho các bài toán đơn giản hoặc môi trường ổn định	Phù hợp cho các bài toán phức tạp hơn, cần ước lượng chính xác

Bảng 1: Bảng so sánh DQN và Double DQN