Création de différentes IA en utilisant une architecture I2A

Dans le contexte du jeu Sokoban

Thomas Guyomard - Jérémy Tremblay

Université du Littoral Côte d'Opale

Encadrant : Jérôme Buisine

2024/03/29

Sommaire

Références

Sébastien Racanière, Théophane Weber, David P. Reichert, Lars Buesing, Arthur Guez, Danilo Rezende, Adria Puigdomènech Badia, Oriol Vinyals, Nicolas Heess, Yujia Li, Razvan Pascanu, Peter Battaglia, Demis Hassabis, David Silver, Daan Wierstra.
Imagination-Augmented Agents for Deep Reinforcement Learning.
DeepMind, arXiv:1707.06203v2 [cs.LG], 14 Feb 2018.

Max-Philipp B. Schrader.
gym-sokoban.
GitHub repository, GitHub, 2018.
`https://github.com/mpSchrader/gym-sokoban`.

Présentation du jeu Sokoban

Jeu de réflexion. Le but : le joueur doit ranger des caisses sur des cases cibles.

- Le personnage a la capacité d'effectuer des déplacement dans chacune des 4 directions possibles. Le niveau est validé une fois toutes les caisses rangées.

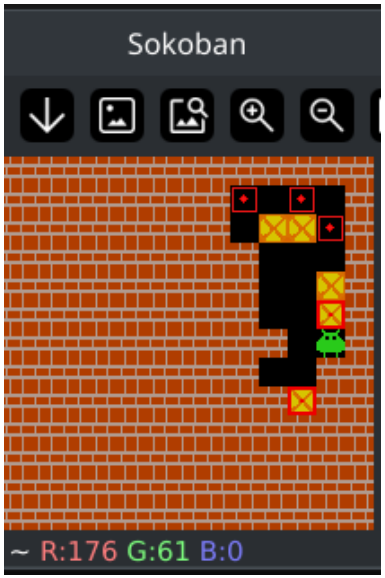

Figure: Capture d'écran du jeu Sokoban.

Agents Augmentés par l'Imagination / (I2A)

Definition

L'apprentissage par renforcement avec augmentation par l'imagination appelée I2A est une approche du domaine de l'intelligence artificielle permettant d'améliorer la prise de décision des agents d'apprentissage. Les agents combinent l'apprentissage automatique avec et sans modèles avec des générations de scénarios pour anticiper des situations inédites et proposer des solutions optimales.

IA Q-learning avec une représentation en tableau

Definitions

L'implémentation combine le Q-learning avec un modèle interne I2A pour résoudre le problème complexe du Sokoban, un jeu de réflexion.

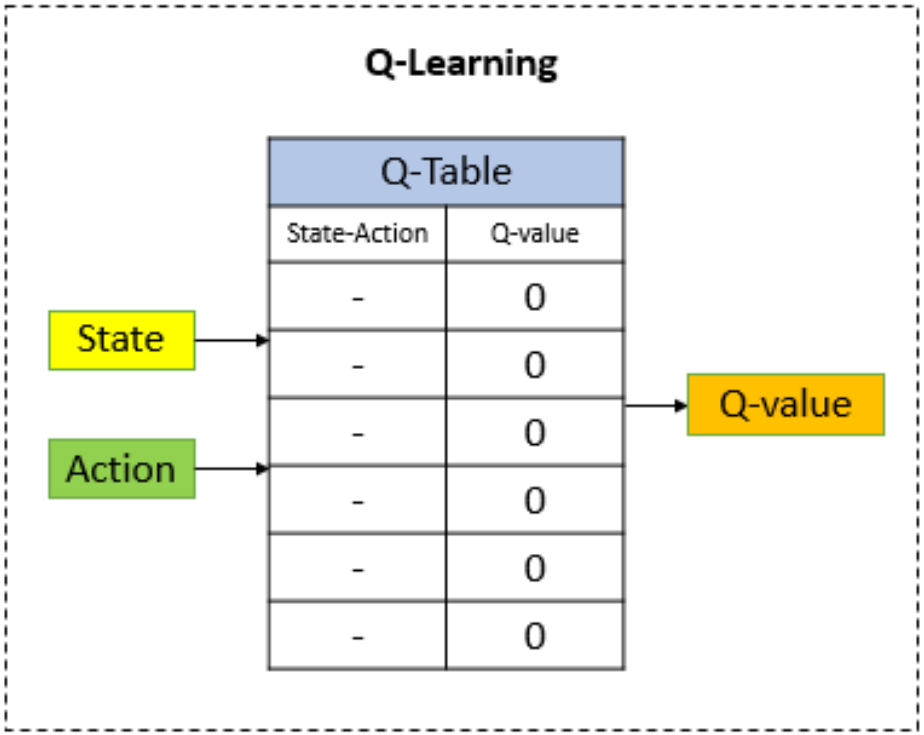

Figure: Schéma du fonctionnement du Q-Learning.

Implémentation du Q-learning

- **Q-table** Contient les états / actions du jeu, elle est initialisée avec des zéros.
- **Boucle d'apprentissage** Itérations sur des épisodes pour mettre à jour la Q-table.
- **Système de récompenses** Intégration des récompenses conformément à la politique définie.
- **Mise à jour de la Q-table** Prise en compte de la récompense immédiate et de l'estimation de la récompense future.
- **Exploration et Exploitation** Equilibre les connaissances actuelles et permet davantage d'apprentissage.

Contenu de la Q-table

```
11468484326015042000:  array([-0.05, 2.77913742,
1.69800543, 3.03067097, 2.54018055, 2.77913742,
2.77913742, 3.03067097, 2.54018055])
```

Résultats du Q-learning (1/6)

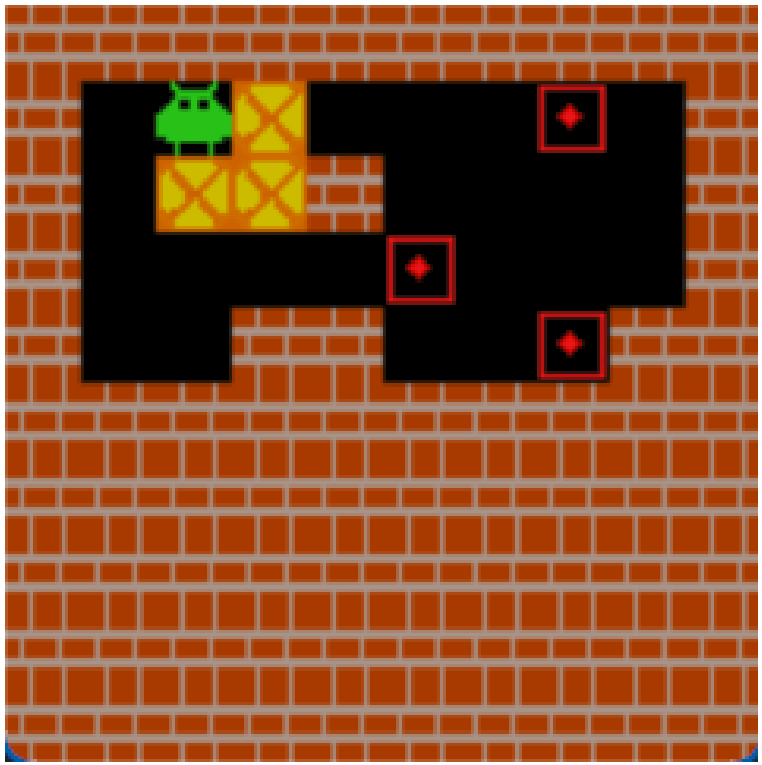

Figure: Évolution des récompenses totales obtenues par épisode.

Contexte de jeu :

- Niveau de taille moyenne
- Jeu de 10x10 cases
- 3 caisses à pousser

Contexte d'exécution :

- 10 000 parties réalisées
- Utilisation d'un algorithme de Q-learning
- 8h30 de temps d'exécution au total

Résultats du Q-learning (2/6)

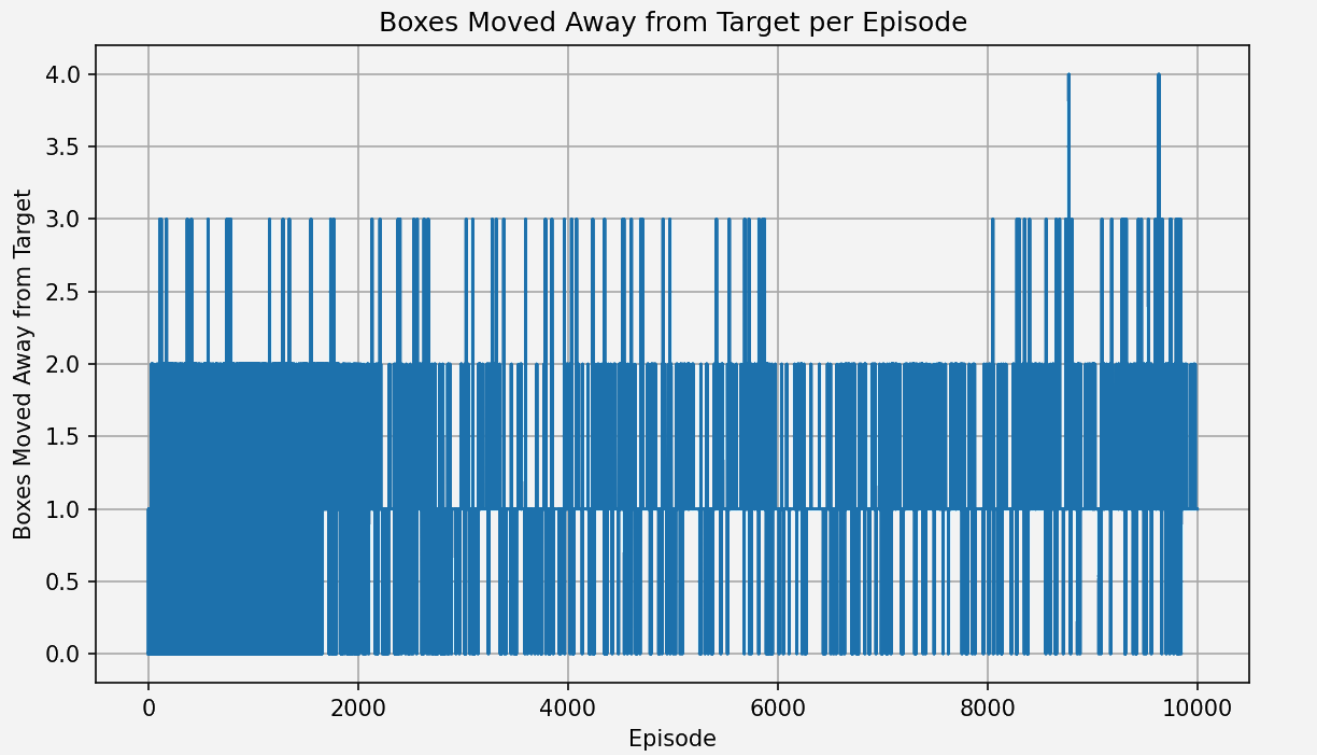

Figure: Évolution du nombre de caisses poussées en dehors des emplacements par épisode.

Observation des résultats

L'agent a tendance à pousser une ou deux caisses en dehors des emplacements par épisode.

Résultats du Q-learning (3/6)

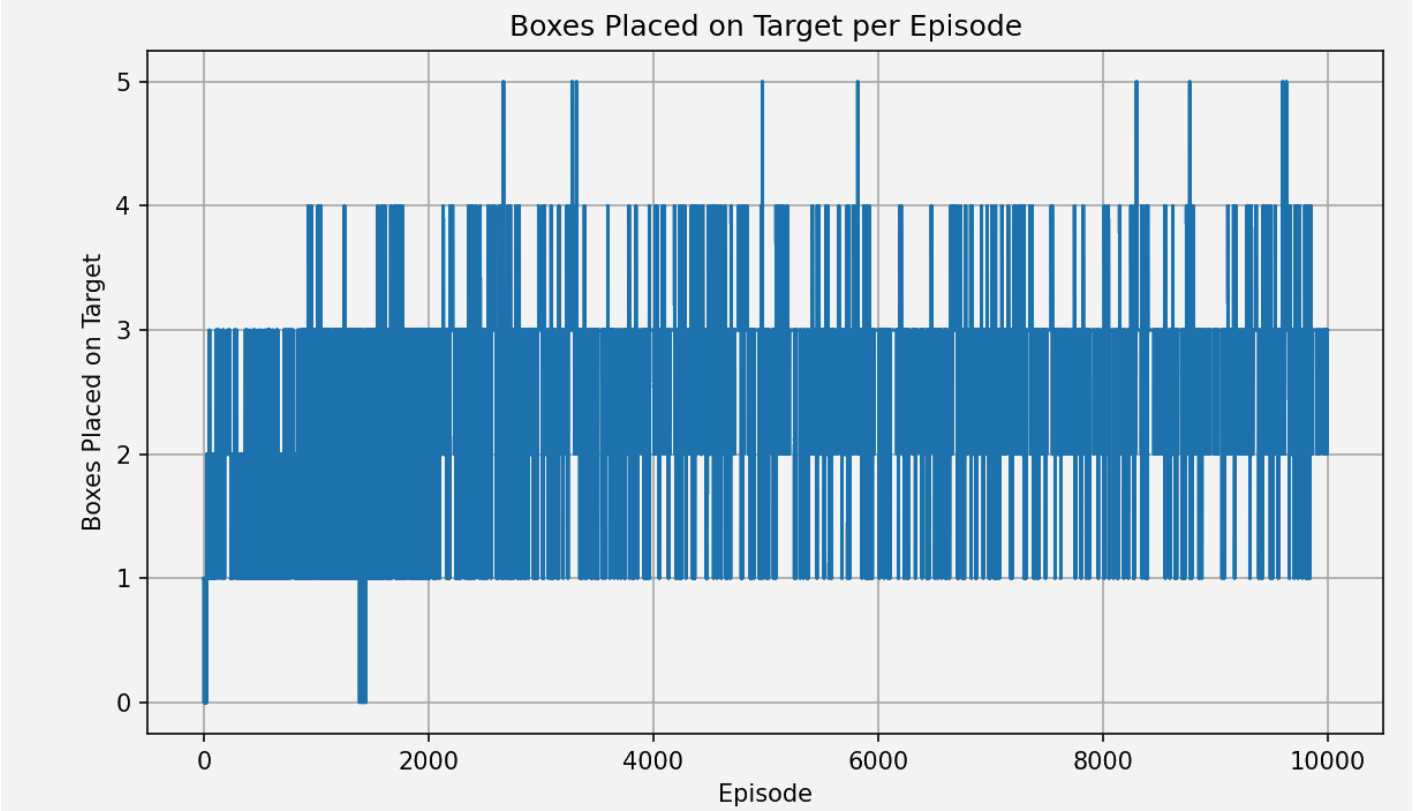

Figure: Évolution du nombre de caisses poussées sur un emplacement par épisode.

Observation des résultats

L'agent a compris qu'il devait pousser des caisses sur un emplacement.

Résultats du Q-learning (4/6)

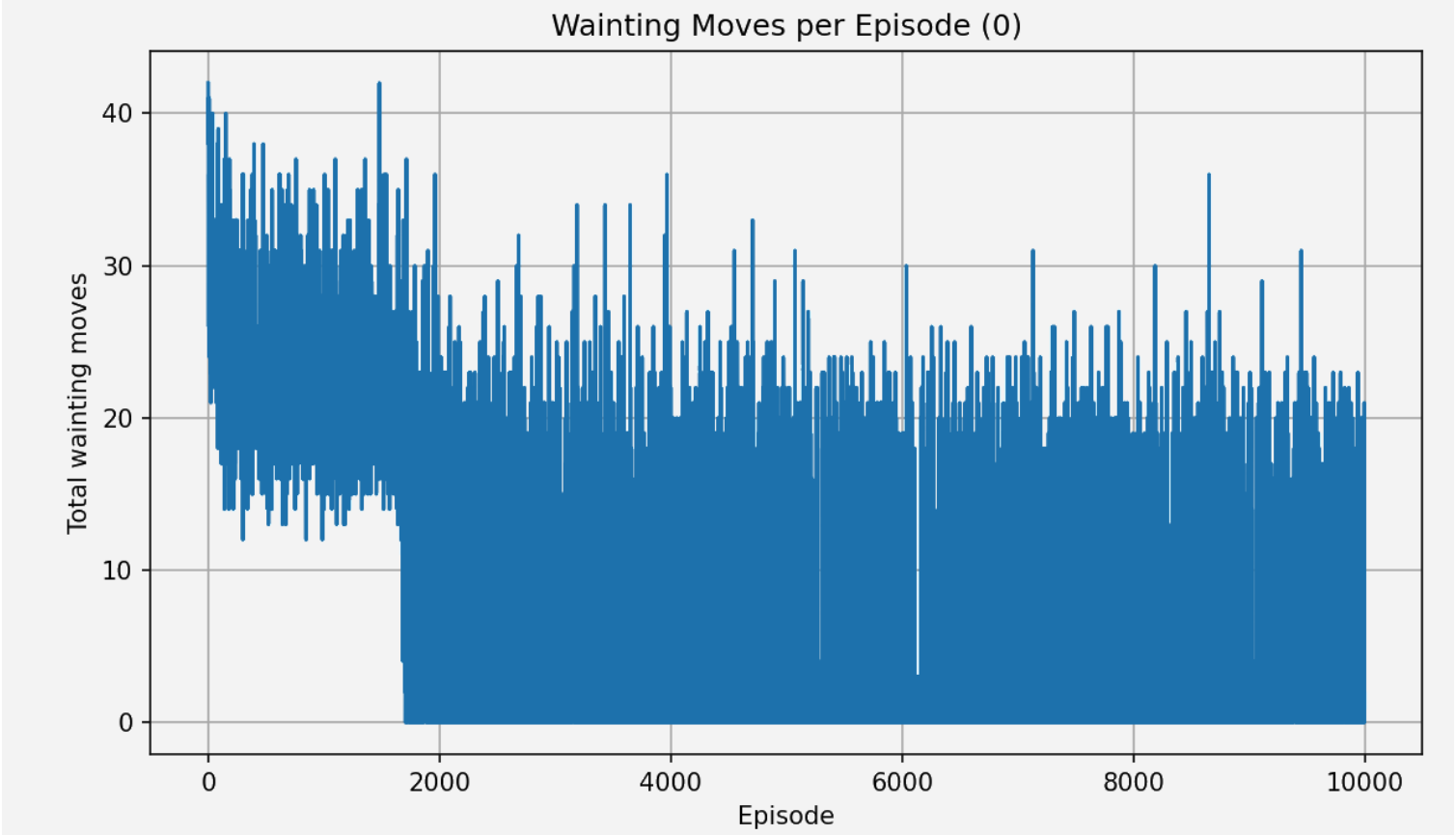

Figure: Évolution du nombre d'inactions par épisode.

Observation des résultats

L'agent a compris que les mouvements blancs étaient inutiles.

Résultats du Q-learning (5/6)

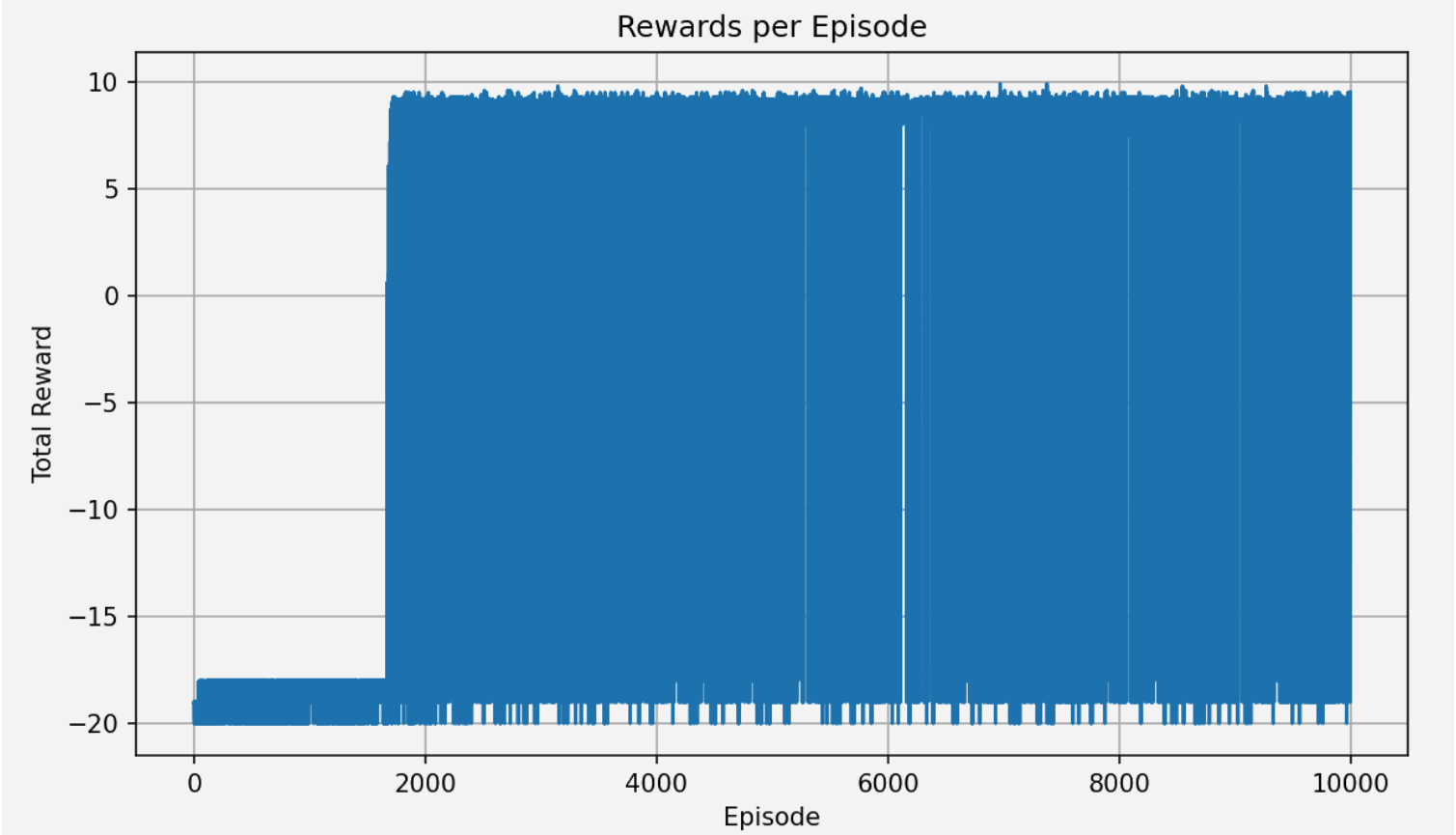

Figure: Évolution des récompenses totales obtenues par épisode.

Observation des résultats

Changement de comportement : Exploration → Exploitation.

Résultats du Q-learning (6/6)

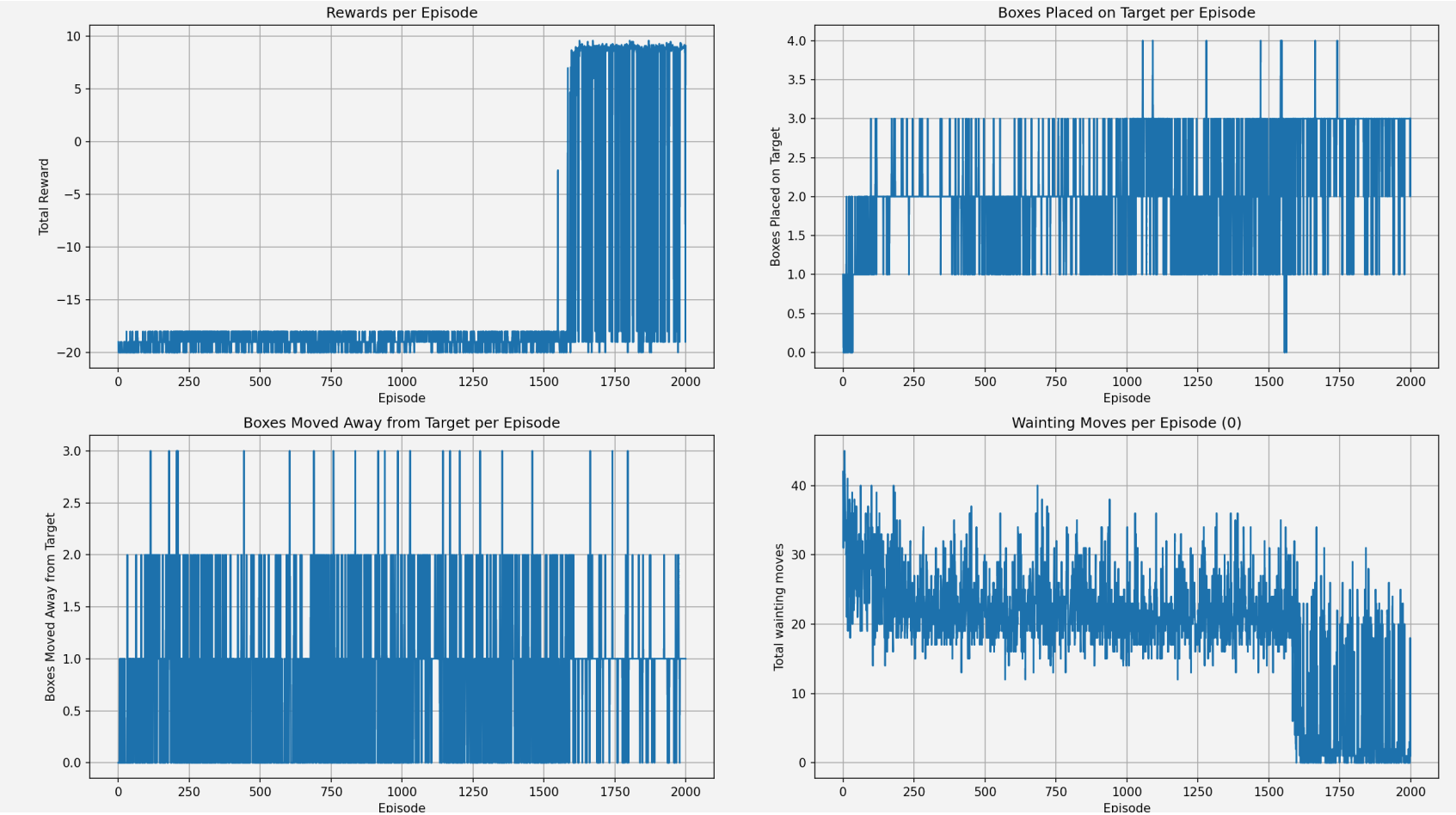

Figure: Résultats du Q-learning sur 2000 épisodes.

Plan de Travail

1	Exploiter notre implémentation du deep Q-learning et comparer les résultats avec ceux du Q-learning.
2	Finaliser le MCTS et analyser les résultats.
3	Explorer la possibilité d'utiliser les représentations visuelles de Gym Sokoban pour enrichir la planification.
4	Tester l'agent sur différents niveaux afin d'évaluer sa capacité de généralisation.
5	Si possible, analyser les performances de chaque approche, en mettant en évidence les forces et les faiblesses dans le contexte de Sokoban.

Merci pour votre attention.