

Assemblée générale MIREs



Optimisation distribuée des systèmes de transport intelligents dans la 5G (OSTI5G)



07 juillet 2025

Présenté par : Ehud MOSSA OCKEGNA (stagiaire de M2)

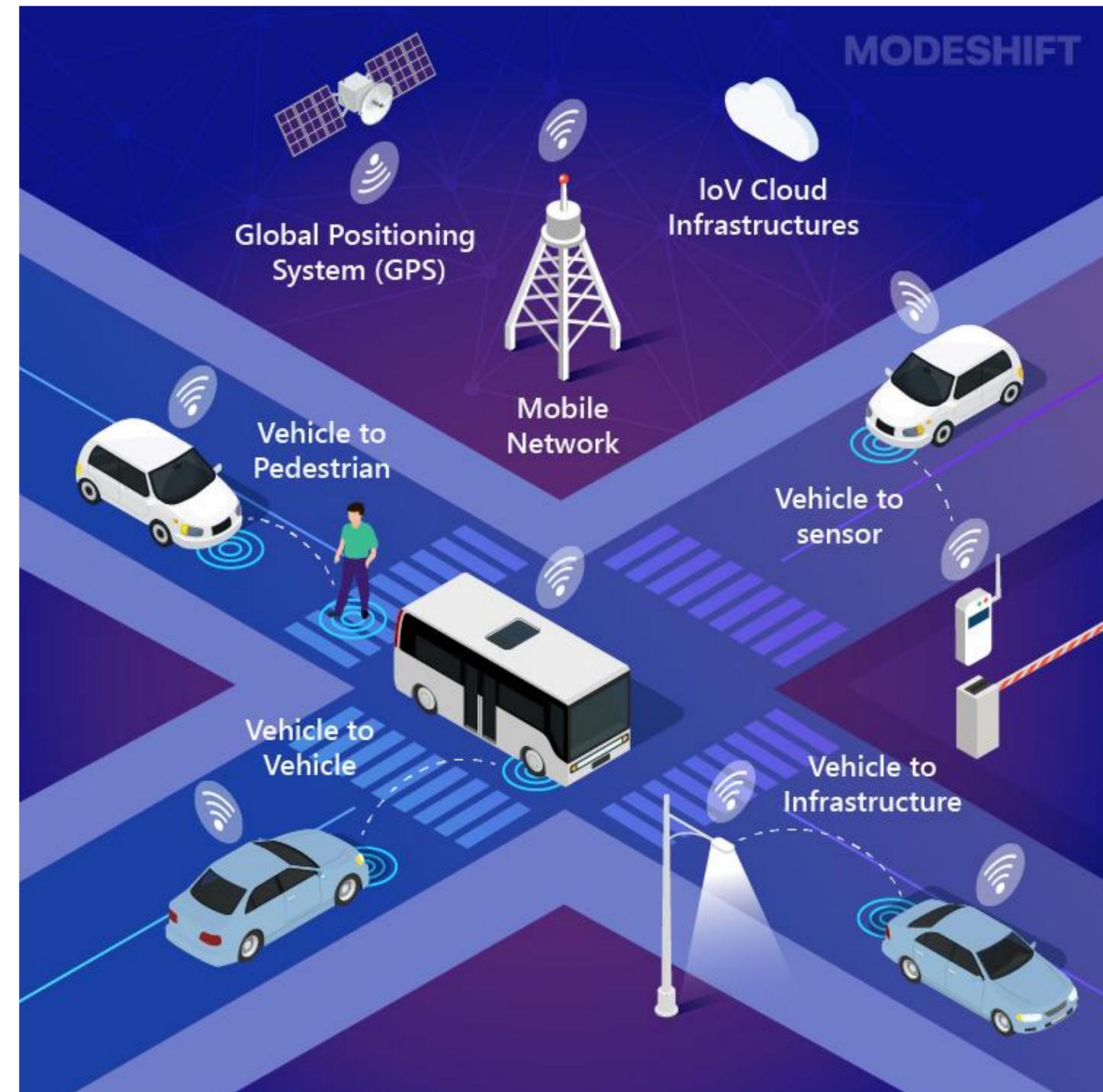
Membres du projet :

Emmanuel MOULAY (coordinateur)

Clency PERRINE

Philippe CARRÉ

Les systèmes de transport intelligents dans la 5G



Vehicle-to-everything (V2X)

- Fluidifier le trafic
 - Feux intelligents
 - Voitures autonomes
- Intégration de la 5G et de l'IA
- Sécurité routière
 - Détection en temps réel
 - Voitures autonomes

Objectifs du projet

Problématique abordée des systèmes de transport intelligents dans la 5G :

- Plannification comportementale des véhicules autonomes
- La gestion des feux de signalisation pour la fluidité routière

Moyens du projet

- Mon stage de Master financé par MIREES
- La thèse d'Edouard Laurent , chercheur chez Google DeepMind :
Apprentissage par renforcement sûr et efficace pour la
plannification comportementale en conduite autonome

Concepts de base de l'apprentissage par renforcement

L'apprentissage par renforcement (RL) est généralement formalisé comme un processus de décision de Markov (MDP)

$$(\mathcal{S}, \mathcal{A}, \mathcal{P}, R, \gamma)$$



Espace des états

$$(s_t)$$



Espace des actions

$$(a_t)$$

$$a_t \sim \pi(a_t | s_t)$$



Dynamique

$$s_{t+1} \sim \mathcal{P}(s_{t+1} | s_t, a_t)$$

$$\mathcal{P} \in \mathcal{M}(\mathcal{S})^{\mathcal{S} \times \mathcal{A}}$$



Récompenses

$$R(s_t, a_t)$$

$$R \in [0, 1]^{\mathcal{S} \times \mathcal{A}}$$

$$\gamma \in [0, 1[$$

Politique : $\pi(a_t | s_t) = \mathbb{P}[A_t = a_t | S_t = s_t]$

Retour sur politique

$$G^\pi = \sum_{t=0}^{+\infty} \gamma^t R(s_t, a_t)$$

Fonction de valeur

$$V^\pi(s) = \mathbb{E}[G^\pi | s_0 = s]$$

$$Q^\pi(s, a) = \mathbb{E}[G^\pi | s_0 = s, a_0 = a]$$

Le but de l'apprentissage par renforcement est de trouver la politique optimale π^*

Équation d'optimalité de Bellman

La fonction de valeur d'action optimale Q^* satisfait l'Équation d'Optimalité de Bellman

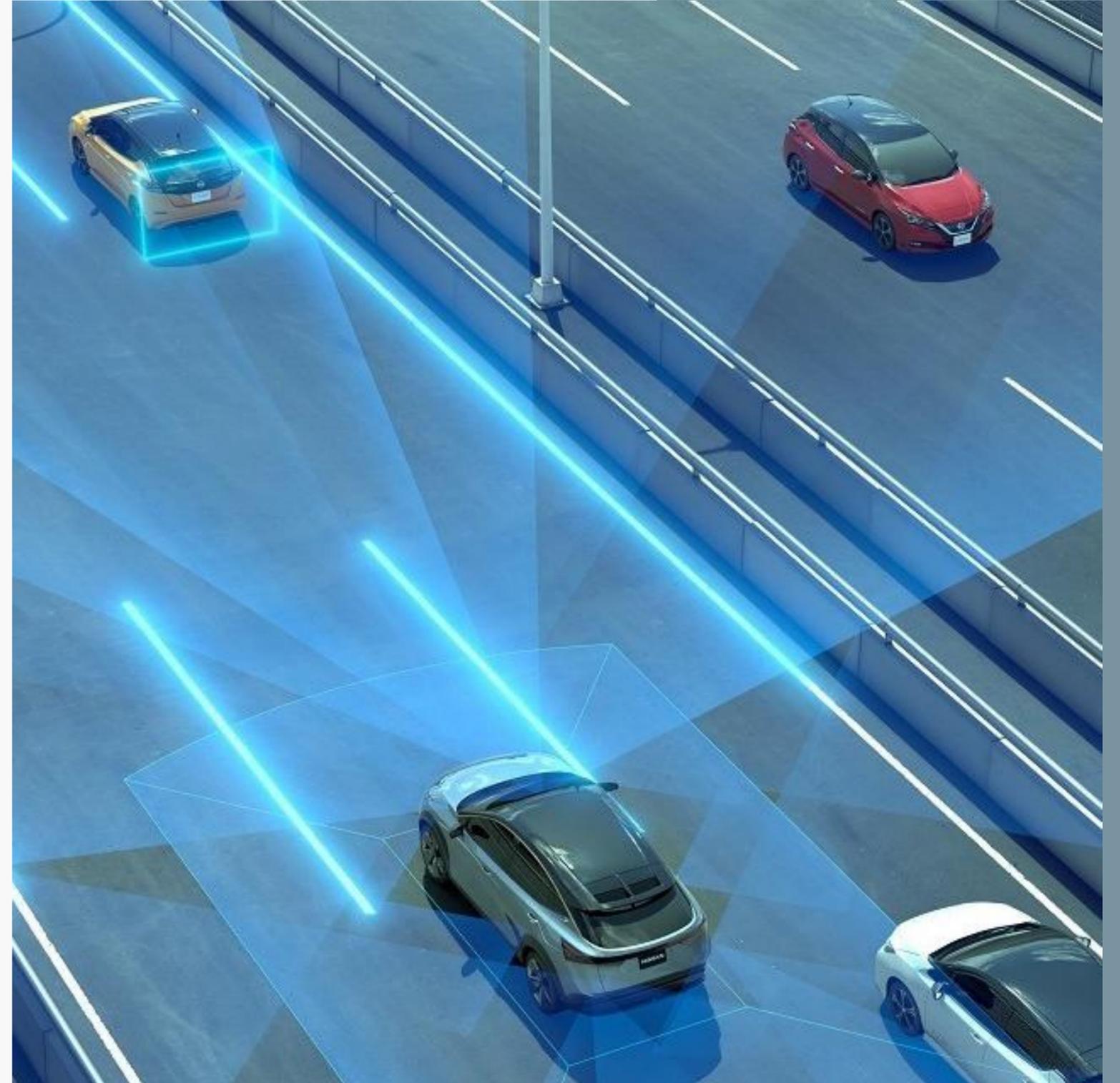
$$Q^*(s, a) = (\mathcal{T}Q^*)(s, a) \triangleq \mathbb{E}_{s' \sim P(\cdot|s, a)} \left[\max_{a' \in \mathcal{A}} (R(s, a) + \gamma Q^*(s', a')) \right].$$

De plus, \mathcal{T} est une γ – contraction pour la $\|\cdot\|_\infty$:

$$\forall Q_1, Q_2 \in \mathbb{R}^{\mathcal{S} \times \mathcal{A}}, \quad \|\mathcal{T}Q_1 - \mathcal{T}Q_2\|_\infty \leq \gamma \|Q_1 - Q_2\|_\infty.$$

Problème 1

Plannification comportementale
des véhicules autonomes



Algorithmes de DRL

Model free

Model based

Social attention

BFTQ

OPD

(Munos,2008)

KL-OLOP

GBOP

DQN

PPO

GBOP-D

Estimation contrôle
et prédiction
minimax

Algorithmes de DRL

Model free

Model based

Social attention

BFTQ

OPD

(Munos,2008)

KL-OLOP

GBOP

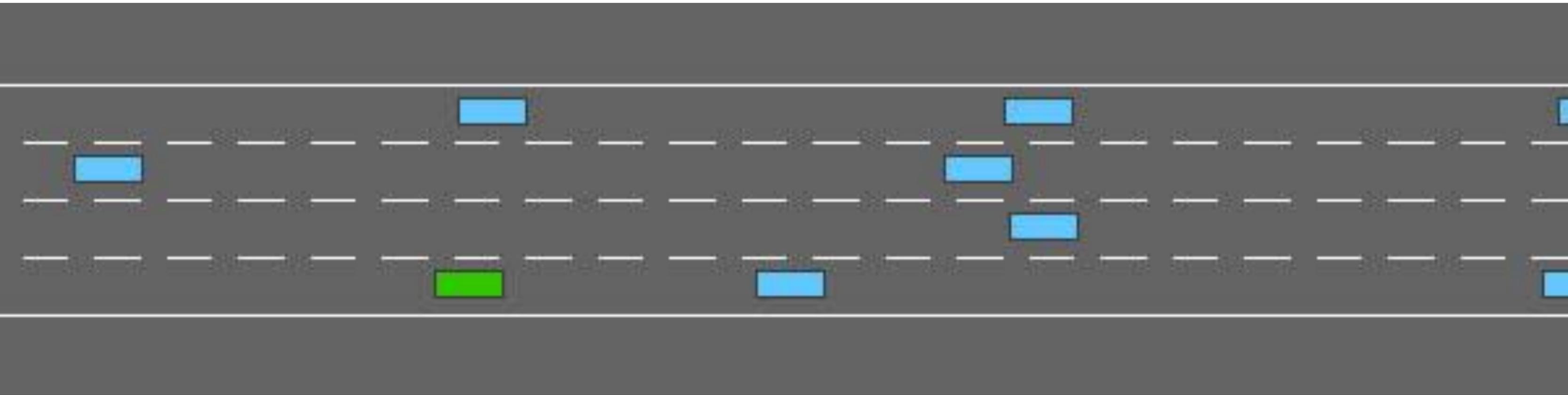
DQN

PPO

GBOP-D

Estimation contrôle
et prédiction
minimax

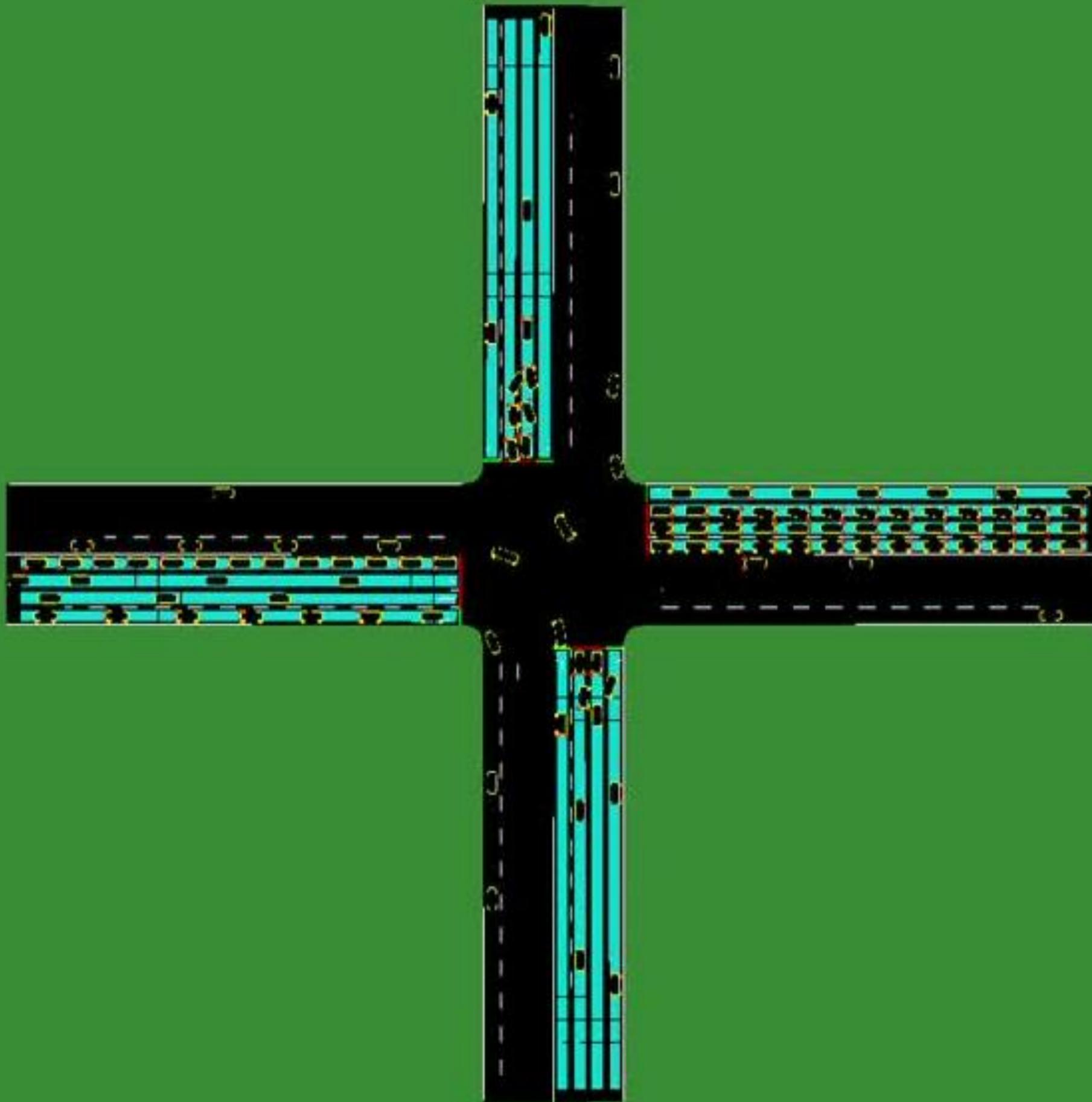
Résultats avec Highway de gymnasium



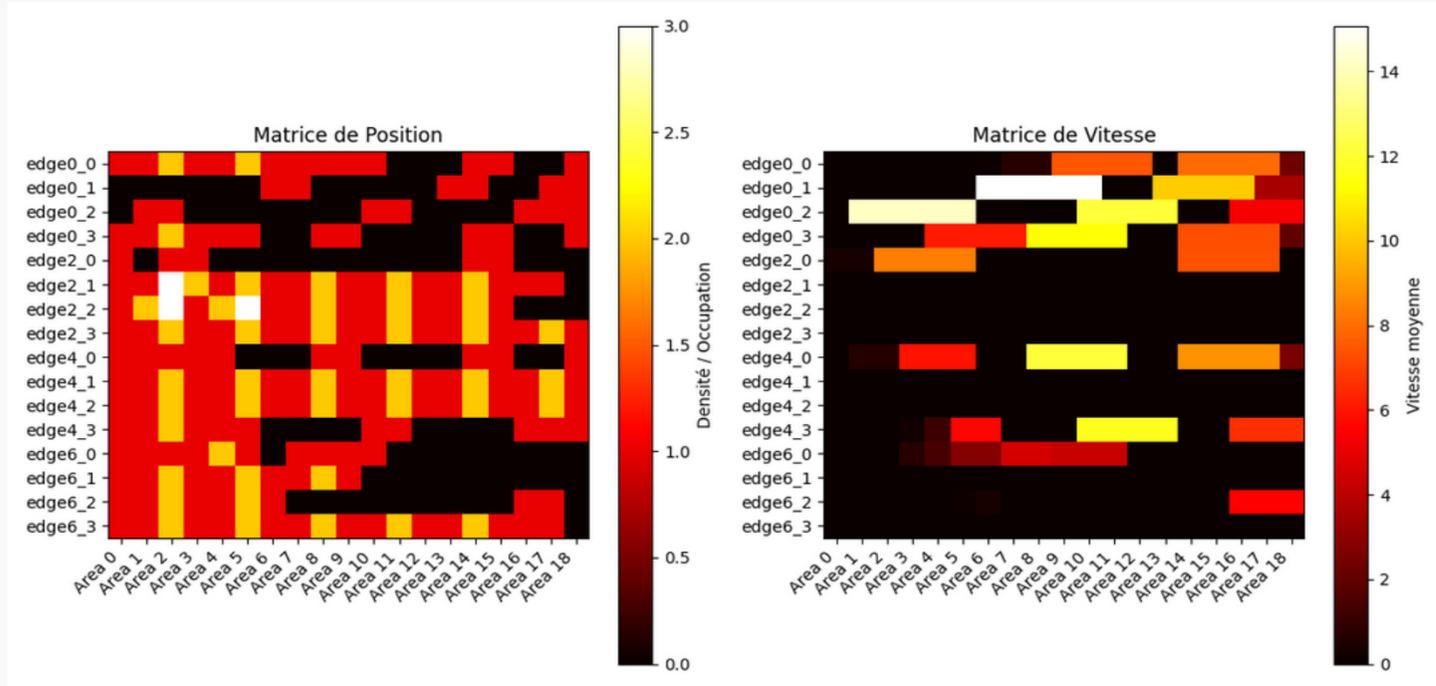
Problème 2

La gestion des feux de signalisation pour la fluidité routière (**contribution**)



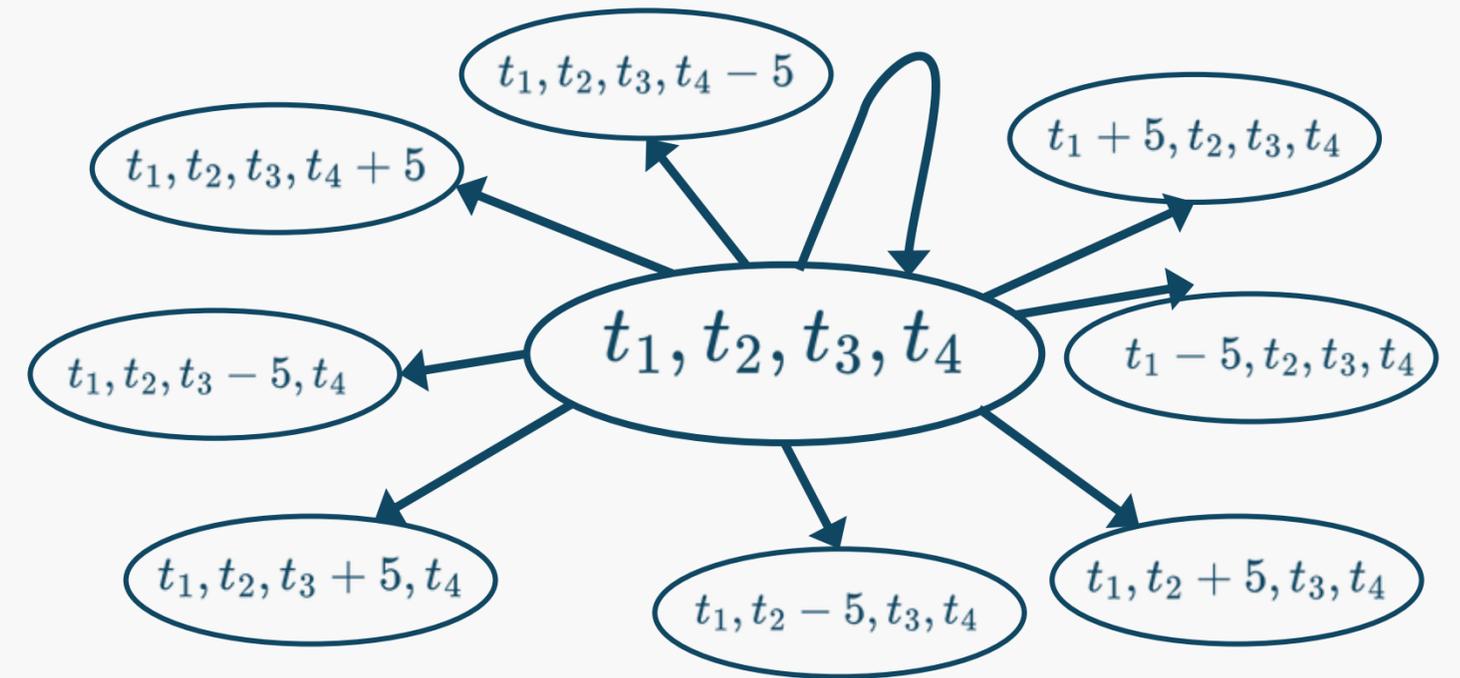


État



Liu, D., & Li, L. (2023). A traffic light control method based on multi-agent deep reinforcement learning algorithm. *Scientific Reports*, 13(1), 9396.

Actions



Liu, D., & Li, L. (2023). A traffic light control method based on multi-agent deep reinforcement learning algorithm. *Scientific Reports*, 13(1), 9396.

Récompense

$$r_{\text{ajusté}} = \begin{cases} 1 & \text{si } r_t > 0 \\ 0.5 & \text{si } r_t = 0 \\ 0 & \text{sinon} \end{cases}$$

Leurent, E. (2020). *Safe and efficient reinforcement learning for behavioural planning in autonomous driving* (Doctoral dissertation, Université de Lille).

Dynamique stochastique

$$\gamma = 0.95$$

Algorithmes de DRL

Model free

Model based

Social attention

BFTQ

OPD

(Munos,2008)

KL-OLOP

GBOP

DQN

PPO

GBOP-D

Estimation contrôle
et prédiction
minimax

Algorithmes de DRL

Model free

Model based

Social attention

BFTQ

OPD

(Munos,2008)

KL-OLOP

GBOP

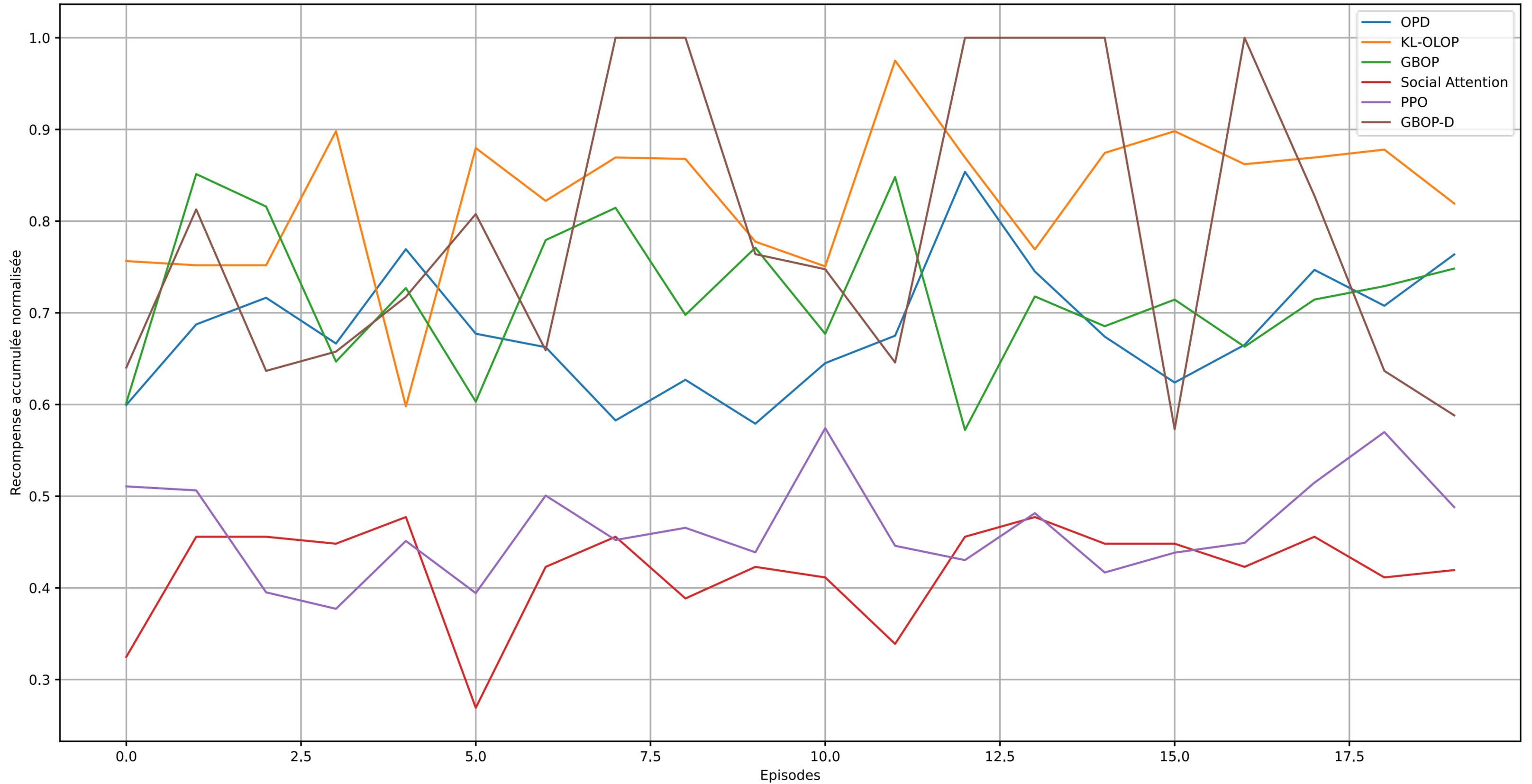
DQN

PPO

GBOP-D

Estimation contrôle
et prédiction
minimax

Récompense accumulée en fonction de chaque agent



Conclusion et perspectives

- Les algorithmes de DRL développés dans la thèse d'Edouard Leurent ont démontrés leur efficacité pour le problème de gestion des feux de signalisation.
- Les meilleurs algorithmes DRL semblent être le KL-OLOP et le GBOP-D
- Publier les résultats obtenus pendant mon stage de M2.
- Prendre en compte une détection réaliste des véhicules en utilisant le modèle d'IA YOLO pour la gestion des feux de signalisation.
- Gérer les feux de signalisation en tenant compte des contraintes liées au chargement des batteries des véhicules électriques.



Merci pour votre
attention

