CEA Comité Español de Automática
XLVII JORNADAS DE AUTOMÁTICA CÓRDOBA 2026
UNIVERSIDADE DA CORUÑA

Jornadas de Automática

Gestión inteligente de semáforos mediante visión y aprendizaje por refuerzo

Diego Caballero[a,*], Jaime Villa[a], Rubén Fernández[a], Mohamed Abderrahim[a], Araceli Sanchis[b], José María Armingol[a]

[a]*Departamento de Ingeniería Eléctrica, Universidad Carlos III de Madrid, Av. de la Universidad 30, 28911 Leganés, Madrid, España.*
[b]*Departamento de Informática, Universidad Carlos III de Madrid, Av. de la Universidad 30, 28911 Leganés, Madrid, España.*

To cite this article: Caballero, D., Villa, J., Fernández, R., Abderrahim, M., Sanchis, A., Armingol, J.M. 2026. Intelligent Traffic Signal Management through Vision and Reinforcement Learning. Jornadas de Automática, 47. https://doi.org/10.17979/ja-cea.2026.47.13792

Resumen

La congestión del tráfico urbano constituye uno de los principales retos de las ciudades modernas. Los sistemas tradicionales de control de semáforos, basados en tiempos fijos, no se adaptan a las variaciones dinámicas del tráfico. Este trabajo presenta un agente autónomo de control de semáforos basado en aprendizaje por refuerzo profundo (DRL) que utiliza únicamente imágenes fotorrealistas de la intersección como entrada, eliminando la dependencia de costosas infraestructuras de sensores. La validación se realiza en una plataforma de co-simulación que combina CARLA, para el realismo visual, y SUMO, para la dinámica microscópica del tráfico. Los resultados muestran reducciones del tiempo de espera de hasta el 22,68 % en la intersección de entrenamiento, una mejora del 23,76 % al reentrenar el agente en una intersección con topología distinta y una sinergia positiva en escenarios multi-agente, donde la cooperación entre intersecciones adyacentes mejora la recompensa hasta un 59,78 %.

Palabras clave: Sistemas de control de tráfico, Control mediante aprendizaje por refuerzo, Sistemas inteligentes de transporte, Sistemas multi-agente, Modelado y simulación de sistemas de transporte, Movilidad urbana.

Intelligent Traffic Signal Management through Vision and Reinforcement Learning

Abstract

Urban traffic congestion remains a major challenge in modern cities. Traditional fixed-time traffic signal controllers cannot adapt to dynamic traffic conditions. This work presents an autonomous traffic signal control agent based on Deep Reinforcement Learning (DRL) that uses only photorealistic images of the intersection as input, removing the dependency on costly physical sensor infrastructure. The validation is performed in a high-fidelity co-simulation platform combining CARLA, for visual realism, and SUMO, for microscopic traffic dynamics. Results show reductions of up to 22.68 % in average waiting time at the training intersection, an improvement of 23.76 % when retraining the agent in an intersection with a different topology, and positive synergy in multi-agent settings, where cooperation between adjacent intersections improves the reward by up to 59.78 %.

Keywords: Traffic control systems, Reinforcement learning control, Intelligent transportation systems, Multi-agent systems, Modeling and simulation of transportation systems, Urban Mobility.

1. Introducción

La congestión del tráfico en áreas urbanas es uno de los grandes retos de las ciudades modernas y un fenómeno con evidente impacto global. Según el último ranking de TomTom International BV (2025), las ciudades más afectadas presentan niveles de congestión cercanos al 70 %, lo que se traduce en horas perdidas, mayor consumo de combustible, mayor contaminación y un creciente impacto sobre la salud pública (Lieberthal et al., 2024; United Nations, 2018). La gestión de los semáforos en intersecciones urbanas es uno de los factores que mayor influencia tiene sobre la fluidez del tráfico (Zaghal et al., 2018). Sin embargo, la mayor parte de los semáforos del mundo siguen operando mediante ciclos de tiempos fijos que,

*Autor para correspondencia: dicaball@inf.uc3m.es

pese a poder variar a lo largo del día, no se ajustan en tiempo real a las condiciones dinámicas del tráfico.

Los sistemas adaptativos comerciales más extendidos, como SCATS (Sims and Dobinson, 1980) o SCOOT (Robertson and Bretherton, 1991), han demostrado mejoras significativas, pero presentan dos limitaciones importantes:

1. Dependen de planes de fases diseñados manualmente.
2. Requieren una infraestructura masiva de sensores físicos cuya implantación es costosa y compleja.

En este contexto, el aprendizaje por refuerzo profundo (*Deep Reinforcement Learning*, DRL) (Arulkumaran et al., 2017) ofrece un paradigma disruptivo, capaz de aprender políticas de control directamente a partir de la interacción con el entorno y, especialmente, a partir de entradas sensoriales de alta dimensión, como imágenes.

Este trabajo presenta el diseño, la implementación y la validación de un agente DRL para la gestión inteligente de semáforos. La principal aportación es que la única información del entorno utilizada por el agente son imágenes de la intersección, eliminando la dependencia de detectores específicos en la vía. Frente a la mayoría de enfoques DRL, que parten de representaciones numéricas del estado (vectores de colas, velocidades o tiempos de espera) o de imágenes de baja fidelidad gráfica, el empleo de un simulador fotorrealista hace que la distribución visual de entrada se aproxime a la de una cámara de tráfico real. Este realismo es un requisito previo para una eventual transferencia del agente a entornos físicos (*sim-to-real*) sin un cambio drástico en las observaciones, algo que los simuladores de baja fidelidad no permiten. El agente se valida en una plataforma de co-simulación que integra CARLA (Dosovitskiy et al., 2017) y SUMO (Lopez et al., 2018), y se evalúa en tres escenarios:

1. En la intersección utilizada para el entrenamiento.
2. En una intersección con topología distinta, para analizar la capacidad de generalización.
3. En un escenario multi-agente, para estudiar las sinergias entre intersecciones adyacentes.

2. Estado del arte

El control de semáforos mediante DRL es una línea de investigación muy activa (Eom and Kim, 2020; Tomar et al., 2022). Algunos de los trabajos más representativos utilizan variantes del algoritmo *Deep Q-Network* (DQN) sobre representaciones numéricas del estado, como vectores con posiciones, velocidades, longitudes de cola o tiempos de espera de los vehículos. Por ejemplo, Liang et al. (2019) introducen un modelo 3DQN que combina *Dueling DQN*, *Double Q-learning* y *Prioritized Experience Replay*, obteniendo mejoras consistentes frente a controladores de tiempos fijos. Bouktif et al. (2023) proponen un enfoque centrado en la coherencia entre estado y recompensa que mejora la estabilidad de la convergencia, mientras que Cai and Wei (2024) incorporan una *noise network* para mejorar la robustez frente a condiciones variables. En el ámbito multi-agente, Benhamza et al. (2024) proponen un esquema de aprendizaje por refuerzo multi-agente con recompensas mutuas para varias intersecciones, aunque sin comparar de forma explícita el desempeño individual frente al multi-agente.

El uso de imágenes como entrada al agente es mucho menos frecuente. Lee et al. (2020) emplean imágenes generadas en el simulador Vissim, cuya fidelidad gráfica es limitada. Este aspecto compromete la transferibilidad a escenarios reales, donde las condiciones visuales son notablemente distintas a las usadas en entrenamiento. Frente a estos trabajos, el agente aquí propuesto se entrena exclusivamente con imágenes fotorrealistas generadas por CARLA y se evalúa, además, en una configuración multi-agente que permite cuantificar la sinergia entre intersecciones controladas por agentes adyacentes.

3. Arquitectura del sistema

El sistema desarrollado integra cuatro componentes principales: el entorno de co-simulación CARLA-SUMO, un módulo de sensores virtuales, los agentes DQN encargados del control de los semáforos y un nodo de infraestructura responsable de la orquestación global. La Figura 1 muestra una visión general de la arquitectura.

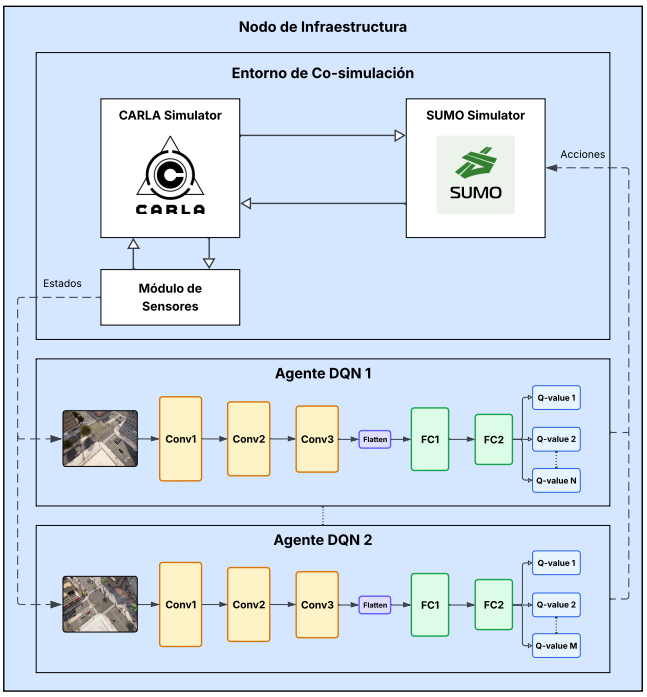


Figura 1: Arquitectura general del sistema propuesto.

3.1. *Entorno de co-simulación*

La hipótesis central del trabajo es que una imagen de la intersección contiene información suficiente para tomar decisiones de control adecuadas. Por ello, el realismo visual del entorno simulado resulta crítico. CARLA (Dosovitskiy et al., 2017) se selecciona como motor de renderizado y de física por su elevada fidelidad gráfica. Su gestor de tráfico nativo, sin embargo, ofrece menos flexibilidad que la de simuladores microscópicos especializados. Para superar esta limitación, se utiliza una co-simulación con SUMO (Lopez et al., 2018), el cual controla la lógica del tráfico (generación de vehículos,

asignación de rutas, estado de los semáforos) mientras que CARLA replica visualmente el estado. La sincronización se realiza con un *time-step* de 0,1 s, el máximo permitido por CARLA para que su motor físico funcione correctamente.

3.2. Diseño del agente DQN

El núcleo del sistema lo constituyen agentes DQN (Mnih et al., 2015) basados en redes de neuronas convolucionales (CNN) cuyo proceso de decisión se formaliza como un proceso de decisión de Markov (MDP) (Sutton and Barto, 2018).

Espacio de estados. El estado se representa mediante una imagen RGB capturada por una cámara virtual situada en una posición elevada sobre la intersección. Cada imagen se redimensiona a 100×100 píxeles, se convierte a escala de grises y se normaliza al rango $[0, 1]$. Esta combinación reduce drásticamente el número de parámetros de la red sin sacrificar información relevante.

Espacio de acciones. Las acciones son discretas y cada una corresponde a la activación de una de las fases de los semáforos de la intersección. Tras seleccionar una acción, la fase verde se mantiene durante un tiempo mínimo $T_G = 7\,\text{s}$. Si la siguiente acción implica cambiar de fase, se intercala una fase de ámbar de $T_Y = 3\,\text{s}$, de modo que las acciones de cambio duran $T_G + T_Y$ segundos. Adicionalmente, se incluye una restricción de seguridad que fuerza un cambio de fase si algún semáforo lleva más de $T_S = 45\,\text{s}$ en rojo.

Función de recompensa. La recompensa en el paso t se define como el negativo del tiempo de espera acumulado de todos los vehículos presentes en los carriles de entrada,

$$r_t = -W_t = -\sum_{i=1}^{N_t} w_{i,t} \tag{1}$$

donde N_t es el número de vehículos en la intersección en el paso t y $w_{i,t}$ es el tiempo de espera del vehículo i.

Sin embargo, para estabilizar el aprendizaje, se aplica al valor almacenado en el búfer de experiencias la siguiente transformación logarítmica:

$$r'_t = \text{sgn}(r_t)\ \log(1 + |r_t|) \tag{2}$$

Esta selección está respaldada por trabajos previos (Li et al., 2020), donde el tiempo de espera se identifica como una de las señales de recompensa más eficaces para el control adaptativo de semáforos.

Arquitectura de la red. La red está compuesta por tres capas convolucionales sin *pooling*, seguidas de dos capas totalmente conectadas (Figura 2). Se prescinde del *pooling* porque la posición espacial de los vehículos en la imagen es informativa para decidir qué semáforo poner en verde. La capa de salida tiene tantas neuronas como acciones posibles y proporciona el valor Q asociado a cada una.

Entrenamiento. El entrenamiento se organiza en episodios de 200 acciones, que equivale aproximadamente a 30 minutos de tráfico simulado. Se emplea una política ϵ-*greedy* con decaimiento exponencial, un búfer de *replay* de 50 000 experiencias, una red objetivo actualizada cada 2 000 acciones y el optimizador Adam (Kingma and Ba, 2014) con tasa de aprendizaje 10^{-4}. Se utiliza la función de pérdida *Huber loss* (Huber, 1964), más robusta a valores atípicos que el error cuadrático medio durante las fases iniciales del entrenamiento.

4. Experimentación y resultados

La intersección utilizada para entrenar el agente, mostrada en la Figura 3, es una intersección de cuatro patas con cuatro semáforos independientes mutuamente excluyentes. El entrenamiento se realiza en condiciones de tráfico saturado con 30 vehículos aleatorios por episodio, lo que asegura que el agente aprende a gestionar situaciones de máxima exigencia. Todos los experimentos se ejecutaron en un servidor con GPU NVIDIA RTX 4090 y CPU AMD Ryzen 9 7950X3D.

Figura 3: Intersección de entrenamiento.

La evaluación se efectúa comparando al agente con un controlador de tiempos fijos en dos densidades de tráfico (30 y 50 vehículos por episodio) y dos condiciones distintas: “Tráfico real”, en el que las intersecciones no controladas operan con tiempos fijos, y “Tráfico saturado”, en el que el resto de semáforos se mantienen en verde, generando un flujo continuo y exigente. Para cada configuración se promedian los resultados de 10 episodios y se analizan tres métricas:

- Recompensa promedio (Avg Reward).
- Tiempo de espera promedio por vehículo (Avg WT).
- Número total de vehículos que atraviesan la intersección (Vehicles).

4.1. Intersección de entrenamiento

La Tabla 1 resume el desempeño del agente en su intersección de entrenamiento. En condiciones de tráfico real con 30 vehículos, el agente reduce el tiempo de espera promedio en un 22,68 % y mejora la recompensa en un 26,06 %. En condiciones saturadas con 50 vehículos, las mejoras se mantienen (12,74 % y 5,68 %) y, además, el agente incrementa el número de vehículos procesados, lo que muestra su capacidad para

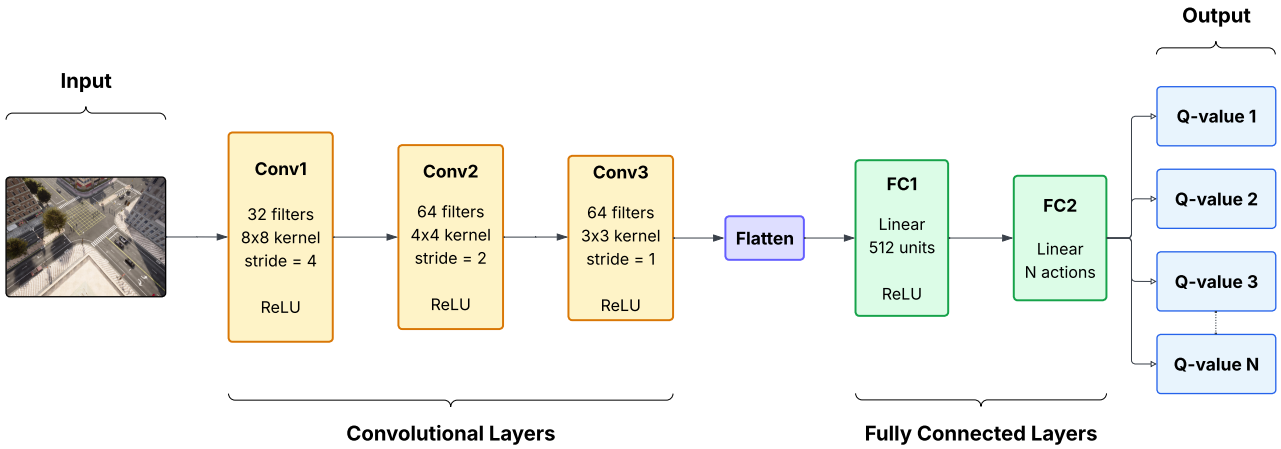


Figura 2: Arquitectura de la red de neuronas del agente DQN.

aumentar la fluidez en escenarios de máxima congestión. El ligero descenso del 0,59 % en el número de vehículos procesados en el escenario de tráfico real con 30 vehículos corresponde a una diferencia de cuatro vehículos (Tabla A.1). A baja densidad, la intersección no llega a congestionarse, de modo que casi todos los vehículos la atraviesan con independencia de la política aplicada. En los escenarios de mayor densidad, donde sí existe congestión y el control marca una diferencia real, el agente incrementa tanto la fluidez como el número de vehículos procesados.

Tabla 1: Comparativa entre el agente DQN y el controlador de tiempos fijos en la intersección de entrenamiento (mejora porcentual).

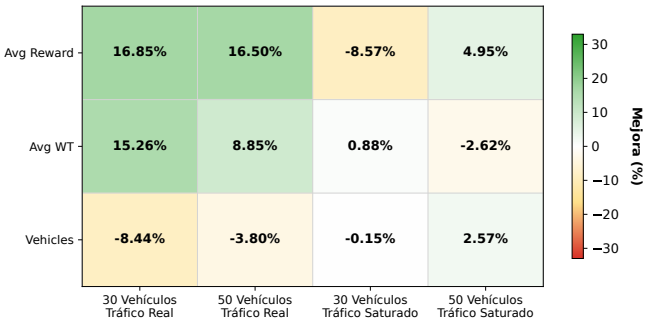
	Tráfico real		Tráfico saturado	
Métrica	30 veh.	50 veh.	30 veh.	50 veh.
Avg Reward	+26,06 %	+14,86 %	+6,80 %	+5,68 %
Avg WT	+22,68 %	+17,07 %	+12,08 %	+12,74 %
Vehicles	-0,59 %	+3,82 %	+4,81 %	+2,13 %

4.2. *Generalización a una nueva intersección*

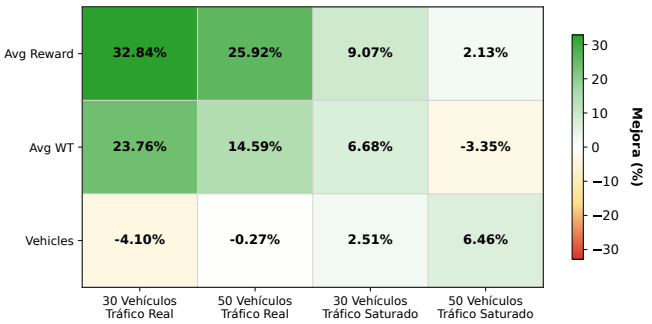
A continuación, se evalúa la capacidad de generalización del agente desplegándolo en una intersección adyacente con tres semáforos en lugar de cuatro (Figura 4). Se aplica un enmascaramiento de acciones para descartar fases inválidas.

Figura 4: Nueva intersección utilizada en el estudio de generalización.

Para aprovechar el conocimiento previo, se aplica una estrategia de *fine-tuning* sobre la nueva intersección, en la que se congelan las capas convolucionales y solo se ajustan las dos capas totalmente conectadas. Como muestra la Figura 5, los resultados son mixtos. En tráfico real el agente reduce el tiempo de espera hasta un 15,26 %, pero en condiciones saturadas degrada algunas métricas.



(a) Resultados de *fine-tuning*



(b) Resultados de reentrenamiento.

Figura 5: Matrices de mejoras en la nueva intersección con la estrategia de *fine-tuning* (a) y de reentrenamiento (b).

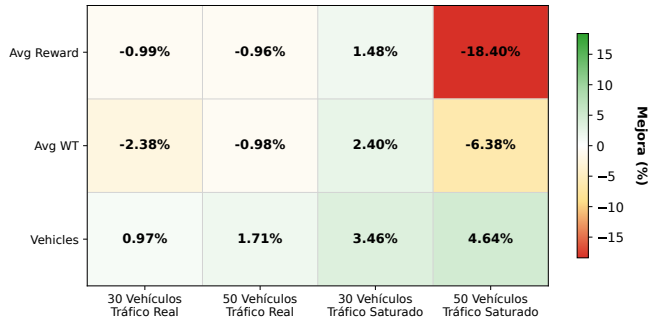
Al reentrenar el agente desde cero se obtienen mejoras claras y robustas: hasta un 23,76 % en el tiempo de espera y un 32,84 % en la recompensa, frente al sistema de tiempos fijos.

Estos resultados sugieren que las representaciones visuales aprendidas son parcialmente transferibles, pero la política de control depende fuertemente de la geometría concreta de la intersección, lo que hace recomendable el reentrenamiento cuando la topología cambia de forma sustancial.

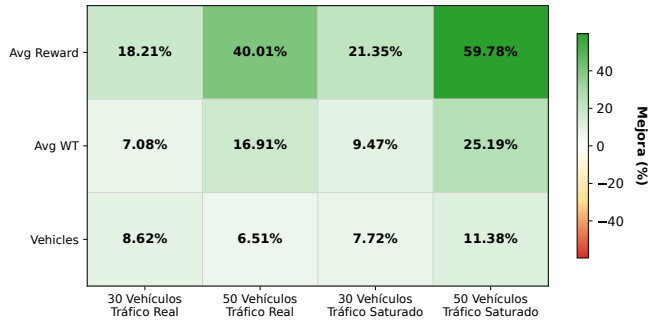
4.3. Escenario multi-agente

Por último, se estudia un escenario multi-agente en el que dos agentes operan simultáneamente sobre las dos intersecciones anteriores, adyacentes entre sí. En este experimento, cada intersección es gestionada por su mejor agente disponible: el agente entrenado sobre la intersección de entrenamiento y el agente reentrenado por completo sobre la nueva intersección.

El objetivo es analizar si la presencia de un agente vecino mejora el rendimiento individual de cada agente. Los resultados, sintetizados en las matrices de mejoras de la Figura 6, revelan un comportamiento marcadamente asimétrico.



(a) Intersección de entrenamiento



(b) Nueva intersección

Figura 6: Matrices de mejoras del esquema multi-agente respecto a la configuración de un único agente en cada intersección.

En la intersección de entrenamiento, la activación del segundo agente apenas modifica el desempeño del primero, salvo una degradación notable de la recompensa en el caso más saturado. Por el contrario, la activación del primer agente provoca una mejora muy notable en la segunda intersección: la recompensa aumenta hasta un 59,78 %, el tiempo de espera promedio se reduce hasta un 25,19 % y el número de vehículos que la atraviesan crece hasta un 11,38 %.

Esta asimetría se explica porque la primera intersección concentra el flujo procedente de gran parte del mapa, mientras que la segunda gestiona principalmente tráfico que se origina en, o se dirige hacia, la primera. Cuando el primer agente regula la salida de vehículos, el segundo recibe un flujo más ordenado que puede optimizar de forma más efectiva. Este efecto sugiere que un despliegue distribuido, con un agente por intersección, podría producir mejoras agregadas significativas a escala de barrio o ciudad.

5. Conclusiones

Este trabajo demuestra la viabilidad de un agente de control de semáforos basado únicamente en visión por ordenador y aprendizaje por refuerzo profundo, sin requerir infraestructura específica de detección. El agente propuesto supera de forma consistente al control clásico de tiempos fijos en su intersección de entrenamiento, con reducciones de hasta el 22,68 % en el tiempo de espera de los vehículos. Además, pese a haberse entrenado en un único régimen de tráfico (saturado y con 30 vehículos), el agente conserva estas mejoras en densidades y condiciones no observadas durante el aprendizaje, evidenciando su robustez frente a escenarios de tráfico diversos. Esta robustez, sin embargo, no se traslada a la topología: la transferencia directa del modelo preentrenado a una intersección con geometría distinta es limitada, pero el reentrenamiento completo permite obtener mejoras de en torno al 23 %, como en el primer caso. En escenarios multi-agente, la coordinación implícita entre agentes que controlan intersecciones adyacentes produce sinergias positivas relevantes. Como líneas futuras se plantea el desarrollo de estrategias avanzadas de coordinación multi-agente y de técnicas de transferencia de conocimiento que reduzcan la necesidad de reentrenamiento al cambiar de intersección.

Agradecimientos

Este trabajo ha sido financiado por la Red Temática ARTIFICIAL INTELLIGENCE FOR INTELLIGENT, AUTONOMOUS, SUSTAINABLE AND SAFE MOBILITY (IA4MOV*), referencia RED2024-153662-T, financiada por el Ministerio de Ciencia, Innovación y Universidades de España, en el marco del Plan Estatal de Investigación Científica, Técnica y de Innovación 2024-2027, así como por las ayudas MCIN/AEI/10.13039/501100011033 con referencias PID2022-140554OB-C32 y PDC2025-165871-C32, y por SEGVAUTO 5Gen-CM (TEC-2024/ECO-277), financiado por la Comunidad de Madrid.

Apéndice A. Tablas de resultados

Las Tablas A.1 a A.3 muestran, para cada configuración evaluada, los valores absolutos de cada métrica promediados en 10 episodios y la correspondiente mejora porcentual.

Apéndice A.1. Intersección de entrenamiento

Tabla A.1: Resultados detallados en la intersección de entrenamiento.

	Tráfico real		Tráfico saturado	
Métrica	30 veh.	50 veh.	30 veh.	50 veh.
Tiempos fijos				
Avg Reward	-41,86	-78,83	-84,48	-136,02
Avg WT	16,39	18,00	18,03	20,59
Vehicles	647,40	964,00	1048,00	1251,50
Agente DQN				
Avg Reward	-30,95	-67,12	-78,74	-128,30
Avg WT	12,68	14,92	15,85	17,97
Vehicles	643,60	1000,80	1098,40	1278,10
Mejora (%)				
Avg Reward	+26,06 %	+14,86 %	+6,80 %	+5,68 %
Avg WT	+22,68 %	+17,07 %	+12,08 %	+12,74 %
Vehicles	-0,59 %	+3,82 %	+4,81 %	+2,13 %

Apéndice A.2. Nueva intersección

Tabla A.2: Resultados detallados en la nueva intersección. Comparación entre sistema de tiempos fijos y agente DQN con estrategia de *fine-tuning* y con reentrenamiento completo.

	Tráfico real		Tráfico saturado	
Métrica	30 veh.	50 veh.	30 veh.	50 veh.
Tiempos fijos				
Avg Reward	-23,20	-114,01	-36,18	-82,00
Avg WT	12,93	18,77	11,71	12,26
Vehicles	555,70	776,30	1147,10	1468,40
Agente fine-tuned				
Avg Reward	-19,30	-95,20	-39,28	-77,95
Avg WT	10,96	17,11	11,61	12,59
Vehicles	508,80	746,80	1145,40	1506,20
Mejora del fine-tuning (%)				
Avg Reward	+16,85 %	+16,50 %	-8,57 %	+4,95 %
Avg WT	+15,26 %	+8,85 %	+0,88 %	-2,62 %
Vehicles	-8,44 %	-3,80 %	-0,15 %	+2,57 %
Agente reentrenado				
Avg Reward	-15,58	-84,47	-32,90	-80,25
Avg WT	9,86	16,03	10,93	12,68
Vehicles	532,90	774,20	1175,90	1563,30
Mejora del reentrenamiento (%)				
Avg Reward	+32,84 %	+25,92 %	+9,07 %	+2,13 %
Avg WT	+23,76 %	+14,59 %	+6,68 %	-3,35 %
Vehicles	-4,10 %	-0,27 %	+2,51 %	+6,46 %

Apéndice A.3. Evaluación multi-agente

Tabla A.3: Resultados detallados de la comparación multi-agente con agente único en ambas intersecciones.

	Tráfico real		Tráfico saturado	
Métrica	30 veh.	50 veh.	30 veh.	50 veh.
Intersección entrenamiento – Agente único				
Avg Reward	-33,27	-72,47	-63,50	-110,97
Avg WT	12,84	14,97	15,28	17,04
Vehicles	681,00	1052,70	944,00	1199,40
Intersección entrenamiento – Multi-agente				
Avg Reward	-33,60	-73,17	-62,55	-131,38
Avg WT	13,14	15,12	14,91	18,13
Vehicles	687,60	1070,70	976,70	1255,10
Intersección entrenamiento – Mejora multi-agente vs único (%)				
Avg Reward	-0,99 %	-0,96 %	+1,48 %	-18,40 %
Avg WT	-2,38 %	-0,98 %	+2,40 %	-6,38 %
Vehicles	+0,97 %	+1,71 %	+3,46 %	+4,64 %
Nueva intersección – Agente único				
Avg Reward	-16,97	-56,82	-37,09	-158,53
Avg WT	10,33	13,54	12,47	17,73
Vehicles	564,90	881,70	800,70	976,10
Nueva intersección – Multi-agente				
Avg Reward	-13,88	-34,09	-29,17	-63,76
Avg WT	9,60	11,25	11,29	13,26
Vehicles	613,60	939,10	862,50	1087,20
Nueva intersección – Mejora multi-agente vs agente único (%)				
Avg Reward	+18,21 %	+40,01 %	+21,35 %	+59,78 %
Avg WT	+7,08 %	+16,91 %	+9,47 %	+25,19 %
Vehicles	+8,62 %	+6,51 %	+7,72 %	+11,38 %

Referencias

Arulkumaran, K., Deisenroth, M. P., Brundage, M., Bharath, A. A., 2017. Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine 34 (6), 26–38.
DOI: 10.1109/MSP.2017.2743240

Benhamza, K., Seridi, H., Agguini, M., Bentagine, A., 2024. A multi-agent reinforcement learning based approach for intelligent traffic signal control. Evolving Systems 15 (6), 2383–2397.
DOI: 10.1007/s12530-024-09622-4

Bouktif, S., Cheniki, A., Ouni, A., El-Sayed, H., 2023. Deep reinforcement learning for traffic signal control with consistent state and reward design approach. Knowledge-Based Systems 267, 110440.
DOI: 10.1016/j.knosys.2023.110440

Cai, C., Wei, M., 2024. Adaptive urban traffic signal control based on enhanced deep reinforcement learning. Scientific Reports 14.
DOI: 10.1038/s41598-024-64885-w

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V., 2017. CARLA: An open urban driving simulator. In: Proceedings of the 1st Annual Conference on Robot Learning. pp. 1–16.

Eom, M., Kim, B.-I., 2020. The traffic signal control problem for intersections: a review. European Transport Research Review 12, 1–20.
DOI: 10.1186/s12544-020-00440-8

Huber, P. J., 1964. Robust estimation of a location parameter. The Annals of Mathematical Statistics 35 (1), 73–101.
DOI: 10.1214/aoms/1177703732

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. CoRR abs/1412.6980.

Lee, J., Chung, J., Sohn, K., 2020. Reinforcement learning for joint control of traffic signals in a transportation network. IEEE Transactions on Vehicular Technology 69, 1375–1387.
DOI: 10.1109/TVT.2019.2962514

Li, D., Wu, J., Xu, M., Wang, Z., Hu, K., 2020. Adaptive traffic signal control model on intersections based on deep reinforcement learning. Journal of Advanced Transportation.
DOI: 10.1155/2020/6505893

Liang, X., Du, X., Wang, G., Han, Z., 2019. A deep reinforcement learning network for traffic light cycle control. IEEE Transactions on Vehicular Technology 68 (2), 1243–1253.
DOI: 10.1109/TVT.2018.2890726

Lieberthal, E. B., Serok, N., Duan, J., Zeng, G., Havlin, S., 2024. Addressing the urban congestion challenge based on traffic bottlenecks. Philosophical Transactions of the Royal Society A 382 (2285), 20240095.
DOI: 10.1098/rsta.2024.0095

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., Wiessner, E., 2018. Microscopic traffic simulation using SUMO. In: 21st International Conference on Intelligent Transportation Systems. pp. 2575–2582.
DOI: 10.1109/ITSC.2018.8569938

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., 2015. Human-level control through deep reinforcement learning. Nature 518 (7540), 529–533.
DOI: 10.1038/nature14236

Robertson, D. I., Bretherton, R. D., 1991. Optimizing networks of traffic signals in real time—the SCOOT method. IEEE Transactions on Vehicular Technology 40 (1), 11–15.
DOI: 10.1109/25.69966

Sims, A. G., Dobinson, K. W., 1980. The Sydney coordinated adaptive traffic (SCAT) system philosophy and benefits. IEEE Transactions on Vehicular Technology 29 (2), 130–137.
DOI: 10.1109/T-VT.1980.23833

Sutton, R. S., Barto, A. G., 2018. Reinforcement Learning: An Introduction, 2nd Edition. The MIT Press.

Tomar, I., Indu, S., Pandey, N., 2022. Traffic signal control methods: Current status, challenges, and emerging trends. In: Proceedings of Data Analytics and Management. Springer Nature Singapore, pp. 151–163.
DOI: 10.1007/978-981-16-6289-8_14

TomTom International BV, 2025. Traffic index ranking. https://www.tomtom.com/traffic-index/ranking/.

United Nations, 2018. World urbanization prospects: The 2018 revision. Tech. rep., United Nations, Department of Economic and Social Affairs.

Zaghal, R., Thabatah, K., Salah, S., 2018. Towards a smart intersection using traffic load balancing algorithm. Proceedings of Computing Conference 2017, 485–491.
DOI: 10.1109/SAI.2017.8252141