

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS

**EVALUACIÓN COMPARATIVA DE LA FAMILIA DE MODELOS YOLO11 PARA
ANALÍTICA DE TRÁFICO URBANO**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

AUTOR

JOSÉ LUIS DE LOS SANTOS MUÑOZ

PROFESOR GUÍA

WERNER KRISTJANPOLLER R.

PROFESOR CORREFERENTE

NICOLÁS JARA C.

VALPARAÍSO, SEPTIEMBRE DE 2026



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Evaluación comparativa de la familia de modelos YOLO11 para analítica de tráfico urbano_____

Nombre del candidato(a): José Luis De los Santos Muñoz _____

Carrera / Grado: Ingeniero Civil Industrial _____

Campus: Casa Central Valparaíso _____ Departamento: Departamento de Industrias _____

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, _____ Werner Kristjanpoller R. _____, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

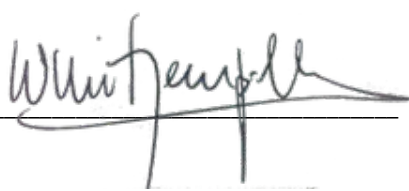
☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 24/09/26 Firma: 

Estudiante o Candidato(a):

Fecha: 17/09/26 Firma: 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Cuando entré a esta Universidad veía todo tan cuesta arriba que pensé: si algún día logro salir de acá, sé que seré capaz de cualquier cosa.

Lo logré.

Agradezco a mis amigos y amigas, en especial a Mila, Ana, Kevin, Seba Doro y Sella, a mis seres queridos, a toda la gente que me ha dado su cariño y a mi papá y a mi mamá.

Dedico esta memoria a Rem, a Sombrita, a mis tíos Pancho, Leo, Nani y Pablo, a mi papá Baltazar y a mi mamá Lily.

Resumen

La gestión del tránsito urbano depende de saber qué circula por cada intersección y en qué momento, y las cámaras ya instaladas en las ciudades pueden entregar ese dato si se les suma un detector de objetos capaz de operar en tiempo real. Elegir cuál detector desplegar, sin embargo, suele resolverse mirando solo la precisión sobre un conjunto de prueba, sin preguntarse cuánto vale esa precisión adicional una vez puesta en la calle.

Esta memoria evalúa los cinco tamaños de la familia YOLO11 (n, s, m, l y x) sobre BDD100K y traduce su desempeño técnico a valor social, con el fin de determinar cuál ofrece el mejor equilibrio entre calidad y costo. Los cinco modelos se ajustaron bajo un protocolo idéntico sobre un subconjunto de 15.000 imágenes de entrenamiento y 3.000 de validación, cada una etiquetada por momento del día, clima y densidad vehicular, lo que permitió medir el desempeño dentro de cada condición en lugar de un promedio que las mezcla. La robustez se estimó por ponderación de probabilidad inversa y por comparaciones con las variables de confusión fijadas, y el costo se midió con telemetría de inferencia sobre GPU. Ese desempeño se llevó a una valorización social sobre las cerca de 3.250 intersecciones semaforizadas de la Región Metropolitana, con precios sociales y una tasa de descuento de 5,5%, valorizando el ahorro de tiempo por menor congestión, los siniestros evitados y las emisiones de CO₂ evitadas, más un análisis de riesgo por simulación de Montecarlo de 100.000 iteraciones sobre 18 parámetros.

El mAP50-95 crece de 0,244 en YOLO11n a 0,338 en YOLO11x y se estanca a partir de YOLO11m, cuyo intervalo de confianza al 95% se solapa con el de los dos modelos mayores. La noche resulta ser la condición crítica para los cinco tamaños, con caídas de entre 22% y 29% del mAP diurno bajo clima y densidad fijos, una magnitud que supera a cualquier diferencia entre condiciones climáticas. En el plano económico, el costo de cómputo representa entre 0,05% y 0,40% del costo total del sistema, frente al 68,2% de la infraestructura de control adaptativo, de modo que la elección del detector no se decide por costo. El sistema es socialmente rentable en todos los escenarios evaluados, con una relación beneficio-costo de entre 2,03 y 2,75 en dólares constantes de 2026, y el valor actual neto de YOLO11m resulta positivo en el 100% de las iteraciones simuladas, con un valor en riesgo al 5% de US\$221 millones.

Se recomienda **YOLO11m**, que captura el 96% del beneficio social máximo alcanzable. El 4% adicional que ofrecen los modelos mayores cabe dentro del margen de incertidumbre de la propia valorización, ya que escalar hasta YOLO11x mejora la relación beneficio-costo en 0,09 puntos mientras esa relación se mueve en una banda de 2,07 puntos con 90% de probabilidad. El aporte metodológico del trabajo es el procedimiento que conecta precisión por condición de operación con valor social, disponible para evaluar los detectores que vengan.

Palabras clave: detección de objetos, YOLO11, visión computacional, analítica de tráfico urbano, semáforos adaptativos, evaluación social de proyectos, simulación de Montecarlo, BDD100K.

Abstract

Urban traffic management depends on knowing what circulates through each intersection and at what time, and the cameras already installed across cities can provide that data once an object detector capable of running in real time is added to them. Choosing which detector to deploy, however, is usually settled by looking only at accuracy on a test set, without asking how much that additional accuracy is worth once the system is on the street.

This thesis evaluates the five model sizes of the YOLO11 family (n, s, m, l and x) on BDD100K and translates their technical performance into social value, in order to determine which one offers the best balance between quality and cost. The five models were fine-tuned under an identical protocol on a subset of 15,000 training images and 3,000 validation images, each labelled by time of day, weather and vehicle density, which made it possible to measure performance within each condition rather than as an average that blends them. Robustness was estimated through inverse probability weighting and through comparisons with the confounding variables held fixed, while cost was measured with GPU inference telemetry. That performance was then carried into a social appraisal covering the roughly 3,250 signalized intersections of the Santiago Metropolitan Region, using social prices and a 5.5% discount rate, valuing the travel time saved through reduced congestion, the road crashes avoided and the CO₂ emissions avoided, together with a risk analysis based on a Monte Carlo simulation of 100,000 iterations over 18 parameters.

The mAP50-95 rises from 0.244 for YOLO11n to 0.338 for YOLO11x and levels off from YOLO11m onwards, whose 95% confidence interval overlaps with those of the two larger models. Night proves to be the critical condition for all five sizes, with drops of between 22% and 29% of the daytime mAP under fixed weather and density, a magnitude that exceeds any difference between weather conditions. On the economic side, computing accounts for between 0.05% and 0.40% of the total system cost, against 68.2% for the adaptive control infrastructure, so the choice of detector is not settled on cost grounds. The system is socially profitable across every scenario evaluated, with a benefit-cost ratio of between 2.03 and 2.75 in constant 2026 dollars, and the net present value of YOLO11m is positive in 100% of the simulated iterations, with a 5% value at risk of US\$221 million.

YOLO11m is recommended, as it captures 96% of the maximum attainable social benefit. The additional 4% offered by the larger models falls within the uncertainty margin of the appraisal itself, since scaling up to YOLO11x improves the benefit-cost ratio by 0.09 points while that ratio moves within a band of 2.07 points with 90% probability. The methodological contribution of this work is the procedure that links accuracy by operating condition to social value, available for evaluating the detectors that follow.

Keywords: object detection, YOLO11, computer vision, urban traffic analytics, adaptive traffic signals, social cost-benefit analysis, Monte Carlo simulation, BDD100K.

Índice general

Resumen	1
Abstract	2
1) Problema de investigación	8
2) Objetivos	11
2.1) Objetivo general	11
2.2) Objetivos específicos	12
2.3) Alcance	12
3) Marco teórico	13
3.1) Visión computacional y detección de objetos	13
3.2) La familia YOLO y YOLO11	14
3.3) Métricas de evaluación en detección de objetos	17
3.4) Robustez de los detectores ante condiciones adversas	18
3.5) Analítica de tráfico urbano, Smart Cities y gemelos digitales	19
3.6) Economía de la inferencia: edge vs cloud y costo total de propiedad	21
3.7) El dataset BDD100K	22
3.8) Estado del arte y brecha	22
4) Metodología	23
4.1) Datos y preparación (Obj. 1)	24
4.2) Entrenamiento (Obj. 2)	26
4.3) Evaluación de precisión y robustez (Obj. 3)	27
4.4) Evaluación operacional y de costo (Obj. 4)	33
4.5) Valorización social del impacto y recomendación (Obj. 5)	35
4.6) Diseño del análisis de riesgo (Obj. 6)	46
5) Resultados	50
5.1) Desempeño global y costo	50
5.2) Robustez por momento del día (Objetivo 3)	55

5.3) Desempeño por clima	57
5.3.1) Análisis climático restringido a imágenes diurnas	61
5.3.2) Comparaciones controladas: momento del día y clima	64
5.4) Desempeño por densidad vehicular	68
5.5) Telemetría y estrés de hardware (Objetivo 4)	70
5.6) Síntesis de hallazgos de precisión y robustez	72
6) Valorización económica y discusión	74
6.1) Estructura del costo y del beneficio	74
6.2) Resultados de la valorización (escenario base)	77
6.3) Análisis de riesgo mediante simulación de Montecarlo	82
7) Discusión y recomendación	88
8) Limitaciones y perspectivas futuras	92
8.1) Limitaciones de los datos	92
8.2) Limitaciones del traslado de parámetros	93
8.3) Limitaciones del modelo económico	94
8.4) Limitaciones del diseño experimental	95
8.5) Perspectivas futuras	95
9) Conclusiones	96
Conclusión general	98
Referencias (APA)	99
Anexo: Curvas Precisión-Recall y Matrices de Confusión	110

Índice de figuras

Figura 1	11
Figura 2	14
Figura 3	17
Figura 4	23
Figura 5	26
Figura 6	52
Figura 7	52
Figura 8	52
Figura 9	56
Figura 10	56
Figura 11	57
Figura 12	60
Figura 13	61
Figura 14	64
Figura 15	64
Figura 16	65
Figura 17	66
Figura 18	67
Figura 19	68
Figura 20	69
Figura 21	70
Figura 22	71
Figura 23	72
Figura 24	78
Figura 25	79
Figura 26	80
Figura 27	81
Figura 28	82
Figura 29	84
Figura 30	85

Figura 31	86
Figura 32	91
Figura 33	110
Figura 34	112

Índice de tablas

Tabla 1	16
Tabla 2	27
Tabla 3	32
Tabla 4	33
Tabla 5	47
Tabla 6	50
Tabla 7	51
Tabla 8	55
Tabla 9	55
Tabla 10	57
Tabla 11	58
Tabla 12	59
Tabla 13	61
Tabla 14	62
Tabla 15	64
Tabla 16	66
Tabla 17	68
Tabla 18	69
Tabla 19	70
Tabla 20	76
Tabla 21	77
Tabla 22	77
Tabla 23	77
Tabla 24	79
Tabla 25	83
Tabla 26	90

1) Problema de investigación

En 2018 la población mundial que vivía en zonas urbanas alcanzaba 55%, y se proyecta que pasará a cerca de 68% al 2050, sumando alrededor de 2.500 millones de habitantes urbanos [1]. Este crecimiento intensifica la congestión, uno de los costos urbanos más significativos, ya que incluso en Estados Unidos, con infraestructura vial consolidada, el conductor promedio perdió 43 horas por congestión durante 2024 [2]. En América Latina el fenómeno es especialmente agudo, puesto que sus ciudades figuran de forma recurrente entre las más congestionadas del mundo. Según el INRIX Global Traffic Scorecard 2019, Bogotá fue la ciudad más congestionada del planeta, en el puesto 1, con 191 horas perdidas por conductor al año, seguida de Río de Janeiro con 190 horas y con Ciudad de México en torno a 158 horas [3]. El Banco Interamericano de Desarrollo estima, por su parte, que la congestión cuesta alrededor del 1,1% del PIB en metrópolis como Buenos Aires, São Paulo y Montevideo, llegando São Paulo a superar los US\$2.000 millones en 2019 [4]. En este contexto, Santiago de Chile registra pérdidas relevantes, dado que en 2019 la demora agregada por congestión alcanzó 194 millones de horas (29 horas por habitante, 76 horas por viajero en automóvil), lo que se tradujo en pérdidas anuales de US\$1.046 millones, equivalentes al 1,04% del PIB de la ciudad [4]. Reducir estos tiempos de viaje, aun de forma marginal, genera un ahorro relevante, ya que una disminución del 4% de la congestión equivaldría a cerca de US\$40 millones anuales para Santiago, y un sistema real de gestión adaptativa del tránsito, Surtrac, en su despliegue piloto de 2012 en Pittsburgh sobre una red que ya operaba con semaforización coordinada y actuada simple, alcanzó reducciones de más de 25% en los tiempos de viaje [5].

La congestión constituye un costo de tiempo y, además, el mismo flujo vehicular que la provoca deteriora el medioambiente y la seguridad vial. En lo ambiental, el transporte motorizado aporta un 33% de las emisiones de contaminantes en la Región Metropolitana, la segunda fuente tras la leña residencial (34%), según el Inventario de Emisiones 2022 [6]. En lo vial, la seguridad sigue siendo crítica, puesto que durante 2024 fallecieron 1.439 personas en siniestros de tránsito

en Chile, cuyas causas principales fueron la imprudencia del conductor, presente en el 69,1% de los siniestros, la desobediencia a la señalización con un 9%, la conducción bajo los efectos del alcohol con un 8,7% y la velocidad imprudente con un 6% [7]. Precisamente sobre esa última causa existe evidencia sistemática de que las cámaras de fotodetección reducen el exceso de velocidad y los siniestros mediante un efecto disuasorio (el conductor modera su conducta al saber que está siendo monitoreado), consistente entre estudios aunque de magnitud variable: revisiones sistemáticas reportan reducciones de 8% a 49% en siniestros cerca de las cámaras (mayoría entre 14% y 25%) y de 17% a 58% en siniestros graves o fatales en áreas más amplias (típicamente 30% a 40%) [8]. Dado que varias de estas causas, como el exceso de velocidad, el no respetar la señalización o los conflictos entre vehículos y peatones, pueden detectarse a partir del video de las cámaras urbanas, se justifica priorizar la detección confiable de vehículos y también de peatones y ciclistas.

La gestión inteligente del tránsito, que abarca el control adaptativo de semáforos y los gemelos digitales urbanos, se apoya cada vez más en la detección automática de objetos a partir del video de las cámaras de tránsito, cuyo conteo vehicular alimenta a ambos. Chile ya cuenta con infraestructura instalada para ello, puesto que en 2020 la Unidad Operativa de Control de Tránsito (UOCT) operaba 3.250 intersecciones semaforizadas en la Región Metropolitana y en 2024 inauguró un sistema integrado de cámaras con inteligencia artificial para la gestión de tránsito en Santiago [9, 10].

El componente central de todos estos sistemas es un modelo de detección de objetos capaz de identificar y contar vehículos, peatones y ciclistas en tiempo real, clasificarlos por tipo, y a partir de ello estimar flujo, densidad y velocidad por carril y detectar eventos anómalos como accidentes o congestión súbita, ya que de su exactitud depende el conteo y la estimación de flujo y densidad que alimentan tanto el control adaptativo como los gemelos digitales. Como toda la cadena de decisión se construye sobre esa medición inicial, si la detección falla el error se propaga aguas abajo y degrada la calidad del control, la planificación y el monitoreo urbano que dependen de ella.

El desempeño del detector cae de noche, con lluvia, niebla o alta densidad vehicular, justo cuando más se necesita una lectura confiable del tránsito, y esta degradación tiene un peso económico relevante: en Estados Unidos cerca del 49% de las muertes de ocupantes de vehículos ocurren de noche pese a que solo alrededor del 25% de la conducción es nocturna, con una tasa de mortalidad aproximadamente tres veces la diurna [11], y a escala agregada los siniestros viales costaron a Estados Unidos US\$340 mil millones en 2019 [12] y en Chile, durante 2024, se registraron 1.439 fallecidos [7]. Por eso la utilidad real de esta detección depende de su robustez ante esas condiciones adversas: como la degradación ocurre justamente de noche y con lluvia, las mismas condiciones que concentran una mortalidad varias veces mayor a la diurna, un modelo que falla ahí falla donde el costo de cada error es más alto, de modo que su robustez por condición concentra un valor marginal alto. Por esta razón, medir la robustez por condición resulta central para evaluar la utilidad real de un detector en la gestión del tránsito.

El análisis del modelo a desplegar tendrá un enfoque tanto técnico como económico, puesto que implica equilibrar precisión, robustez y costo de operación a escala de una red de cámaras. Existe una comparación de YOLO en tráfico [13], que evalúa YOLO11 en detección vehicular y encuentra mejoras frente a YOLOv8 y YOLOv10 en la detección de vehículos pequeños y ocluidos, y un benchmark comparativo de toda la familia YOLO de v3 a v12 [14], que ubica a YOLO11 con el mejor equilibrio precisión-eficiencia de la serie, sin embargo, ambas comparan versiones distintas o variantes aisladas, ya que es menos frecuente una comparación entre precisión y costo a lo largo de los cinco tamaños, de n a x , que reporte de forma conjunta mAP, FPS, parámetros y latencia sobre un mismo dataset de tráfico urbano. Ese vacío constituye, precisamente, el problema de investigación de esta memoria.

Preguntas de investigación:

1. ¿Cuál es el costo operacional de cada modelo y cómo se traduce en un costo por inferencia comparable entre los cinco tamaños?
2. ¿Qué modelo ofrece el mejor desempeño por unidad de cómputo, medido como precisión

(mAP) por FLOP y por milisegundo de inferencia?

3. ¿Qué tanto se degrada el desempeño de cada modelo ante condiciones adversas (noche, clima y alta densidad vehicular) y cuál resulta más robusto?
4. ¿Qué modelo ofrece el mejor beneficio-costos para un sistema urbano de gestión de tránsito basado en detección por video en el Gran Santiago y qué valor social, ambiental y económico (en términos de retorno social de la inversión) se estima para su despliegue?

2) Objetivos

2.1) Objetivo general

Evaluar técnicamente los cinco modelos YOLO11 (n, s, m, l y x) sobre el conjunto de datos BDD100K en términos de precisión, robustez por condición y costo, y traducir ese desempeño técnico a su valor social, ambiental y económico, expresado como retorno social de la inversión, en el contexto de un sistema urbano de gestión de tránsito que integra el conteo vehicular, la estimación de flujo, el apoyo a los semáforos adaptativos y la detección de incidentes, con el fin de determinar el modelo que ofrece el mejor equilibrio entre calidad y costo para la toma de decisiones de gestión de tránsito en una ciudad inteligente.

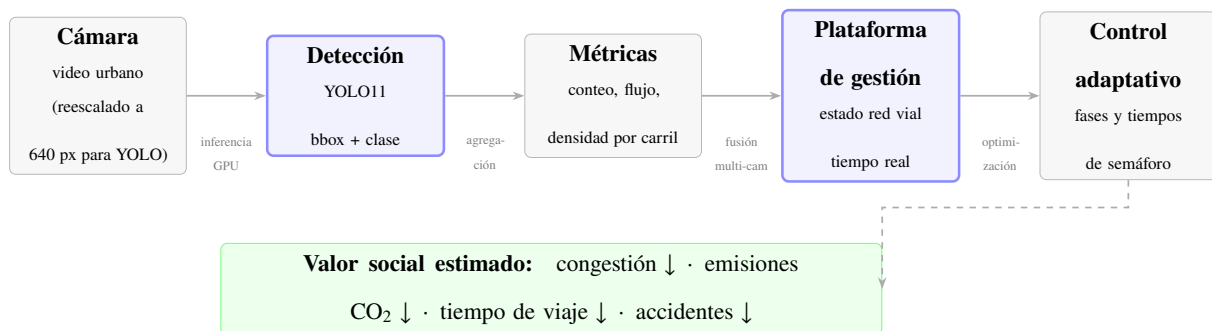


Figura 1: Arquitectura del sistema propuesto: desde la captura de video urbano hasta el control adaptativo de semáforos y su impacto social estimado. Con borde azul, los componentes analizados en esta memoria (detector YOLO11 y plataforma de gestión). La flecha discontinua representa el flujo de valor al sistema. Fuente: elaboración propia.

2.2) Objetivos específicos

1. Construir un subconjunto de imágenes de BDD100K, la base de datos de conducción urbana de Berkeley, con cada imagen etiquetada según su condición de operación (día, noche, clima y densidad de tráfico), de modo de permitir un análisis desagregado del desempeño.
2. Realizar el fine-tuning de los cinco modelos YOLO11 (n, s, m, l y x) con los datos de BDD100K, bajo un protocolo de entrenamiento comparable.
3. Evaluar la precisión y la robustez de cada modelo por condición, cuantificando la degradación del desempeño ante condiciones adversas.
4. Medir el costo computacional y operacional de cada modelo y derivar un costo por inferencia comparable entre los cinco tamaños.
5. Realizar una valorización económica social del despliegue de los modelos en un sistema urbano de gestión de tránsito basado en detección por video, traduciendo a valor económico el ahorro de tiempo por menor congestión, el valor de los accidentes evitados, el beneficio ambiental por menores emisiones y las vidas salvadas en el área urbana del Gran Santiago, en Santiago de Chile.
6. Cuantificar la incertidumbre de la valorización social mediante un análisis de riesgo por simulación de Montecarlo sobre las distribuciones de parámetros de decisión.

2.3) Alcance

El estudio abarca la evaluación técnica y económica de los cinco modelos YOLO11 para la detección de objetos de tráfico urbano. El entrenamiento y la evaluación del desempeño se realizan sobre BDD100K, un conjunto de imágenes extraídas de videos de conducción reales, grabados desde el parabrisas de vehículos circulando por ciudades y carreteras de Estados Unidos, principalmente Nueva York y el área de la bahía de San Francisco, en las que cada objeto de tránsito

(vehículos, peatones, ciclistas, señales y semáforos, entre otros) está anotado con una caja delimitadora, y donde además cada imagen trae una etiqueta de la condición de la escena (momento del día, clima y tipo de entorno) [15]. Se trata de un conjunto grande y diverso, con 100.000 imágenes y cerca de 1,84 millones de objetos anotados [15], lo que permite medir la robustez ante día, noche y clima sin etiquetado manual adicional. BDD100K se emplea como referencia extrapolable del desempeño del detector, dado que la detección de vehículos, peatones y ciclistas comparte patrones visuales comunes entre entornos urbanos.

La valorización social se sitúa en el área urbana del Gran Santiago, en Santiago de Chile, tomando referencias y casos locales, como el costo de la congestión, la siniestralidad vial, las emisiones y la infraestructura ya instalada, de modo que los resultados sirvan a la toma de decisiones de una entidad pública, como una municipalidad o un ministerio. La incertidumbre se acota con un análisis de riesgo mediante simulación de Montecarlo sobre las distribuciones documentadas de cada uno.

Queda fuera del alcance el reentrenamiento sobre datos locales de Santiago, la implementación productiva del sistema de gestión de tránsito y la modificación de las arquitecturas de YOLO11, ya que el foco es la comparación de los cinco modelos y la valorización de su despliegue, no el desarrollo de un nuevo modelo.

3) Marco teórico

3.1) Visión computacional y detección de objetos

La detección de objetos es la tarea de visión computacional que consiste en localizar y clasificar simultáneamente los objetos presentes en una imagen, donde cada objeto se localiza con un recuadro rectangular que lo encierra, conocido como bounding box. A diferencia de la clasificación de imágenes, que asigna una etiqueta global, la detección entrega posición y clase de múltiples objetos por escena, lo que la hace central para aplicaciones de conducción autónoma y analítica de tráfico. Los enfoques modernos se basan en redes neuronales convolucionales

profundas y se agrupan en tres familias: las arquitecturas de dos etapas, como R-CNN, Fast R-CNN y Faster R-CNN, que primero proponen regiones candidatas y luego las clasifican, con alta precisión a costa de mayor lentitud; las de una etapa, como SSD, RetinaNet y la familia YOLO, que detectan en una sola pasada y resultan más rápidas [16, 14]; y las basadas en transformers, como DETR [17] y sus variantes en tiempo real RT-DETR [18]. Entre ellas, la familia YOLO es una de las más utilizadas para la detección en tiempo real, tanto en la industria como en la investigación, gracias a su buen balance entre precisión y velocidad, a un ecosistema maduro en torno a Ultralytics y a su facilidad de despliegue en dispositivos de borde con cómputo y memoria limitados. Estas características la han extendido a aplicaciones tan diversas como el control de calidad en líneas de producción, la agricultura de precisión y el monitoreo de tránsito [19, 20].



Figura 2: Diferencia entre clasificación (un recuadro global para toda la escena), detección (una caja por cada objeto) y segmentación (máscaras que siguen la silueta de cada objeto). Fuente: elaboración propia.

3.2) La familia YOLO y YOLO11

YOLO introdujo la detección en una sola pasada de la red, dividiendo la imagen en una grilla y prediciendo simultáneamente cajas y probabilidades de clase, lo que permite detección en tiempo real [21]. Desde entonces la familia ha evolucionado a través de múltiples versiones

hasta YOLO11, lanzado por Ultralytics el 10 de septiembre de 2024 [20]. Posteriormente, Ultralytics lanzó YOLO26 el 14 de enero de 2026 [22, 23], que en su variante n reporta hasta un 43% más de velocidad de inferencia en CPU respecto de YOLO11n y elimina el paso de supresión de no-máximos (NMS) del post-procesamiento [24]. Esta memoria utiliza YOLO11 y no YOLO26 por un criterio de madurez declarado por este trabajo: al ejecutarse el entrenamiento y la comparación entre mayo y junio de 2026, YOLO11 llevaba cerca de 20 meses disponible, contados desde su publicación el 10 de septiembre de 2024 [20], frente a los cinco meses de YOLO26 [24]. Ese margen se traduce en literatura académica ya publicada que analiza su arquitectura [25] y en evaluaciones comparativas independientes que cubren sus cinco escalas [14], respaldo del que YOLO26 todavía no disponía.

Como las redes de detección modernas, YOLO11 se organiza en un backbone, que es la red base que recorre la imagen y extrae sus características visuales como bordes, texturas y formas, y un neck, que es la etapa intermedia que combina esas características a distintas escalas antes de la detección final. Respecto de versiones anteriores, YOLO11 mejora ambas partes con tres bloques: el bloque C3k2, que extrae características con convoluciones más pequeñas y así usa menos parámetros que el bloque C2f previo; el bloque SPPF (Spatial Pyramid Pooling Fast), que resume la información a varias escalas para reconocer objetos de tamaños distintos; y el módulo de atención C2PSA, que ayuda a la red a concentrarse en las zonas más relevantes de la imagen [25, 20]. Se ofrece en cinco tamaños (n, s, m, l, x) que escalan precisión y costo.

Los cinco modelos de la familia YOLO11, según los *benchmarks* oficiales sobre COCO [20]:

Modelo	mAP@0.5:0.95	Parámetros	FLOPs	Velocidad T4 TensorRT	Velocidad CPU ONNX
YOLO11n	39,5	2,6 M	6,5 B	1,5 ms	56,1 ms
YOLO11s	47,0	9,4 M	21,5 B	2,5 ms	90,0 ms
YOLO11m	51,5	20,1 M	68,0 B	4,7 ms	183,2 ms
YOLO11l	53,4	25,3 M	86,9 B	6,2 ms	238,6 ms
YOLO11x	54,7	56,9 M	194,9 B	11,3 ms	462,8 ms

Tabla 1: Métricas oficiales de la familia YOLO11 (n, s, m, l, x) sobre el conjunto de validación de COCO: precisión (mAP@0.5:0.95), parámetros, FLOPs y velocidad de inferencia en GPU (T4 con TensorRT) y CPU (ONNX).

Fuente: [20].

Es decir, al pasar de YOLO11n a YOLO11x la precisión sube unos 15 puntos de mAP, pero a costa de cerca de 22× los parámetros y 30× los FLOPs, y de una latencia en CPU unas 8× mayor. Este trade-off precisión-costo a lo largo de los cinco tamaños es lo que la memoria busca cuantificar en el contexto de tráfico urbano.

Cada tamaño conviene a un escenario distinto. Los modelos pequeños, como YOLO11n, son adecuados para el despliegue en el borde, el procesamiento en tiempo real, el bajo consumo y la operación de muchas cámaras simultáneas, aunque presentan menor precisión [20, 26] y se desempeñan peor en los objetos pequeños y lejanos, efecto que esta memoria verifica sobre sus propios resultados en §5.3.2. En el otro extremo, para las tareas más demandantes, como las escenas con muchos objetos pequeños, conviene recurrir a los modelos grandes, como YOLO11l y YOLO11x, que alcanzan la máxima precisión a cambio de un alto costo de cómputo y energía y de requerir una GPU potente [20, 26].

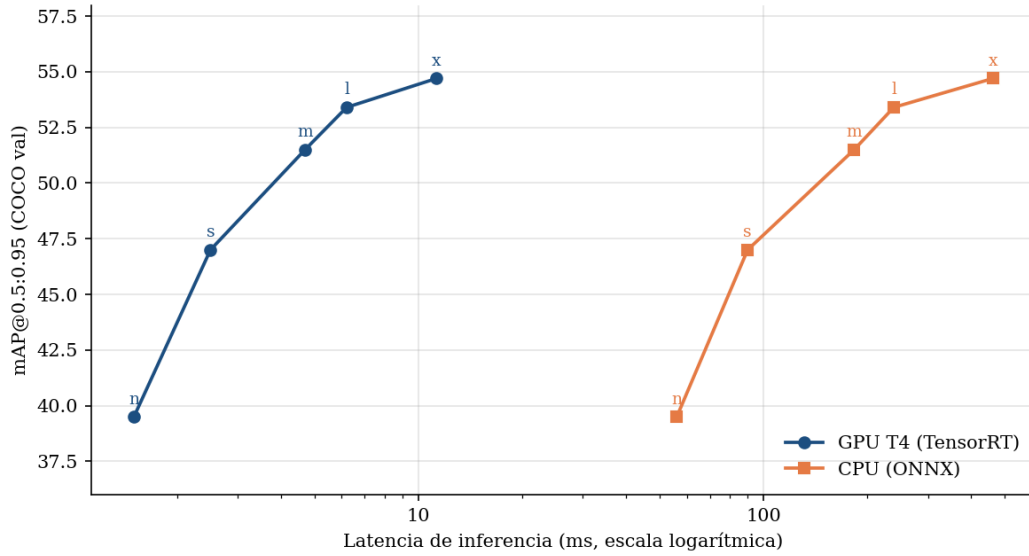


Figura 3. Precisión (mAP@0.5:0.95 sobre COCO val) en función de la latencia de inferencia para las cinco variantes de YOLO11, medida en GPU T4 con TensorRT y en CPU con ONNX (escala logarítmica). La curva se aplana a partir de YOLO11m mientras la latencia sigue creciendo. Fuente: elaboración propia a partir de los benchmarks oficiales de Ultralytics [20], los mismos de la tabla anterior.

3.3) Métricas de evaluación en detección de objetos

El desempeño en detección se mide con métricas basadas en la superposición entre la caja predicha y la real, cuantificada por el índice IoU (Intersection over Union), que es el cociente entre el área en que ambas cajas se solapan y el área total que cubren en conjunto:

$$\text{IoU} = \frac{\text{área de solapamiento}}{\text{área de unión}}$$

Una detección se considera correcta, o verdadero positivo (VP), cuando su IoU con una caja real supera un umbral, si no corresponde a ningún objeto real es un falso positivo (FP), y un objeto real que no se detecta es un falso negativo (FN). Con estas cantidades se definen las métricas:

- **Precisión:** fracción de detecciones correctas sobre el total de detecciones.

$$\text{Precisión} = \frac{VP}{VP + FP}$$

- **Recall:** fracción de objetos reales efectivamente detectados.

$$\text{Recall} = \frac{VP}{VP + FN}$$

- **F1-score:** media armónica de precisión y recall, que equilibra ambas en un solo número.

$$F1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Recall}}{\text{Precisión} + \text{Recall}}$$

- **mAP (mean Average Precision):** promedio, sobre las N clases, del área bajo la curva precisión-recall de cada clase (la precisión promedio o AP).

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N AP_i$$

Se reporta a un umbral IoU de 0,5 (mAP@0.5) [27] y promediado en el rango de umbrales de 0,5 a 0,95 (mAP@0.5:0.95), métrica estándar del protocolo de evaluación de COCO [28].

Para aplicaciones en tiempo real se complementan con métricas operacionales: **FPS** (cuadros por segundo), **latencia** por inferencia, **tamaño del modelo** y **uso de memoria/VRAM**.

3.4) Robustez de los detectores ante condiciones adversas

La degradación del desempeño de las redes neuronales ante corrupciones visuales comunes es un fenómeno sistemático y medible [29]. En tráfico urbano esto es crítico, puesto que la lluvia, la niebla y la noche son condiciones habituales de operación las cuales producen degradación

siendo un problema abierto con literatura propia [30]. Para estudiar el problema se han creado datasets específicos como DAWN, con 1.000 imágenes reales de niebla, lluvia, nieve y tormenta de arena [31], y Foggy Cityscapes [32].

Estudios sobre escenas nocturnas muestran caídas concretas: en cámaras de red urbanas, la confianza de detección de un mismo vehículo estacionado puede bajar de 0,8 a las 20:00 h, cuando aún queda luz natural, a 0,64 a las 21:00 h, con el IoU cayendo de 0,67 a 0,36, hasta la no-detección a las 22:00 h, cuando la escena queda iluminada solo por el alumbrado público (Tung et al. [33]). En escenarios de niebla la evidencia disponible es indirecta, dado que una variante de YOLO11 rediseñada específicamente para imágenes con niebla (DB-YOLO) supera a la línea base YOLO11n en 8,1 puntos porcentuales de mAP50 en el dataset de niebla real RTTS (de 48,5% a 56,6%) y en 2,5 puntos en el sintético Foggy Cityscapes, con aumentos por clase en RTTS de 10 puntos de AP en peatones y 7 en automóviles [34], lo que indica que el modelo genérico deja desempeño sin capturar en esa condición.

3.5) Analítica de tráfico urbano, Smart Cities y gemelos digitales

La detección de objetos es la capa de percepción de los sistemas de gestión inteligente de tránsito, dado que a partir del conteo y clasificación de vehículos se estima densidad y flujo, insumos para el **control adaptativo de semáforos** [35, 20] y para gemelos digitales que simulan políticas de tránsito sin afectar la operación real ni la privacidad. Un sistema reciente de gestión de tránsito basado en YOLOv11 reporta un mAP de 92,4% y un IoU de 0,85 en la identificación de vehículos, y plantea que combinar el control adaptativo de semáforos con el monitoreo de tráfico en tiempo real mejora la movilidad urbana y la sostenibilidad [35]. Un caso real es Surtrac (Pittsburgh), un sistema de control adaptativo de semáforos que optimiza en tiempo real las fases de cada intersección de forma descentralizada, recalculando su agenda hasta una vez por segundo y coordinándose con las intersecciones vecinas, reportó reducciones de más de 25% en los tiempos de viaje y de más de 40% en los de espera [5]. Un gemelo digital urbano lleva esta capa de percepción un paso más allá, ya que mapea los vehículos y peatones como nodos

en una representación en tiempo real de la red vial, sobre la cual se pueden analizar patrones de velocidad y movimiento para anticipar incidentes y evaluar políticas de tránsito. El sistema valorizado en esta memoria corresponde a la capa de percepción y control adaptativo (conteo vehicular, estimación de flujo, apoyo semafórico y detección de incidentes) y no incluye esa capa de simulación de políticas, por lo que en adelante se denomina **sistema de gestión de tránsito basado en detección por video**, el gemelo digital completo es su extensión natural, no su punto de partida. En el plano local esta tecnología ya tiene adopción en Chile. En Las Condes, el programa municipal *Barrio Protegido* dotó de análisis con inteligencia artificial a más de 1.400 cámaras de vigilancia distribuidas en 139 barrios, que detectan eventos en tiempo real y envían alertas inmediatas a la central de seguridad [36], dentro de una plataforma municipal que integra más de 20.000 dispositivos conectados [37]. Vitacura, por su parte, integra sus cámaras al Sistema Integrado de Televigilancia con Inteligencia Artificial (SITIA), una red que interconecta cámaras de varios municipios, de Carabineros y de privados bajo análisis centralizado con inteligencia artificial, que reconoce automáticamente placas patente contra la base oficial de vehículos con encargo de Carabineros y pone las cámaras en una plataforma de visualización centralizada, lo que permite monitorear el territorio en tiempo real y coordinar mejor a los equipos de seguridad [38, 39]. Estos despliegues muestran que la capa de detección de objetos sobre video ya está instalada en el país y que la discusión relevante es qué modelo la sustenta.

Desde una perspectiva de madurez tecnológica, las 34 comunas de la Región Metropolitana que conforman la red UOCT [40] se pueden ordenar en tres niveles según su grado de adopción de IA en gestión de tráfico y vigilancia. El nivel más avanzado corresponde a **Las Condes**, cuya red de cámaras con análisis de IA ya descrita constituye el precedente local más próximo al modelo de despliegue analizado en esta memoria. Un nivel intermedio lo representan las comunas integradas al **SITIA**, que partió en agosto de 2023 como piloto sobre las 14 comunas del territorio de la Fiscalía Centro Norte de la Región Metropolitana [41], en abril de 2025 anunció la incorporación de 1.500 cámaras e integró a once municipios, entre ellos Vitacura [42], y a 2026

declara presencia en 36 comunas de la región, con más de 2.000 cámaras unificadas y más de 550 lectores de patente integrados [39]. El nivel base lo forman las comunas restantes, que aún no cuentan con una capa propia de analítica con IA y cuya gestión de tráfico descansa exclusivamente en la operación centralizada de semáforos. La infraestructura común a los tres niveles es la red de la **UOCT**, que administra de forma centralizada 3.250 intersecciones semaforizadas distribuidas en esas 34 comunas [9], a las que en 2024 se añadió el sistema integrado de cámaras con inteligencia artificial inaugurado por el Gobierno Regional Metropolitano [10]. Dentro de esa infraestructura, las 567 intersecciones bajo **SCOOT** en corredores arteriales de alto volumen (que ya operan con control adaptativo de ciclos, según el último conteo publicado por la UOCT, correspondiente a 2016 [43]) son el punto de entrada natural para el sistema propuesto, puesto que la optimización de fases ya existe y la ganancia incremental del modelo de detección de IA se concentra en la lectura nocturna, donde los detectores de video que hoy alimentan esos sistemas pierden precisión, ya que la evaluación de campo de nueve sistemas de detección en una intersección semaforizada encontró un aumento notorio de detecciones falsas de los sistemas de video al pasar de día a noche [44]. En las restantes intersecciones sin control adaptativo, del orden de 2.700, el potencial es mayor, dado que la mejora no es incremental sino inaugural, ya que el sistema activaría por primera vez la optimización de ciclos en función del estado real del tráfico, capturando la ganancia plena del ASCT (*Adaptive Signal Control Technology*, tecnología de control adaptativo de semáforos). Esta diferencia entre intersecciones SCOOT y no-SCOOT es la que fundamenta la ponderación del parámetro r_c (el porcentaje de reducción de congestión que el sistema alcanza en la red) en la valorización del §4.5.

3.6) Economía de la inferencia: edge vs cloud y costo total de propiedad

Desplegar detección sobre una red de cámaras implica decidir entre procesar en el **borde** (edge, hardware en/junto a la cámara) o en la **nube**. Para cargas de inferencia continua como videovigilancia, el procesamiento en el borde desplaza el costo desde cargos recurrentes de suscripción, almacenamiento y transferencia de datos hacia una inversión inicial en hardware, de modo que

el punto de equilibrio frente a la nube depende del volumen y la permanencia de la carga [45]. La magnitud de ese efecto para el caso de esta memoria se cuantifica en §6.1 con su propio modelo de costos. Como marco formal de costeo se utiliza el **TCO** (CAPEX + OPEX) y métricas normalizadas tipo **LCOAI** (*Levelized Cost of AI*), que expresan el costo por unidad de inferencia válida, de forma análoga al costo nivelado de la energía [46]. Este marco permite traducir las diferencias de precisión y costo entre los cinco tamaños en un número comparable por inferencia, en línea con los objetivos de costo e impacto económico de la memoria.

3.7) El dataset BDD100K

BDD100K (*Berkeley DeepDrive*) es uno de los datasets de conducción más grandes y diversos: 100.000 videos de ~40 s a 720p y 30 fps, anotados para múltiples tareas de percepción [15]. Para detección incluye clases directamente relevantes al tráfico urbano: *car, bus, truck, person, rider, bike, motor, train, traffic light, traffic sign*. De forma crucial para esta memoria, **provee anotación a nivel de imagen de seis condiciones climáticas** (*clear, overcast, partly cloudy, rainy, snowy, foggy*) y **tres momentos del día** (*daytime, night, dawn/dusk*), con una proporción aproximadamente equilibrada entre material diurno y nocturno [15]. Esto permite desagregar el desempeño por condición sin etiquetado manual adicional, habilitando directamente los objetivos 1 y 3.

3.8) Estado del arte y brecha

La literatura confirma tres cosas.

- (a) YOLO11 es un detector en tiempo real de excelente precisión, velocidad y eficiencia [20].
- (b) El clima adverso degrada la detección, lo que ha motivado el desarrollo de arquitecturas específicas para condiciones difíciles [47], y la noche la degrada de forma independiente [33].
- (c) Existen numerosos trabajos de gestión de tránsito con YOLO en Smart Cities [35].

Existen además benchmarks que comparan precisión, tiempo de inferencia, FLOPs y tamaño de modelo entre detectores, de modo que el trade-off precisión-costo no está inexplorado. Lo que se observa con menos frecuencia es que esa comparación se realice de forma integrada y orientada a una decisión concreta de despliegue. Ahí se sitúa el aporte de esta memoria, que compara las **cinco escalas oficiales de YOLO11 bajo un protocolo único**, en **condiciones operacionales relevantes** (día, noche, clima y densidad), sobre el **hardware objetivo**, y cierra con una **regla explícita de decisión técnico-económica** que traduce el desempeño en una recomendación operacional y en un impacto social.

4) Metodología

La metodología es de tipo **experimental, cuantitativo y reproducible**, y se deriva del marco teórico (§3). Consta de cinco etapas encadenadas, donde la salida de cada una alimenta a la siguiente, tal como resume el diagrama de flujo a continuación. Primero se construye un subconjunto de BDD100K con la condición de cada imagen etiquetada; sobre él se hace el fine-tuning de los cinco modelos YOLO11 con un protocolo idéntico; luego se evalúa su precisión y su robustez de forma global y desagregada por condición; en paralelo se mide el costo computacional y operacional de cada modelo; y, por último, ese desempeño y ese costo se traducen a valor social mediante una valorización con criterio de decisión beneficio-costos. Cada etapa corresponde a un script reproducible (`prepare_bdd100k.py`, `train.py` y `evaluate_by_condition.py`) y todo el proceso usa una semilla fija (42) para ser repetible.

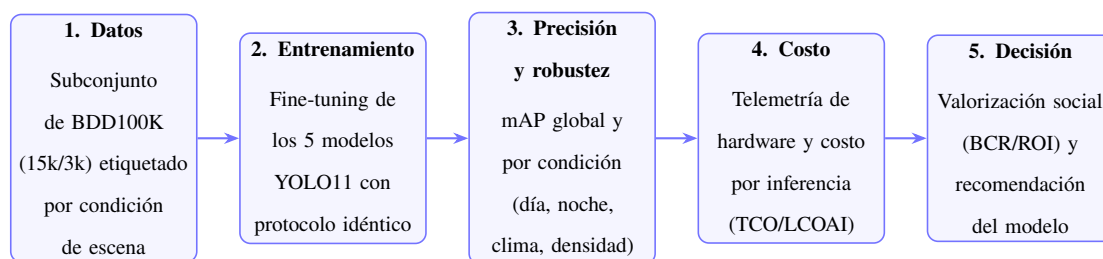


Figura 4: Diagrama de flujo de la metodología en cinco etapas, desde la preparación de los datos y el entrenamiento hasta la evaluación y la valorización social. Cada etapa alimenta a la siguiente. Fuente: elaboración propia.

4.1) Datos y preparación (Obj. 1)

Se utiliza BDD100K (Berkeley DeepDrive), un conjunto de 100.000 imágenes de conducción urbana con anotaciones de detección capturadas en distintas ciudades, horarios y climas [15]. Se eligió porque cada imagen trae como atributos su momento del día (día, noche, amanecer/atardecer) y su clima (despejado, lluvia, niebla, nublado, nieve), lo que permite medir la degradación del desempeño por condición sin etiquetado manual adicional, que es el diferenciador de esta memoria. El dataset viene dividido oficialmente en train (70.000 imágenes), val (10.000) y test (20.000), como las etiquetas del test no son públicas, ya que Berkeley las reserva para su leaderboard al igual que COCO o Cityscapes, se reporta sobre el conjunto de validación, que cumple aquí la función de conjunto de prueba siguiendo la convención del dataset [15].

A partir de las etiquetas oficiales, el script `prepare_bdd100k.py` construye el subconjunto de trabajo: convierte las anotaciones al formato YOLO (un archivo de texto por imagen con clase `cx cy w h` normalizado), filtra a las siete clases de tráfico urbano relevantes (car, bus, truck, person, rider, bike, motor) y registra en un `metadata.csv` la condición de cada imagen (momento del día, clima y densidad, esta última derivada del número de objetos: baja con cinco o menos, media hasta quince y alta sobre quince). Ese metadata es lo que después habilita la evaluación desagregada por condición.

```
# Clases de trafico urbano (indice = id de clase YOLO)
CLASSES = ["car", "bus", "truck", "person", "rider", "bike", "motor"]

def box2d_to_yolo(box): # pixeles -> (cx,cy,w,h) normalizado
    x1,y1,x2,y2 = box["x1"],box["y1"],box["x2"],box["y2"]
    return ((x1+x2)/2/IMG_W, (y1+y2)/2/IMG_H,
            abs(x2-x1)/IMG_W, abs(y2-y1)/IMG_H)

random.seed(42); random.shuffle(eligible) # submuestreo reproducible
eligible = eligible[:max_items] # 15000 train / 3000 val

# Condicion de cada imagen -> metadata.csv (habilita el analisis por condicion)
meta_rows.append({"name": name, "timeofday": attr.get("timeofday"),
                  "weather": attr.get("weather"), "density": density_bucket(n_obj)})
```

Se trabaja con un subconjunto de 15.000 imágenes de entrenamiento y 3.000 de validación, muestreado con semilla fija. La razón es que el objetivo es comparar los cinco modelos bajo un protocolo idéntico, no maximizar el mAP absoluto, ya que un subconjunto representativo permite mantener un conjunto de validación grande (3.000 imágenes) para una estimación de baja varianza. Como el submuestreo se aplica por igual a los cinco modelos, no introduce sesgo en la comparación, que es relativa.

Sobre la estrategia de validación, se usa hold-out (división fija train/val) y no validación cruzada k-fold. La razón es de costo y de tamaño, puesto que en aprendizaje profundo con datasets grandes el hold-out entrega una estimación estable a una fracción del costo, mientras que k-fold, que entrena k veces rotando los pliegues, se reserva para datasets de menor tamaño relativo [48], sin que exista un umbral numérico único en la literatura. Con 3.000 imágenes de validación la varianza extra del hold-out es pequeña, puesto que el mAP promedia sobre miles de detecciones distribuidas en esas imágenes, lo que estabiliza la estimación por encima del nivel de las diferencias entre modelos que se busca detectar, como verifica el remuestreo pareado de §5.1, que compara los cinco modelos sobre las mismas imágenes, de este modo no se justifica el costo de entrenar cinco veces cada uno de los cinco modelos. Además, el diseño no sufre el sesgo optimista de selección, ya que no se hace búsqueda de hiperparámetros sobre validación (los cinco modelos comparten un protocolo fijo), por lo que el conjunto de validación funciona legítimamente como prueba.

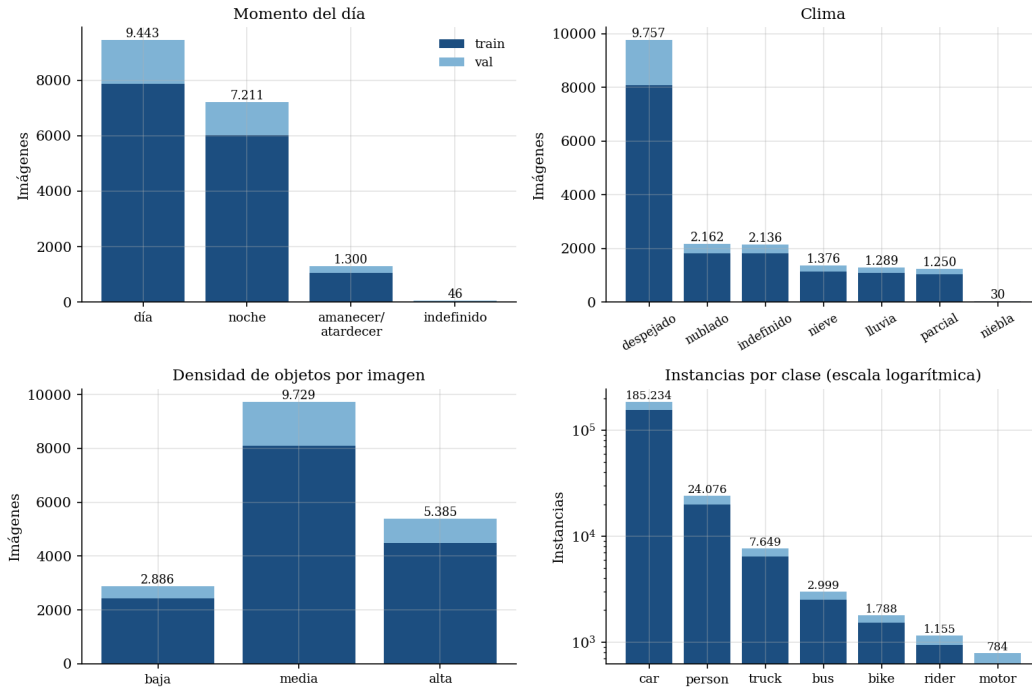


Figura 5. Distribución del subconjunto utilizado (15.000 imágenes de entrenamiento y 3.000 de validación de BDD100K) por momento del día, clima, densidad de objetos por imagen e instancias por clase. El panel de clases usa escala logarítmica dado el fuerte desbalance hacia la clase *car*. Fuente: elaboración propia a partir de las anotaciones de BDD100K [15].

4.2) Entrenamiento (Obj. 2)

Se hace fine-tuning de los cinco modelos YOLO11 (n, s, m, l, x) partiendo de pesos preentrenados en COCO, con el *framework* Ultralytics, plataforma de código abierto que desarrolla y mantiene la familia YOLO y que provee una API unificada de entrenamiento, evaluación y exportación, sobre una GPU NVIDIA RTX PRO 2000 Blackwell con 16 GB de VRAM. El transfer learning, proceso en que se parte de una red ya entrenada en un dataset amplio, en este caso COCO con 118.000 imágenes, y se ajusta sobre el dominio objetivo, en este caso BDD100K, acelera la convergencia y mejora el desempeño frente a entrenar desde cero, y es la práctica estándar en detección. La decisión central es usar exactamente los mismos hiperparámetros para los cinco modelos, de modo que cualquier diferencia de desempeño se atribuya al tamaño de la arquitectura y no al ajuste, lo que aísla la variable de interés.

Hiperparámetro	Valor	Por qué
Pesos iniciales	COCO	El transfer learning desde un dataset grande acelera la convergencia frente a entrenar desde cero.
Épocas	50	Techo del entrenamiento. El early stopping corta antes si el mAP de validación deja de mejorar, de modo que las curvas de §5.1 muestran dónde convergió cada modelo.
Tamaño de imagen	640	Resolución estándar de YOLO11 en los benchmarks oficiales [20], comparable con la literatura.
Batch	16	Valores mayores no aceleraron el entrenamiento en la GPU utilizada, ya que el cuello de botella no es el cómputo (§5.5).
Paciencia	15	Early stopping: corta tras 15 épocas sin mejora en validación.
Semilla	42	Reproducibilidad del submuestreo y la inicialización.

Tabla 2: Hiperparámetros de entrenamiento, idénticos para los cinco modelos YOLO11. Fuente: elaboración propia.

```
from ultralytics import YOLO
for w in ["yolo11n.pt", "yolo11s.pt", "yolo11m.pt", "yolo11l.pt", "yolo11x.pt"]:
    model = YOLO(w) # pesos preentrenados en COCO
    model.train(data="bdd100k_yolo/data.yaml",
                epochs=50, imgsz=640, batch=16,
                patience=15, seed=42, plots=True) # mismos hiperparametros (5 modelos)
```

El batch se fijó en 16, dado que valores mayores no aceleraron el entrenamiento en la GPU utilizada (el diagnóstico del cuello de botella se reporta en §5.5). El early stopping con paciencia 15 corta el entrenamiento cuando el mAP@0.5:0.95 de validación no mejora durante quince épocas consecutivas, lo que ahorra cómputo y reduce el riesgo de sobreajuste.

4.3) Evaluación de precisión y robustez (Obj. 3)

Para cada modelo entrenado se calculan mAP@0.5, mAP@0.5:0.95, precisión, recall y F1, tanto de forma global sobre las 3.000 imágenes de validación como desagregada por cada condición (momento del día, clima y densidad). El script `evaluate_by_condition.py` agrupa las imágenes de validación por condición usando el metadata, arma para cada grupo un `data.yaml`

temporal cuyo conjunto de validación apunta solo a esas imágenes, y corre la evaluación sobre él, de modo que todas las métricas provienen de una misma rutina y son comparables entre modelos.

```
# Para cada condicion se arma una lista de imagenes y se corre model.val()
groups = defaultdict(list)
for row in metadata: # solo split == "val"
    for dim in ("timeofday", "weather", "density"):
        groups[(dim, row[dim])].append(img_path)

res = model.val(data=yaml_cond) # data.yaml cuyo val apunta a esa lista
fila = {"mAP50-95": float(res.box.map),
        "precision": float(res.box.mp), "recall": float(res.box.mr)}
# Robustez = caida relativa de mAP50-95 respecto a daytime / clear
```

La métrica principal es el mAP@0.5:0.95, que promedia el desempeño sobre diez umbrales de IoU (de 0,5 a 0,95) y por eso es más estricta y completa que el mAP@0.5 [28]. Por su naturaleza compuesta, un valor de 0,30 indica que el modelo logra una precisión media del 30% al combinar simultáneamente la detección de objetos y la calidad de su localización espacial, siendo penalizado con fuerza cuando la caja predicha no coincide bien con la real a umbrales de solapamiento estrictos, no significa que el modelo detecte solo el 30% de los objetos. La robustez se mide como la caída relativa del mAP@0.5:0.95 frente a la condición base (día y despejado), de modo que un modelo es más robusto cuanto menor es esa caída. Esta degradación tiene peso económico, ya que la noche concentra el mayor costo de error, dado que en Estados Unidos cerca del 49% de las muertes de ocupantes ocurre de noche pese a que solo cerca del 25% de la conducción es nocturna (NHTSA, varghese y Shankar [11]), de modo que un modelo que se degrada de noche falla justo donde cada error cuesta más. A falta de una medición equivalente para Chile, ese patrón se traslada como supuesto declarado al modelo de valorización (§4.5).

Para controlar la correlación entre condiciones, dado que el clima y el momento del día están correlacionados en BDD100K puesto que lluvia y nieve ocurren casi exclusivamente de día en el subconjunto, las métricas por condición se estiman mediante ponderación inversa de prob-

abilidad (IPW). El método pondera cada condición usando la distribución de las demás en la categoría de referencia (día para hora, despejado para clima, distribución global para densidad), lo que equivale a comparar las condiciones como si tuviesen la misma composición de las restantes:

$$\widehat{\text{mAP}}(t) = \sum_{c \in C} \hat{P}(c \mid t_0) \cdot \text{mAP}(t, c) \quad (1)$$

donde t es la condición evaluada, t_0 es la condición de referencia, C es el conjunto de celdas cruzadas válidas (al menos 15 imágenes), y $\hat{P}(c \mid t_0)$ es la proporción empírica de la celda cruzada c en la condición de referencia. El script `evaluate_weighted_conditions.py` evalúa cada celda del cruce de las tres dimensiones (hora, clima, densidad) de forma independiente con `model.val()` y aplica esta ponderación para cada dimensión. Las 329 imágenes de validación con `weather = undefined` se excluyen por corresponder a una etiqueta de clima no asignada en los metadatos de BDD100K [15], y no a una condición meteorológica observable, quedando 2.671 imágenes efectivas. Las tablas de §§5.2 a 5.4 reportan los valores con esta corrección.

```
# evaluate_weighted_conditions.py -- IPW por dimension (extracto)
WEATHER_EXCLUDE = {"undefined"} # etiqueta de anotacion fallida (Yu et al. 2020)
MIN_IMGS = 15 # minimo por celda para estimacion estable

# Peso de cada celda cruzada segun la distribucion de referencia (ej. timeofday=daytime)
for key, n_ref in counts_3d_tod_ref.items():
    confound = {k: v for k, v in key.items() if k != "timeofday"}
    w_ref[confound] = w_ref.get(confound, 0) + n_ref / n_total_ref

# mAP ponderado IPW para la condicion t (ej. timeofday="night")
w_num, w_den = 0.0, 0.0
for conf_key, weight in w_ref.items():
    cell = norm_key({**conf_key, "timeofday": t})
    m = cell_map.get(cell) # mAP50-95 de la celda; None si n < MIN_IMGS
    if m is not None:
        w_num += weight * m; w_den += weight
w_map = w_num / w_den if w_den > 0 else None
```

El análisis del efecto del clima se realiza en tres pasadas complementarias, cuyos resultados se reportan en §5.3, §5.3.1 y §5.3.2 respectivamente. La primera es el IPW general recién descrito, que controla simultáneamente las tres dimensiones. La segunda es un **IPW restringido a las imágenes diurnas**, que aísla el efecto genuino del clima de la distorsión provocada por la distribución desigual de imágenes nocturnas, ponderando solo por densidad vehicular:

$$\widehat{\text{mAP}}_{\text{día}}(w) = \sum_{d \in D} \hat{P}(d \mid \text{diurno}) \cdot \text{mAP}(w, \text{diurno}, d) \quad (2)$$

donde $D = \{\text{baja, media, alta}\}$ y los pesos $\hat{P}(d \mid \text{diurno})$ se estiman sobre el subconjunto diurno. La tercera son las comparaciones controladas que fijan las variables de confusión en un valor común, descritas al final de esta sección. Al usar tres métodos con supuestos distintos, la coincidencia de sus conclusiones actúa como control de robustez del análisis climático.

Esta memoria usa dos procedimientos de **remuestreo** (*bootstrap*) con propósitos distintos. El remuestreo consiste en volver a extraer, con reemplazo, muestras del propio conjunto de validación y recalculan la métrica en cada una, de modo que la dispersión de esos valores estima cuánta incertidumbre tiene el resultado.

El **remuestreo pareado entre modelos**, llamado pareado porque en cada iteración los cinco modelos se evalúan sobre exactamente las mismas imágenes, de modo que ninguna diferencia entre ellos puede atribuirse a que a uno le tocan imágenes más fáciles, (Tabla 4, §5.1) verifica que las diferencias de mAP entre los cinco tamaños son estadísticamente significativas y no artefactos del tamaño del conjunto de validación. Para $B = 1.000$ iteraciones se remuestrea con reemplazo el conjunto de validación completo y se recalcula el mAP:

$$\hat{\mu}_b(m) = \text{mAP}(\text{remuestra}_b, \text{modelo } m), \quad b = 1, \dots, B \quad (3)$$

El IC 95% de cada modelo es $[\hat{\mu}_{(25)}, \hat{\mu}_{(975)}]$, es decir, los percentiles 2,5 y 97,5 de las 1.000 estimaciones, entonces si los IC de dos modelos no se solapan, su diferencia es estadísticamente

significativa.

El **remuestreo por celda de condición** (integrado en `evaluate_weighted_conditions.py`) cuantifica la incertidumbre estadística en la estimación de mAP de cada celda del cruce hora $\times$ clima $\times$ densidad. Para una celda c con n imágenes se remuestrea con reemplazo ese subconjunto en 1.000 iteraciones:

$$\hat{\mu}_b^{(c)} = \text{mAP}(\{x_{i_1}, \dots, x_{i_n}\}_b), \quad i_j \sim \mathcal{U}\{1, \dots, n\} \text{ con reemplazo} \quad (4)$$

Este procedimiento no requiere correr la red neuronal de nuevo: durante la validación se guardan, imagen por imagen, las predicciones de la red y los objetos de referencia. Calcular el mAP de cualquier remuestreo de esas estadísticas es solo aritmética sobre las listas ya almacenadas (*seed*=42). El IC 95% resultante indica cuánta incertidumbre tiene la estimación de mAP de esa celda, la que decrece con el número de imágenes disponibles. Los anchos de IC observados por tamaño de celda se reportan junto a las tablas de resultados (§5.3.1).

Análisis de distribución de tamaño de objetos por condición climática. Para verificar empíricamente si la diferencia de mAP entre condiciones climáticas se debe, en parte, a diferencias en el tamaño de los objetos anotados, dado que esa hipótesis está planteada en la literatura sobre calidad de anotación en condiciones adversas [49], se calcula el área en píxeles de cada bounding box en el conjunto de validación y se agrupa por condición climática. El área se obtiene a partir de las anotaciones en formato YOLO (coordenadas normalizadas c_x, c_y, w_n, h_n relativas al ancho y alto de la imagen de 1.280×720 píxeles):

$$A_{\text{px}} = w_n \times 1280 \times h_n \times 720 \quad (5)$$

Se aplican los umbrales COCO [28]: **pequeño** $A < 32^2 = 1.024 \text{ px}^2$; **mediano** $1.024 \leq A < 96^2 = 9.216 \text{ px}^2$; **grande** $A \geq 9.216 \text{ px}^2$. Los resultados (porcentaje de objetos por categoría de tamaño, área mediana y media) se reportan en §5.3.1. Se excluyen las condiciones `undefined`

y foggy (sin masa estadística suficiente en validación).

Comparaciones controladas de momento del día y clima (§5.3.2). Como complemento al análisis IPW, se realizan dos comparaciones que eliminan la necesidad de ponderación fijando las variables de confusión en un valor común.

La **comparación de momentos del día** fija clima = despejado y densidad = alta. La densidad alta es la única que garantiza paridad exacta entre los grupos ($n_{\text{día}} = n_{\text{noche}} = 195$; $n_{\text{amanecer}} = 43$, por encima del umbral $\text{MIN_IMGS} = 15$). Con densidad media, la noche aporta 74% más imágenes que el día (558 vs. 320), haciendo menos limpia la comparación directa dado que los grupos tendrían tamaños muy distintos y diferente representatividad de escenas. Con densidad baja, la desproporción se amplía a 4,1 veces (noche = 239, día = 58) y el amanecer queda excluido ($n = 10$). La tabla siguiente resume los tamaños de muestra por densidad:

Momento del día	density = baja	density = media	density = alta
Día (daytime)	58	320	195
Amanecer (dawn/dusk)	10*	60	43
Noche (night)	239	558	195
Ratio noche / día	4,1×	1,74×	1,0×

Tabla 3: Tamaños de muestra por momento del día y nivel de densidad en el subconjunto de validación. * $n < \text{MIN_IMGS} = 15$, excluido del análisis. Fuente: metadata BDD100K, elaboración propia.

La **comparación de condiciones climáticas** fija momento del día = diurno y densidad = media. La tabla siguiente muestra la cobertura de cada condición por nivel de densidad, con la lluvia como caso crítico por ser la condición menos frecuente en el subconjunto diurno:

Clima	density = baja	density = media	density = alta
Despejado	58	320	195
Nublado	24	141	129
Parcialmente nublado	19	90	80
Lluvia	10*	65	32
Nieve	15	66	46

Tabla 4: Cobertura de cada condición climática por nivel de densidad en el subconjunto diurno de validación. * $n = 10 < \text{MIN_IMGS} = 15$, lluvia excluida en densidad baja. Fuente: metadata BDD100K, elaboración propia.

La densidad baja excluye lluvia ($n = 10$), impidiendo comparar las cinco condiciones climáticas. La densidad alta incluye lluvia pero con solo $n = 32$: dado que el error estándar del mAP estimado es proporcional a $1/\sqrt{n}$, esa estimación es $\sqrt{65/32} \approx 1,43$ veces más variable que con $n = 65$, lo que reduce la precisión de la comparación. La densidad media es la única que cubre los cinco climas garantizando al menos $n = 65$ en el grupo más pequeño (lluvia), y es por eso la elección adoptada. Los resultados se reportan en §5.3.2.

4.4) Evaluación operacional y de costo (Obj. 4)

Además de la precisión, se registra por modelo la telemetría que determina el costo de operar a escala de una red de cámaras: FPS, latencia de inferencia, consumo en watts, uso de VRAM, número de parámetros y FLOPs (resultados en §5.5). La telemetría se mide corriendo inferencia real sobre 500 imágenes de validación, una por una (batch 1, emulando una cámara en streaming), en la GPU de prueba NVIDIA RTX PRO 2000 Blackwell (16 GB). La energía por inferencia se reporta en milijulios, métrica con la que se comparan los detectores en los benchmarks sobre dispositivos de borde [26], dado que combina la potencia demandada y la duración de cada inferencia, de modo que un modelo más rápido gasta menos energía por imagen aunque consuma una potencia similar. Estas cifras se traducen a un costo total de propiedad (TCO) y a un costo por inferencia comparable entre los cinco tamaños mediante el siguiente procedimiento. A partir de

los FPS de cada modelo y de los FPS que requiere una cámara para conteo y control de tránsito, del orden de 10 a 15 fps (suficiente sin necesidad de la tasa completa de video), se calcula cuántas cámaras atiende un equipo de inferencia y, por lo tanto, cuántos equipos se necesitan para cubrir la red. Con eso se arma el CAPEX (equipos más instalación) y el OPEX anual (energía y mantenimiento), y el costo total se descuenta a la tasa social sobre un horizonte de cinco años, que es la vida útil supuesta del hardware.

```
# Cuantas camaras atiende un equipo, segun el FPS del modelo (telemetria 5.5)
camaras_por_disp = int(fps_modelo // fps_requerido) # fps_requerido = 15
n_disp = ceil(n_camaras / camaras_por_disp) # n_camaras = 3250

capex = n_disp * (precio_disp + instalacion)
energia_anio = n_disp * (potencia_W/1000) * 8760 * precio_kwh
# Costo de computo descontado a la tasa social (5,5%), horizonte 5 anios
costo_computo = capex + sum(energia_anio/(1+0.055)**t for t in range(1,6))
lcoai = costo_computo / total_inferencias # costo por inferencia comparable
```

El costo por inferencia, análogo al costo nivelado de la energía y denominado LCOAI (Levelized Cost of AI), es el TCO dividido por el total de inferencias del horizonte, y permite comparar entre los cinco tamaños el costo de cómputo que corresponde a cada alternativa [46]. Este es el único componente del costo que varía según el modelo, dado que el resto de la infraestructura (cámaras de la red existente, plataforma y control adaptativo por intersección) es común a los cinco y no entra en la comparación de costo diferencial. El cálculo asume además el **mismo tipo de dispositivo de borde para los cinco modelos**, de modo que un modelo más pesado se traduce en más unidades y no en unidades más potentes, ya que el número de dispositivos se deriva de cuántas cámaras alcanza a procesar cada uno según los FPS medidos. Este supuesto mantiene el hardware comparable entre alternativas y su efecto sobre el resultado es acotado, puesto que el cómputo representa una fracción menor del costo total (§6.2).

4.5) Valorización social del impacto y recomendación (Obj. 5)

El criterio para seleccionar el modelo que se recomienda no es solo su precisión técnica, sino el valor social neto que genera su despliegue. El caso de uso central es un sistema de gestión de tránsito basado en detección por video que integra el conteo vehicular, la estimación de flujo, el apoyo a los semáforos adaptativos y la detección de incidentes viales (accidentes, vehículos detenidos en calzada, peatones en zonas de riesgo y congestión anómala), se usa como marco para valorar los beneficios integrados. La valorización social se evalúa desde la perspectiva de una Municipalidad o un Ministerio, con precios sociales y una tasa social de descuento del 5,5% [50], sobre Santiago a escala de ciudad, dimensionado sobre las cerca de 3.250 intersecciones semaforizadas de la UOCT [9].

El beneficio social se desglosa en cuatro mecanismos. Cada uno se construye anclando un agregado oficial, escalándolo por la fracción que el sistema puede capturar y ponderándolo por la calidad de detección del modelo. De dónde sale cada número:

- **Ahorro de tiempo por menor congestión.** Se parte del costo de la congestión de Santiago, del orden de US\$1.353 millones al año en USD constantes de 2026 (US\$1.046 millones de 2019, medidos con datos de movilidad Waze por el BID, actualizados con IPC-US BLS 2019→2026 $\times 1,294$) [4], cifra que ya valoriza el tiempo perdido a partir del valor social del tiempo de viaje [51]. Sobre él se aplica r_c , el porcentaje de reducción del retardo por congestión que el control adaptativo de semáforos lograría en la red. El orden de magnitud de ese porcentaje está documentado, ya que los sistemas ASCT reportan reducciones de tiempo de viaje de 9,36% en Florida y de más de 25% con Surtrac en Pittsburgh [52, 5].

Ahora bien, esa ganancia no se reparte por igual en toda la red. Un semáforo de una calle secundaria y otro de un eje troncal aportan lo mismo si se los cuenta, pero no si se mide el retardo que cada uno acumula, dado que el tránsito urbano se concentra en pocas vías: el 20% superior de las calles absorbe cerca del 80% del flujo [53]. Por eso la ganancia

de cada tipo de intersección se pondera por su **participación en el retardo** y no por su número.

El escenario base de esta memoria utiliza **6,25%**, resultado de aplicar esa ponderación sobre tres tipos de intersección. El primero corresponde a las intersecciones que ya operan con control adaptativo dinámico bajo **SCOOT** (*Split, Cycle and Offset Optimization Technique*), sistema de control adaptativo de ciclos largos (100 a 120 s). El último conteo público de la propia UOCT es de 2016 y registra **567 intersecciones en control dinámico SCOOT**, agrupadas en 32 redes, sobre 2.900 intersecciones integradas al sistema de control de Santiago, quedando las 2.323 restantes con planes de tiempo prefijados [43]. La serie muestra un crecimiento sostenido de la cobertura adaptativa, desde 236 intersecciones habilitadas para control dinámico en 1999 [54] y 400 registradas como hito de 2010 en la cronología institucional de la UOCT [55], hasta esas 567 de 2016, el sistema sigue operativo, dado que la UOCT publicó en 2023 su procedimiento vigente de calibración de arcos SCOOT [56]. La antigüedad del último conteo se declara como limitación y no altera el modelo, puesto que la ponderación usa la participación en el retardo y no el conteo, y esa participación se sensibiliza en §4.6. SCOOT se instala en los corredores de mayor volumen, al igual que en ciudades como Seattle (los corredores de mayor tránsito) [57], Monterey (todas las arterias) [58] y Toronto (ocho corredores arteriales seleccionados) [59], por reunir los corredores de mayor demanda se les asigna un 40% del retardo abordable. En ellas el control adaptativo ya existe y la ganancia incremental de agregar detección IA es menor y se concentra en el 25% del tráfico nocturno, dado que los detectores de video que alimentan SCOOT pierden precisión de noche, condición que YOLO11 mitiga precisamente [44], se estima $g_{\text{SCOOT}} = 2,5\%$ incremental, con mínimo 0%, puesto que la evaluación piloto de Virginia documentó mejoras nulas en corredores que ya funcionaban bien (Fontaine et al. [60]) y la revisión de NCHRP encontró beneficios menores cuando las señales previas se retimaban con frecuencia (Stevanovic [61]). El segundo tipo son las intersecciones arteriales y colectoras sin control adaptativo, a las que se asigna un

45% del retardo abordable y la ganancia plena del ASCT, $g_{ASCT} = 10\%$ [62, 63], valor que se adopta como piso conservador, dado que la FHWA reporta una mejora promedio superior a ese 10%. El tercer tipo son las intersecciones de bajo flujo, con un 15% del retardo abordable y una ganancia media $g_{low} = 5\%$, la mitad de la plena, dado que la revisión de NCHRP reporta resultados inconsistentes del ASCT en vías con menos de 35.000 vehículos diarios (Stevanovic [61]). El promedio ponderado por participación en el retardo es:

$$r_c = 0,40 \times 2,5\% + 0,45 \times 10\% + 0,15 \times 5\% = 6,25\%$$

Las participaciones de la tipología son supuestos declarados de este trabajo, dado que no se encontró una medición chilena de la distribución del retardo por jerarquía vial, y por eso se sensibilizan junto a las ganancias en el análisis de riesgo (§4.6): la ganancia plena se muestrea de una Beta-PERT(7%; 10%; 25%), cuya moda se fija de forma conservadora en el piso reportado por la FHWA, que documenta una mejora promedio del tiempo de viaje superior a 10% [62] y cuyo extremo inferior cubre los despliegues documentados en la base de evaluaciones ITS del USDOT (7,0% a 8,6% en Oakland County y 13% en Los Ángeles) [64] y cuyo extremo superior corresponde al 25% del piloto de Surtrac en Pittsburgh [5], cifra que por tratarse de un piloto de 2012, de más de una década de antigüedad, se usa aquí únicamente como cota superior optimista y no como resultado esperado típico. La ganancia incremental sobre SCOOT se muestrea de una PERT(0%; 2,5%; 5%), la de bajo flujo de una PERT(0%; 5%; 10%), y las participaciones de distribuciones triangulares en torno a sus valores base. El 6,25% se multiplica además por la fracción de la congestión abordable por la red semaforizada: un 36,8%, el cual se calcula como el producto de dos mediciones. La primera acota cuánto del retardo urbano queda fuera de las autopistas, puesto que el Urban Mobility Report sitúa cerca del 54% del retardo de las áreas urbanas grandes de Estados Unidos en ellas [65], de modo que el

46% restante ocurre en calles. Ese 46% es una cota y no el valor buscado, ya que no toda calle está semaforizada. La segunda medición resuelve precisamente esa fracción, dado que la jerarquía vial urbana concentra el tránsito en pocos ejes, con el 20% superior de las calles absorbiendo cerca del 80% del flujo [53], y ese tramo superior es justamente el que opera semaforizado bajo control de la UOCT. El producto de ambas, $46\% \times 80\% = 36,8\%$, es el valor que se adopta. La derivación es conservadora, puesto que Jiang mide flujo y no retardo, y el retardo crece más que proporcionalmente con la razón entre volumen y capacidad, de modo que su concentración en los ejes principales es mayor o igual que la del flujo.

- **Accidentes evitados.** Se parte del costo social de los siniestros de tránsito en Chile, US\$4.364 millones en USD constantes de 2026 ([66], actualizado $\times 1,05$ IPC-US BLS 2024 $\rightarrow$ 2026), equivalente al 1,31% del PIB en el año de referencia, con 75.653 siniestros y 1.439 fallecidos [7], y se toma la participación de la Región Metropolitana en ese costo, del 22,9%, derivada en esta memoria a partir del desglose de víctimas por gravedad de la Región Metropolitana y de los costos unitarios por gravedad de CONASET [67, 66], lo que define $B_{\text{accid, RM}} = 4.364 \text{ M} \times 0,229 = 998$ millones de USD 2026 al año, y se aplica el porcentaje de reducción de siniestros (r_a) por la implementación del control adaptativo con vigilancia visual. El valor base $r_a = 7,8\%$ proviene de Kodi et al. [68], quienes cuantifican ese efecto en una red urbana de Florida bajo ASCT mediante un CMF (*Crash Modification Factor*, el factor que se multiplica por la tasa de siniestros sin la intervención para estimar la tasa con ella, un valor bajo 1 indica reducción) de 0,922, equivalente a una reducción de 7,8% ($1 - 0,922$), y el rango completo en la literatura oscila entre 7,8% y 17% según el diseño del estudio y las características de la red, con el extremo superior medido mediante un análisis Empirical Bayes en los corredores piloto de Virginia [60]. Gran Santiago, como red urbana densa de velocidades moderadas, se clasifica en el extremo inferior de ese rango, donde el efecto disuasorio de las cámaras [8] ya está incorporado. La incertidumbre de r_a se trata en el análisis de riesgo (§4.6) mediante una

distribución Beta-PERT(6,2%; 7,8%; 17%), cuyo máximo corresponde al extremo superior publicado en la literatura [60] y cuyo mínimo extiende el valor base un 20% a la baja como supuesto de sensibilización declarado de este trabajo, en línea con la recomendación de la guía de análisis beneficio-costos del USDOT de sensibilizar los parámetros de mayor incertidumbre [69].

- **Vidas salvadas.** Se reportan de forma destacada como subcomponente físico de los accidentes evitados, sin sumarse aparte para no duplicar el beneficio: las 1.439 muertes anuales por la participación de la Región Metropolitana y por el porcentaje de reducción de siniestros.
- **Impacto ambiental.** La menor congestión reduce las emisiones dado que los vehículos ralentizan menos. Wu et al. cuantifican en datos reales de China que el control adaptativo de semáforos reduce en un 11% los tiempos de viaje en hora punta, con una reducción directa de 31,73 millones de toneladas de CO₂ al año en las 100 ciudades más congestionadas [70]. El porcentaje de reducción de emisiones es proporcional al de reducción de tiempo de viaje. La reducción se valora al precio social del carbono de US\$66,6 por tonelada de CO₂ en USD constantes de 2026 [50] (actualizado $\times 1,05$ IPC-US BLS 2024→2026), aplicado sobre los 10,62 millones de toneladas de CO₂ al año que el inventario oficial de emisiones del sector transporte de la Región Metropolitana atribuye a 2022 [71].

El enlace entre lo técnico y lo económico es un factor de calidad de detección por modelo: el mAP de cada condición se pondera (la noche pesa más por su mayor costo de error [11]) y se normaliza respecto del mejor modelo, de modo que el modelo más robusto realiza una fracción mayor del beneficio y la robustez entra de forma directa en la función de valor.

```
# Beneficio social anual por mecanismo, escalado por el factor de calidad:
# q^tr (fcal) en tiempo y ambiental, q^seg (fcal_seg) en accidentes y vidas
b_tiempo = 1353e6 * red_congestion * cobertura * fcal # BID 2021 x1.294 IPC -> 2026 USD
b_accidentes = (4364e6 * rm_share_costo) * red_accidentes * fcal_seg # CONASET 2024 x1.05 IPC
```

```

co2_evitado = co2_rm * red_congestion * cobertura # co2_rm = 10618074 ton (MMA 2023)
b_ambiental = co2_evitado * 66.6 * fcal # MDS 2024 x1.05 IPC -> 2026 USD
vidas = 1439 * rm_share_vidas * red_accidentes * fcal_seg
# Escenario base: red_congestion = 0.0625 ; red_accidentes = 0.078 ; cobertura = 0.368
# rm_share_costo = 0.2287 ; rm_share_vidas = 0.2168
b_total = vp(b_tiempo + b_accidentes + b_ambiental) # valor presente 5 años
BCR = b_total / costo
ROI = (b_total - costo) / costo

```

Las expresiones formales distinguen dos factores de calidad según el mecanismo de beneficio. En ambas, el subíndice $k \in \{n, s, m, l, x\}$ identifica la variante de YOLO11 evaluada. El **factor de calidad de tráfico** q_k^{tr} pondera el mAP por la exposición del tránsito según el tramo horario (75% diurno, 25% nocturno) y normaliza al modelo de referencia. Ambos pesos provienen de una única fuente y de un único criterio, ya que la medición de NHTSA sobre kilómetros recorridos (VMT) define el tramo nocturno por hora de reloj, entre las 18:00 y las 5:59, y le atribuye cerca del 25% del VMT [11]. Es decir que no se considera un tiempo de crepúsculo, por lo que las imágenes de amanecer y atardecer de BDD100K quedan fuera del promedio ponderado, que se calcula solo sobre los bloques diurno y nocturno, decisión respaldada por el resultado empírico de esta misma memoria (§5.3.2), donde la iluminación de transición muestra una caída media de 2,5% frente al día, un orden de magnitud menor que la nocturna, de modo que su exclusión no altera el factor de calidad en más de 0,002. Se prefirió esta convención antes que introducir un peso crepuscular tomado del conteo de víctimas fatales, puesto que mezclar una participación en kilómetros recorridos con una participación en siniestralidad haría que los tres pesos no fueran homogéneos entre sí.

$$q_k^{\text{tr}} = \frac{0,75 \widehat{\text{mAP}}_k^d + 0,25 \widehat{\text{mAP}}_k^n}{\max_{k'}(0,75 \widehat{\text{mAP}}_{k'}^d + 0,25 \widehat{\text{mAP}}_{k'}^n)} \quad (6)$$

El **factor de calidad de seguridad** q_k^{seg} incorpora que la tasa de mortalidad por milla nocturna es $\eta_n = 49/25 \div 51/75 = 2,882$ veces la diurna (medición de NHTSA para Estados Unidos trasladada como supuesto declarado [11], dado que el 49% de las fatalidades ocurren en el 25%

del VMT nocturno, mientras el 51% restante se distribuye en el 75% diurno), puesto que la detección nocturna tiene un valor social mayor dado que los errores cuestan más vidas. Los pesos se ponderan por severidad y se renormalizan:

$$q_k^{\text{seg}} = \frac{0,75 \widehat{\text{mAP}}_k^d + 0,25 \eta_n \widehat{\text{mAP}}_k^n}{\max_{k'}(0,75 \widehat{\text{mAP}}_{k'}^d + 0,25 \eta_n \widehat{\text{mAP}}_{k'}^n)}, \quad \eta_n = \frac{49/25}{51/75} = 2,882 \quad (7)$$

Con $\eta_n = 2,882$, los pesos efectivos de q_k^{seg} son $\approx (0,51, 0,49)$ para (día, noche).

Los **beneficios anuales** emplean el factor correspondiente a cada mecanismo:

$$B_{\text{tiempo},k} = B_{\text{cong}} \cdot r_c \cdot \varphi \cdot q_k^{\text{tr}} \quad (8)$$

$$B_{\text{accid},k} = B_{\text{accid,RM}} \cdot r_a \cdot q_k^{\text{seg}} \quad (9)$$

$$B_{\text{amb},k} = E_{\text{CO}_2} \cdot r_c \cdot \varphi \cdot p_{\text{CO}_2} \cdot q_k^{\text{tr}} \quad (10)$$

$$B_{\text{total},k} = B_{\text{tiempo},k} + B_{\text{accid},k} + B_{\text{amb},k} \quad (11)$$

El **costo de cómputo** incluye CAPEX de equipos y OPEX de energía descontada, el **costo total** suma el componente común (infraestructura, plataforma y mantenimiento). Los símbolos de las ecuaciones siguientes son:

N	intersecciones desplegadas (3.250)	c_{disp}	costo por dispositivo de inferencia (USD)
$N_{\text{disp},k}$	dispositivos necesarios para el modelo k	c_{inst}	instalación por dispositivo (USD)
FPS_k	velocidad de inferencia del modelo k (fps)	P_k	potencia media del modelo k (W)
f_{req}	FPS mínimos por cámara (15 fps)	H	horas de operación anuales (8.760)
c_{infra}	costo de control adaptativo por intersección (USD)	p_{kWh}	precio de la energía (USD/kWh)
c_{plat}	plataforma central de gestión (USD)	f_{mant}	fracción de mantenimiento anual sobre CAPEX
T	horizonte de evaluación (años)	i	tasa social de descuento

$$C_{\text{comp},k} = N_{\text{disp},k} \cdot (c_{\text{disp}} + c_{\text{inst}}) + \sum_{t=1}^T \frac{N_{\text{disp},k} \cdot (P_k/1000) \cdot H \cdot p_{\text{kWh}}}{(1+i)^t}, \quad N_{\text{disp},k} = \left\lceil \frac{N}{\lfloor \text{FPS}_k / f_{\text{req}} \rfloor} \right\rceil \quad (12)$$

$$C_k = C_{\text{comp},k} + N \cdot c_{\text{infra}} + c_{\text{plat}} + \sum_{t=1}^T \frac{f_{\text{mant}} \cdot (N \cdot c_{\text{infra}} + c_{\text{plat}} + C_{\text{comp},k})}{(1+i)^t} \quad (13)$$

De los parámetros de costo, la plataforma central de gestión no proviene de una fuente única sino de una **estimación por analogía** sobre los comparables documentados en la base de costos ITS del USDOT [72], actualizados a dólares de 2026 con el IPC-US del BLS. La ficha 2008-sc00162 reporta que el software central de gestión con su integración costó entre US\$0,25 millones, correspondientes a una solución comercial de catálogo (COTS, *commercial off-the-shelf*) en el Distrito 4 de la agencia de transporte de Florida, y US\$2 millones en el Distrito 6 de FDOT, ambos en dólares de 2008, equivalentes a US\$0,38 y US\$3,07 millones de 2026 al aplicar el factor 1,5361. La ficha 2025-sc00573 de la misma base estima la modernización del ATMS de Chicago en US\$5 millones de 2022, equivalentes a US\$5,65 millones de 2026 con factor 1,1301. Sobre esos tres valores normalizados se construye una estimación de tres puntos [73]. La **moda** se fija en el Distrito 6 (US\$3,07 millones), ya que se trata de un costo ejecutado por una agencia pública, corresponde exactamente al alcance de c_{plat} (software central más integración, sin equipamiento de terreno) y proviene de una red de escala semejante a la de Santiago. El **mínimo** se fija en la opción COTS (US\$0,38 millones), que representa el escenario en que la UOCT adopta una plataforma de catálogo en vez de una integración a medida. El **máximo** se fija en Chicago (US\$5,65 millones), dado que su cifra corresponde a un paquete de once componentes más amplio que el requerido aquí, incluyendo medidas automatizadas de desempeño de semáforos, soporte a vehículos conectados e integración de datos de estacionamiento, de modo que acota por arriba antes que describir el caso central. De ahí resulta $c_{\text{plat}} = \text{US\$3,07 millones}$ (pago único), valor cuya media Beta-PERT asociada es US\$3,05 millones, prácticamente coincidente con la moda. El valor cubre únicamente el software de plataforma y sus servicios asociados, desglosado en licenciamiento del software de gestión de

tránsito ($\approx$ US\$1,85 millones), middleware e integración de APIs con los sistemas de la UOCT ($\approx$ US\$0,72 millones) y servicios de implementación y capacitación ($\approx$ US\$0,50 millones), la infraestructura física se contabiliza por separado en c_{infra} .

El parámetro c_{infra} recoge el costo de dotar de control adaptativo a una intersección ya semaforizada, es decir el controlador, el gabinete, la detección y el licenciamiento del software de control, sin incluir las cámaras, el equipo de inferencia, la plataforma central ni el mantenimiento, que se contabilizan por separado. Su valor base es de US\$40.781 por intersección en USD de 2026, obtenido del promedio de US\$28.725 que la base de conocimiento del USDOT reporta para la implementación de control adaptativo a partir de una encuesta a siete agencias operadoras, excluidos los valores atípicos [74], actualizado con el IPC-US del BLS entre 2013 y 2026 (factor 1,4197). Se prefirió esta fuente porque reporta el costo de implementar control adaptativo por intersección, alcance compatible con el del parámetro. La ficha no precisa si ese costo corresponde a la adaptación de semáforos existentes o a una instalación nueva, de modo que se asume lo primero como supuesto declarado de este trabajo, puesto que es el escenario propio de las agencias operadoras encuestadas, que ya contaban con red semaforizada.

La amplitud de su distribución de riesgo, PERT(-15% ; base; $+100\%$), merece justificación explícita, puesto que el techo duplica el valor base. La evidencia disponible muestra que el costo por intersección depende fuertemente del tamaño del despliegue, ya que en la evaluación de ocho corredores de Florida el costo unitario osciló entre US\$30.000 y US\$96.400 de 2021, y el máximo correspondió al corredor de East Van Fleet Drive en Bartow, de apenas cinco intersecciones y 1,1 millas [75]. El techo de la distribución se ancla en ese caso, dado que actualizarlo a USD de 2026 da US\$117.656 y descontar de esa cifra la mantención, que las cifras de Florida incluyen pero el modelo contabiliza aparte mediante f_{mant} , deja un costo de capital equivalente de US\$82.449, es decir aproximadamente el doble del valor base (factor 2,02 $\times$). Ese descuento asume, como supuesto declarado de este trabajo, que la mantención incluida en las cifras de Florida se acumuló bajo el mismo horizonte de 5 años y la misma tasa de 5,5% del modelo, puesto que la ficha no precisa el período sobre el que se acumuló. De este modo

la simulación admite el escenario en que Santiago no captura ninguna economía de escala y termina pagando por cada una de sus 3.250 intersecciones lo que pagó el despliegue más caro documentado por el USDOT.

Existe además una referencia todavía más alta, puesto que el piloto de control adaptativo con inteligencia artificial de Maricopa County, en Arizona, se cotizó en US\$115.810 de capital por intersección más US\$10.050 anuales de operación [76]. No se adoptó como techo por dos razones. La primera es de alcance, ya que esa cifra empaqueta las cámaras y el cómputo de inteligencia artificial dentro del mismo monto, componentes que este modelo contabiliza por separado y que en el caso de Santiago tienen un costo marginal muy inferior, dado que la red de cámaras de la UOCT ya está instalada y el equipo de inferencia se estima en US\$700. Incorporarla entera en c_{infra} implicaría contabilizar dos veces esos componentes. La segunda es de naturaleza del dato, puesto que se trata de una cotización del proveedor y no de un costo ejecutado y verificado, a diferencia de las encuestas a agencias operadoras que sustentan las otras dos cifras. Con todo, su magnitud ya queda cubierta por el techo adoptado, dado que el corredor de Bartow, el más pequeño de la muestra de Florida, arroja US\$117.656 de 2026, prácticamente el mismo valor que el piloto de Maricopa, de modo que el escenario de costo unitario alto está representado en la distribución aunque esta cotización no se incorpore.

Los **criterios de decisión** empleados son:

$$\text{BCR}_k = \frac{\sum_{t=1}^T \frac{B_{\text{total},k}}{(1+i)^t}}{C_k}, \quad \text{VAN}_k = \sum_{t=1}^T \frac{B_{\text{total},k}}{(1+i)^t} - C_k, \quad \text{ROI}_k = \frac{\text{VAN}_k}{C_k} \quad (14)$$

El **tiempo de recuperación del capital** (TRC) es el tiempo T_k en que el beneficio anual nominal iguala el costo total de la inversión, sin descontar, obtenido como:

$$T_k = 12 \times \frac{C_k}{b_k} \quad (15)$$

donde b_k es el beneficio social anual del modelo k y el resultado se expresa en meses. Se reporta sin descuento por su lectura directa para el decisor (meses hasta recuperar la inversión al ritmo de beneficio corriente), el criterio de decisión formal de este trabajo es el VAN y el BCR, que sí incorporan la tasa social de descuento.

La decisión se toma con la relación beneficio-costos social (BCR) y el retorno de la inversión (ROI), los dos criterios que el propio USDOT emplea para evaluar proyectos de control adaptativo de semáforos [77], tratando el riesgo mediante una simulación de Montecarlo [78] que muestrea simultáneamente los 18 parámetros más inciertos del modelo desde distribuciones documentadas (§4.6), entre ellos los porcentajes de reducción, los agregados macro de congestión y siniestralidad, los costos unitarios y el precio del kWh. Los flujos se expresan en moneda constante, lo que se declara de forma explícita. El rango de referencia de BCR de 5,4 a 12,8 proviene de la evaluación de un sistema real de control adaptativo de semáforos del Departamento de Transporte de Estados Unidos [52], donde el extremo inferior de 5,4 corresponde a los beneficios sin contar seguridad y el superior de 12,8 a los mismos beneficios incluyéndola, sirve para situar los resultados de la memoria dentro de lo que han logrado despliegues comparables. La recomendación final del modelo surge de cruzar este valor social con el costo de cómputo de cada tamaño. El despliegue se modela en el borde (edge), con un equipo de inferencia junto a las cámaras, de modo que el número de dispositivos y la energía escalan con el tamaño del modelo. En el modelo de borde adoptado, con cómputo en el dispositivo junto a la cámara (como en Las Condes), el costo de transmisión corresponde solo a los resultados de la inferencia (conteos, alertas), no al flujo de video completo, lo que mantiene el costo de comunicaciones bajo, si el dispositivo de borde fuese un nodo intermedio que recibe video de varias cámaras por red, habría un costo de ancho de banda adicional, aunque ese costo es común a todos los modelos y no afecta la comparación relativa entre ellos.

4.6) Diseño del análisis de riesgo (Obj. 6)

Un análisis de escenarios puntuales responde qué pasaría si los parámetros tomaran valores extremos, pero no cuantifica cuán probable es cada resultado. Para incorporar esa dimensión, el riesgo del modelo económico se evalúa mediante una simulación de Montecarlo [78] de 100.000 iteraciones con semilla fija, de modo que el ejercicio completo es reproducible. En cada iteración se muestrean simultáneamente los 18 parámetros más inciertos del modelo desde las distribuciones de la tabla siguiente y se recalcula el flujo completo (beneficios anuales, valor presente a 5 años al 5,5%, VAN, BCR y vidas salvadas), los cinco modelos YOLO11 se evalúan con la misma muestra de parámetros, de manera que cada iteración representa un mismo estado del mundo y la comparación entre modelos es pareada. Los parámetros con rango publicado en la literatura se modelan con distribuciones Beta-PERT(mín; moda; máx), estándar en análisis cuantitativo de riesgo de proyectos porque suaviza los extremos frente a la triangular [73], mientras que los agregados oficiales puntuales sin rango publicado se modelan con distribuciones triangulares simétricas cuya amplitud ($\pm 10\%$ a $\pm 20\%$ según el parámetro) es un supuesto de sensibilización declarado de este trabajo, en línea con la guía de análisis beneficio-costos del USDOT [69], que recomienda sensibilizar los parámetros de mayor incertidumbre sin prescribir una amplitud específica. En todos los casos la moda coincide con el valor del escenario base.

Parámetro	Base	Distribución	Sustento del rango
Ganancia plena ASCT (arteriales), g_{ASCT}	10%	PERT(7%; 10%; 25%)	Piso conservador FHWA EDC-1 (2011), que reporta mejora promedio superior a 10% [62], extremos según despliegues USDOT ITS y Surtrac [64, 5]
Ganancia incremental sobre SCOOT, g_{SCOOT}	2,5%	PERT(0%; 2,5%; 5%)	Mín. 0%: mejora nula en corredores que ya funcionan bien [60, 61]
Ganancia en bajo flujo, g_{low}	5%	PERT(0%; 5%; 10%)	Resultados inconsistentes con AADT < 35.000 [61], moda mitad de la plena, supuesto declarado
Share del retardo en corredores SCOOT, s_{SCOOT}	40%	Triangular(30%; 40%; 50%)	Supuesto declarado anclado en la concentración Pareto del flujo [53]
Share del retardo en bajo flujo, s_{low}	15%	Triangular(10%; 15%; 20%)	Supuesto declarado

Parámetro	Base	Distribución	Sustento del rango
Reducción de accidentes, r_a	7,8%	PERT(6,2%; 7,8%; 17%)	Máx. del rango publicado [60], mín. = base -20%, supuesto declarado
Cobertura de congestión abordable, φ	36,8%	PERT(27,6%; 36,8%; 46%)	Derivado: 46% del retardo fuera de autopistas [65] $\times$ 80% del flujo en el tramo superior de la jerarquía vial [53], máx. en la cota medida y banda simétrica $\pm 9,2$
Costo congestión Santiago (MUSD 2026/año)	1.353	Triangular $\pm 20\%$	Transferencia BID [4] a 2026, amplitud supuesto declarado
Costo siniestros nacional (MUSD 2026/año)	4.364	Triangular $\pm 10\%$	Cifra oficial CONASET [66]; supuesto declarado de valoración
Share RM del costo social, s_{RM}	22,9%	Triangular(20,6%; 22,9%; 25,2%)	Derivado del desglose de víctimas por gravedad de la RM y de los costos unitarios por gravedad [67, 66]; banda $\pm 10\%$ por variación interanual
Share RM de fallecidos, s_{RM}^v	21,7%	Triangular(19,5%; 21,7%; 23,8%)	312 de 1.439 fallecidos del país en 2024 [67]
Emisiones CO ₂ transporte RM (Mton/año)	10,62	Triangular $\pm 15\%$	Incertidumbre típica de inventarios de GEI, supuesto declarado
Precio social del carbono (USD 2026/ton)	66,6	Triangular $\pm 20\%$	Valor normativo MDS [50]; amplitud supuesto declarado
Costo infraestructura por intersección (USD 2026)	40.781	PERT(-15%; base; +100%)	Moda USDOT ITS KRS [74]; techo anclado al despliegue más caro documentado [75]
Plataforma central de gestión (MUSD 2026)	3,07	PERT(0,38; 3,07; 5,65)	Mín. COTS FDOT D4; moda FDOT D6 Miami-Dade; máx. ATMS Chicago; los tres deflactados a 2026 [72]
Mantenimiento anual (% del CAPEX)	10%	PERT(5%; 10%; 15%)	Rango habitual O&M de sistemas ITS, supuesto declarado
Dispositivo edge más instalación (USD 2026)	700	PERT(400; 700; 1.200)	Rango de equipos clase Jetson, supuesto declarado
Precio energía (USD 2026/kWh)	0,20	Triangular $\pm 20\%$	Variabilidad tarifaria comercial RM, supuesto declarado

Tabla 5: Distribuciones de los 18 parámetros muestreados en la simulación de Montecarlo. La moda de cada distribución coincide con el valor del escenario base (§4.5 y §6.1). La reducción de congestión r_c no se muestrea directamente sino que se deriva en cada iteración como $s_{SCOOT} \cdot g_{SCOOT} + (1 - s_{SCOOT} - s_{low}) \cdot g_{ASCT} + s_{low} \cdot g_{low}$, preservando la lógica de la tipología del modelo determinista (§4.5). Las Beta-PERT usan $\lambda = 4$ [73]. Fuente: elaboración propia en base a las fuentes indicadas.

Quedan fijos los parámetros que no representan incertidumbre del entorno. La tasa social de descuento de 5,5% es una convención normativa del Sistema Nacional de Inversiones [50], el

horizonte de 5 años es una decisión de diseño atada a la vida útil del hardware [69], las 3.250 intersecciones son un dato administrativo del alcance del proyecto [9], y los pesos de exposición por tramo horario junto al factor de severidad nocturna son mediciones empíricas de NHTSA en Estados Unidos trasladadas como supuesto declarado (§4.5) [11], al igual que los 1.439 fallecidos de 2024 son registro oficial de CONASET [7]. Los factores de calidad q^{tr} y q^{seg} también se mantienen fijos. Son cocientes de mAP de los cinco modelos evaluados sobre exactamente las mismas 3.000 imágenes de validación, de modo que la comparación es pareada, ya que las imágenes difíciles penalizan a los cinco modelos a la vez y el cociente entre modelos resulta más estable que cada mAP por separado. Perturbarlos con ruido independiente ignoraría ese emparejamiento y exageraría su variabilidad, por lo que se excluyen del muestreo y esa exclusión se declara como limitación del análisis.

Sobre este diseño se planifican tres verificaciones de robustez, cada una dirigida a un supuesto distinto, cuyos resultados se reportan en §6.3.

La primera revisa la amplitud asignada al precio social del carbono. El Ministerio de Desarrollo Social lo publica como valor normativo puntual [50], sin rango de incertidumbre asociado, de modo que el $\pm 20\%$ con que se sensibiliza es un supuesto declarado de este trabajo, en línea con la recomendación de la guía de análisis beneficio-costos del USDOT de sensibilizar los parámetros de mayor incertidumbre [69].

La segunda revisa el supuesto de independencia entre las ganancias del control adaptativo (g_{ASCT} y g_{SCOOT}) y la reducción de accidentes (r_a). Como la literatura no reporta una estimación conjunta de cuánto se mueven juntas, se repite la simulación completa imponiendo entre ellas una cópula gaussiana de $\rho = 0,5$ mediante reordenamiento de rangos [79], un valor deliberadamente exigente que opera como prueba de estrés antes que como estimación. El procedimiento preserva exactamente las distribuciones marginales y solo altera con qué muestra de un parámetro se empareja cada muestra del otro.

La tercera evalúa de forma determinista los **vértices** del espacio de parámetros, es decir los pun-

tos en que los 18 parámetros toman simultáneamente su extremo adverso o su extremo favorable. Sirve para verificar si el soporte del modelo puede contener VAN negativos y para dimensionar el rango completo posible.

La segunda verificación merece un desarrollo aparte, puesto que antes de imponer una correlación hay que descartar que el vínculo entre ambos parámetros sea un artefacto de las fuentes. La sospecha de partida es razonable, ya que la ganancia de congestión y la reducción de accidentes dependen ambas de qué tan bien quedó implementado el control adaptativo en la red concreta, de manera que una implementación buena debería mover las dos hacia arriba y una deficiente, las dos hacia abajo.

Revisadas las fuentes, no existe doble conteo muestral entre ellas. Las que sostienen las ganancias de congestión miden exclusivamente indicadores de movilidad, esto es tiempo de viaje, demora o detenciones, y nunca siniestros. La distribución de g_{ASCT} se construye sobre mediciones de tiempo de viaje del programa EDC-1 de la FHWA [62] y de los despliegues de Oakland County y Los Ángeles [64], mientras que la de g_{SCOOT} se construye sobre el hallazgo de mejora nula en corredores que ya funcionan bien [60, 61]. La fuente de r_a , en cambio, mide exclusivamente conteos y tasas de siniestros mediante CMF, y nunca tiempo de viaje, puesto que se construye sobre 20 intersecciones de Florida evaluadas con un antes-después full-Bayes (CMF de siniestros totales = 0,922, verificado directamente contra el artículo original) [68] y sobre las 47 intersecciones de 10 corredores piloto de Virginia en que se hizo el análisis de seguridad [60].

La correlación posible es entonces estructural y no muestral, dado que no proviene de reutilizar los mismos datos sino de que ambos efectos dependen de la calidad de una misma implementación. La evaluación de Virginia entrega evidencia directa de esa co-ocurrencia, ya que midió en los mismos corredores tanto mejoras de tiempo de viaje en la vía principal como una reducción de 17% de los siniestros, como dos métricas separadas del mismo despliegue [60].

5) Resultados

5.1) Desempeño global y costo

Modelo	Precisión	Recall	mAP50	mAP50-95	Parámetros	FLOPs
YOLO11n	0,586	0,392	0,430	0,251	2,6 M	6,5 B
YOLO11s	0,621	0,470	0,503	0,297	9,4 M	21,5 B
YOLO11m	0,651	0,506	0,554	0,330	20,1 M	68,0 B
YOLO11l	0,673	0,510	0,561	0,332	25,3 M	86,9 B
YOLO11x	0,673	0,529	0,571	0,340	56,9 M	194,9 B

Tabla 6: Comparación global (conjunto de validación, 3.000 imágenes). Fuente: elaboración propia.

Hallazgo del trade-off precisión-costo: la precisión crece con fuerza en los primeros saltos y después se estanca. El mAP50-95 sube de 0,251 (n) a 0,297 (s) y a 0,330 (m), con incrementos de +0,046 y +0,033, pero a partir de YOLO11m la mejora se aplana, ya que YOLO11l aporta solo +0,002 (0,332) y YOLO11x suma +0,008 adicional (0,340). En contraste, el costo sigue creciendo de forma multiplicativa, puesto que pasar de YOLO11m a YOLO11x multiplica por cerca de 2,8 los parámetros (de 20,1 a 56,9 M) y por cerca de 2,9 los FLOPs (de 68,0 a 194,9 B) para ganar apenas 0,010 de mAP50-95. Esto evidencia rendimientos decrecientes marcados a partir de YOLO11m, dado que los modelos pequeños y medianos capturan casi todo el desempeño alcanzable en este subset, mientras que los grandes añaden costo sin una mejora proporcional de precisión. Este es el criterio económico central para decidir qué modelo desplegar (ver §6/Discusión).

Para verificar que las diferencias entre modelos son estadísticamente significativas y no artefactos del tamaño del conjunto de validación, se aplicó remuestreo pareado con 1.000 iteraciones sobre las 3.000 imágenes de validación (ver §4.3).

Modelo	mAP50-95 (remuestreo)	IC 95% inferior	IC 95% superior
YOLO11n	0,244	0,234	0,258
YOLO11s	0,296	0,283	0,311
YOLO11m	0,326	0,312	0,343
YOLO11l	0,330	0,317	0,347
YOLO11x	0,338	0,324	0,355

Tabla 7: Intervalos de confianza al 95% para mAP50-95, estimados por remuestreo pareado (1.000 iteraciones, 3.000 imágenes val). El punto central puede diferir en $\leq 0,007$ del mAP50-95 oficial de la tabla anterior, dado que el remuestreo recalcula la métrica desde las predicciones cacheadas para evitar 1.000 corridas del modelo (§4.3), lo que introduce una diferencia menor frente a la evaluación oficial de Ultralytics, lo relevante para la significancia estadística es el ancho y solapamiento de los IC, no la coincidencia exacta del punto central. Fuente: elaboración propia.

Los IC muestran que n y s se distinguen de forma clara (rangos sin solapamiento), al igual que s respecto de m. Por el contrario, los IC de YOLO11m, YOLO11l y YOLO11x se solapan de forma considerable, lo que confirma que las diferencias entre esos tres modelos no son estadísticamente significativas y respalda la afirmación de rendimientos decrecientes a partir de YOLO11m.

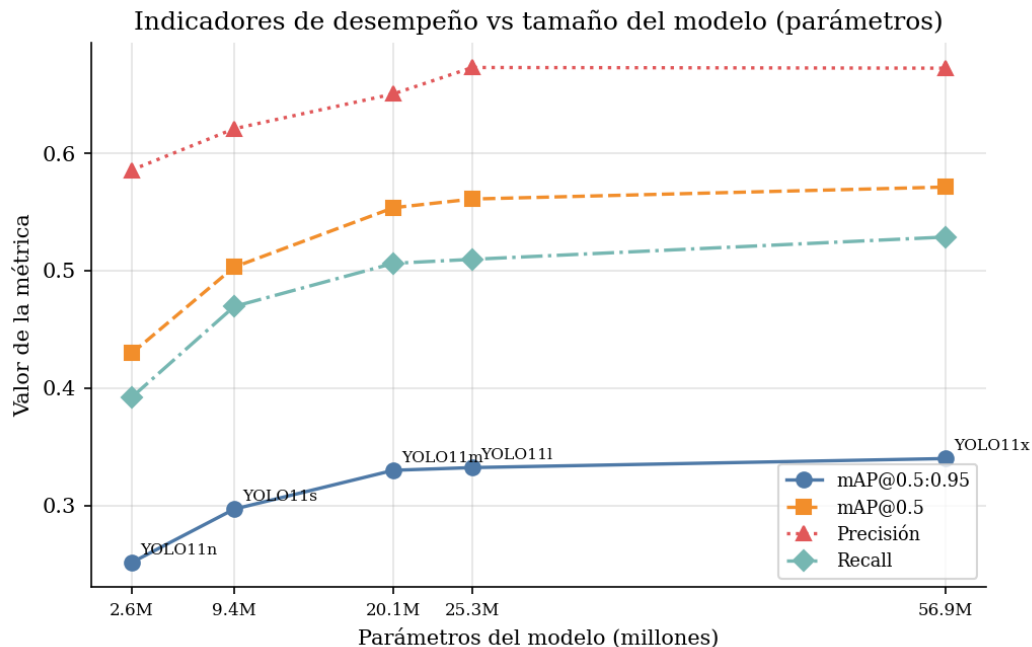


Figura 6. Indicadores de desempeño (mAP@0.5:0.95, mAP@0.5, Precisión, Recall) en función del número de parámetros del modelo (escala logarítmica). La curva de mAP@0.5:0.95 ilustra el estancamiento a partir de YOLO11m. Fuente: elaboración propia.

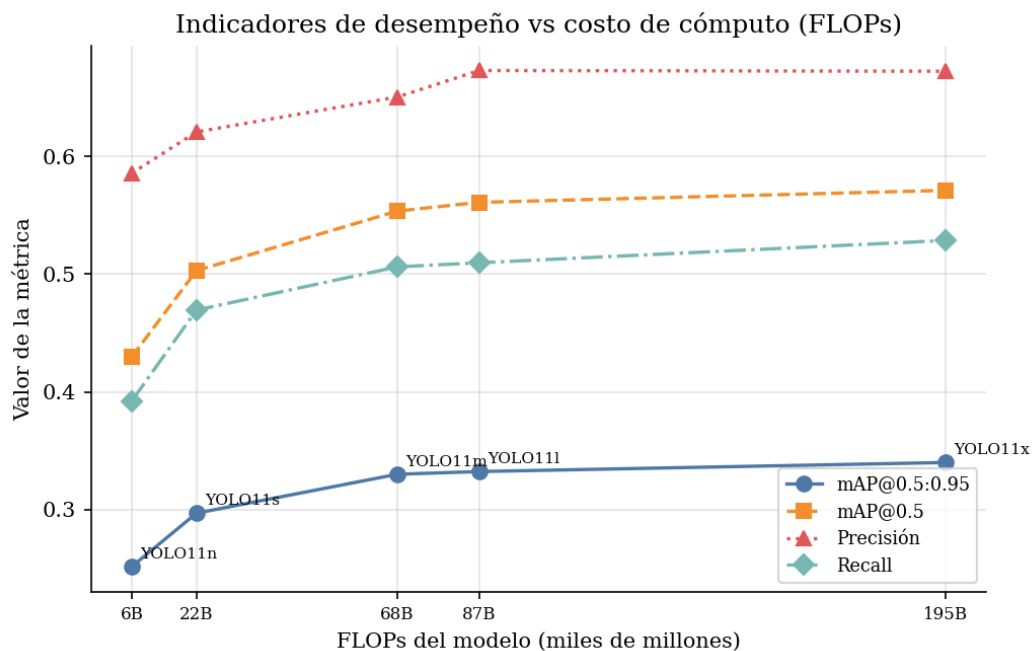


Figura 7. Indicadores de desempeño en función del costo de cómputo (FLOPs, escala logarítmica). Muestra que el costo crece de forma multiplicativa mientras la ganancia de precisión se aplana, confirmando los rendimientos decrecientes a partir de YOLO11m. Fuente: elaboración propia.

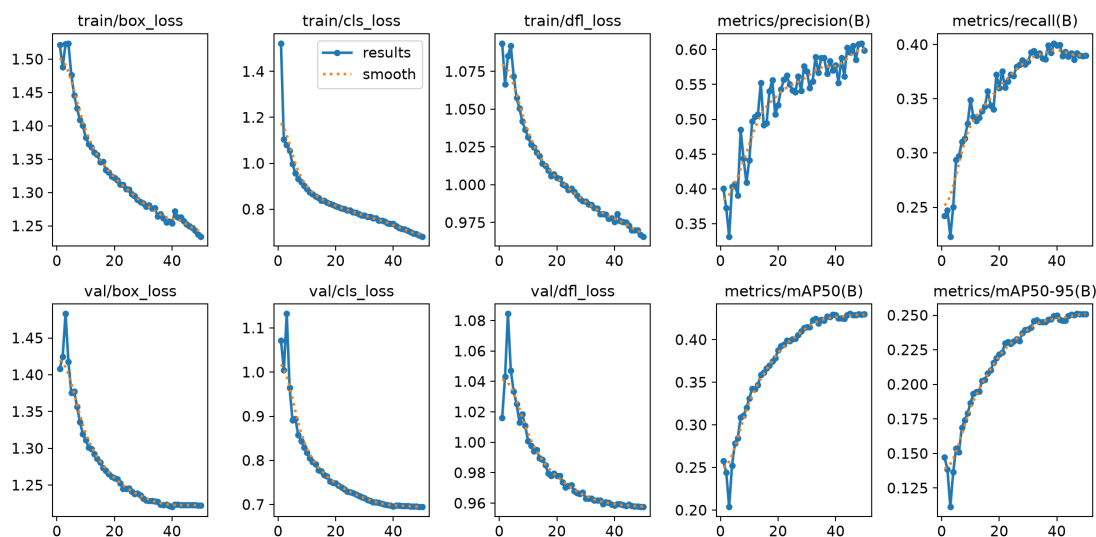


Figura 8a. Curvas de entrenamiento YOLO11n: evolución de las pérdidas de caja (box loss), clasificación (cls loss) y distribución focal (dfl loss) en entrenamiento y validación, junto con mAP@0.5 y mAP@0.5:0.95 por época. Fuente: elaboración propia.

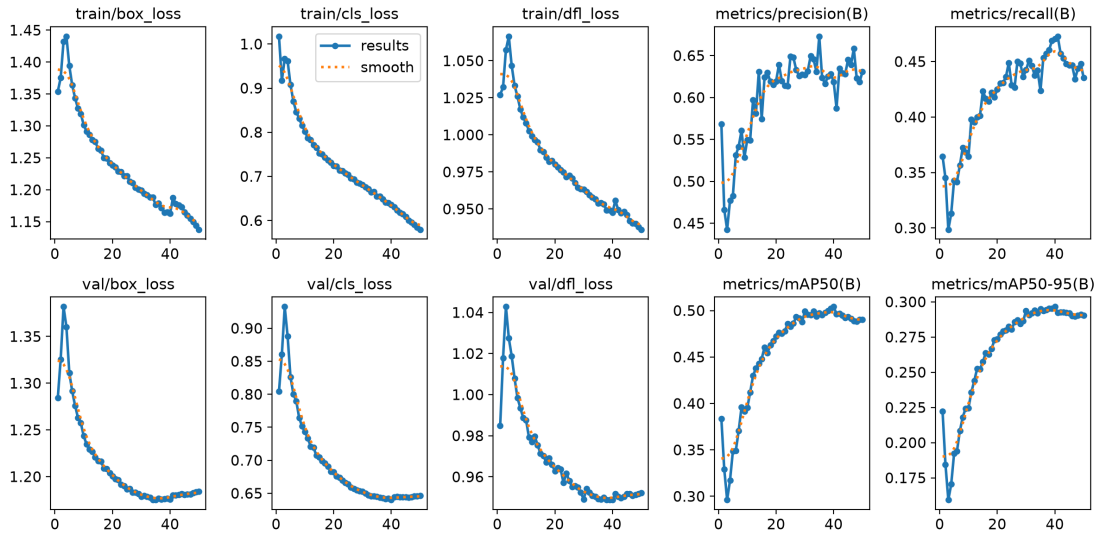


Figura 8b. Curvas de entrenamiento YOLO11s: ídem figura anterior. La convergencia es más rápida que YOLO11n y la pérdida de validación se estabiliza antes. Fuente: elaboración propia.

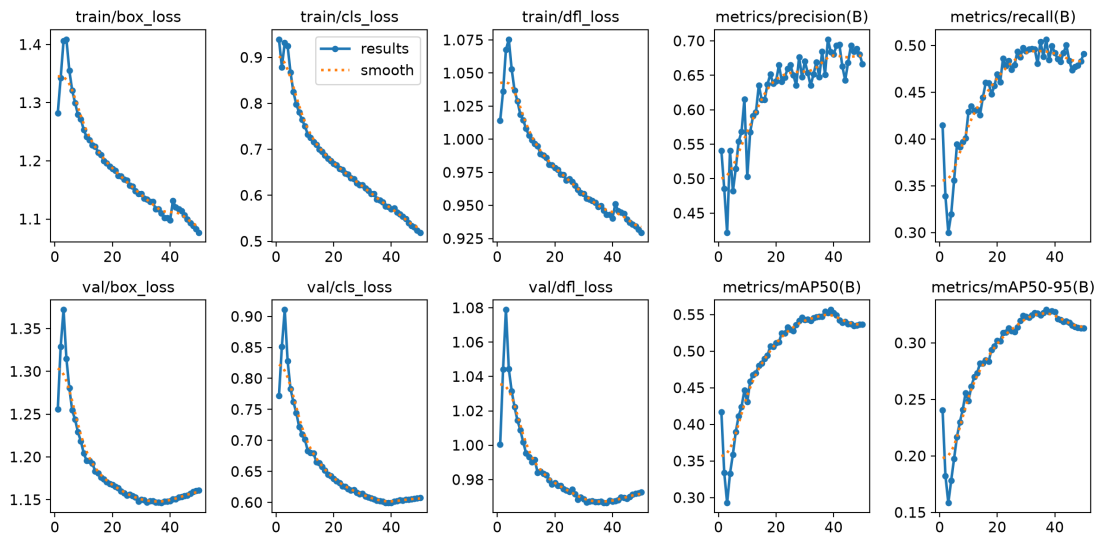


Figura 8c. Curvas de entrenamiento YOLO11m: el mAP@0.5:0.95 de validación alcanza su máximo relativo entre los modelos de esta familia para este subset, la pérdida de validación diverge levemente respecto al entrenamiento en las últimas épocas, consistente con el early stopping activado. Fuente: elaboración propia.

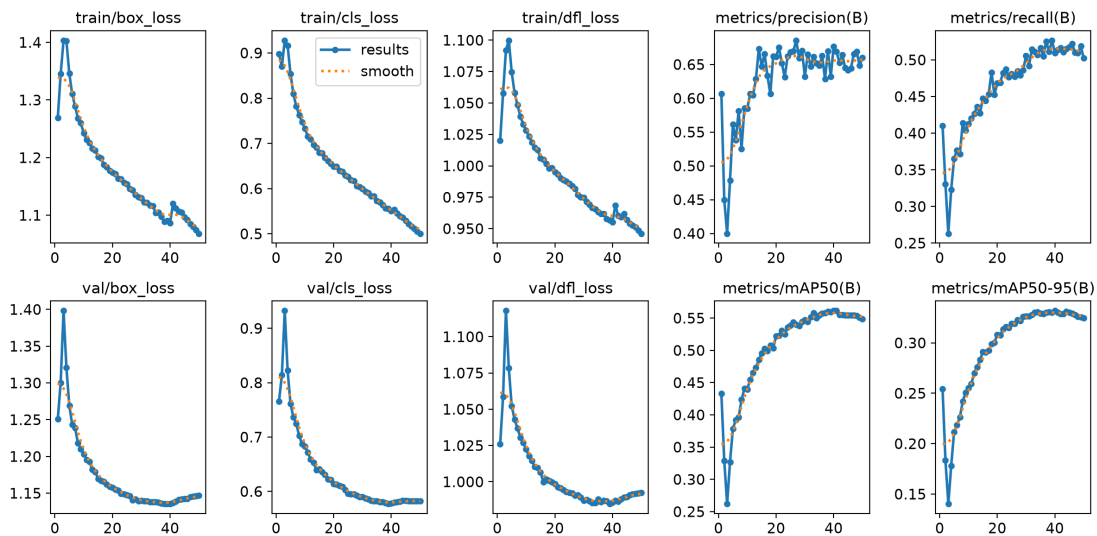


Figura 8d. Curvas de entrenamiento YOLO11l: la mejora marginal de mAP@0.5:0.95 respecto a YOLO11m es mínima (+0,002), evidenciando el estancamiento de la precisión con mayor capacidad de modelo. Fuente: elaboración propia.

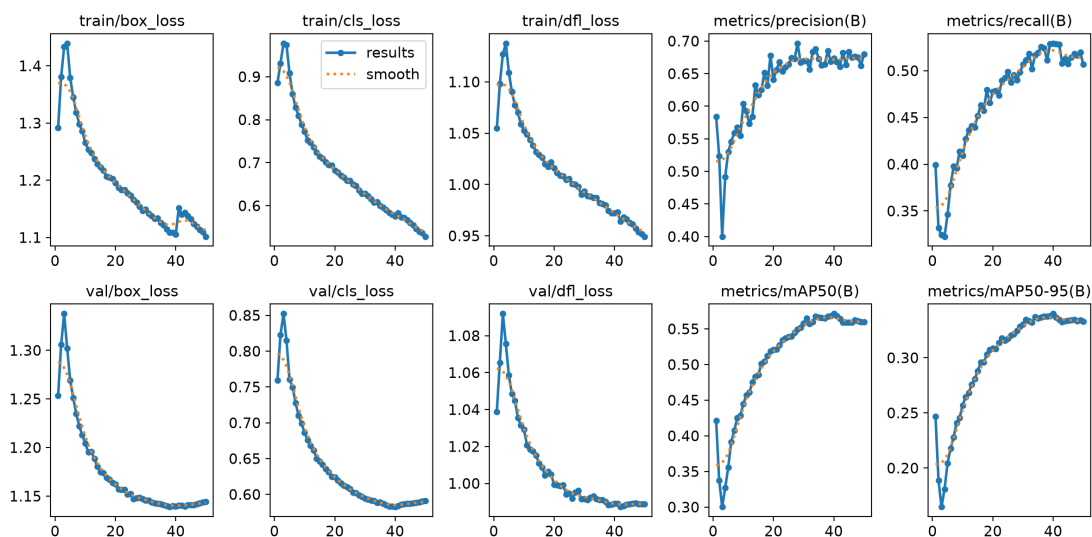


Figura 8e. Curvas de entrenamiento YOLO11x: modelo de mayor capacidad, la ganancia de mAP@0.5:0.95 sobre YOLO11m es de solo +0,010 a cambio de multiplicar por ~2,8 los parámetros y por ~2,9 los FLOPs. Fuente: elaboración propia.

5.2) Robustez por momento del día (Objetivo 3)

Condición	n_img	YOLO11n	YOLO11s	YOLO11m	YOLO11l	YOLO11x
Día (daytime)	1.575	0,264	0,309	0,343	0,346	0,355
Amanecer/atardecer	235	0,247	0,305	0,334	0,330	0,338
Noche (night)	1.182	0,221	0,260	0,295	0,301	0,308
Caída día→noche (sin ponderar)		-16,2%	-16,0%	-14,1%	-13,0%	-13,4%

Tabla 8: mAP50-95 por momento del día, sin ponderar (comparación directa). Fuente: elaboración propia.

Condición	n_img	YOLO11n	YOLO11s	YOLO11m	YOLO11l	YOLO11x
Día (daytime)	1.575	0,285	0,330	0,363	0,372	0,375
Amanecer/atardecer	235	0,278	0,319	0,366	0,352	0,357
Noche (night)	1.182	0,215	0,258	0,294	0,303	0,309
Caída día→noche (IPW)		-24,3%	-21,8%	-19,1%	-18,6%	-17,6%

Tabla 9: mAP50-95 ponderado por momento del día, controlando por clima y densidad (IPW). Fuente: elaboración propia.

Hallazgo central: **la noche es la condición crítica para los cinco modelos**, ya que todos pierden desempeño respecto del día. La caída relativa se reduce de forma leve al escalar el modelo, pasando de cerca de 24% en YOLO11n a cerca de 18% en YOLO11l y YOLO11x, lo que indica que un modelo mayor aporta algo de robustez nocturna pero no elimina la brecha. La degradación es mayor a la que sugiere la comparación directa sin la ponderación (entre 13% y 16%, primera tabla), ya que las imágenes nocturnas de BDD100K tienden a ocurrir en condiciones de menor densidad vehicular, lo que enmascaraba parte del efecto nocturno. El problema nocturno es por lo tanto estructural y no se resuelve solo con más capacidad de cómputo.

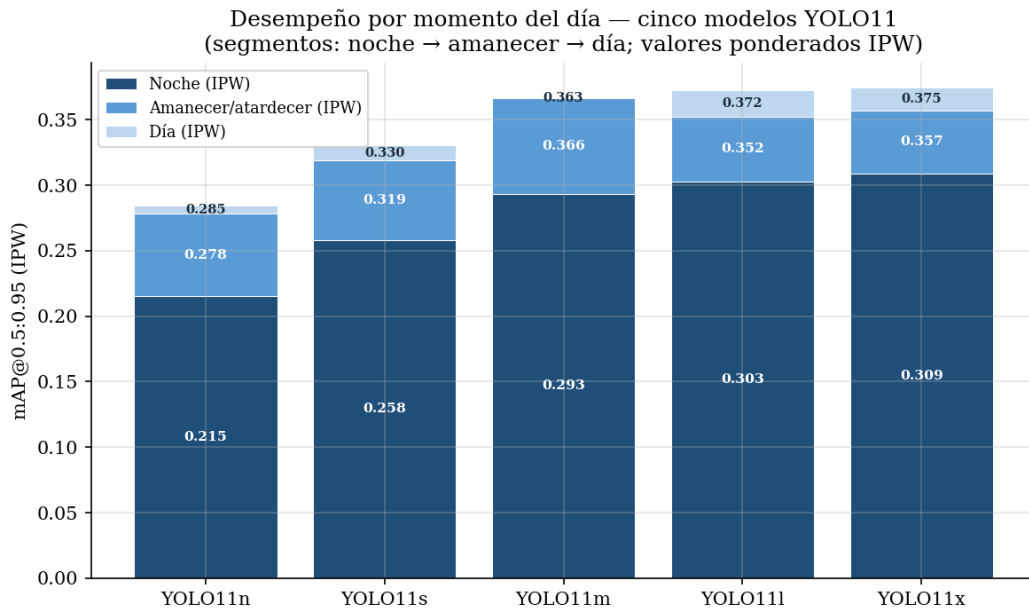


Figura 9. mAP@0.5:0.95 ponderado IPW por modelo, segmentado por momento del día (controlando por clima y densidad). Cada barra muestra los segmentos noche (base), amanecer (medio) y día (cima), los valores dentro de cada segmento son el mAP IPW de esa condición. Fuente: elaboración propia.

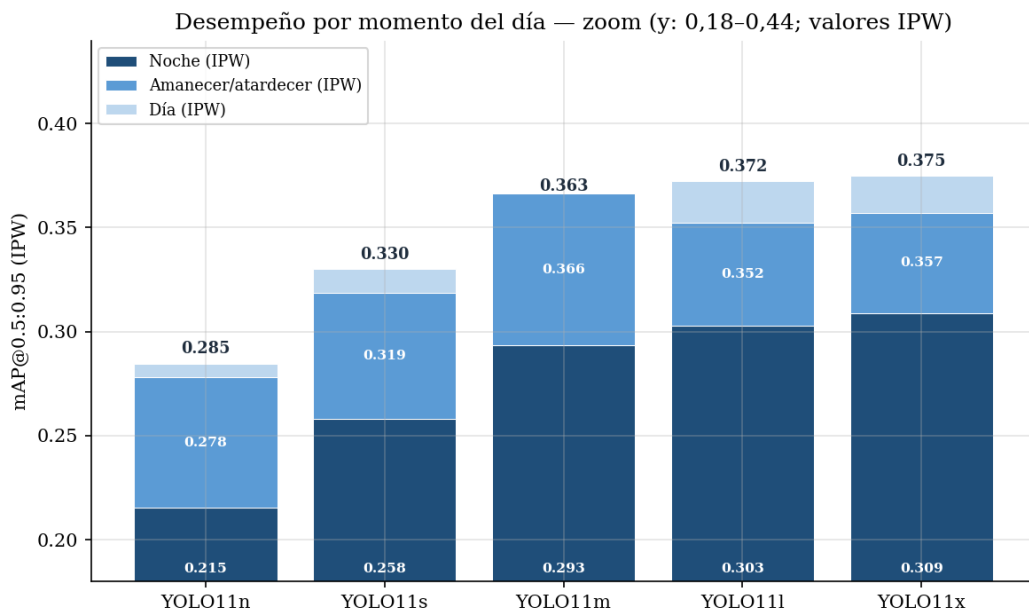


Figura 10. Ídem figura anterior, con eje y recortado (0,18 a 0,44) para visualizar mejor las diferencias, los valores IPW de noche, amanecer y día están etiquetados en cada barra. Fuente: elaboración propia.

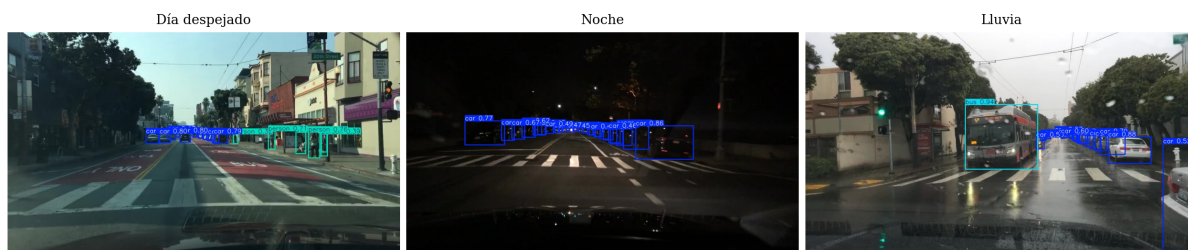


Figura 11. Detecciones del mismo modelo (YOLO11m entrenado en esta memoria, umbral de confianza 0,25) sobre tres imágenes de validación de BDD100K en día despejado, noche y lluvia. De día se detectan vehículos y peatones con confianzas altas; de noche los peatones dejan de detectarse y las confianzas de los vehículos caen; bajo lluvia la detección se mantiene pese a las gotas en el lente. Fuente: elaboración propia mediante inferencia sobre BDD100K [15].

5.3) Desempeño por clima

Clima	n_img	YOLO11n	YOLO11s	YOLO11m	YOLO11l	YOLO11x
Despejado (clear)	1.678	0,240	0,285	0,323	0,330	0,339
Nublado (overcast)	347	0,264	0,310	0,342	0,339	0,345
Parcialmente nublado	213	0,263	0,328	0,348	0,370	0,372
Lluvia (rainy)	202	0,284	0,307	0,329	0,349	0,324
Nieve (snowy)	227	0,297	0,356	0,376	0,348	0,368
Niebla (foggy)	4	n/d	n/d	n/d	n/d	n/d

Tabla 10: mAP50-95 por clima, sin ponderar (comparación directa). Fuente: elaboración propia.

Clima	n_img	YOLO11n	YOLO11s	YOLO11m	YOLO11l	YOLO11x
Despejado (clear)	1.678	0,236	0,288	0,328	0,336	0,346
Nublado (overcast)	347	0,305	0,339	0,380	0,376	0,365
Parcialmente nublado	213	0,313	0,378	0,401	0,435	0,429
Lluvia (rainy)	202	0,234	0,264	0,279	0,302	0,297
Nieve (snowy)	227	0,280	0,304	0,341	0,323	0,315
Niebla (foggy)	4	n/d	n/d	n/d	n/d	n/d

Tabla 11: mAP50-95 ponderado por clima, controlando por hora del día y densidad vehicular (IPW). Fuente: elaboración propia.

La comparación directa (primera tabla) muestra lluvia y nieve con mAP similar o superior al de despejado en la mayoría de los modelos. Este patrón es un artefacto de la correlación entre condiciones: las escenas lluviosas y nevadas en BDD100K ocurren casi exclusivamente de día y en zonas de baja densidad vehicular, ambas condiciones que favorecen la detección. Tras controlar la distribución de hora del día y densidad (segunda tabla), la lluvia cae por debajo del despejado en todos los modelos, con una caída de entre el 1% (YOLO11n, donde la diferencia es marginal) y el 15% (YOLO11m). La lluvia pasa así a ser la peor condición climática tras la corrección IPW en los cinco modelos. El despejado, por su parte, queda por debajo de nublado y parcialmente nublado también en los cinco, lo que lo deja como segunda peor condición en YOLO11n, YOLO11s y YOLO11m, y como tercera en YOLO11l y YOLO11x, donde la nieve queda por debajo de él. Que una condición sin obstáculos de visibilidad rinda peor que el cielo nublado es el resultado que requiere explicación. Hay dos efectos que explican esto, uno de distribución del dataset y otro genuino de la condición.

El efecto de distribución IPW: la ponderación busca que todas las condiciones climáticas se evalúen sobre la misma mezcla de horario y densidad, tomando como referencia la composición del clima despejado. Cuando una celda cruzada no alcanza el mínimo de 15 imágenes se omite y el promedio se renormaliza sobre las celdas restantes, de modo que la mezcla efectiva deja de ser la misma entre condiciones. La tabla siguiente cuantifica esa desviación.

Clima	Imágenes nocturnas (val)	Peso de la referencia que sobrevive al umbral	Peso nocturno efectivo tras renormalizar
Despejado (clear)	992	99,4%	59,5%
Lluvia (rainy)	84	78,2%	60,7%
Nieve (snowy)	81	67,4%	49,3%
Nublado (overcast)	3	40,3%	0,0%
Parcialmente nublado	1	34,1%	0,0%

Tabla 12: Composición efectiva de cada condición climática tras la ponderación IPW. La distribución de referencia es la del clima despejado, con un 59,5% de peso en celdas nocturnas. La segunda columna indica qué fracción de ese peso corresponde a celdas que superan el mínimo de 15 imágenes, y la tercera, el peso nocturno que queda una vez renormalizado sobre las celdas sobrevivientes. Fuente: elaboración propia.

El despejado se evalúa con un 59,5% de peso en celdas nocturnas, mientras que el nublado y el parcialmente nublado lo hacen con un 0%, puesto que sus tres y una imagen nocturna respectivamente quedan bajo el umbral. Ambos números pasaron por el mismo procedimiento, pero describen mezclas distintas, una mayoritariamente nocturna frente a una enteramente diurna. La lluvia y la nieve, en cambio, conservan un peso nocturno de 60,7% y 49,3%, comparable al 59,5% del despejado, de manera que la comparación entre despejado y lluvia sí queda controlada y el hallazgo de que la lluvia rinde peor se sostiene. La cautela aplica específicamente al contraste entre despejado y las dos condiciones nubladas, y por eso se complementa con la restricción a imágenes diurnas de la sección siguiente.

El efecto de tamaño de objetos: la visibilidad reducida en lluvia y nieve limita el rango de anotación, de modo que los objetos registrados son más cercanos y más grandes. El análisis empírico de los bounding boxes de validación (§5.3.1) confirma que lluvia y nieve tienen solo un 39% de objetos pequeños ($A < 1.024 \text{ px}^2$) y un área mediana de 1.685 a 1.780 px^2 , frente al 43 a 46% de pequeños y 1.185 a 1.424 px^2 de mediana en el resto de condiciones. Un sesgo de anotación de esta forma está documentado en auditorías de etiquetado bajo baja visibilidad, donde los objetos pequeños o distantes son precisamente los que quedan sin anotar [49] y explica parte de la ventaja de lluvia y nieve en el mAP bruto, sin embargo, no explica la diferencia entre

despejado y nublado/parcialmente nublado, donde la proporción de objetos pequeños es similar o incluso mayor en estas últimas.

Por la combinación de ambos efectos, el análisis IPW de clima debe interpretarse con cautela: la diferencia clear vs. overcast/partly cloudy no refleja únicamente el efecto del clima sobre la detección, sino también la distribución desigual de horario nocturno entre condiciones. Una comparación más limpia requiere restringir el análisis a imágenes diurnas, controlando solo por densidad, esto se presenta en §5.3.1. La condición niebla se excluye del análisis, dado que cuenta con solo 4 imágenes en validación.

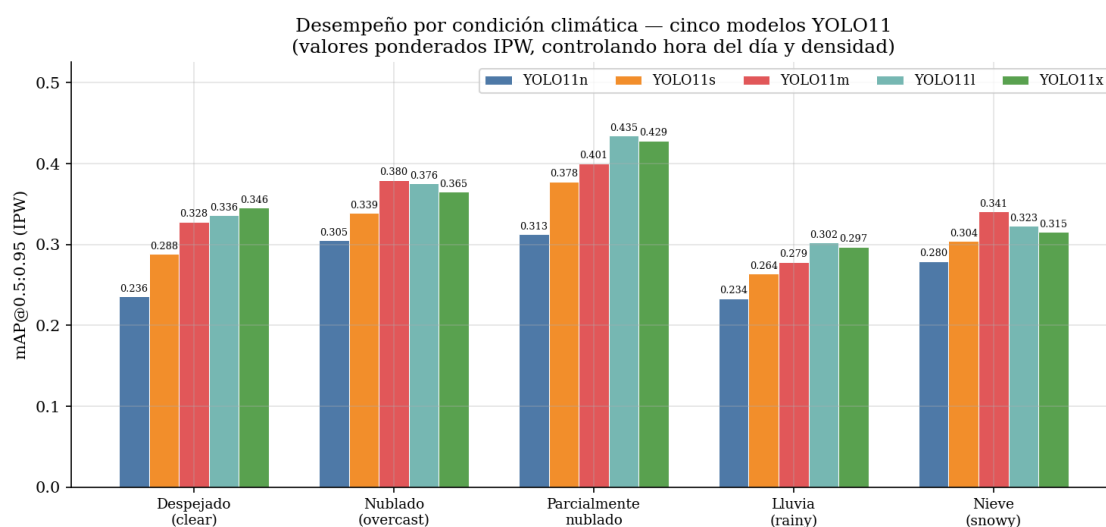


Figura 12. mAP@0.5:0.95 ponderado IPW por condición climática para los cinco modelos YOLO11 (controlando por hora del día y densidad). La lluvia muestra el mAP IPW más bajo en los cinco modelos, y el despejado queda por debajo de nublado y parcialmente nublado por su alta proporción de imágenes nocturnas (59%) en la distribución de referencia. Fuente: elaboración propia.

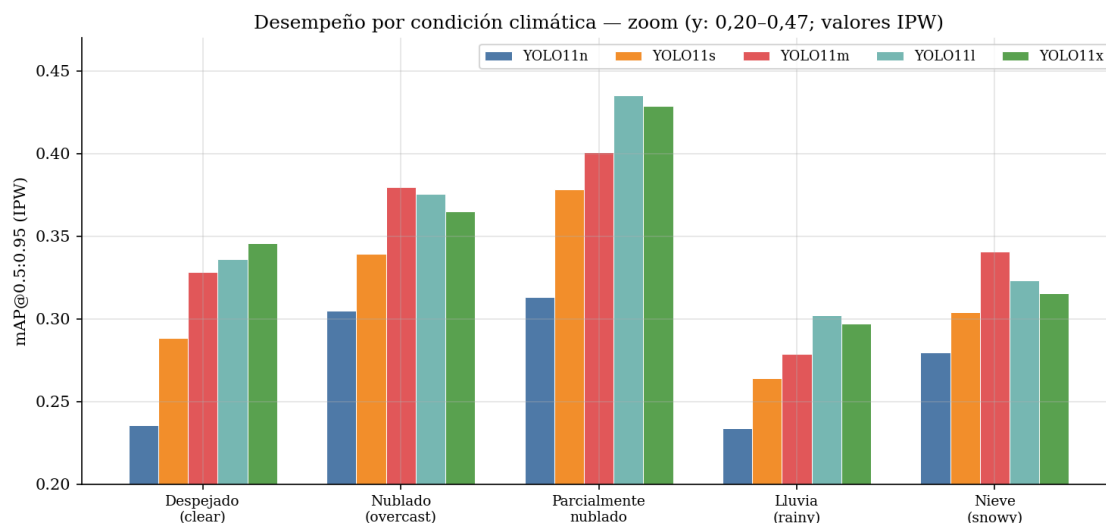


Figura 13. Ídem figura anterior, con eje y recortado (0,20 a 0,47) para visualizar mejor las diferencias entre condiciones climáticas (valores IPW). Fuente: elaboración propia.

5.3.1) Análisis climático restringido a imágenes diurnas

Siguiendo la segunda pasada del diseño climático (§4.3), se restringe el análisis a las 1.290 imágenes de validación con `timeofday = daytime` (excluyendo `undefined` y `foggy`) y se pondera solo por densidad vehicular, con pesos estimados sobre el subconjunto diurno: baja = 0,098; media = 0,529; alta = 0,374 ($n = 1.290$). La condición de lluvia excluye la celda de densidad baja ($n = 10 < 15$), los pesos se renormalizan sobre media y alta.

Clima	n (diurnas)	YOLO11n	YOLO11s	YOLO11m	YOLO11l	YOLO11x
Despejado (clear)	573	0,266	0,312	0,351	0,361	0,376
Nublado (overcast)	294	0,302	0,335	0,375	0,373	0,362
Parcialmente nublado	189	0,309	0,373	0,396	0,430*	0,423*
Lluvia (rainy)	107 [†]	0,298	0,340	0,339	0,363	0,333
Nieve (snowy)	127	0,298	0,344	0,371	0,353	0,373

Tabla 13: mAP50-95 ponderado IPW restringido a imágenes diurnas (controlando solo por densidad). * La celda parcialmente nublado × baja densidad se estima sobre $n = 19$, con intervalos de confianza al 95% muy anchos, de 0,60 a 0,87 en YOLO11l y de 0,62 a 0,86 en YOLO11x. [†] Lluvia excluye densidad baja ($n = 10$); pesos renormalizados. Fuente: elaboración propia.

El resultado principal es que, al eliminar el efecto nocturno, **el despejado deja de ser uniformemente la peor condición**, puesto que solo lo es en YOLO11n y YOLO11s, mientras que en YOLO11m y YOLO11x la peor pasa a ser la lluvia y en YOLO11l, la nieve. Lo que sí persiste en los cinco modelos es que el despejado queda por debajo de parcialmente nublado, y en cuatro de los cinco también por debajo de nublado, con YOLO11x como única excepción. Para interpretar este resultado se analizó la distribución del tamaño de los objetos anotados en validación (procedimiento descrito en §4.3), según los umbrales COCO: pequeño $A < 1.024 \text{ px}^2$, mediano $A < 9.216 \text{ px}^2$, grande $A \geq 9.216 \text{ px}^2$.

Clima	Objetos	% pequeño	% mediano	% grande	Mediana (px ²)	Media (px ²)
Despejado (clear)	19.480	42,6	39,1	18,3	1.424	9.274
Nublado (overcast)	5.015	45,7	37,8	16,5	1.233	8.577
Parcialmente nublado	2.971	46,2	38,3	15,5	1.185	7.989
Lluvia (rainy)	2.342	39,0	39,9	21,1	1.685	11.897
Nieve (snowy)	2.871	38,9	40,4	20,6	1.780	11.842

Tabla 14: Distribución de tamaño de objetos (umbrales COCO) en el conjunto de validación, por condición climática. Imagen 1.280 × 720 px. Excluidas undefined y foggy. Fuente: elaboración propia.

Los datos confirman **parcialmente** la hipótesis de sesgo de tamaño. Lluvia y nieve tienen la menor proporción de objetos pequeños (39%) y las mayores áreas mediana y media (1.685 a 1.780 y 11.842 a 11.897 px² respectivamente), lo que es coherente con que la visibilidad reducida acorta el rango de anotación y eleva el tamaño promedio de los objetos detectados. Sin embargo, **despejado no es la condición con más objetos pequeños**: nublado (45,7%) y parcialmente nublado (46,2%) tienen proporciones de pequeños mayores que despejado (42,6%) y áreas medianas menores. Por lo tanto, el sesgo de tamaño de objetos explica la ventaja de lluvia y nieve respecto del resto, pero no explica por qué despejado tiene mAP más bajo que nublado o parcialmente nublado en el análisis diurno. Ese diferencial puede deberse a características de la escena propias del clima despejado, como contraste extremo, deslumbramiento o reflejos en asfalto seco, que no se capturan en la distribución de tamaños.

La condición lluvia converge con nieve en el análisis diurno (ambas entre 0,298 y 0,373 según

el modelo), sin la penalización artificial que el IPW completo le imponía por su alta proporción de escenas diurnas. Esa convergencia no es uniforme entre modelos: en YOLO11n y YOLO11s, lluvia supera a despejado por 0,028 a 0,032 puntos, pero en YOLO11m y YOLO11l la relación se invierte (lluvia queda 0,012 y 0,043 puntos por debajo de despejado, respectivamente), mientras que en YOLO11l ambas condiciones quedan prácticamente empatadas (diferencia de 0,002). Esa inversión en los modelos grandes admite una explicación de datos antes que de arquitectura, puesto que el subconjunto de entrenamiento contiene 1.087 imágenes de lluvia y 1.149 de nieve, alrededor de un 7% del total cada una frente al 54% del despejado, un volumen con el que un modelo de 56,9 millones de parámetros no alcanza a explotar su capacidad en esas condiciones. El efecto se desarrolla en §5.6.

Parcialmente nublado obtiene el mAP más alto en los cinco modelos. En YOLO11l y YOLO11x su celda de densidad baja alcanza valores de 0,833 y 0,766 que elevan el promedio ponderado. Sin embargo, al descartar la celda por completo y renormalizar sobre densidad media y alta, parcialmente nublado sigue siendo el mejor clima, con 0,386 frente a 0,361 del segundo en YOLO11l y 0,385 frente a 0,376 en YOLO11x. Refuerza esa lectura que sea además la condición con mayor proporción de objetos pequeños (46,2%) y menor área mediana (1.185 px²), es decir la más exigente según el criterio de tamaño.

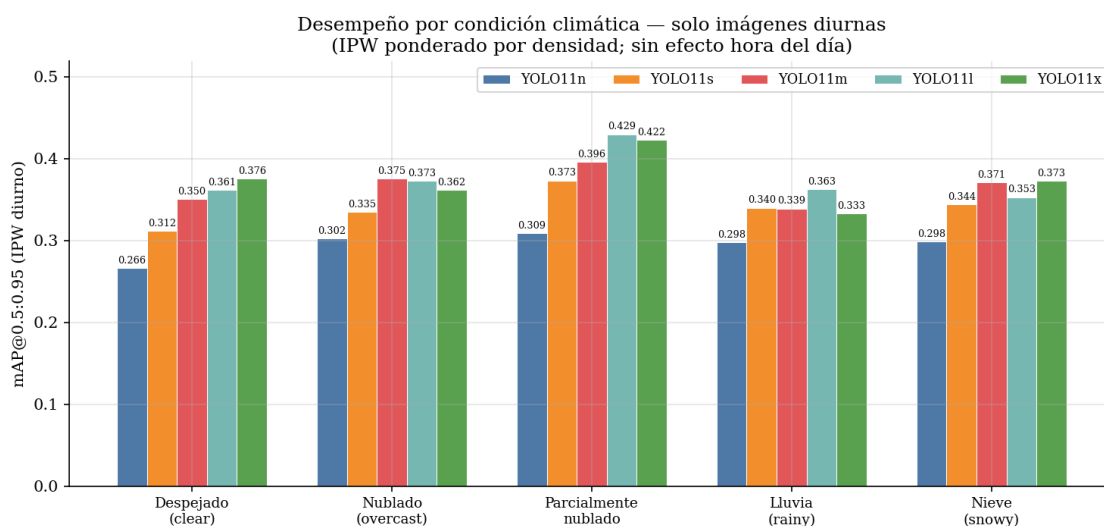


Figura 14. mAP@0.5:0.95 ponderado IPW por condición climática, restringido a imágenes diurnas (controlando solo por densidad vehicular). Despejado queda por debajo de nublado y parcialmente nublado en todos los modelos salvo YOLO11x, mientras que en los modelos grandes la lluvia y la nieve pasan a ser las condiciones más exigentes. Fuente: elaboración propia.

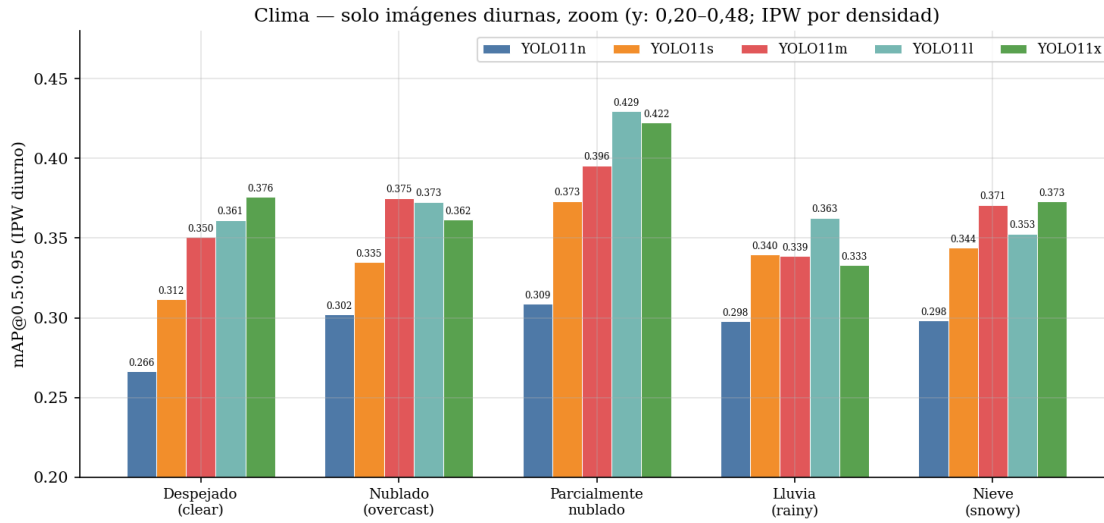


Figura 15. Ídem figura anterior, con eje y recortado (0,20 a 0,48). Fuente: elaboración propia.

5.3.2) Comparaciones controladas: momento del día y clima

Las comparaciones controladas descritas en §4.3 fijan las variables de confusión directamente, sin ponderación, lo que permite leer los resultados de cada condición sin supuestos sobre distribuciones marginales.

Momento del día (weather = clear, density = alta). La tabla siguiente muestra el mAP@0.5:0.95 para los cinco modelos segmentado por momento del día bajo condiciones fijas de clima despejado y densidad vehicular alta ($n_{\text{día}} = n_{\text{noche}} = 195$; $n_{\text{amanecer}} = 43$).

Momento del día	n	YOLO11n	YOLO11s	YOLO11m	YOLO11l	YOLO11x
Día (daytime)	195	0,252	0,305	0,347	0,355	0,369
Amanecer (dawn/dusk)	43	0,244	0,301	0,333	0,332	0,379
Noche (night)	195	0,196	0,230	0,246	0,264	0,286
Caída día → noche		−22,4%	−24,6%	−29,1%	−25,6%	−22,6%

Tabla 15: mAP@0.5:0.95 por momento del día en comparación controlada (weather = clear, density = alta, $n_{\text{día}} = n_{\text{noche}} = 195$). La caída porcentual se calcula como $(mAP_{\text{día}} - mAP_{\text{noche}})/mAP_{\text{día}}$. Fuente: elaboración propia.

La caída nocturna oscila entre el 22,4% (YOLO11n) y el 29,1% (YOLO11m), con una media de 24,9% para los cinco modelos. El amanecer muestra una degradación leve pero consistente en su dirección, puesto que el mAP cae en cuatro de los cinco modelos, entre 1,3% en YOLO11s y 6,5% en YOLO11l, y solo mejora en YOLO11x, con un 2,7%. La caída media es de 2,5%, un orden de magnitud menor que la nocturna, y se estima sobre 43 imágenes, de modo que la iluminación de transición sí degrada el desempeño respecto del día, aunque en una magnitud muy inferior a la nocturna. La degradación nocturna, en cambio, es el hallazgo más robusto del análisis, dado que se mantiene constante con distintos umbrales de densidad y clima y su magnitud supera a cualquier diferencia entre condiciones climáticas (§5.3 y figura siguiente).

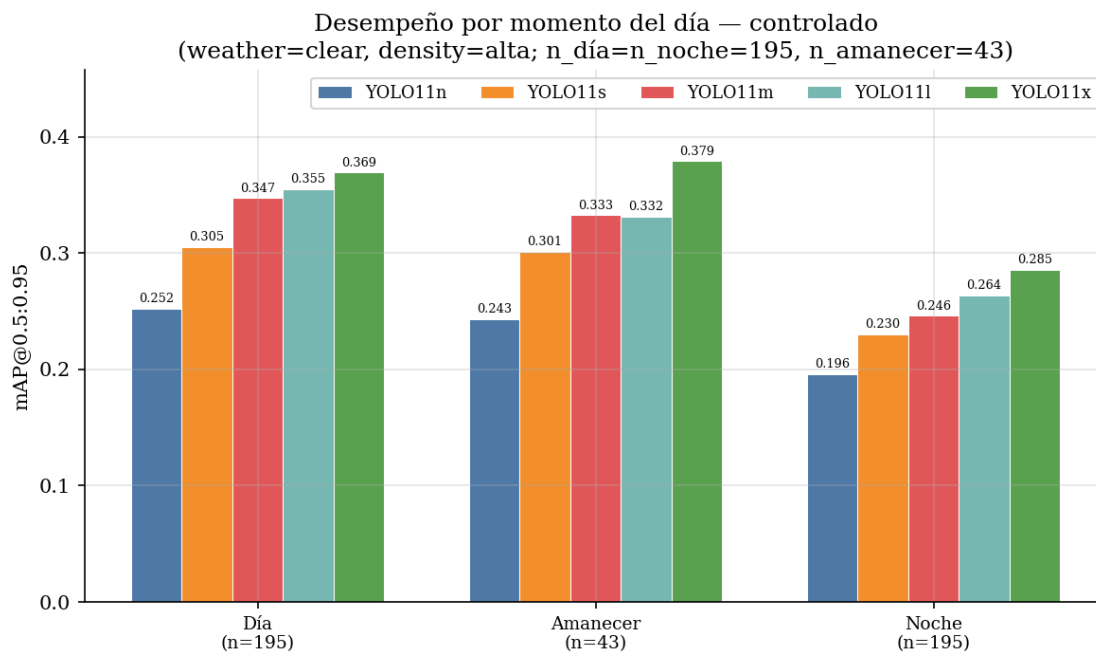


Figura 16. mAP@0.5:0.95 por momento del día en comparación controlada (weather=clear, density=alta). La noche es la condición más exigente en todos los modelos, con caídas de entre el 22% y el 29% respecto al día.

Fuente: elaboración propia.

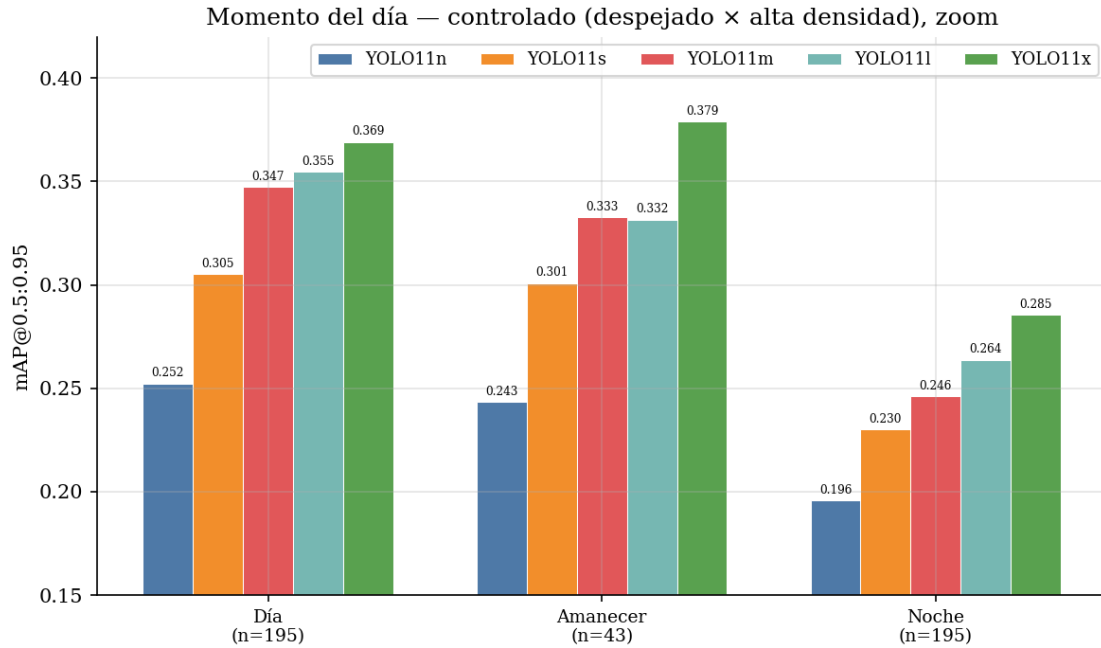


Figura 17. Ídem figura anterior, con eje y recortado (0,15 a 0,42) para apreciar diferencias entre modelos dentro de cada condición. Fuente: elaboración propia.

Clima (timeofday = daytime, density = media). La tabla siguiente muestra el mAP@0.5:0.95 para los cinco modelos segmentado por condición climática bajo condiciones fijas de horario diurno y densidad media.

Clima	n	YOLO11n	YOLO11s	YOLO11m	YOLO11l	YOLO11x
Despejado (clear)	320	0,268	0,305	0,347	0,365	0,378
Nublado (overcast)	141	0,319	0,350	0,384	0,394	0,360
Parcialmente nublado	90	0,348	0,406	0,423	0,438	0,449
Lluvia (rainy)	65	0,317	0,382	0,384	0,425	0,357
Nieve (snowy)	66	0,306	0,343	0,368	0,355	0,379

Tabla 16: mAP@0.5:0.95 por condición climática en comparación controlada (timeofday = daytime, density = media; min $n = 65$). Fuente: elaboración propia.

El patrón climático presenta una diferencia cualitativa entre modelos pequeños y grandes. Para YOLO11n, s y m, **despejado es la condición de menor mAP en todos los casos**, dado que

el tiempo claro favorece la presencia de vehículos distantes y pequeños que estos modelos no detectan bien (efecto corroborado en §5.3.1). Para YOLO11l y YOLO11x el ordenamiento cambia: nieve es la condición más exigente para YOLO11l (0,355) y lluvia para YOLO11x (0,357), mientras que despejado queda en posición intermedia. Este **cruce de los modelos más grandes** indica que, a mayor capacidad del modelo, la penalización por objetos pequeños se reduce y las condiciones de baja visibilidad (lluvia, nieve) emergen como la limitación predominante. Parcialmente nublado es el mejor clima en los cinco modelos, con mAP de 0,348 a 0,449.

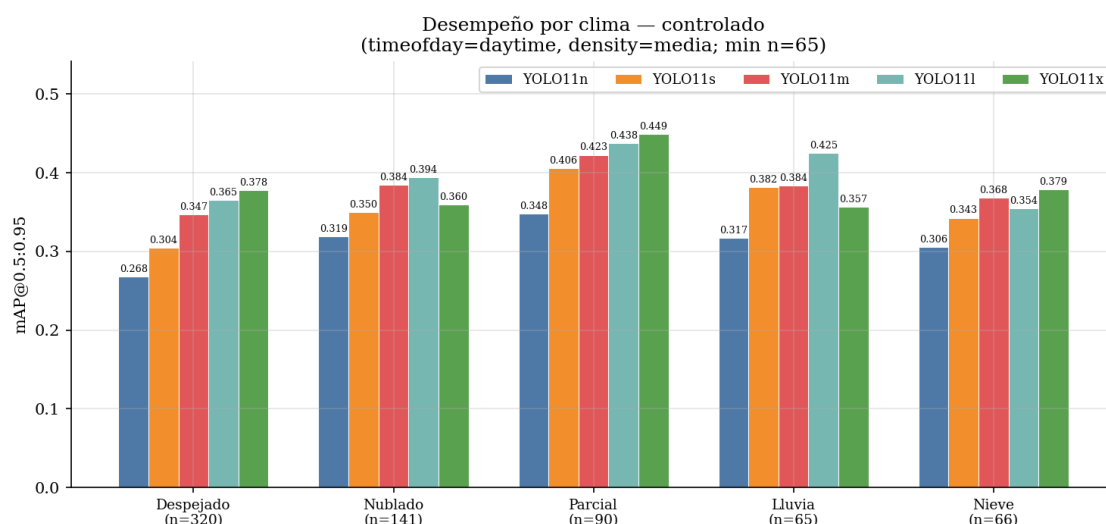


Figura 18. mAP@0.5:0.95 por condición climática en comparación controlada (timeofday=daytime, density=media). Para los modelos más pequeños (n, s, m), despejado es el peor clima, para los más grandes (l, x), lluvia o nieve emergen como las condiciones más exigentes. Fuente: elaboración propia.

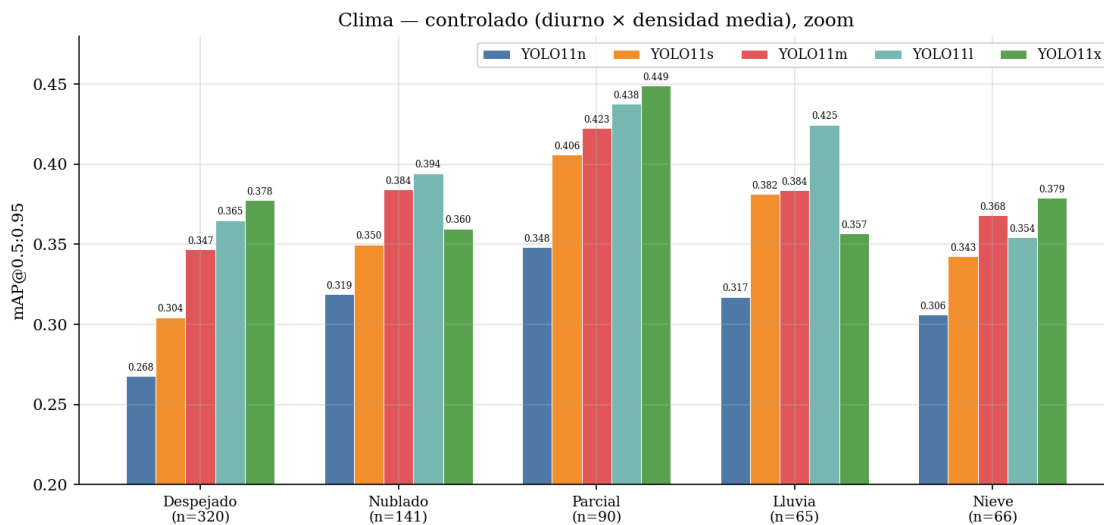


Figura 19. Ídem figura anterior, con eje y recortado (0,20 a 0,48). Fuente: elaboración propia.

Los tres hallazgos principales, la noche como condición más exigente, el clima parcialmente nublado como el más favorable y la mayor penalización relativa de los modelos pequeños en condiciones despejadas, convergen entre el análisis IPW de §5.2 y §5.3, que repondera las 2.671 imágenes válidas controlando por confundidores, y las comparaciones controladas de esta sección, que fijan clima y densidad directamente.

5.4) Desempeño por densidad vehicular

Densidad	n_img	YOLO11n	YOLO11s	YOLO11m	YOLO11l	YOLO11x
Baja (≤ 5)	458	0,245	0,312	0,374	0,389	0,364
Media (6 a 15)	1.638	0,264	0,305	0,341	0,348	0,350
Alta (> 15)	904	0,237	0,287	0,318	0,314	0,332
Caída baja→alta (sin ponderar)		−3,3%	−8,0%	−15,0%	−19,3%	−8,8%

Tabla 17: mAP50-95 por densidad vehicular (nº de objetos por imagen), sin ponderar (comparación directa). Fuente: elaboración propia.

Densidad	n_img	YOLO11n	YOLO11s	YOLO11m	YOLO11l	YOLO11x
Baja (≤ 5)	458	0,297	0,368	0,410	0,419	0,423
Media (6 a 15)	1.638	0,264	0,307	0,346	0,354	0,352
Alta (>15)	904	0,228	0,271	0,296	0,304	0,321
Caída baja→alta (IPW)		-23,3%	-26,4%	-27,8%	-27,4%	-24,1%

Tabla 18: mAP50-95 por densidad vehicular (nº de objetos por imagen), ponderado (IPW) controlando por momento del día y clima. Fuente: elaboración propia.

La densidad alta degrada a los cinco modelos respecto de la densidad baja, con una caída de entre 23% y 28% al controlar por hora del día y clima (segunda tabla). Este efecto es más pronunciado que el observado sin la ponderación (entre 3% y 19%, primera tabla): las escenas de baja densidad de BDD100K tienden a ocurrir en condiciones de hora y clima más adversas, lo que deprimía su mAP aparente y enmascaraba la magnitud real del gradiente. La caída no sigue una tendencia monotonía con el tamaño del modelo: YOLO11n y YOLO11x muestran una degradación de alrededor del 23 a 24%, mientras que YOLO11s, YOLO11m y YOLO11l exhiben caídas de entre el 26% y el 28%, lo que indica que el tamaño del modelo no determina por sí solo la robustez ante alta densidad.

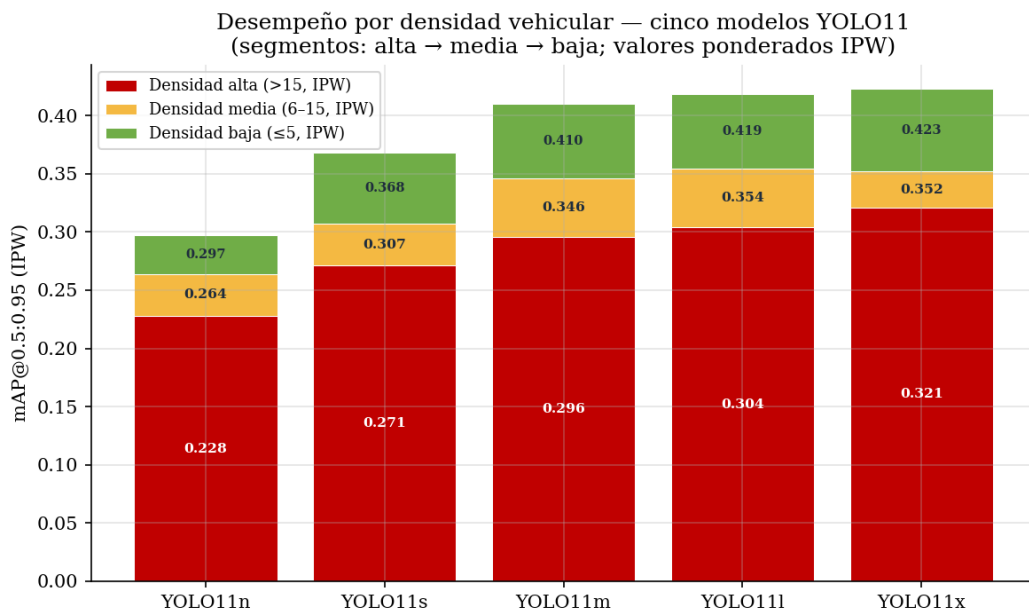


Figura 20. mAP@0.5:0.95 ponderado IPW por modelo, segmentado por densidad vehicular (controlando por

hora del día y clima). Segmentos: alta densidad (base), media (medio), baja (cima), el gradiente densidad-desempeño es más pronunciado en IPW que en la comparación directa. Fuente: elaboración propia.

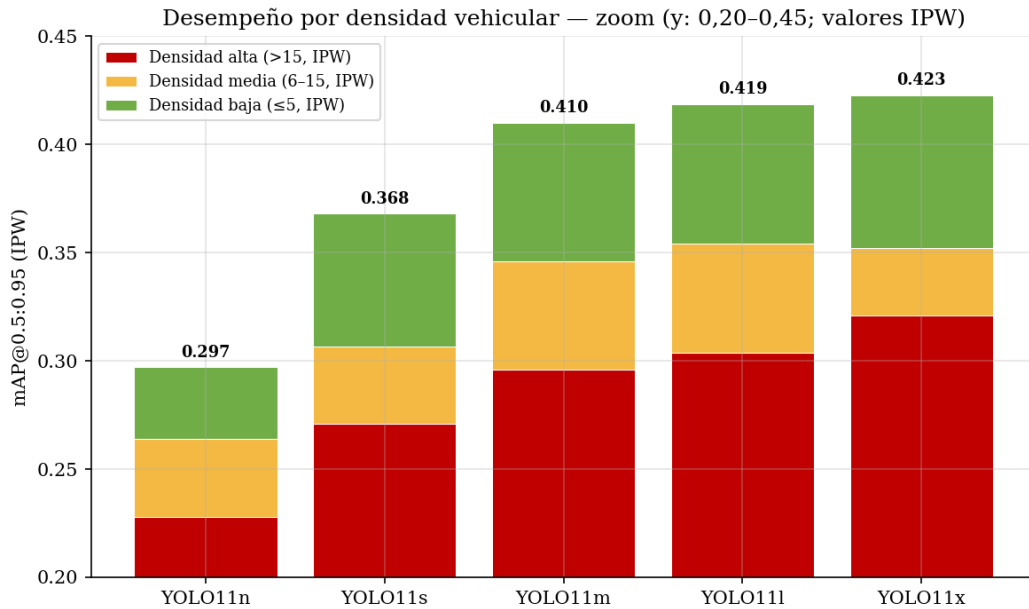


Figura 21. Ídem figura anterior, con eje y recortado (0,20 a 0,45) para visualizar mejor las diferencias entre niveles de densidad (valores IPW). Fuente: elaboración propia.

5.5) Telemetría y estrés de hardware (Objetivo 4)

La telemetría de cada modelo, medida según el protocolo definido en la metodología (§4.4: 500 imágenes de validación, batch 1, GPU RTX PRO 2000), es la siguiente.

Modelo	Latencia inferencia (ms)	FPS	VRAM pico (MiB)	Util. GPU media (%)	Potencia media (W)	Temp. media (°C)	Energía por inferencia (mJ)
YOLO11n	2,22	450,7	63,5	21,9	32,4	40,6	72
YOLO11s	2,35	425,9	93,9	34,4	38,8	43,3	91
YOLO11m	4,63	216,2	157,2	48,2	53,0	47,5	245
YOLO11l	6,26	159,8	184,7	62,7	58,4	50,9	366
YOLO11x	11,92	83,9	335,6	79,0	64,9	54,2	773

Tabla 19: Telemetría por modelo (inferencia, batch 1, 640 px, GPU RTX PRO 2000). Fuente: elaboración propia. El costo computacional crece de forma pronunciada con el tamaño del modelo. Al pasar de YOLO11n a YOLO11x la latencia de inferencia se multiplica por cerca de 5,4, la energía por in-

ferencia por cerca de 10,7, la VRAM pico por cerca de 5,3 y la utilización media de la GPU sube de 22% a 79%, lo que refleja un nivel de estrés de hardware mucho mayor. Con la definición de energía por inferencia de la metodología (§4.4) y los valores de la tabla, YOLO11n ($32,4 \text{ W} \times 2,22 \text{ ms} = 72 \text{ mJ}$) y YOLO11x ($64,9 \text{ W} \times 11,92 \text{ ms} = 773 \text{ mJ}$) difieren $2,0\times$ en potencia pero $10,7\times$ en energía por inferencia, puesto que la latencia amplifica la brecha, el wattage solo no la revelaría. En la GPU de prueba los cinco modelos corren en tiempo real, ya que incluso YOLO11x supera los 80 FPS, de modo que la diferencia de costo importa sobre todo al definir cuántas cámaras puede atender un mismo equipo y al desplegar en hardware de borde más limitado. Leída junto con el estancamiento de la precisión a partir de YOLO11m (§5.1), esta telemetría confirma que los modelos pequeños y medianos ofrecen una relación desempeño-coste muy superior, lo que constituye el insumo central de la valorización económica (§6). En el entrenamiento se observó un patrón de subutilización análogo, ya que a resolución 640 el cuello de botella es la aumentación de datos en CPU y no la GPU, razón por la que subir el batch sobre 16 o precargar imágenes en memoria no aceleró el proceso de forma apreciable (decisión de diseño en §4.2).

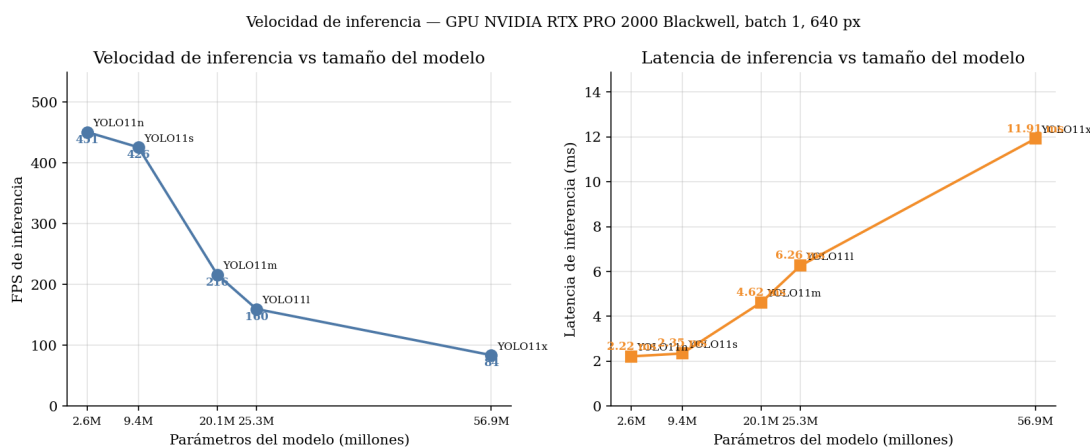


Figura 22. FPS y latencia de inferencia en función del número de parámetros (batch 1, 640 px, GPU RTX PRO 2000 Blackwell). Al escalar de YOLO11n (2,6 M) a YOLO11x (56,9 M) los FPS caen de 451 a 84 y la latencia se multiplica por $\sim 5,4$. Las etiquetas de YOLO11m y YOLO11x pueden diferir en 0,01 ms de la Tabla 15, dado que ambas se redondean de forma independiente desde el mismo dato de origen (4,625 y 11,915 ms) y el punto medio

exacto redondea de forma distinta según la implementación. Fuente: elaboración propia.

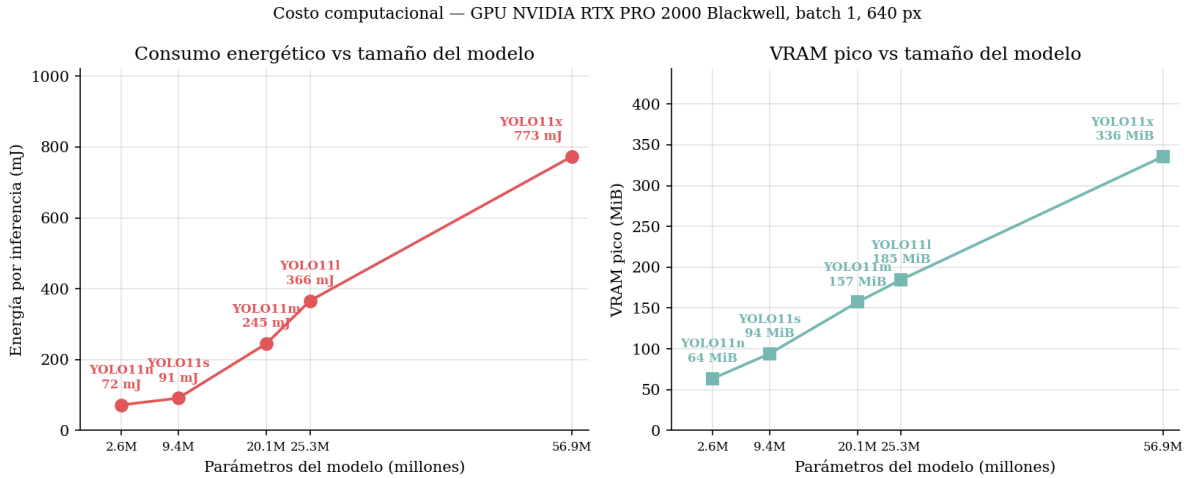


Figura 23. Energía por inferencia (J) y VRAM pico (MiB) en función del número de parámetros (batch 1, 640 px, GPU RTX PRO 2000 Blackwell). La energía se multiplica por $\sim 10,7$ y la VRAM por $\sim 5,3$ al escalar de YOLO11n a YOLO11x, determinantes del costo de despliegue en el borde. Fuente: elaboración propia.

5.6) Síntesis de hallazgos de precisión y robustez

Dos hallazgos técnicos transversales cierran el capítulo de resultados, puesto que condicionan la selección de modelo según el contexto de despliegue antes de entrar a la valorización económica.

Un hallazgo relevante de este experimento es la **convergencia entre YOLO11l y YOLO11x** en condiciones climáticas minoritarias. En el análisis diurno (§5.3.1), YOLO11l supera a YOLO11x en lluvia (0,363 vs. 0,333) y nublado (0,373 vs. 0,362). Este comportamiento no es una propiedad general de la arquitectura: en condiciones con suficientes datos de entrenamiento, un modelo mayor generaliza igual o mejor. El patrón observado es específico del tamaño del subconjunto de entrenamiento utilizado en esta memoria (15k imágenes), en el que las condiciones minoritarias cuentan con 1.087 imágenes de lluvia y 1.149 de nieve, alrededor de un 7% del conjunto cada una, lo que es insuficiente para que YOLO11x, con 56,9 millones de parámetros, explote su capacidad en esas condiciones. Con el conjunto completo de BDD100K (70k imágenes de entrenamiento, con varios miles de imágenes adversas), la ventaja

de YOLO11x sobre YOLO11l en condiciones minoritarias sería previsiblemente mayor. En el contexto de este estudio, sin embargo, YOLO11l alcanza un equilibrio donde su capacidad (25,3 millones de parámetros) es suficiente para la tarea sin sesgar el aprendizaje hacia las condiciones mayoritarias, ya que el clima despejado concentra el 53,9% de las imágenes de entrenamiento y su combinación con horario nocturno, un 33,0%. Este efecto no altera la recomendación general, puesto que la diferencia de BCR entre ambos es de 0,06 (2,69 frente a 2,75) y el costo de cómputo representa menos del 0,5% del costo total (\$6.2), pero sí es relevante si el despliegue opera en contextos con lluvia o nieve frecuentes y el modelo se reentrena sobre un corpus mayor.

La comparación controlada de §5.3.2 confirma que **la noche es la condición crítica con independencia del clima o la densidad**: con clima y densidad fijos, la caída nocturna alcanza entre el 22% y el 29% del mAP diurno, una magnitud que supera a cualquier diferencia entre condiciones climáticas. El amanecer se ubica entre ambos regímenes, con una caída media de 2,5% frente al día, muy por debajo de la nocturna. En cuanto al clima, la comparación controlada revela un **cruce entre modelos**: para YOLO11n, s y m el clima despejado es la condición más exigente (por la mayor proporción de objetos pequeños distantes), mientras que para YOLO11l y YOLO11x la lluvia o la nieve emergen como los peores escenarios, dado que la capacidad adicional del modelo para resolver objetos pequeños y distantes reduce la penalización por tamaño de objeto, pero no puede compensar plenamente la degradación de la señal visual por precipitación. Este cruce tiene implicaciones concretas para la selección de modelo según el clima de despliegue. En climas con lluvia o nieve frecuentes, el modelo de mayor mAP absoluto en precipitación es YOLO11l (0,425 en la comparación controlada con densidad media), y por esa razón es el modelo recomendado para despliegues en zonas con precipitación habitual. YOLO11x queda por debajo de YOLO11m (0,384) y YOLO11s (0,382) en lluvia con densidad media (0,357), si bien el mAP mínimo absoluto corresponde a YOLO11n (0,317). Ese comportamiento responde al mismo efecto de tamaño del conjunto de entrenamiento descrito más arriba. En climas predominantemente despejados con alta visibilidad y objetos lejanos, los

modelos pequeños (n, s, m) son los más perjudicados porque la visibilidad larga expone más objetos pequeños y distantes que los modelos de menor capacidad no alcanzan a resolver, en ese escenario, los modelos grandes (YOLO11l o YOLO11x) son los menos penalizados por esa limitación y ofrecen la mayor ganancia relativa, por lo que son los recomendados para despliegues en climas de alta visibilidad con horizonte largo.

6) Valorización económica y discusión

El valor central de la memoria es traducir el desempeño técnico a valor económico para la toma de decisiones públicas. Se realiza una **valorización social**, adoptando la perspectiva de una Municipalidad o un Ministerio, sobre un **sistema de gestión de tránsito basado en detección por video con IA** que integra el conteo vehicular, la estimación de flujo, el apoyo a los semáforos adaptativos y la detección de incidentes viales (accidentes, vehículos detenidos, peatones en calzada y congestión anómala), dimensionado sobre las ~3.250 intersecciones semaforizadas de la UOCT en la Región Metropolitana [9]. El detalle metodológico y la derivación de cada parámetro con sus fuentes se presentan en la metodología (§4.5), esta sección reporta los resultados. El valor económico se usa como criterio de decisión mediante la relación beneficio-costos (BCR) y el retorno de la inversión (ROI), con un análisis de riesgo por simulación de Montecarlo (§6.3). Los flujos se expresan en moneda constante y se descuentan a la **tasa social de descuento de 5,5%** [50], sobre un horizonte de 5 años, plazo atado a la vida útil del activo principal, ya que la guía de evaluación beneficio-costos del USDOT exige dimensionar el horizonte según la vida útil del activo [69] y el hardware de cómputo de borde tiene un ciclo de reemplazo de 5 a 7 años. De ese rango se adopta el extremo inferior como supuesto conservador.

6.1) Estructura del costo y del beneficio

El **costo** se separa en el costo de cómputo de IA, que es el único que cambia según el modelo, y un costo común del sistema (infraestructura de control adaptativo por intersección, estimada en US\$40.781 por intersección en USD de 2026 a partir del promedio de US\$28.725 que reporta el

USDOT sobre siete agencias operadoras [74], más la plataforma central de gestión y el mantenimiento). El **beneficio social** se desglosa en los cuatro mecanismos derivados en la metodología (§4.5): el ahorro de tiempo por menor congestión sobre el costo de congestión de Santiago de US\$1.353 millones al año (US\$1.046 millones de 2019 [4], actualizado a USD 2026), los accidentes evitados sobre el costo social regional de US\$998 millones al año ($B_{\text{accid,RM}}$, derivado en §4.5), el impacto ambiental valorando el CO₂ evitado al precio social del carbono de US\$66,6 por tonelada [50], y las vidas salvadas como subcomponente de los accidentes, sobre los 1.439 fallecidos de 2024 [7]. Todos los montos monetarios están expresados en **USD constantes de 2026**. El beneficio realizado por cada modelo se pondera por un **factor de calidad de detección** ponderado por condición, que penaliza la degradación nocturna [11], de modo que la robustez entra en la función de valor.

La Tabla siguiente resume los parámetros macroeconómicos con sus fuentes de origen, mientras que la derivación de cada uno, incluida la plataforma central de gestión, se presenta en la metodología (§4.5).

Parámetro	Símbolo	Valor	Fuente
Costo congestión Santiago (USD 2026/año)	B_{cong}	1.353 M	BID [4] $\times$ 1,294 IPC-US
Fracción de congestión abordable por red semaforizada	φ	36,8%	46% UMR/TTI [65] $\times$ 80% Jiang [53]
Reducción de congestión aplicable	r_c	6,25%	Tipología ponderada por retardo (§4.5); [62, 53, 61]
Costo siniestros viales RM (USD 2026/año)	$B_{\text{accid,RM}}$	998 M	CONASET [7, 66] $\times$ 1,05 IPC-US $\times s_{\text{RM}}$
Reducción de accidentes aplicable	r_a	7,8%	[68]
Emisiones CO ₂ transporte RM (ton/año)	E_{CO_2}	10.618.074	Inventario MMA-SEREMI RM [71], Tabla 8-4; mecanismo: [70]
Precio social del carbono (USD 2026/ton)	p_{CO_2}	66,6	MDS [50] $\times$ 1,05 IPC-US
Pesos de exposición (q^{tr}): día / noche	w_d, w_n	0,75 / 0,25	VMT por tramo horario, NHTSA EE.UU. [11], trasladado como supuesto declarado
Factor severidad nocturna (q^{seg})	η_n	$49/25 \div 51/75 = 2,882$	NHTSA EE.UU. [11], supuesto trasladado
Tasa social de descuento	i	5,5%	MDS [50]

Parámetro	Símbolo	Valor	Fuente
Horizonte de análisis	T	5 años	USDOT BCA [69]
Intersecciones semaforizadas RM	N	3.250	MTT/UOCT [9]
Costo infraestructura adaptativa (USD 2026/intersección)	c_{infra}	40.781	USDOT ITS KRS [74] $\times$ 1,4197 IPC-US
Plataforma central de gestión (USD 2026, lump sum)	c_{plat}	3.072.000	USDOT ITS Costs [72] $\times$ 1,5361 IPC-US

Tabla 20: Parámetros del modelo económico (todos los montos en USD constantes de 2026). La derivación de cada parámetro se presenta en la metodología (§4.5) y en el texto de esta sección. Fuente: elaboración propia en base a las fuentes indicadas.

Las expresiones formales del factor de calidad, los beneficios por mecanismo, los costos y los criterios de decisión se presentan en la metodología (§4.5). A modo de ilustración, el cálculo completo para YOLO11m en el escenario base es:

$$q_m^{\text{tr}} = \frac{0,75 \times 0,343 + 0,25 \times 0,295}{0,3433} = \frac{0,3311}{0,3433} = 0,964$$

$$q_m^{\text{seg}} = \frac{0,75 \times 0,343 + 0,25 \times 2,882 \times 0,295}{0,4881} = \frac{0,4698}{0,4881} = 0,962 \quad (\eta_n = \frac{49/25}{51/75} = 2,882)$$

$$B_{\text{tiempo},m} = \$1.353 \text{ M} \times 6,25\% \times 36,8\% \times 0,965 = \$30,0 \text{ M/año}$$

$$B_{\text{accid},m} = \$998 \text{ M} \times 7,8\% \times 0,963 = \$74,9 \text{ M/año}$$

$$B_{\text{amb},m} = 10,62 \text{ M ton} \times 6,25\% \times 36,8\% \times \$66,6/\text{ton} \times 0,965 = \$15,7 \text{ M/año}$$

$$B_{\text{total},m} = 30,0 + 74,9 + 15,7 = 120,6 \text{ M/año} \implies \text{VP}_5 = \$515,1 \text{ M}$$

$$\text{BCR}_m = \frac{515,1}{193,8} = 2,66 \quad \text{VAN}_m = 515,1 - 193,8 = 321,3 \text{ M}$$

6.2) Resultados de la valorización (escenario base)

Modelo	q^{tr}	q^{seg}	C. cómputo	C. total	Benef. VP	VAN	BCR	ROI	TRC	Vidas/año
YOLO11n	0,74	0,73	0,10	193,7	392	199	2,03	1,03	25 m	18
YOLO11s	0,87	0,86	0,12	193,7	460	267	2,38	1,38	21 m	21
YOLO11m	0,96	0,96	0,26	193,8	515	321	2,66	1,66	18 m	23
YOLO11l	0,98	0,98	0,37	194,0	522	328	2,69	1,69	18 m	24
YOLO11x	1,00	1,00	0,77	194,5	535	340	2,75	1,75	18 m	24

Tabla 21: Criterios de decisión por modelo (escenario base, horizonte 5 años). Montos en MUSD. q^{tr} = factor de calidad de tráfico (pesos de exposición 75/25, [11]); q^{seg} = factor de calidad de seguridad (pesos efectivos $\approx 51/49$, con $\eta_n = 2,882$). TRC = tiempo de recuperación del capital (sin descuento). Fuente: elaboración propia.

La tabla siguiente desglosa cómo se computa el factor de calidad q_k para cada modelo a partir de las métricas por momento del día:

Modelo	mAP día ($\widehat{\text{mAP}}^d$)	mAP amanecer ($\widehat{\text{mAP}}^a$)	mAP noche ($\widehat{\text{mAP}}^n$)	Ponderado ($q_k \cdot q_{\text{max}}$)	Factor q_k
YOLO11n	0,264	0,247	0,221	0,253	0,737
YOLO11s	0,309	0,305	0,260	0,297	0,865
YOLO11m	0,343	0,334	0,295	0,331	0,964
YOLO11l	0,346	0,330	0,301	0,335	0,976
YOLO11x	0,355	0,338	0,308	0,343	1,000

Tabla 22: Factor de calidad q_k^{tr} por modelo: mAP50-95 por momento del día y resultado de la ecuación con pesos de exposición $w_d = 0,75$ y $w_n = 0,25$ por tramo horario ([11]), la columna de amanecer se informa como resultado empírico y no entra en la ponderación, que usa solo los bloques diurno y nocturno (§4.5). YOLO11x es la referencia ($q_k = 1,000$). Fuente: elaboración propia.

El desglose de beneficio social por mecanismo en el escenario base es el siguiente:

Modelo	Ahorro tiempo (M/año)	Accidentes evitados (M/año)	CO ₂ evitado (M/año)	Total (M/año)	VP 5 años (M)	Vidas/año
YOLO11n	22,9	56,9	12,0	91,8	392	17,8
YOLO11s	26,9	66,8	14,1	107,8	460	20,9
YOLO11m	30,0	74,9	15,7	120,6	515	23,4
YOLO11l	30,4	76,1	15,9	122,4	522	23,8
YOLO11x	31,1	77,9	16,3	125,3	535	24,3

Tabla 23: Desglose de beneficio social por mecanismo (escenario base, USD millones constantes de 2026). VP 5 años descontado a $i = 5,5\%$. Las vidas salvadas son subcomponente de accidentes evitados y no se suman al total. Fuente: elaboración propia.

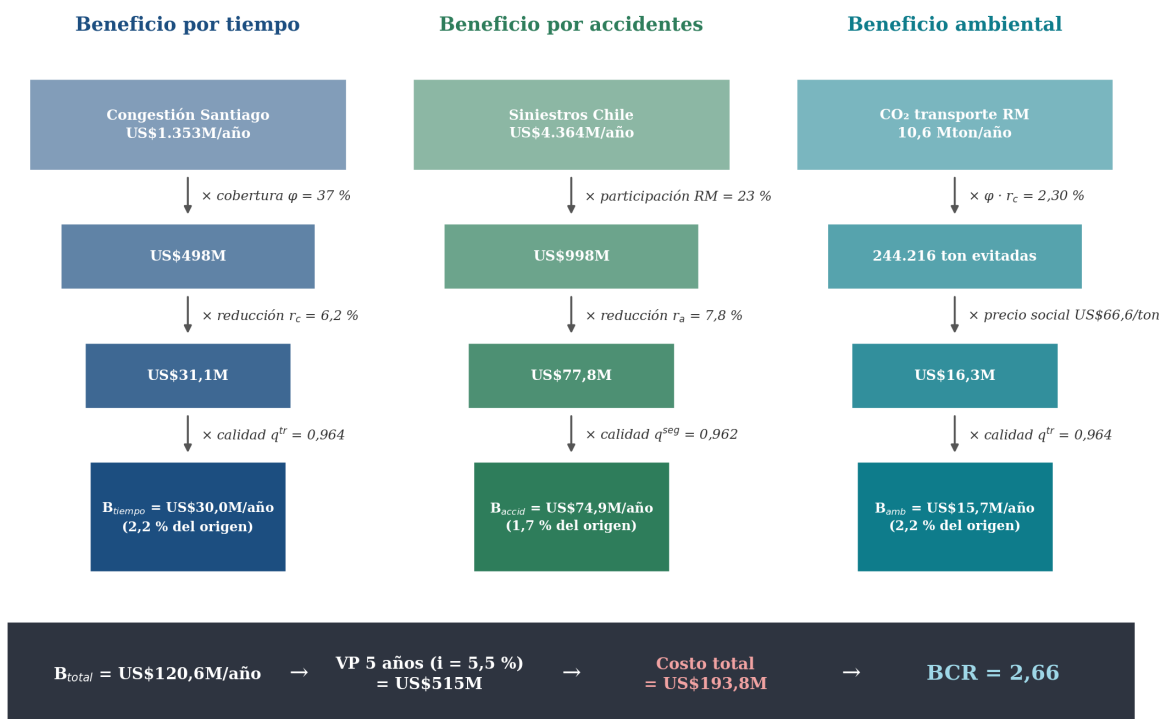


Figura 24: Cascada de valor económico y costo para YOLO11m (USD constantes de 2026). Cada panel muestra cómo el agregado de origen se reduce sucesivamente mediante los parámetros del modelo hasta llegar al beneficio o costo efectivo. Los porcentajes son relativos al valor inicial de cada panel. Fuente: elaboración propia.

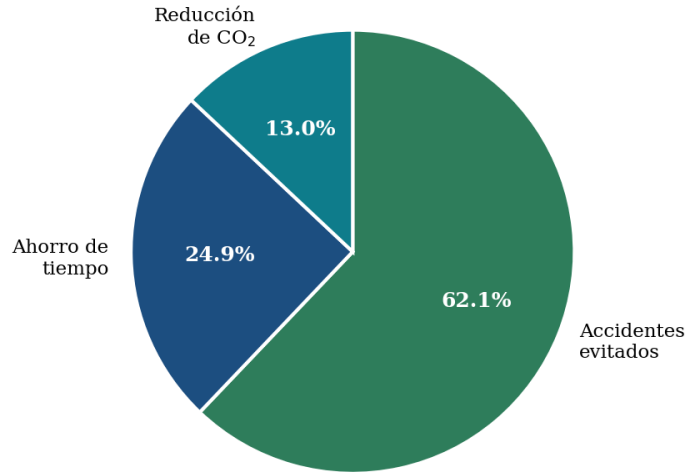


Figura 25: Desglose porcentual del beneficio social anual (YOLO11m, escenario base, USD 2026). Los accidentes evitados concentran el 62,1% del valor, el ahorro de tiempo aporta el 24,9% y la reducción de CO₂ el 13,0%. Fuente: elaboración propia.

En el escenario base (USD constantes de 2026), el beneficio social anual del sistema se compone, por ejemplo para YOLO11m, de cerca de US\$30 millones por ahorro de tiempo, US\$75 millones por accidentes evitados y US\$16 millones por reducción de emisiones, con unas 23 vidas salvadas al año en la Región Metropolitana. El costo de cómputo de IA, que es el único componente que varía entre modelos, se desglosa como sigue:

Modelo	N_{disp}	cám/disp	CAPEX cómputo (kUSD)	OPEX energía VP (kUSD)	Cómputo total (kUSD)
YOLO11n	109	30	76	26	103
YOLO11s	117	28	82	34	116
YOLO11m	233	14	163	92	255
YOLO11l	325	10	228	142	369
YOLO11x	650	5	455	316	771

Tabla 24: Desglose del costo de cómputo de IA por modelo (miles de USD). $N_{\text{disp}} = \lceil 3.250 / \lfloor \text{FPS}/15 \rfloor \rceil$; CAPEX = $N_{\text{disp}} \times \$700/\text{equipo}$; OPEX = VP(energía 5 años, $i = 5,5\%$, $p_{\text{kWh}} = \$0,20$). El costo común (infraestructura, plataforma y mantenimiento) va de US\$193,55 M a US\$193,71 M según el modelo, dado que el mantenimiento se calcula sobre el CAPEX total. Fuente: elaboración propia.

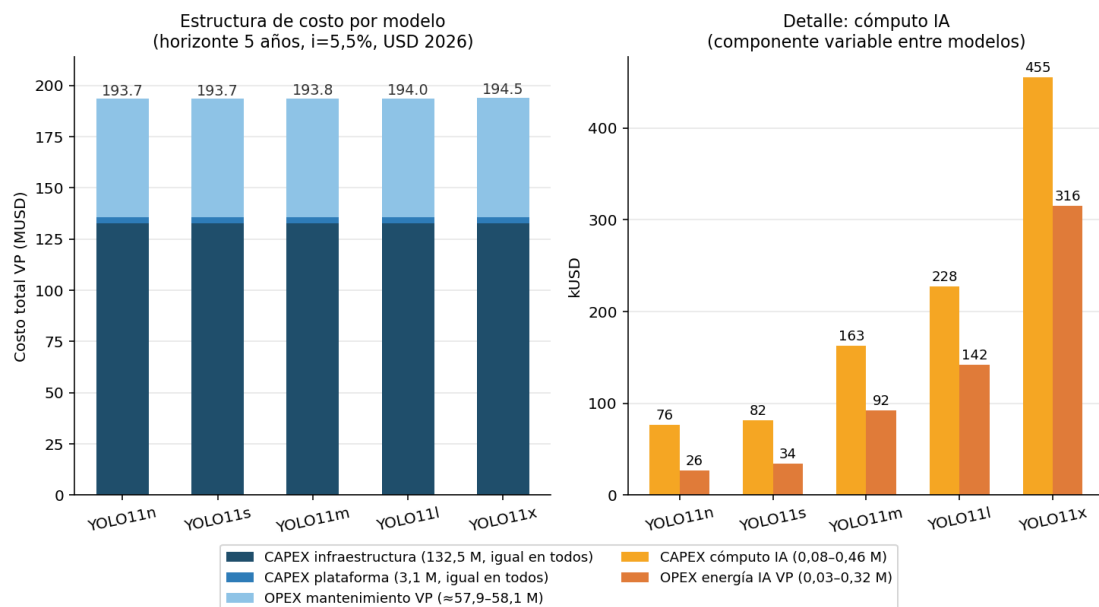
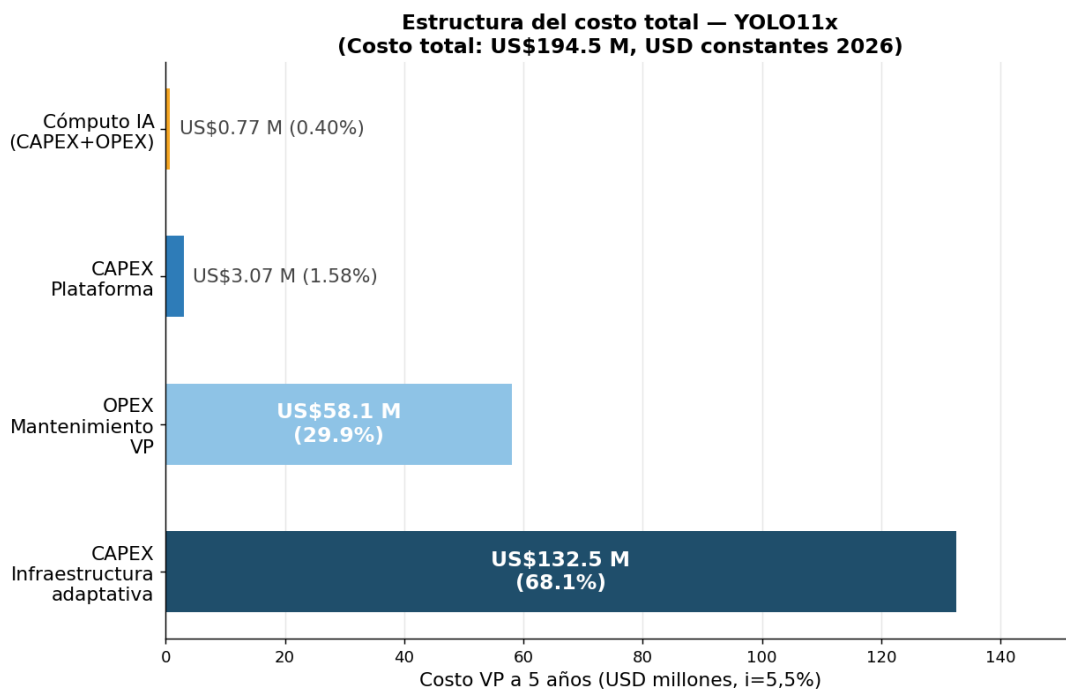


Figura 26: Estructura de costo (CAPEX + OPEX descontado a 5 años) por modelo, US\$ constantes de 2026. Los segmentos comunes (infraestructura de control adaptativo y mantenimiento) son idénticos entre modelos y concentran prácticamente la totalidad del costo, los segmentos que dependen del modelo (equipos de cómputo y energía) resultan visualmente imperceptibles frente a ellos, lo que confirma que la elección del tamaño de modelo no altera el costo total del sistema. Fuente: elaboración propia.



El cómputo de IA (CAPEX = 650 disp. × USD 700 + OPEX energía VP) representa solo el 0,40% del costo total, lo que hace que la elección del modelo no sea determinante para la rentabilidad del sistema.

Figura 27: Estructura del costo total del sistema para YOLO11x (US\$194,5 M, USD constantes de 2026). La infraestructura de control adaptativo (68,2%) y el mantenimiento (29,9%) concentran el 98,1% del costo, el cómputo de IA representa una fracción marginal (0,40%) incluso en el modelo más exigente computacionalmente, lo que confirma que la elección del modelo no altera la rentabilidad del sistema. Fuente: elaboración propia.

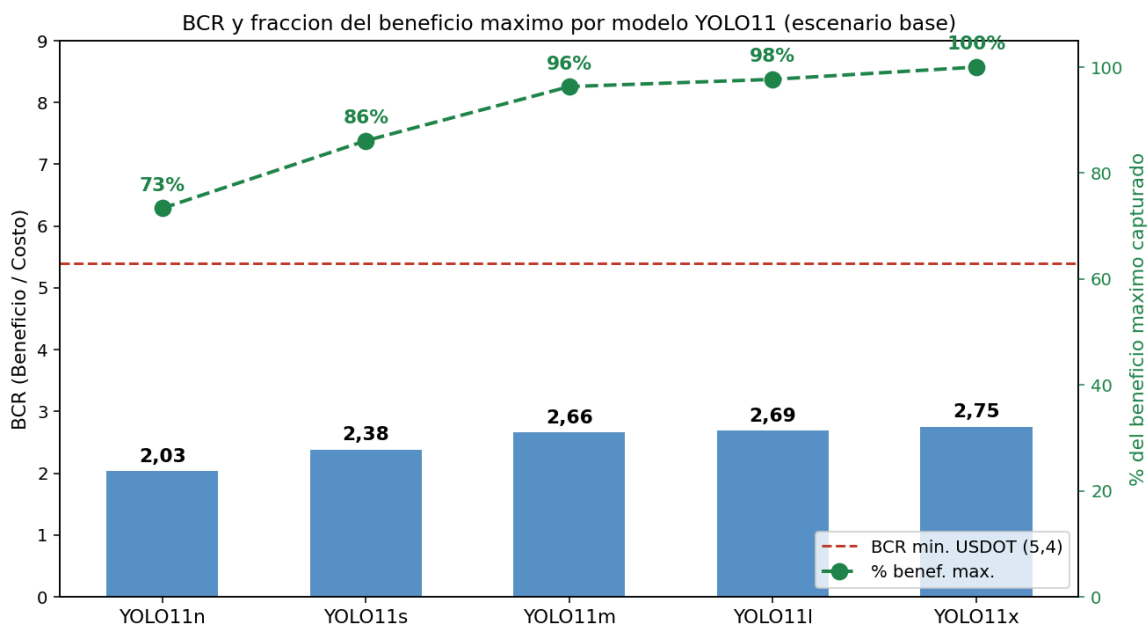


Figura 28: BCR acumulado y fracción del beneficio máximo capturado por modelo (escenario base, $r_c = 6,25\%$ ponderado por tipología, $r_a = 7,8\%$). El BCR crece rápido de n a m y se estabiliza en un plateau de $\approx +0,05$ por paso a partir de YOLO11m, mientras que ese modelo ya captura el 96% del beneficio teórico máximo (BCR de referencia USDOT: 5,4). Fuente: elaboración propia.

6.3) Análisis de riesgo mediante simulación de Montecarlo

La simulación se ejecutó según el diseño de la metodología (§4.6), con 100.000 iteraciones de los 18 parámetros inciertos y muestra común para los cinco modelos. Los resultados confirman y refuerzan la conclusión del escenario base. El VAN es positivo en el 100% de las 100.000 iteraciones para YOLO11m, YOLO11l y YOLO11x, mientras que en los dos modelos de menor precisión aparecen combinaciones adversas aisladas, 48 iteraciones con VAN negativo en YOLO11n (99,95% positivo) y 2 en YOLO11s (99,998%), de modo que la rentabilidad social es un resultado robusto para el modelo recomendado y prácticamente robusto para el resto. Para YOLO11m el VAN tiene mediana de US\$367 millones y percentil 5 de US\$221 millones (USD constantes de 2026), lo que constituye un valor en riesgo al 5%, es decir, con 95% de confianza el VAN supera US\$221 millones. La probabilidad de que YOLO11m supere la BCR de referencia de USDOT (5,4, el extremo inferior del rango de despliegues evaluados, corre-

spondiente a los beneficios sin contar seguridad) [52] es de 0,03%, y las vidas salvadas por año se sitúan entre 20 y 38 con 90% de probabilidad (mediana 26). La media simulada del VAN de YOLO11m (US\$375 millones) queda por sobre el caso base determinista (US\$321 millones), que corresponde al percentil 32 de la distribución simulada, esto ocurre porque los rangos publicados de los parámetros de efectividad son asimétricos hacia arriba (la ganancia ASCT llega a 25% en los despliegues documentados por el USDOT [64] y la reducción de accidentes a 17% [60]) mientras la moda se ancló en el valor base, de manera que el escenario base es conservador por construcción.

Modelo	VAN P5 (VaR 5%)	VAN P50	VAN P95	BCR P5	BCR P50	BCR P95	P(BCR≥5,4)	Vidas/año P50
YOLO11n	107	228	375	1,40	2,07	2,98	0,0%	20
YOLO11s	171	306	474	1,65	2,43	3,49	0,0%	23
YOLO11m	221	367	555	1,84	2,72	3,91	0,03%	26
YOLO11l	228	375	565	1,87	2,75	3,96	0,04%	27
YOLO11x	238	389	583	1,91	2,81	4,04	0,06%	27

Tabla 25: Estadísticos de la simulación de Montecarlo por modelo (100.000 iteraciones, muestra común de parámetros). VAN en MUSD constantes de 2026, p5, P50 y P95 son percentiles de la distribución simulada y el P5 del VAN corresponde al valor en riesgo al 5%. $P(VAN > 0) = 100\%$ para YOLO11m, YOLO11l y YOLO11x, 99,95% para YOLO11n y 99,998% para YOLO11s. Fuente: elaboración propia.

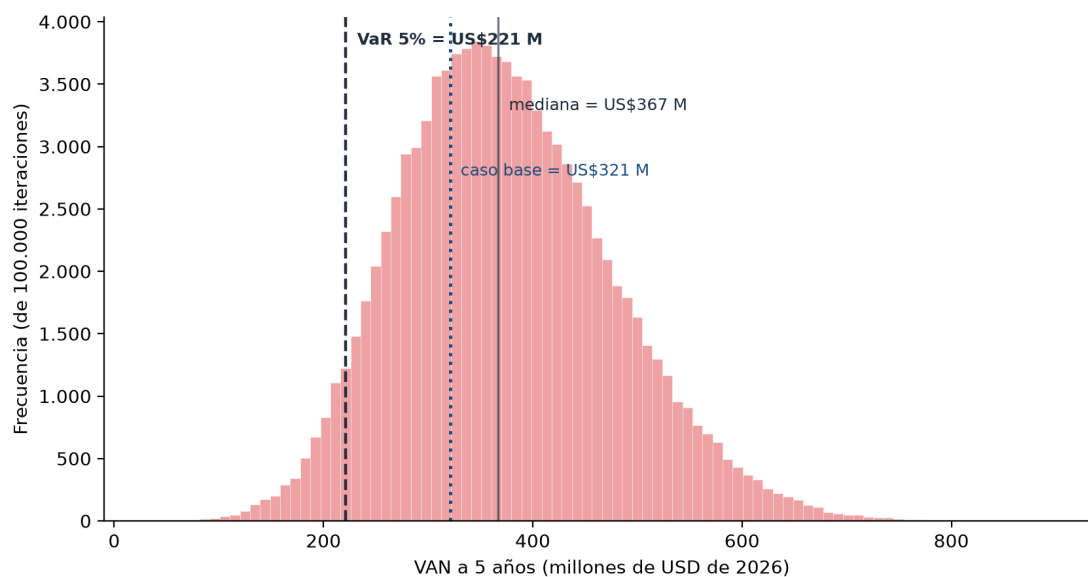


Figura 29: Distribución simulada del VAN de YOLO11m (100.000 iteraciones, MUSD constantes de 2026), con el valor en riesgo al 5% (US\$221 M) y el caso base determinista (US\$321 M, percentil 32) marcados. El caso base resulta conservador porque los rangos publicados de los parámetros de efectividad son asimétricos hacia arriba y la moda se ancló en el valor base. Fuente: elaboración propia.

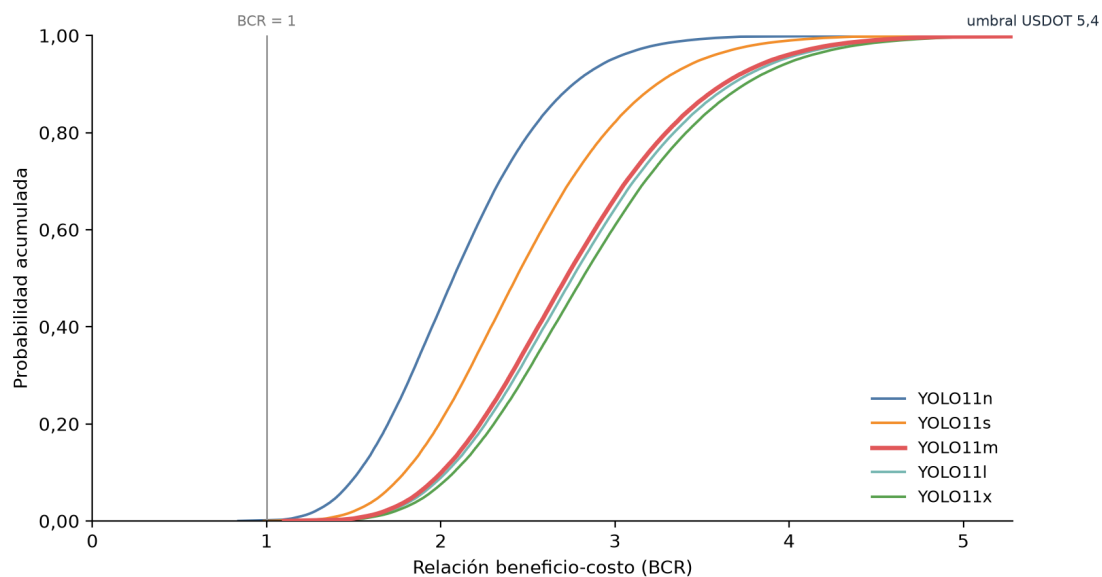


Figura 30: Función de distribución acumulada del BCR de los cinco modelos bajo la muestra común de parámetros, con los umbrales de rentabilidad ($BCR = 1$) y de referencia USDOT ($BCR = 5,4$) [52]. La masa de probabilidad queda a la derecha de la unidad en los cinco modelos salvo por una fracción marginal de YOLO11n (0,05% de las iteraciones) y de YOLO11s (0,002%), mientras que la probabilidad de superar 5,4 no llega a 0,1% en ningún modelo, de 0,0% en YOLO11n a 0,03% en YOLO11m y 0,06% en YOLO11x. Fuente: elaboración propia.

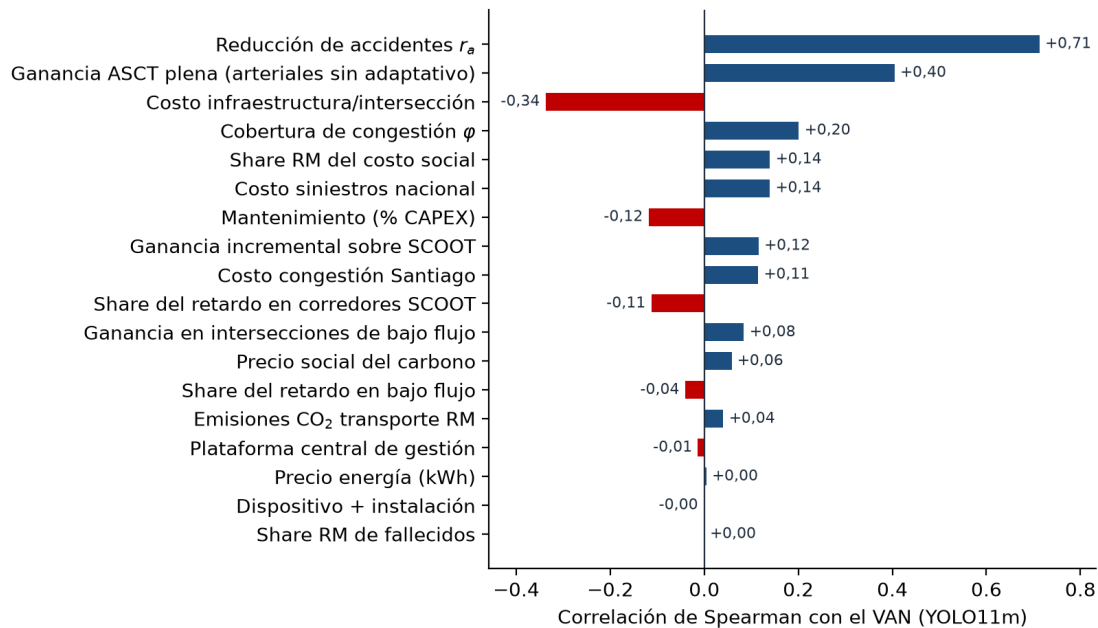


Figura 31: Diagrama de tornado para YOLO11m, con la correlación de rangos de Spearman de cada parámetro muestreado contra el VAN (100.000 iteraciones). Los parámetros de efectividad (reducción de accidentes y ganancia ASCT) dominan el riesgo del proyecto, entre los de costo, solo el costo de infraestructura por intersección alcanza magnitud relevante ($\rho = -0,338$), mientras el costo de cómputo, la plataforma y la energía permanecen marginales. Fuente: elaboración propia.

El diagrama de tornado ordena los parámetros por su correlación de rangos de Spearman con el VAN de YOLO11m. Tres de ellos concentran el riesgo y conviene detenerse en por qué.

La **reducción de accidentes** ($\rho = +0,713$) domina por un margen amplio, ya que supera casi cuatro veces al parámetro que le sigue. Su peso combina dos cosas. Por una parte, los siniestros evitados aportan el 62,1% del beneficio social, de modo que cualquier variación se transmite casi completa al resultado. Por otra, su rango publicado es de los más anchos, ya que la literatura va de 6,2% a 17% según el diseño del estudio [68, 60], un techo que multiplica por 2,2 el valor base. No es el rango más ancho del modelo, puesto que la ganancia del control adaptativo llega a 2,5 veces su valor base, pero sí es el que actúa sobre la porción mayor del beneficio y sin diluirse en otros parámetros. La consecuencia práctica es que la incertidumbre del proyecto está en cuánto reduce realmente la siniestralidad el control adaptativo, y no en ningún parámetro de costo.

La **ganancia plena del control adaptativo** (+0,405) ocupa el segundo lugar pese a tener el rango más ancho de todo el modelo, con un máximo de 25% medido en el piloto de Surtrac [5] frente a una moda de 10%. Queda detrás porque su efecto llega diluido, dado que entra en la reducción de congestión r_c repartándose con las otras dos ganancias de la tipología, y r_c alimenta solo el ahorro de tiempo y las emisiones evitadas, que juntos son el 37,9% del beneficio. El contraste con la reducción de accidentes ilustra que en un tornado pesa tanto la amplitud del rango como la fracción del resultado sobre la que actúa.

El **costo de infraestructura por intersección** (−0,338) es el tercero, y es el único parámetro de costo que importa. Su signo es negativo por tratarse de un costo. Pesa tanto porque representa el 68,2% del costo total del sistema y porque su techo, anclado en el despliegue más caro documentado por el USDOT (§4.5), alcanza el doble del valor base. Esto precisa dónde está el riesgo económico, es decir en cuánto cuesta llevar el control adaptativo a cada intersección.

El resto se ordena en tres grupos con lecturas distintas. Los **agregados de valoración**, es decir la cobertura de la red semaforizada, la participación regional en el costo de siniestros y el costo nacional de los siniestros, se sitúan entre 0,139 y 0,201: sus bandas son estrechas por provenir de agregados oficiales. Los **parámetros que reparten el efecto entre tipos de intersección**, junto con el mantenimiento, quedan por debajo de 0,120, puesto que cada uno gobierna solo una porción de la red y ninguno alcanza a mover el total por sí solo. Los **parámetros de cómputo** son los últimos, con $|\rho| \leq 0,015$, lo que era esperable, dado que la plataforma, la energía y los dispositivos suman menos del 2% del costo total (§6.2), de manera que ni duplicándolos se movería el VAN de forma perceptible.

Un caso merece explicación aparte. La **participación regional en los fallecidos** aparece con correlación nula, y no es un error de la simulación, sino que refleja el diseño del modelo, ya que las vidas salvadas se reportan como indicador físico y no se suman al VAN para no contabilizar dos veces el mismo beneficio (§4.5). Ese parámetro mueve las vidas reportadas, no el resultado económico.

Las tres verificaciones de robustez diseñadas en la metodología (§4.6) entregan los siguientes resultados. Primero, el supuesto de amplitud del precio social del carbono no compromete las conclusiones, dado que el beneficio ambiental representa el 13% del beneficio total y la correlación del precio del carbono con el VAN en el tornado es marginal ($\rho = +0,06$), y un precio distinto movería la magnitud del VAN sin cambiar su signo ni el orden de los modelos, y ese movimiento sería además al alza, puesto que los referentes internacionales superan al valor chileno, como la estimación de la EPA de 2023 de aproximadamente US\$190 por tonelada [80], casi tres veces los US\$66,6 que este modelo adopta como valor base, y se mantiene el anclaje en el precio social chileno por consistencia con el Sistema Nacional de Inversiones [50]. Segundo, al imponer la cópula gaussiana con $\rho = 0,5$ entre g_{ASCT} y r_a , el valor en riesgo al 5% del VAN de YOLO11m pasa de US\$221 a US\$206 millones y la probabilidad de superar la BCR de 5,4 sube levemente de 0,03% a 0,10%, con lo que las conclusiones son robustas al supuesto de independencia. Tercero, en el vértice desfavorable, con los 18 parámetros simultáneamente en su extremo adverso, el VAN de YOLO11m es de US\$–201 millones con BCR de 0,55 (US\$–259 millones y BCR 0,42 para YOLO11n), de modo que el soporte del modelo sí admite valores negativos y la rentabilidad observada no es una propiedad estructural sino una consecuencia de lo improbable que resulta esa combinación. El vértice queda muy por debajo del mínimo simulado (US\$34 millones para YOLO11m en 100.000 iteraciones), lo que es esperable por construcción, puesto que la probabilidad de que los 18 parámetros caigan a la vez cerca de su extremo es ínfima, ya que caer cada uno dentro de su cuarto más desfavorable ocurre con probabilidad $0,25^{18} \approx 1,5 \times 10^{-11}$, una vez cada 69 mil millones de iteraciones. En el vértice favorable el VAN alcanza US\$1.570 millones con BCR de 12,43 para YOLO11m, lo que dimensiona el rango completo posible del modelo.

7) Discusión y recomendación

El sistema es **socialmente rentable en todos los escenarios deterministas analizados y para todos los modelos**. El BCR del escenario base se sitúa entre 2,03 (YOLO11n) y 2,75

(YOLO11x) (USD constantes de 2026), en el escenario determinista ningún modelo alcanza la BCR de referencia de USDOT (5,4) [52], aunque todos superan varias veces el umbral de rentabilidad ($BCR = 1$), lo que confirma la solidez del sistema incluso con la variante de menor precisión nocturna. La simulación de Montecarlo (§6.3) matiza esta lectura determinista, dado que al incorporar la asimetría hacia arriba de los parámetros de efectividad la probabilidad de que YOLO11m supere esa referencia es de 0,03% (0,06% para YOLO11x), de modo que, una vez corregida la participación regional del costo de siniestros, alcanzar el desempeño del referente de Florida deja de ser un resultado plausible para este despliegue. El VAN de YOLO11m permanece positivo en el 100% de las 100.000 combinaciones muestreadas de los 18 parámetros inciertos, con un valor en riesgo al 5% de US\$221 millones. En otras palabras, por cada dólar invertido el retorno social estimado es de al menos 1,10 dólares incluso en la peor de esas combinaciones, y con 95% de confianza supera los 1,84 dólares por dólar invertido. En YOLO11n, la variante de menor precisión, ese mínimo baja a 0,84 dólares por dólar, si bien el VAN resulta negativo en solo 48 de las 100.000 iteraciones.

El hallazgo central para la decisión es que **el costo de cómputo de IA es una fracción mínima del costo total** (entre US\$0,1 y US\$0,8 millones frente a unos US\$194 millones del sistema completo, USD 2026), de modo que el costo lo domina la infraestructura común y no la elección del modelo. Por lo tanto conviene el modelo que maximiza el beneficio por su precisión. La tabla siguiente muestra el BCR acumulado por modelo y el incremento marginal de cada escalado, incluyendo el incremento en costo total (que ya incorpora CAPEX, energía y mantenimiento de cómputo):

Modelo	BCR (base)	VAN (MUSD)	% benef. máx.	ΔBCR	ΔBenef. VP (MUSD)	ΔCosto total (MUSD)
YOLO11n	2,03	199	73%	—	—	—
YOLO11s	2,38	267	86%	+0,35	+68	+0,02
YOLO11m	2,66	321	96%	+0,28	+55	+0,17
YOLO11l	2,69	328	98%	+0,04	+7	+0,15
YOLO11x	2,75	340	100%	+0,06	+13	+0,49

Tabla 26: Criterios acumulados y ganancia marginal al escalar un nivel de modelo (escenario base). ΔCosto total incluye CAPEX, OPEX de energía y el incremento de mantenimiento, descontados a $i = 5,5\%$, el BCR ya refleja todos estos costos. Fuente: elaboración propia.

Nótese que el BCR ya incorpora el costo completo de cómputo y energía de cada modelo: la columna ΔCosto total muestra el incremento real al escalar, incluyendo CAPEX de equipos, OPEX de energía y su efecto sobre el mantenimiento. Dicho esto, el BCR sigue mejorando al escalar (de 2,66 en YOLO11m a 2,75 en YOLO11x) porque el beneficio adicional supera al costo adicional. La razón para proponer YOLO11m como óptimo no es que los modelos mayores sean más caros en términos absolutos, sino que la ganancia marginal de BCR cae drásticamente a partir de m (+0,04 y +0,06 por paso frente a +0,35 y +0,28 en las transiciones $n \rightarrow s$ y $s \rightarrow m$) y es inferior a la incertidumbre del modelo de beneficio, puesto que la simulación de Montecarlo (§6.3) muestra que el BCR de YOLO11m se mueve entre 1,84 y 3,91 con 90% de probabilidad (percentiles 5 y 95), una banda de 2,07 puntos dentro de la cual la diferencia mediana de 0,10 entre YOLO11m y YOLO11x no es discernible. En la práctica, YOLO11m captura el 96% del beneficio teórico máximo y el 4% restante queda absorbido por la incertidumbre de los parámetros de beneficio. Este hallazgo es consistente con los resultados de la evaluación comparativa exhaustiva de la familia YOLO [14], donde YOLO11m alcanza un mAP50-95 de 0,795 y YOLO11l uno de 0,794 sobre el conjunto de señales de tránsito, es decir una diferencia de una milésima entre dos modelos de tamaño consecutivo, y donde los autores identifican a YOLO11m como el modelo de mejor desempeño global al considerar a la vez precisión, tamaño y eficiencia. La lectura de la serie completa de tamaños en el gráfico de resultados de ese trabajo, reproducido a continuación, muestra además que las ganancias de mAP50-95 entre modelos adyacentes decrecen al crecer el tamaño, con los mayores saltos en las transiciones $n \rightarrow s$ y $s \rightarrow m$ y

una estabilización pronunciada a partir de m, mientras que las diferencias de latencia y consumo entre m, l y x sí crecen de forma sustancial. Esa lectura de la serie completa es propia de esta memoria, puesto que los autores reportan los valores por modelo y destacan el estancamiento entre m y l, pero no enuncian el patrón monótono como un hallazgo en esos términos.

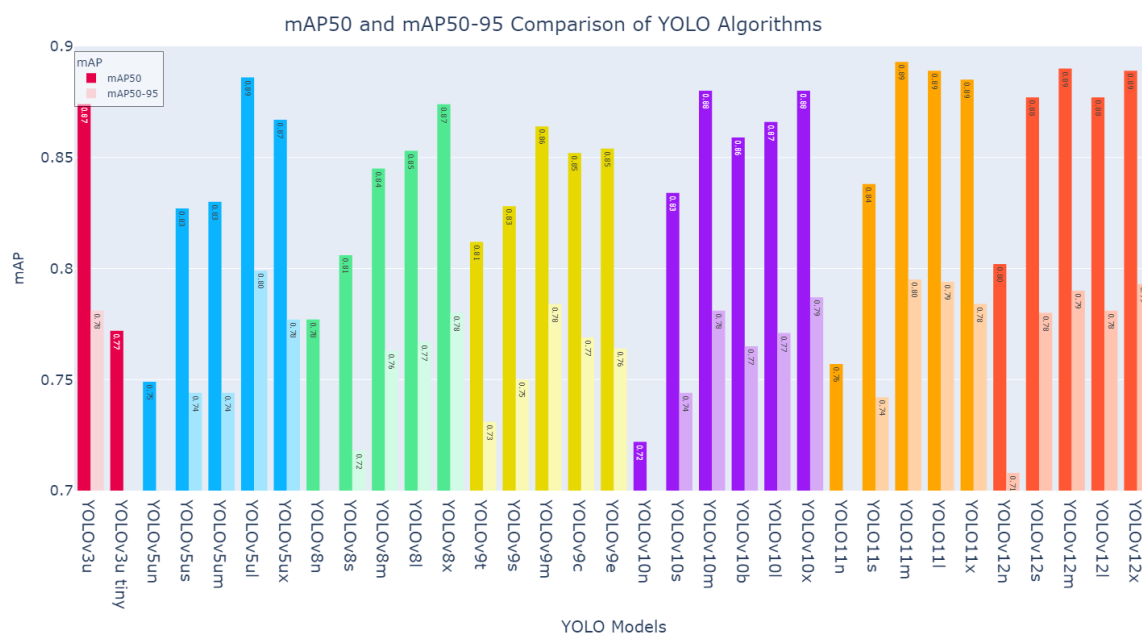


Figura 32: Comparación de mAP50 y mAP50-95 de las familias YOLO sobre el conjunto de señales de tránsito. Cada modelo se representa con dos barras, la de la izquierda para mAP50 y la de la derecha para mAP50-95. En la serie de YOLO11 el mAP50-95 sube de 0,74 en YOLO11s a 0,80 en YOLO11m y desde ahí no vuelve a crecer, con 0,79 en YOLO11l y 0,78 en YOLO11x. Reproducida de la Figura 13 de Jegham et al. (2024) [14].

En aplicaciones de gestión de tráfico urbano en tiempo real, YOLO11m combina esa estabilización del mAP con inferencia suficientemente rápida para operar en servidores locales o en nube, capacidad que sistemas de gestión de tránsito basados en YOLO11 ya han demostrado en operación, alcanzando un mAP de 92,4% y un IoU de 0,85 al alimentar el control adaptativo de semáforos [35].

Valor adicional no contabilizado: detección automática de incidentes (AID). El modelo base no incluye los beneficios de la detección automática de incidentes, aunque el mismo hard-

ware puede habilitarla sin costo incremental significativo. Sistemas basados en aprendizaje automático alcanzan tiempos medios de detección de incidentes de entre 0,9 y 1,6 minutos (ElSahly y Abdelfatah [81]), frente a los múltiples minutos que implican los reportes de terceros, ya que, según la experiencia documentada por la FHWA en Estados Unidos, cada minuto adicional en que el incidente primario sigue constituyendo un peligro incrementa en 2,8% la probabilidad de un siniestro secundario [82]. Más allá de su valor propio, esta detección puede servir de base a otros proyectos que reutilicen la misma red de cámaras y el mismo cómputo. En ese escenario el costo de inferencia se prorroga entre iniciativas y el beneficio agregado crece, de modo que un modelo como YOLO11l podría justificarse, puesto que su mayor precisión pasaría a pesar sobre una base de beneficio más amplia mientras su costo marginal se reparte.

8) Limitaciones y perspectivas futuras

Los resultados de esta memoria descansan sobre un conjunto acotado de datos y sobre parámetros que en varios casos debieron trasladarse desde contextos distintos al chileno. Esta sección reúne esas restricciones, que a lo largo del documento se declararon donde correspondía, y deriva de ellas las líneas de trabajo que quedan abiertas.

8.1) Limitaciones de los datos

La restricción de fondo es que **BDD100K registra conducción en Estados Unidos y no en Santiago** [15]. El parque vehicular, la señalización, la geometría vial y las pautas de conducción difieren de las chilenas, de modo que las cifras de mAP reportadas caracterizan el desempeño sobre ese dominio y no sobre las cámaras de la UOCT. La comparación entre los cinco modelos sí es trasladable, puesto que todos enfrentan el mismo dominio, pero el nivel absoluto de precisión que se obtendría en Santiago queda sin medir.

El **subconjunto de 15.000 imágenes de entrenamiento** impone una segunda restricción. Las condiciones minoritarias quedan con 1.087 imágenes de lluvia y 1.149 de nieve, alrededor de un 7% del conjunto cada una, frente al 53,9% del clima despejado. Ese desbalance es la causa del

cruce documentado en §5.6, donde YOLO11l supera a YOLO11x en lluvia y nublado, ya que un modelo de 56,9 millones de parámetros no alcanza a explotar su capacidad con ese volumen de ejemplos adversos. La nieve agrega una limitación de pertinencia, dado que es prácticamente inexistente en Santiago, de manera que aporta a la comparación técnica entre modelos pero no al caso de despliegue local.

Una tercera restricción es que **algunas combinaciones de condiciones no existen en el conjunto de validación**. El clima nublado cuenta con tres imágenes nocturnas y el parcialmente nublado con una, de modo que esas celdas quedan bajo el mínimo de 15 imágenes y se descartan. La consecuencia está cuantificada en §5.3, donde el despejado se evalúa con un 59,5% de peso nocturno efectivo mientras el nublado lo hace con un 0%. Ninguna técnica de ponderación puede corregir eso, puesto que no se trata de un sesgo sino de ausencia de observaciones. La niebla, con cuatro imágenes en validación, se excluye por completo del análisis. El propio umbral de 15 imágenes es un criterio declarado de este trabajo y admite revisión.

8.2) Limitaciones del traslado de parámetros

Varios parámetros del modelo económico provienen de mediciones extranjeras, ya que no se encontró una equivalente para Chile. Los **pesos de exposición por tramo horario** (75% diurno y 25% nocturno) y el **factor de severidad nocturna** ($\eta_n = 2,882$) son mediciones de NHTSA sobre kilómetros recorridos y víctimas fatales en Estados Unidos [11], trasladadas como supuesto declarado. El **costo de infraestructura por intersección** se construye sobre encuestas a agencias operadoras estadounidenses [74], y se asume además que corresponde a la adaptación de semáforos existentes y no a instalaciones nuevas, puesto que la ficha de origen no lo precisa. La **fracción de congestión abordable** φ combina una medición del reparto del retardo urbano en áreas metropolitanas de Estados Unidos [65] con una regularidad de concentración del flujo vial [53].

A lo anterior se suman los supuestos propios de este trabajo. La **participación de cada tipo de intersección en el retardo abordable** (40%, 45% y 15%) corresponde a estimaciones propias

de este trabajo, dado que no se encontró un estudio chileno de la distribución del retardo por jerarquía vial, y por eso se sensibiliza en el análisis de riesgo. El **conteo de intersecciones con control dinámico SCOOT** corresponde al último dato público de la UOCT, de 2016 [43], de modo que la cobertura actual podría ser mayor. Y las **amplitudes de las distribuciones triangulares**, de $\pm 10\%$ a $\pm 20\%$ según el parámetro, son supuestos de sensibilización declarados, ya que las fuentes publican valores puntuales sin rango asociado.

8.3) Limitaciones del modelo económico

El modelo valoriza **tres mecanismos de beneficio**, que son el ahorro de tiempo, los siniestros evitados y las emisiones evitadas. Quedan fuera otros usos que la misma infraestructura habilitaría, en particular la detección automática de incidentes discutida en §7, la fiscalización automatizada y el uso de los conteos para planificación de infraestructura. El resultado es por lo tanto una **cota inferior** del valor social del sistema.

El **horizonte de cinco años** se ata a la vida útil del hardware de cómputo, que es el activo de reemplazo más frecuente, y no a la de la infraestructura de control adaptativo, que es considerablemente mayor. Esa elección deprime el beneficio acumulado sin reducir la inversión inicial, de manera que opera como supuesto conservador, aunque podría significar que el flujo de beneficios se trunque antes de que el activo principal termine su vida.

El **despliegue se modela únicamente en el borde**, con un equipo de inferencia junto a las cámaras. Una arquitectura centralizada en servidor tendría una estructura de costos distinta, con menor desperdicio de capacidad de cómputo y un costo de transmisión de video que el modelo actual no contabiliza. Finalmente, la valorización adopta precios sociales y la tasa social de descuento del Sistema Nacional de Inversiones [50], de modo que sus resultados responden a la pregunta de conveniencia pública y no son comparables con una evaluación privada de rentabilidad.

8.4) Limitaciones del diseño experimental

Cada modelo se entrenó **una sola vez, con semilla fija**. Los intervalos de confianza reportados en §5.1 provienen de remuestreo pareado sobre las predicciones ya almacenadas, de modo que capturan la variabilidad debida a qué imágenes componen el conjunto de validación, pero no la variabilidad del entrenamiento en sí. Repetir el entrenamiento con semillas distintas permitiría acotar esa segunda fuente, que es la relevante cuando dos modelos quedan separados por milésimas de mAP, como ocurre entre YOLO11l y YOLO11x.

La **telemetría se midió sobre una única GPU**, una NVIDIA RTX PRO 2000 Blackwell, con lotes de una imagen. Los valores absolutos de FPS, energía y VRAM serían distintos sobre el hardware de borde que el modelo económico supone, del tipo Jetson, aunque el ordenamiento relativo entre modelos, que es lo que alimenta el cálculo de dispositivos necesarios, debería mantenerse.

8.5) Perspectivas futuras

Las restricciones anteriores señalan con bastante precisión el trabajo que sigue.

La línea de mayor retorno es **reentrenar sobre el conjunto completo de BDD100K**, con cerca de 70.000 imágenes de entrenamiento y varios miles de escenas adversas, para verificar si el cruce entre YOLO11l y YOLO11x desaparece. Si desaparece, la recomendación de modelo para climas con precipitación frecuente cambiaría, y con ella parte del análisis de §5.6.

La segunda es **validar sobre imágenes de las propias cámaras de la UOCT**, lo que permitiría medir la brecha de dominio entre BDD100K y Santiago y, con ello, cuánto se degrada el factor de calidad al pasar al despliegue real.

En el plano económico, **valorizar la detección automática de incidentes** convertiría en beneficio contabilizado lo que hoy es una fuente reconocida pero no cuantificada, y **medir la distribución chilena del retardo por jerarquía vial** reemplazaría por evidencia el supuesto que hoy

sostiene la ponderación de la tipología de intersecciones, que es además uno de los parámetros que el análisis de riesgo identifica con peso apreciable.

Finalmente, la comparación puede extenderse a **arquitecturas más recientes**, como YOLO26 o YOLOv12, esta última ya presente en la evaluación comparativa reproducida en §7, bajo el mismo protocolo y el mismo marco de valorización, de modo que el aporte metodológico de esta memoria, que es traducir precisión por condición en valor social, quede disponible para evaluar la siguiente generación de detectores sin rehacer el andamiaje.

9) Conclusiones

Esta memoria se propuso evaluar los cinco modelos YOLO11 sobre tráfico urbano y traducir ese desempeño técnico a valor social, con el fin de determinar cuál ofrece el mejor equilibrio entre calidad y costo para un sistema de gestión de tránsito. Los seis objetivos específicos se cumplieron y sus resultados se resumen a continuación.

1) ¿Permite el conjunto construido analizar el desempeño de forma desagregada por condición de operación?

Sí. Se preparó un subconjunto de BDD100K de 15.000 imágenes de entrenamiento y 3.000 de validación, cada una etiquetada por momento del día, clima y densidad vehicular (§4.1). Ese etiquetado es lo que hizo posible todo el análisis posterior, puesto que permitió medir el desempeño dentro de cada condición en vez de un promedio global que las mezcla.

2) ¿Son comparables entre sí los cinco modelos entrenados?

Sí. Los cinco se ajustaron bajo un protocolo idéntico, con los mismos hiperparámetros, la misma semilla y el mismo conjunto de datos (§4.2). Esa igualdad de trato es la que permite atribuir las diferencias observadas al tamaño del modelo y no al procedimiento de entrenamiento.

3) ¿Cuánto se degrada la precisión de cada modelo ante condiciones adversas?

El mAP50-95 global estimado por remuestreo crece de 0,244 en YOLO11n a 0,338 en

YOLO11x, aunque con un estancamiento marcado a partir de YOLO11m, cuyo intervalo de confianza al 95% se solapa con los de YOLO11l y YOLO11x (§5.1). El hallazgo más robusto del análisis por condición es que **la noche es la condición crítica**, con caídas de entre 22% y 29% del mAP diurno bajo clima y densidad fijos, una magnitud que supera a cualquier diferencia entre condiciones climáticas (§5.3.2). Ese resultado se sostiene bajo dos métodos con supuestos distintos, la ponderación por probabilidad inversa y las comparaciones con confusores fijados, lo que le da respaldo metodológico. El análisis por clima entrega un segundo hallazgo, que es un **cruce entre modelos**, dado que los tamaños pequeños se penalizan más con cielo despejado, donde la visibilidad larga expone objetos pequeños y distantes, mientras que en los tamaños grandes la lluvia y la nieve emergen como las condiciones más exigentes (§5.6).

4) ¿Cuánto cuesta operar cada modelo y cómo se compara ese costo entre tamaños?

La telemetría muestra que escalar de YOLO11n a YOLO11x multiplica por 10,7 la energía por inferencia y divide por 5,4 los cuadros por segundo (§5.5). Sin embargo, al llevar esas magnitudes a un costo total de propiedad sobre 3.250 intersecciones, **el cómputo de IA resulta ser una fracción marginal del costo**, entre 0,05% y 0,40% del total según el modelo, frente al 68,2% que representa la infraestructura de control adaptativo (§6.2). La elección del detector no es, por lo tanto, una decisión de costo.

5) ¿Cuál es el valor social del despliegue y cómo se distribuye entre mecanismos?

El sistema es socialmente rentable en todos los escenarios deterministas evaluados, con una relación beneficio-costos de entre 2,03 para YOLO11n y 2,75 para YOLO11x en USD constantes de 2026 (§6.2). Ningún modelo alcanza la referencia de 5,4 que el USDOT reporta para despliegues comparables, si bien todos superan varias veces el umbral de rentabilidad. El beneficio se concentra en los siniestros evitados, que aportan el 62,1% del total, seguidos por el ahorro de tiempo con un 24,9% y las emisiones evitadas con un 13,0%.

6) ¿Cuánta incertidumbre tiene esa valorización?

La simulación de Montecarlo de 100.000 iteraciones sobre 18 parámetros muestra que **el VAN de**

YOLO11m es positivo en el 100% de las combinaciones muestreadas, con un valor en riesgo al 5% de US\$221 millones (§6.3). El riesgo del proyecto está dominado por la efectividad y no por el costo, ya que la reducción de siniestros explica por sí sola la mayor parte de la varianza del resultado. El único parámetro de costo con peso apreciable es la infraestructura por intersección, que es precisamente la partida ajena a la elección del modelo.

Conclusión general

El modelo recomendado es **YOLO11m**, que captura el **96% del beneficio social máximo alcanzable**. Sus factores de calidad de detección llegan a 0,96 tanto en el mecanismo de tráfico como en el de seguridad, de modo que realiza casi todo el beneficio que rendiría el mejor modelo de la familia. El 4% adicional que aportan YOLO11l y YOLO11x cabe dentro del margen de incertidumbre del propio modelo de valorización, puesto que escalar de YOLO11m a YOLO11x mejora la relación beneficio-costos en 0,09 puntos mientras esa relación se mueve en una banda de 2,07 puntos con 90% de probabilidad.

El comportamiento bajo condiciones degradadas refuerza esa elección, ya que es ahí donde un sistema de gestión de tránsito se juega su utilidad. La noche es la condición crítica para los cinco tamaños por igual, con caídas de entre 22% y 29% del mAP diurno a clima y densidad fijos, de manera que crecer en tamaño alivia el problema sin resolverlo. El clima, en cambio, sí distingue entre modelos. Los tamaños pequeños se penalizan sobre todo con cielo despejado, donde la visibilidad larga expone objetos lejanos que su capacidad no alcanza a resolver, mientras que en los tamaños grandes la lluvia y la nieve pasan a ser lo más exigente. El cruce entre ambos regímenes ocurre justo después de YOLO11m, dado que ya en YOLO11l y YOLO11x la precipitación desplaza al despejado como peor condición. En una ciudad como Santiago, donde la lluvia es estacional y la nieve prácticamente inexistente, ese cruce juega a favor de la variante intermedia.

Referencias (APA)

- [1] Naciones Unidas, UN DESA, División de Población. (2018). *World Urbanization Prospects: The 2018 Revision*. <https://population.un.org/wup/assets/WUP2018-Report.pdf>
- [2] INRIX. (2025). *2024 Global Traffic Scorecard*. <https://inrix.com/press-releases/2024-global-traffic-scorecard-us/>
- [3] INRIX. (2019). *Global Traffic Scorecard 2019*. <https://inrix.com/press-releases/2019-traffic-scorecard-us/>
- [4] Calatayud, A., Sánchez González, S., Bedoya-Maya, F., Giráldez Zúñiga, F., & Márquez, J. M. (2021). *Congestión urbana en América Latina y el Caribe: características, costos y mitigación*. Banco Interamericano de Desarrollo. <https://doi.org/10.18235/0003149>
- [5] Smith, S. F., Barlow, G. J., Xie, X.-F., & Rubinstein, Z. B. (2013). *Smart Urban Signal Networks: Initial Application of the SURTRAC Adaptive Traffic Signal Control System*. ICAPS (AAAI). <https://doi.org/10.1609/icaps.v23i1.13594>
- [6] Superintendencia del Medio Ambiente [SMA]. (2026). *Autoridades lanzan Gestión de Episodios Críticos (GEC) 2026 para enfrentar la mala calidad del aire en la Región Metropolitana*, citando el Inventario de Emisiones 2022. <https://portal.sma.gob.cl/index.php/autoridades-lanzan-gestion-de-episodios-criticos-gec-2026-para-enfrentar-la-mala-calidad-del-aire-en-la-region-metropolitana/>

- [7] CONASET. (2025). *Informe nacional de siniestros de tránsito en Chile 2024*. <https://conaset.cl/wp-content/uploads/2025/07/Informe-nacional-de-siniestros-en-Chile-2024.pdf>
- [8] Wilson, C., Willis, C., Hendrikz, J. K., Le Brocque, R., & Bellamy, N. (2010). *Speed cameras for the prevention of road traffic injuries and deaths*. Cochrane Database of Systematic Reviews, (11), CD004607. <https://doi.org/10.1002/14651858.CD004607.pub4>
- [9] Unidad Operativa de Control de Tránsito (UOCT), MTT Chile. <https://mtt.gob.cl/archivos/25154>
- [10] Gobierno Regional Metropolitano de Santiago. (2024, 30 de abril). *Gobernador de Santiago y delegada lanzan el primer Sistema integrado de cámaras de televigilancia y gestión de tránsito con uso de inteligencia artificial de la RM*. <https://www.gobiernosantiago.cl/gobernador-de-santiago-y-delegada-lanzan-el-primer-sistema-integrado-de-camaras-de-televigilancia-y-gestion-de-transito-con-uso-de-inteligencia-artificial-de-la-rm/>
- [11] Varghese, C., & Shankar, U. (2007). *Passenger Vehicle Occupant Fatalities by Day and Night: A Contrast* (Traffic Safety Facts, Research Note DOT HS 810 637). National Highway Traffic Safety Administration. <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/810637>
- [12] National Highway Traffic Safety Administration [NHTSA]. (2019). *The Economic and Societal Impact of Motor Vehicle Crashes, 2019*. <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813403.pdf>
- [13] Alif, M. A. R. (2024). *YOLOv11 for Vehicle Detection: Advancements, Performance, and Applications in Intelligent Transportation Systems*. arXiv:2410.22898. <https://arxiv.org/abs/2410.22898>

- [14] Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. (2024). *Evaluating the Evolution of YOLO (You Only Look Once) Models: A Comprehensive Benchmark Study of YOLO11 and Its Predecessors*. arXiv:2411.00201. <https://arxiv.org/abs/2411.00201>
- [15] Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., & Darrell, T. (2020). *BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning*. CVPR. <https://doi.org/10.1109/CVPR42600.2020.00271>
- [16] Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2019). *Object Detection with Deep Learning: A Review*. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3212–3232. <https://doi.org/10.1109/TNNLS.2018.2876865>
- [17] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). *End-to-End Object Detection with Transformers*. En Computer Vision – ECCV 2020, Lecture Notes in Computer Science, vol. 12346, pp. 213–229. Springer. https://doi.org/10.1007/978-3-030-58452-8_13
- [18] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., & Chen, J. (2024). *DETRs Beat YOLOs on Real-time Object Detection*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024), pp. 16965–16974. https://openaccess.thecvf.com/content/CVPR2024/papers/Zhao_DETRs_Beat_YOLOs_on_Real-time_Object_Detection_CVPR_2024_paper.pdf
- [19] Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). *A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS*. Machine Learning and Knowledge Extraction, 5(4), 1680–1716. <https://doi.org/10.3390/make5040083>
- [20] Ultralytics. (2024). *YOLO11 Docs*. <https://docs.ultralytics.com/models/yolo11>
- [21] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. CVPR. <https://doi.org/10.1109/CVPR.2016.91>

- [22] Jocher, G. et al. (2026). *Ultralytics YOLO26: Unified Real-Time End-to-End Vision Models*. arXiv:2606.03748.
- [23] Ultralytics. (2026, 14 de enero). *Ultralytics YOLO26: Architecture, Benchmarks & Edge Deployment* (blog oficial). <https://www.ultralytics.com/blog/ultralytics-yolo26-the-new-standard-for-edge-first-vision-ai>
- [24] Ultralytics. (2026). *YOLO26: documentación oficial del modelo* (fecha de publicación 14 de enero de 2026; benchmarks de inferencia en CPU ONNX). <https://docs.ultralytics.com/models/yolo26/>
- [25] Khanam, R., & Hussain, M. (2024). *YOLOv11: An Overview of the Key Architectural Enhancements*. arXiv:2410.17725. <https://arxiv.org/abs/2410.17725>
- [26] Kong, C., Li, F., Yan, X., Yang, J., Mo, P., Luo, Q., & Mao, R. (2026). *Object detection on low-compute edge SoCs: a reproducible benchmark and deployment guidelines*. Scientific Reports. <https://doi.org/10.1038/s41598-026-36862-y>
- [27] Lin, T.-Y., et al. (2014). *Microsoft COCO: Common Objects in Context*. ECCV. https://doi.org/10.1007/978-3-319-10602-1_48
- [28] COCO Consortium. (2024). *COCO Detection Evaluation* (protocolo oficial de evaluación, implementado en `pycocotools/cocoeval.py` de la COCO API). <https://github.com/cocodataset/cocoapi/blob/master/PythonAPI/pycocotools/cocoeval.py>
- [29] Hendrycks, D., & Dietterich, T. (2019). *Benchmarking Neural Network Robustness to Common Corruptions and Perturbations*. ICLR. <https://danhendrycks.com/robustness/>
- [30] Kumar, D., & Muhammad, N. (2023). *Object Detection in Adverse Weather for Autonomous Driving through Data Merging and YOLOv8*. Sensors, 23(20), 8471. <https://doi.org/10.3390/s23208471>

- [31] Kenk, M. A., & Hassaballah, M. (2020). *DAWN: Vehicle Detection in Adverse Weather Nature Dataset*. arXiv:2008.05402. <https://arxiv.org/abs/2008.05402>
- [32] Sakaridis, C., Dai, D., & Van Gool, L. (2018). *Semantic Foggy Scene Understanding with Synthetic Data*. IJCV. <https://doi.org/10.1007/s11263-018-1072-8>
- [33] Tung, C., Ortiz, M. R., Goel, A., Thiruvathukal, G. K., & Lu, Y.-H. (2019). *Large-Scale Object Detection of Images from Network Cameras in Variable Ambient Lighting Conditions*. IEEE MIPR 2019. <https://arxiv.org/abs/1812.11901>
- [34] Luan, D., Zhang, Z., Zong, P., Zhou, J., Xie, L., Liu, H., & Zhu, J. (2025). *DB-YOLO: An Enhanced YOLOv11 Approach for Vehicle and Pedestrian Detection in Foggy Scenarios*. IET Image Processing, 19, e70246. <https://doi.org/10.1049/ipr2.70246>
- [35] Talaat, F. M., El-Balka, R. M., Sweidan, S., et al. (2025). *Smart traffic management system using YOLOv11 for real-time vehicle detection and dynamic flow optimization in smart cities*. Neural Computing and Applications, 37, 19957–19974. <https://doi.org/10.1007/s00521-025-11434-9>
- [36] Portal Innova. (2023, 6 de septiembre). *Videovigilancia inteligente reduce un 60% de los robos en domicilios de Las Condes* (programa municipal *Barrio Protegido*; solución desarrollada por SONDA junto a la startup VSaaS). <https://portalinnova.cl/videovigilancia-inteligente-reduce-un-60-de-los-robos-en-domicilios-de-las-condes/>
- [37] Municipalidad de Las Condes. (2023, 20 de julio). *Con sensores IoT e inteligencia artificial: las nuevas tecnologías de Las Condes para ser una ciudad inteligente*. <https://www.lascondes.cl/con-sensores-iot-e-inteligencia-artificial-las-nuevas-tecnologias-de-las-condes-para-ser-una-ciudad-inteligente/>
- [38] Ministerio del Interior y Seguridad Pública. (2024). *Sistema Integrado de Televigilancia con Inteligencia Artificial (SITIA): integración de cámaras de municipios, Carabineros y*

- privados*. <https://www.interior.gob.cl/noticias/2024/05/07/gobierno-presenta-avances-de-sistema-que-integra-camaras-de-televigilancia-de-carabineros-municipios-y-centros-comerciales/>
- [39] SITIA, Subsecretaría de Prevención del Delito. (2026). *SITIA: Sistema Integrado de Teleprotección con Inteligencia Artificial de Chile* (sitio oficial; cifras de cobertura y Portal de Patentes). <https://www.sitia.gob.cl/>
- [40] Emol. (2023, 25 de abril). *UOCT inicia reconexión de 50 cruces semaforizados en ocho ejes de la RM*. El Mercurio On-Line. <https://www.emol.com/noticias/Nacional/2023/04/25/1093174/transportes-reconexion-semaforos-rm.html>
- [41] SITIA, Subsecretaría de Prevención del Delito. (2023, 5 de agosto). *Gobierno, empresas y la academia dan el vamos a piloto de sistema de teleprotección con inteligencia artificial*. <https://www.sitia.gob.cl/>
- [42] SITIA, Subsecretaría de Prevención del Delito. (2025, 11 de abril). *Anuncian incorporación de 1.500 cámaras al programa de Sistema Integrado de Teleprotección con Inteligencia Artificial*; copia local citas_pdf/sitia_camaras2025.html. <https://www.sitia.gob.cl/>
- [43] Jofré, F., Unidad Operativa de Control de Tránsito [UOCT]. (2016). *Avances y desafíos en la gestión, modelación y programación de semáforos*. Segundo Plenario de la Sociedad Chilena de Ingeniería de Transporte (SOCHITRAN), 22 de julio de 2016. <https://sochitran.cl/wp-content/uploads/2022/07/Avances-y-Desafios-en-la-Gestion-Modelacion-y-Programacion-de-Semaforos-F.-Jofre-UOCT.pdf>
- [44] Abdel-Rahim, A., Chang, K., & Zender, R. (2018). *Evaluation of Vehicle Detection Systems for Traffic Signal Operations*. Journal of Transportation Engineering, Part A: Systems, 144(2). <https://doi.org/10.1061/JTEPBS.0000112>

- [45] Jyoti, R. (2026). *Edge vs. cloud TCO: The strategic tipping point for AI inference*. CIO. <https://www.cio.com/article/4109609/edge-vs-cloud-tco-the-strategic-tipping-point-for-ai-inference.html>
- [46] Curcio, E. (2026). *Evaluating the Lifecycle Economics of AI: The Levelized Cost of Artificial Intelligence (LCOAI)*. Information Systems, 136, 102634. <https://www.sciencedirect.com/science/article/abs/pii/S0306437925001206>
- [47] Yuan, Y., Dong, W., Yang, S., & Wu, T. (2026). *AWD-YOLO: Enhancing autonomous driving perception reliability in adverse weather*. Scientific Reports, 16, artículo 338. <https://doi.org/10.1038/s41598-025-29575-1>
- [48] Raschka, S. (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. arXiv:1811.12808. <https://arxiv.org/abs/1811.12808>
- [49] Zreiqat, D. Z., Nguyen, V., & Rawashdeh, N. A. (2026). *Label quality as a bottleneck for YOLO detection in all-weather autonomous driving using the MUSES dataset*. Discover Electronics. <https://doi.org/10.1007/s44291-026-00244-5>
- [50] Ministerio de Desarrollo Social y Familia, Sistema Nacional de Inversiones. (2024). *Precios Sociales 2024 (tasa social de descuento 5,5%; precio social del carbono US\$63,4/ton CO2)*. https://sni.gob.cl/storage/docs/Informe_precios_sociales_2024_SNI-Chile.pdf
- [51] Ministerio de Desarrollo Social y Familia / SECTRA-MTT. (2023). *Valor Social del Tiempo de Viaje en el contexto del Sistema Nacional de Inversiones*. https://sni.gob.cl/wp-content/uploads/Informe_Final_VST_Etapa_2_2023.pdf
- [52] U.S. Department of Transportation, ITS Deployment Evaluation. (2021). *Implementation of Adaptive Signal Control Technology in Florida Resulted in an Overall Travel Time Reduction of 9.36 Percent for Eight Corridors* (ficha 2021-B01579; estudio primario: Elef-

- eriadou, L. et al., University of Florida Transportation Institute para Florida DOT, 2019).
<https://www.itskrs.its.dot.gov/2021-b01579>
- [53] Jiang, B. (2009). Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science*, 23(8), 1033–1048. <https://arxiv.org/abs/0802.1284>
- [54] Dirección de Presupuestos [DIPRES]. (2000). *Evaluación de Programa N°19: Sistema de Control de Área de Tráfico para Santiago (SCAT)* — Síntesis ejecutiva. Gobierno de Chile. https://www.dipres.gob.cl/597/articles-140975_r_ejecutivo_institucional.pdf
- [55] Transporte Informa RM / UOCT. *Nuestra Historia*. <http://www.transporteinforma.cl/quienes-somos/nuestra-historia>
- [56] Unidad Operativa de Control de Tránsito [UOCT], Región Metropolitana. (2023, mayo). *Minuta Procedimiento General de Calibración de Arcos SCOOT*. <https://www.transporteinforma.cl/content/uploads/2023/05/Minuta-Procedimiento-General-de-Calibracion-de-Arcos-SCOOT.pdf>
- [57] Yunex Traffic. (2024). *SCOOT Adaptive Signal Control: Deployments in North America*. <https://us.yunextraffic.com/portfolio/adaptive-technologies/scoot/>
- [58] Western Systems. (2019). *SCOOT Expansion Brings Smarter Traffic to Monterey, CA*. <https://www.westernsystems-inc.com/news/scoot-monterey-ca-expansion/>
- [59] City of Toronto. (2024). *SCOOT System for Traffic Signal Coordination*. Transportation Services (sistema legado, en migración a SCATS). <https://www.toronto.ca/services-payments/streets-parking-transportation/traffic-management/traffic-signals-street-signs/traffic-signals-in-toronto/scoot-system/>

- [60] Fontaine, M. D., Ma, J., & Hu, J. (2015). *Evaluation of the Virginia Department of Transportation Adaptive Signal Control Technology Pilot Project* (Informe VCTIR 15-R24). Virginia Center for Transportation Innovation and Research. <https://vtrc.virginia.gov/media/vtrc/vtrc-pdf/vtrc-pdf/15-R24.pdf>
- [61] Stevanovic, A. (2019). *NCHRP 20-07/Task 414: Benefits of Adaptive Traffic Control Deployments. A Review of Evaluation Studies*. Transportation Research Board, AASHTO Standing Committee on Highways. https://onlinepubs.trb.org/Onlinepubs/nchrp/docs/NCHRP20-07_Task414FinalReport.pdf
- [62] Federal Highway Administration [FHWA]. (2011). *Adaptive Signal Control Technology (ASCT) — Every Day Counts Initiative (EDC-1)*. <https://www.fhwa.dot.gov/innovation/everydaycounts/edc-1/asct.cfm>
- [63] Federal Highway Administration [FHWA]. (2018). *Research and Technology Evaluation Final Report: Adaptive Signal Control* (FHWA-HRT-17-007). <https://www.fhwa.dot.gov/publications/research/randt/evaluations/17007/17007.pdf>
- [64] U.S. Department of Transportation, ITS Joint Program Office. *ITS Deployment Evaluation Database*, entradas de beneficios de control adaptativo de semáforos: 2000-b00138 (Oakland County, Michigan: reducción de tiempos de viaje de 7,0%–8,6% en horas punta) y 2007-b00316 (Los Ángeles: reducción de 13% en tiempos de viaje, 31% en detenciones y 21% en demora). <https://www.itskrs.its.dot.gov/2000-b00138>
- [65] Schrank, D., Eisele, B., & Lomax, T. (2019). *2019 Urban Mobility Report*. Texas A&M Transportation Institute. <https://static.tti.tamu.edu/tti.tamu.edu/documents/umr/archive/mobility-report-2019.pdf>
- [66] Comisión Nacional de Seguridad de Tránsito [CONASET]. (2025). *Costo social de los siniestros de tránsito en Chile 2024*. <https://www.conaset.cl/wp-content/uploads/2025/12/Costos-Siniestros-2024.pdf>

- [67] Comisión Nacional de Seguridad de Tránsito [CONASET]. (2025). *Diagnóstico de sinistros de tránsito: Región Metropolitana 2024*. Ministerio de Transportes y Telecomunicaciones. <https://www.conaset.cl/wp-content/uploads/2025/10/Metropolitana-2024.pdf>
- [68] Kodi, J. H., Kitali, A. E., Ali, M. S., Alluri, P., & Sando, T. (2021). Estimating safety impacts of adaptive signal control technology using a full Bayesian approach. *Transportation Research Record*, 2675(11), 1168–1179. <https://doi.org/10.1177/03611981211025281>
- [69] U.S. Department of Transportation [USDOT]. (2022). *Benefit-Cost Analysis Guidance for Discretionary Grant Programs*. [https://www.transportation.gov/sites/dot.gov/files/2022-03/Benefit%20Cost%20Analysis%20Guidance%202022%20\(Revised\).pdf](https://www.transportation.gov/sites/dot.gov/files/2022-03/Benefit%20Cost%20Analysis%20Guidance%202022%20(Revised).pdf)
- [70] Wu, K., Ding, J., Lin, J., Zheng, G., Sun, Y., Fang, J., Xu, T., Zhu, Y., & Gu, B. (2025). *Big-data empowered traffic signal control could reduce urban carbon emission*. *Nature Communications*, 16, article 2013. <https://doi.org/10.1038/s41467-025-56701-4>
- [71] Ministerio del Medio Ambiente, Subsecretaría del Medio Ambiente y SEREMI del Medio Ambiente Región Metropolitana. (2023). *Inventario de Emisiones Sector Transporte en la Región Metropolitana*, Informe Final, 7 de diciembre de 2023 (ID licitación 611134-4-LE23). Tabla 8-4, inventario 2022. https://airerm.mma.gob.cl/wp-content/uploads/2024/03/2023_Inventario-Transporte.pdf
- [72] U.S. Department of Transportation, ITS Joint Program Office. *ITS Costs Database*, fichas 2025-sc00573 (modernización del ATMS de Chicago, ≈US\$5 M) y 2008-sc00162 (software central de gestión e integración de los distritos de FDOT, US\$0,25 M–2 M de 2008; despliegue ITS de San Antonio). <https://www.itskrs.its.dot.gov/costs>
- [73] Vose, D. (2008). *Risk Analysis: A Quantitative Guide* (3.^a ed.). John Wiley & Sons.

- [74] U.S. Department of Transportation, ITS Deployment Evaluation. (2013). *The average cost to implement Adaptive Signal Control Technology is \$28,725 per intersection* (ficha 2013-SC00278; estudio primario: Lodes, M. y Benekohal, R. F., *Safety Benefits of Implementing Adaptive Signal Control Technology: Survey Results*, Illinois Center for Transportation, University of Illinois at Urbana-Champaign). <https://itskrs.its.dot.gov/2013-sc00278>
- [75] U.S. Department of Transportation, ITS Deployment Evaluation. (2021). *Overall Cost to Implement Adaptive Signal Control Technology in Florida Ranged from \$30,000 to \$96,400 per Intersection* (ficha 2021-SC00484). <https://itskrs.its.dot.gov/2021-sc00484>
- [76] U.S. Department of Transportation, ITS Deployment Evaluation. (2026). *The Capital Cost of an AI-Driven Adaptive Signal Control Was Estimated at \$115,810 per Intersection with \$10,050 in Annual Operating Expenses* (ficha 2026-SC00586, piloto de Maricopa County, Arizona). <https://itskrs.its.dot.gov/2026-sc00586>
- [77] U.S. Department of Transportation, ITS Deployment Evaluation. (2022). *Use Case: Adaptive Signal Control Benefit-Cost Analysis*. https://www.itskrs.its.dot.gov/sites/default/files/doc/Adaptive_Traffic_Signals_BCA_Use_Case_20220708_Final_508.pdf
- [78] Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, 44(247), 335–341. <https://doi.org/10.1080/01621459.1949.10483310>
- [79] Iman, R. L., & Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics – Simulation and Computation*, 11(3), 311–334. <https://doi.org/10.1080/03610918208812265>

- [80] U.S. Environmental Protection Agency [EPA]. (2023). *Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances*. https://www.epa.gov/system/files/documents/2023-12/epa_scghg_2023_report_final.pdf
- [81] ElSahly, O. y Abdelfatah, A. (2024). Developing a Machine-Learning-Based Automatic Incident Detection System for Traffic Safety: Promises and Limitations. *Infrastructures*, 9(10), 170. <https://doi.org/10.3390/infrastructures9100170>
- [82] Federal Highway Administration [FHWA]. (2010). *Focus States Initiative: Traffic Incident Management Performance Measures Final Report*, FHWA-HOP-10-009, Apéndice B. <https://ops.fhwa.dot.gov/publications/fhwahop10010/appb.htm>

Anexo: Curvas Precisión-Recall y Matrices de Confusión

Las siguientes figuras complementan los resultados de §5.1. Las curvas Precisión-Recall muestran el desempeño por clase a lo largo de todos los umbrales de confianza; las matrices de confusión muestran los errores de clasificación entre clases para el umbral de confianza por defecto. El mAP@0,5 que etiqueta la curva "all classes" de cada gráfico puede diferir en $\leq 0,001$ del valor de la tabla de §5.1, por ser una ejecución de validación independiente sujeta a la variabilidad de punto flotante propia de la inferencia en GPU.

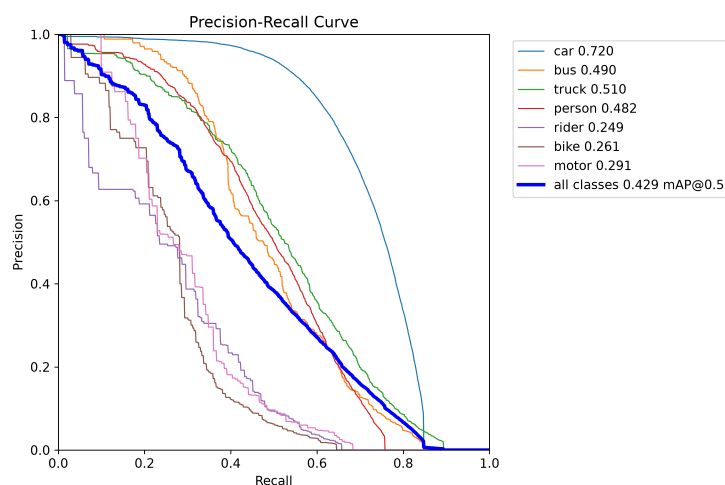


Figura 33a. Curva Precisión-Recall por clase, YOLO11n: cada curva corresponde a una clase (car, bus, truck, person, rider, bike, motor); el área bajo la curva (AP) indica la calidad de detección de esa clase. Fuente: elaboración propia.

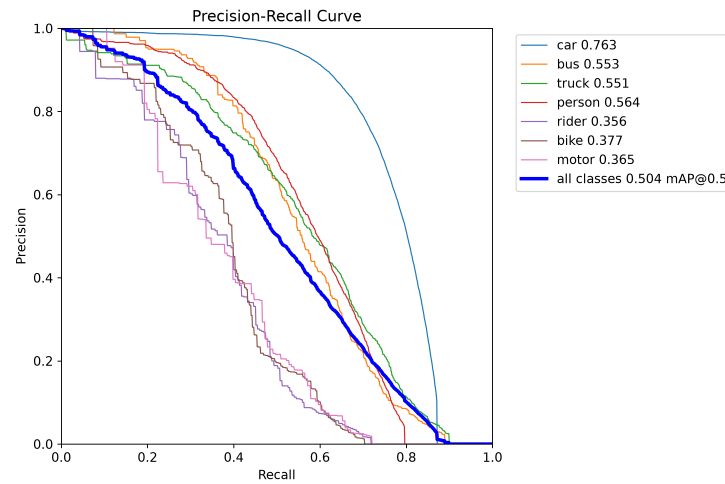


Figura 33b. Curva Precisión-Recall por clase, YOLO11s. Fuente: elaboración propia.

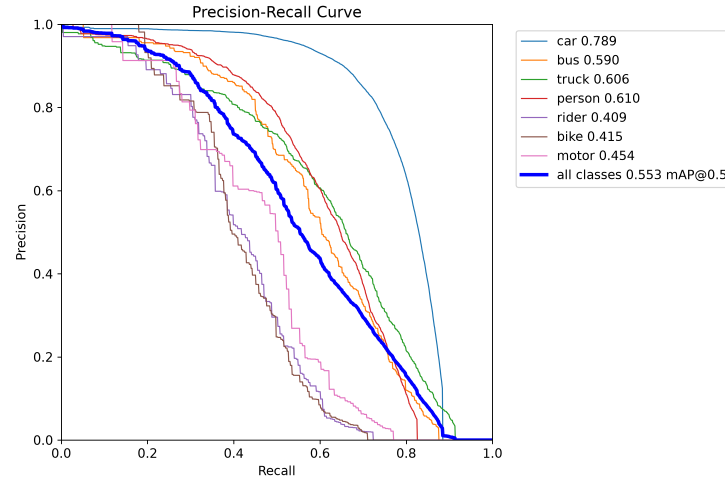


Figura 33c. Curva Precisión-Recall por clase, YOLO11m. Fuente: elaboración propia.

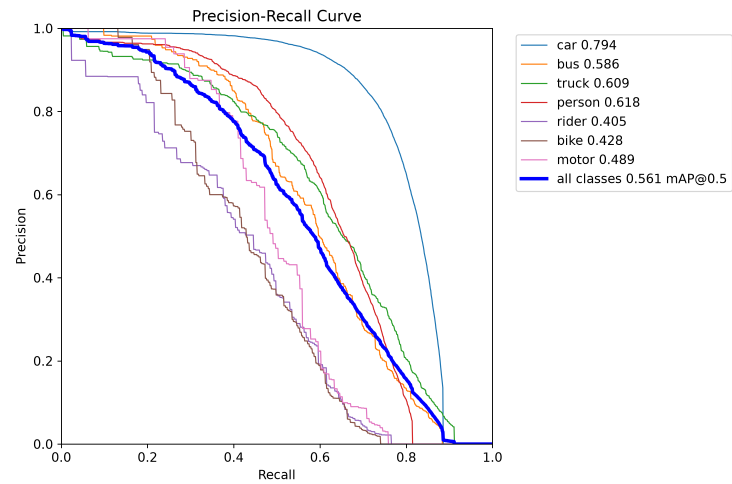


Figura 33d. Curva Precisión-Recall por clase, YOLO11l. Fuente: elaboración propia.

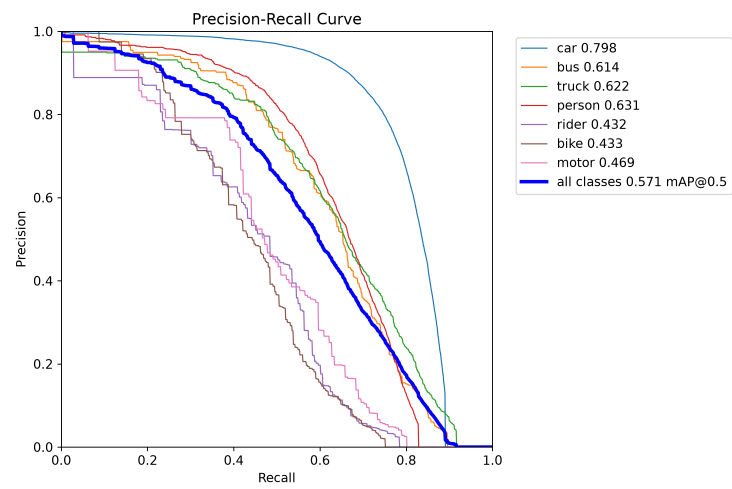


Figura 33e. Curva Precisión-Recall por clase, YOLO11x. Fuente: elaboración propia.

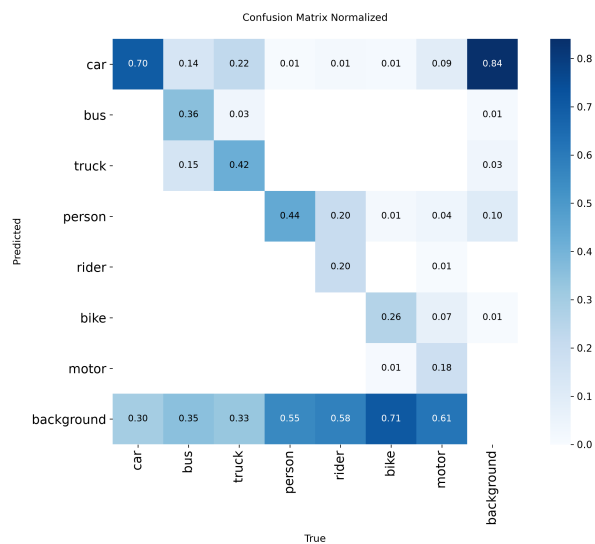


Figura 34a. Matriz de confusión normalizada, YOLO11n: filas = clase predicha, columnas = clase real; la diagonal principal muestra la tasa de detección correcta por clase; la fila “background” representa los falsos negativos. Fuente: elaboración propia.



Figura 34b. Matriz de confusión normalizada, YOLO11s. Fuente: elaboración propia.

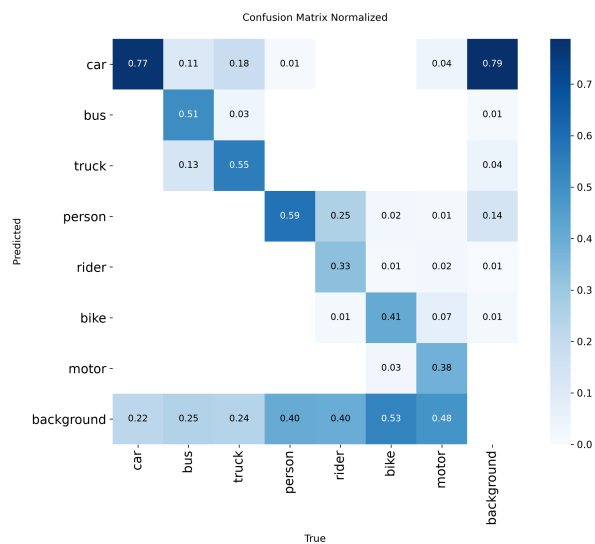


Figura 34c. Matriz de confusión normalizada, YOLO11m. Fuente: elaboración propia.

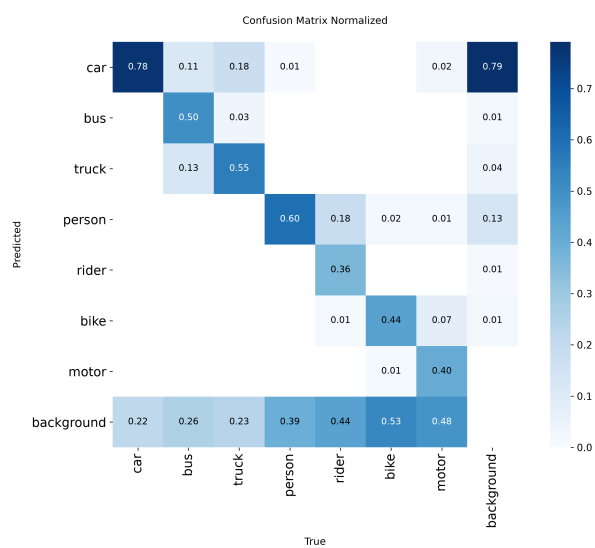


Figura 34d. Matriz de confusión normalizada, YOLO11l. Fuente: elaboración propia.



Figura 34e. Matriz de confusión normalizada, YOLO11x. Fuente: elaboración propia.