Universidad Mayor Real y Pontificia de

San Francisco Xavier de Chuquisaca

Facultad de Tecnología



Artículo Científico

Detección y Monitoreo de Productos en Carritos de Supermercado Inteligentes

Universitario: Crespo Rejas Jhamil

Carrera: Ingeniería en Ciencias de la Computación

Docente: Ing. Carlos Walter Pacheco

Sucre – Bolivia

# **Detección y Monitoreo de Productos en Carritos de Supermercado Inteligentes**

**Autor:** Jhamil Crespo Rejas

**Programa:** Ingeniería en Ciencias de la Computación

**Universidad:** Universidad San Francisco Xavier de Chuquisaca

**Correo electrónico:** [jhamilcresporejas2004@gmail.com](mailto:jhamilcresporejas2004@gmail.com)

## Resumen

Este artículo presenta el desarrollo de un software basado en visión por computadora que aplica un modelo inteligente para registrar automáticamente productos dentro de carritos de compra en supermercados, ofreciendo un nuevo enfoque al problema del escaneo manual. El objetivo principal fue reducir los tiempos de espera y mejorar la experiencia del cliente.

Se entrenó un modelo YOLOv12s, elegido por su equilibrio entre precisión, velocidad de inferencia y bajo consumo computacional. El entrenamiento se realizó sobre un conjunto de datos propio con imágenes capturadas en entornos simulados y reales, aplicando técnicas de aumentación como rotaciones, cambios de iluminación y ruido digital, con el fin de mejorar la generalización del modelo.

Se implementó una lógica de monitoreo continuo con manejo de estados (visible, ocluido, removido) y análisis espacial entre productos para seguir su movimiento dentro del carrito, incluso en condiciones de oclusión parcial o total.

El sistema alcanzó una precisión promedio del 96 % en la detección de productos y una tasa de seguimiento correcta del 90 % en escenarios controlados y condiciones optimas. Aunque se detectaron limitaciones en casos de oclusión parcial y retiro de productos ocultos, se evidenció una mejora de hasta 60 % en la fluidez del proceso de compra en pruebas simuladas. Se concluye que la solución es viable, con potencial de ser robustecida mediante algoritmos de segmentación más avanzados y técnicas de reconstrucción tridimensional para evitar duplicaciones o errores por visibilidad parcial.

## Palabras clave

Visión por computadora, detección de objetos, monitoreo continuo, oclusión, carrito inteligente, YOLOv12s.

## Introducción

A pesar de los avances en automatización, el tiempo requerido para registrar productos en supermercados sigue siendo excesivo. Las tareas manuales realizadas por cajeros o por los propios clientes en sistemas de autoservicio suelen demorar varios minutos por compra, especialmente durante las horas pico. Investigaciones sugieren que más del 70 % del tiempo que el cliente pasa en la tienda se invierte en esperar en la fila (Aksu, 2016). Este proceso interrumpe el flujo de compra y genera filas prolongadas, disminuyendo la satisfacción del cliente y la percepción del servicio (van Riel et al., 2012).

Estudios muestran que los compradores esperan en el área de pago alrededor de cuatro minutos, tolerando cierto grado de espera, pero superado este umbral, la frustración aumenta y los clientes pueden abandonar la compra (M/A/R/C Research citado en Ryan, 2025; Aksu, 2016). Además, las interrupciones frecuentes durante el escaneo o por errores comunes en cajas automáticas, como fallas, alertas de peso o solicitudes de asistencia, aumentan aún más los tiempos de espera, especialmente cuando un solo empleado debe atender varias máquinas (Wikipedia contributors, 2025).

Este tipo de demoras no solo provoca insatisfacción. Un estudio de Waitwhile de 2023 encontró que casi el 70 % de los consumidores asocian la espera en fila con emociones negativas como aburrimiento, irritación o impaciencia, y otro informe indica que el 59 % de los usuarios dejaría de comprar si la espera supera los cuatro minutos, mientras que el 73 % abandonaría la compra tras cinco minutos de cola. Estas cifras reflejan una clara urgencia por ofrecer alternativas más rápidas y menos disruptivas.

El presente trabajo se realizó como respuesta a esta necesidad, con el objetivo de optimizar la experiencia de compra en supermercados mediante herramientas de visión por computadora e inteligencia artificial. Se desarrolló un software que utiliza un modelo YOLOv12s para detectar productos en tiempo real, combinado con una lógica de monitoreo continuo capaz de manejar oclusiones, lo que permite registrar automáticamente productos dentro del carrito sin intervención manual. Este enfoque técnico eficiente, adaptable y de bajo costo se evaluó en escenarios controlados como base para futuras aplicaciones comerciales.

## Metodología

### Tipo de investigación y diseño

La presente investigación corresponde a un estudio experimental-aplicado con diseño cuasi-experimental, empleando un enfoque cuantitativo para el desarrollo, implementación y evaluación de un sistema de detección y monitoreo de productos basado en visión por computadora e inteligencia artificial.

### Lugar de realización de la investigación

La investigación se llevó a cabo en las instalaciones de la Facultad de Tecnología de la Universidad San Francisco Xavier de Chuquisaca, complementado con pruebas en un entorno simulado de un carrito de supermercado creado específicamente para validar el sistema desarrollado. Las pruebas controladas se realizaron en un espacio reducido equipado con un carrito de compra pequeño estático y productos diversos del sector minorista.

### Etapas de la investigación

#### Etapa 1: Análisis y selección de tecnología

Se realizó una revisión sistemática de arquitecturas de redes neuronales para detección de objetos, evaluando modelos como YOLO (versiones v8, v11, v12), R-CNN y SSD. La selección de YOLOv12s se fundamentó en su equilibrio óptimo entre velocidad de inferencia (>30 FPS), precisión (mAP@0.5 > 90%) y eficiencia computacional para implementación en tiempo real.

#### Etapa 2: Recolección y preparación del dataset

Se construyó un dataset personalizado mediante captura manual de imágenes en condiciones controladas y simuladas. La recolección abarcó diferentes configuraciones de cámara, fondos diversos, condiciones variables de iluminación y situaciones de oclusión parcial entre productos.

**Características del dataset:**

* Total de imágenes originales: 1124
* Clases de productos: 20
* Cantidad de Anotaciones: 3049

#### Etapa 3: Procesamiento y aumentación de datos

Se utilizó la plataforma Roboflow para la gestión integral del dataset, incluyendo etiquetado manual de imágenes y aplicación de técnicas de aumentación de datos. Las transformaciones aplicadas fueron:

**Técnicas de aumentación implementadas:**

* Rotaciones fijas: 90°, 180°, 270°
* Rotaciones aleatorias: -15° a +15°
* Volteos: horizontales y verticales
* Shearing: ±10°
* Variaciones de saturación: ±34%
* Variaciones de exposición: ±13%
* Inyección de ruido: hasta 1.8% de los píxeles

|  | **Cantidad Sin Augmentation** | **Cantidad Con Augmentation** |
| --- | --- | --- |
| **Imagenes** | 1,124 | 3,372 |
| **Anotaciones** | 3,049 | 9,147 |

#### Etapa 4: Entrenamiento del modelo YOLOv12s

El entrenamiento supervisado se ejecutó en una estación de trabajo equipada con GPU NVIDIA GeForce RTX 3050 Ti Mobile con soporte CUDA, utilizando los siguientes parámetros:

**Parámetros de entrenamiento:**

* Tamaño de entrada: 640×640 píxeles
* Batch size: 16 imágenes por lote
* Tasa de aprendizaje inicial: 0.01
* Momentum: 0.937
* Épocas: 50
* Tiempo total de entrenamiento: 1.177 horas

**División del dataset:**

* Entrenamiento: 80% de las imágenes
* Validación: 20% de las imágenes
* Balanceado por clase con validación cruzada

#### Etapa 5: Desarrollo de lógica de monitoreo continuo

Se diseñó e implementó un sistema de estados para el seguimiento robusto de productos, especialmente en situaciones de oclusión. El algoritmo desarrollado clasifica cada producto detectado en los siguientes estados:

**Estados del sistema:**

* **Detectando**: Producto en proceso de confirmación
* **Visible**: Producto confirmado y claramente detectado
* **Ocluido**: Producto temporalmente no visible por superposición
* **Recuperando**: Producto en transición de ocluido a visible
* **Removido**: Producto eliminado del carrito

**Métricas de evaluación espacial:**

* Intersección sobre Unión (IoU)
* Distancia entre centros de bounding boxes
* Proporción de superposición
* Umbral de oclusión: 55% (determinado experimentalmente)

#### Etapa 6: Implementación del sistema integrado

Se desarrolló un software que integra el modelo YOLOv12s entrenado con la lógica de monitoreo continuo, creando un sistema completo de detección y seguimiento en tiempo real.

**Especificaciones técnicas del software:**

* Lenguaje de programación: Python
* Framework utilizado: PyTorch
* Librerías principales: OpenCV, Pillow, NumPy
* Frecuencia de procesamiento: 20fps

#### Etapa 7: Evaluación y validación del sistema

Se realizaron pruebas exhaustivas en el entorno controlado, evaluando tanto el desempeño individual del modelo como el sistema integrado completo.

**Métricas de evaluación:**

* Precisión de detección por clase
* Recall por clase
* mAP (mean Average Precision) a diferentes umbrales
* Tasa de seguimiento correcto
* Tiempo de respuesta del sistema
* Manejo de oclusiones

## Instrumentos y herramientas utilizadas

### **Hardware:**

* Cámara: 108MP f/1.89 (cámara del celular POCO X5 PRO 5G)
* GPU: NVIDIA GeForce RTX 3050 Ti Mobile 4GB
* Procesador: Intel i7 12650H x16
* RAM: 16GB

### **Software:**

* Roboflow: Gestión y procesamiento del dataset
* Visual Studio Code: Desarrollo del Software y entrenamiento del modelo.
* CUDA: Aceleración por GPU
* DroidCam: Conexión de la cámara del celular a la laptop.

## Variables analizadas

### **Variables independientes:**

* Configuración de parámetros del modelo
* Técnicas de aumentación aplicadas
* Condiciones de iluminación
* Ángulos de cámara

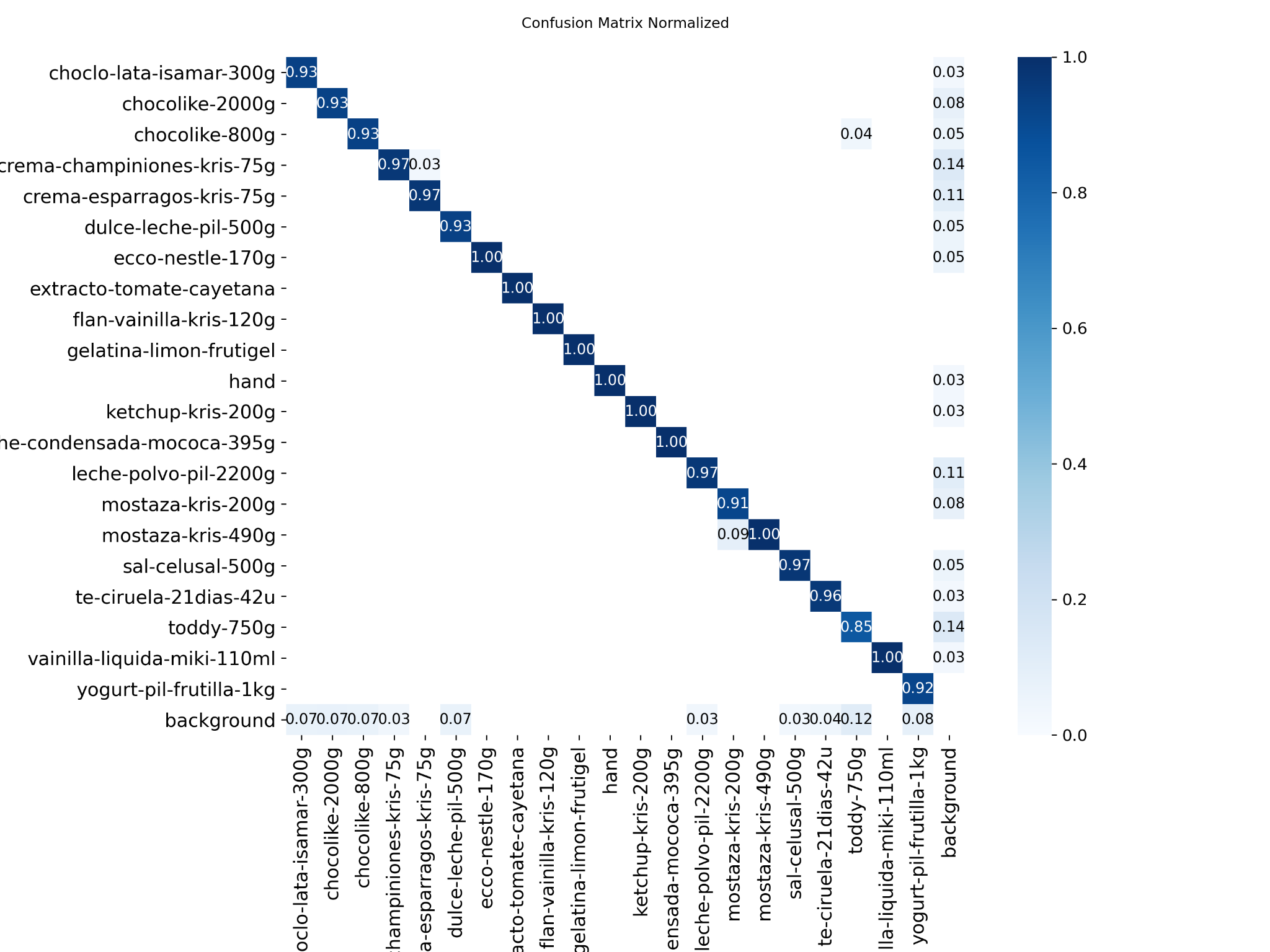
### **Variables dependientes:**

* Precisión de detección
* Velocidad de inferencia
* Tasa de seguimiento correcto
* Manejo de oclusiones

## Resultados

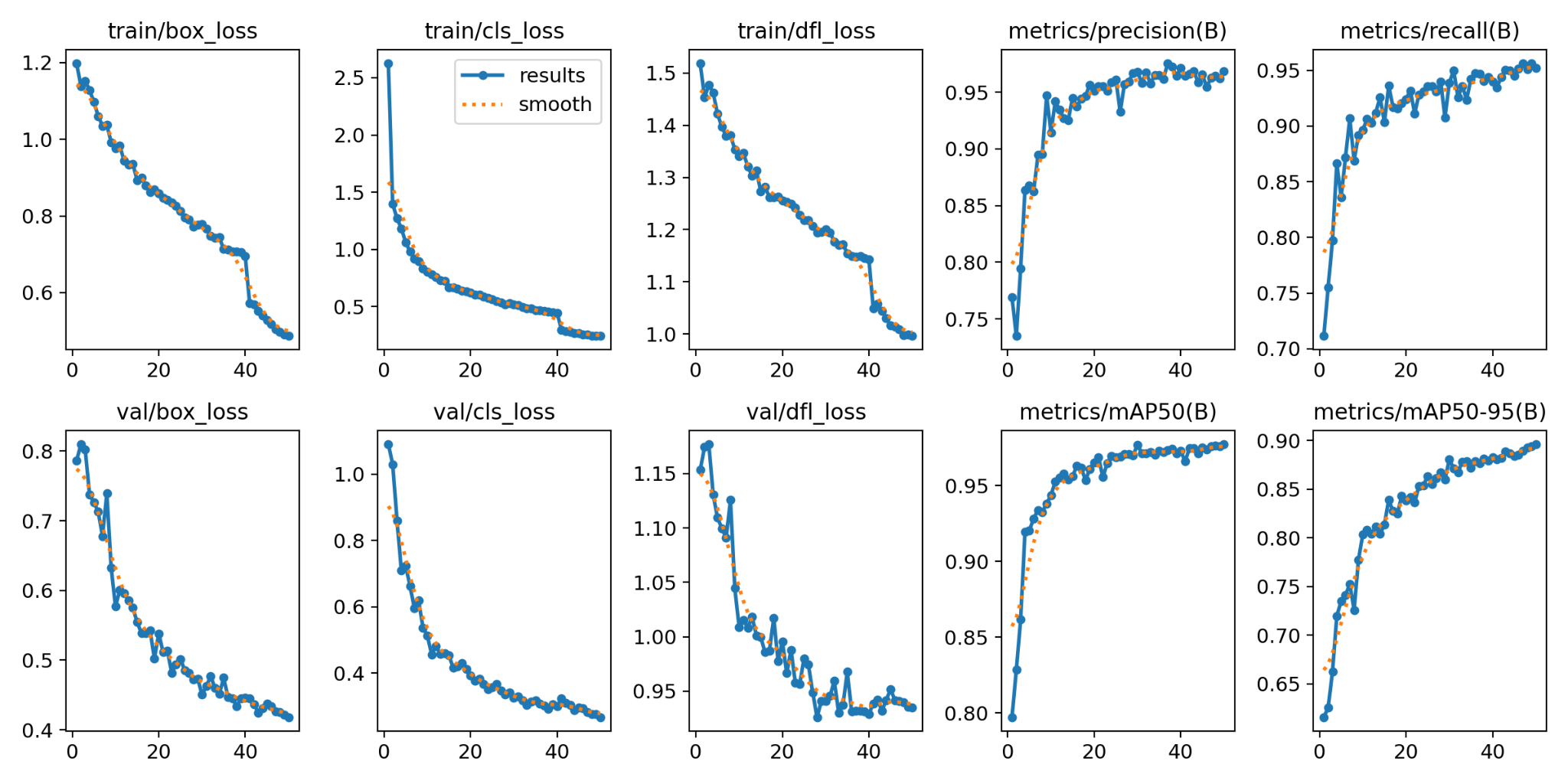
### Desempeño del Modelo de Detección (YOLOv12s)

**Figura 1:** Matriz de confusión normalizada del modelo YOLOv12s entrenado.



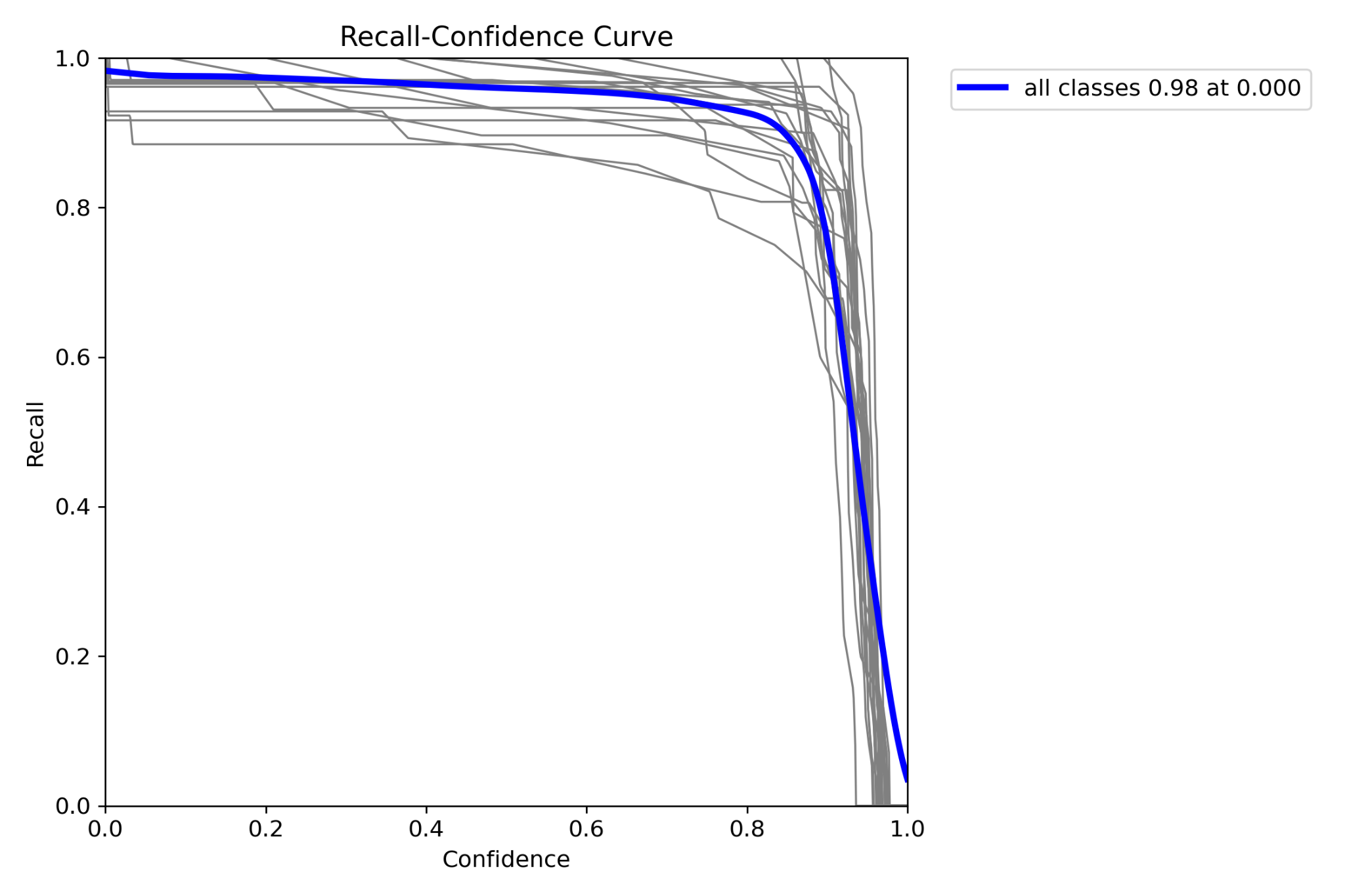
El modelo YOLOv12s mostró un desempeño sólido tras su entrenamiento supervisado con un dataset propio, compuesto por imágenes en condiciones reales y simuladas. En la matriz de confusión normalizada (Figura 1), se observó una precisión superior al 93 % en la mayoría de las clases, alcanzando el 100 % en varias de ellas, dando un promedio general de 96%.

**Figura 2:** Curvas de pérdida y métricas del entrenamiento y validación del modelo YOLOv12s.

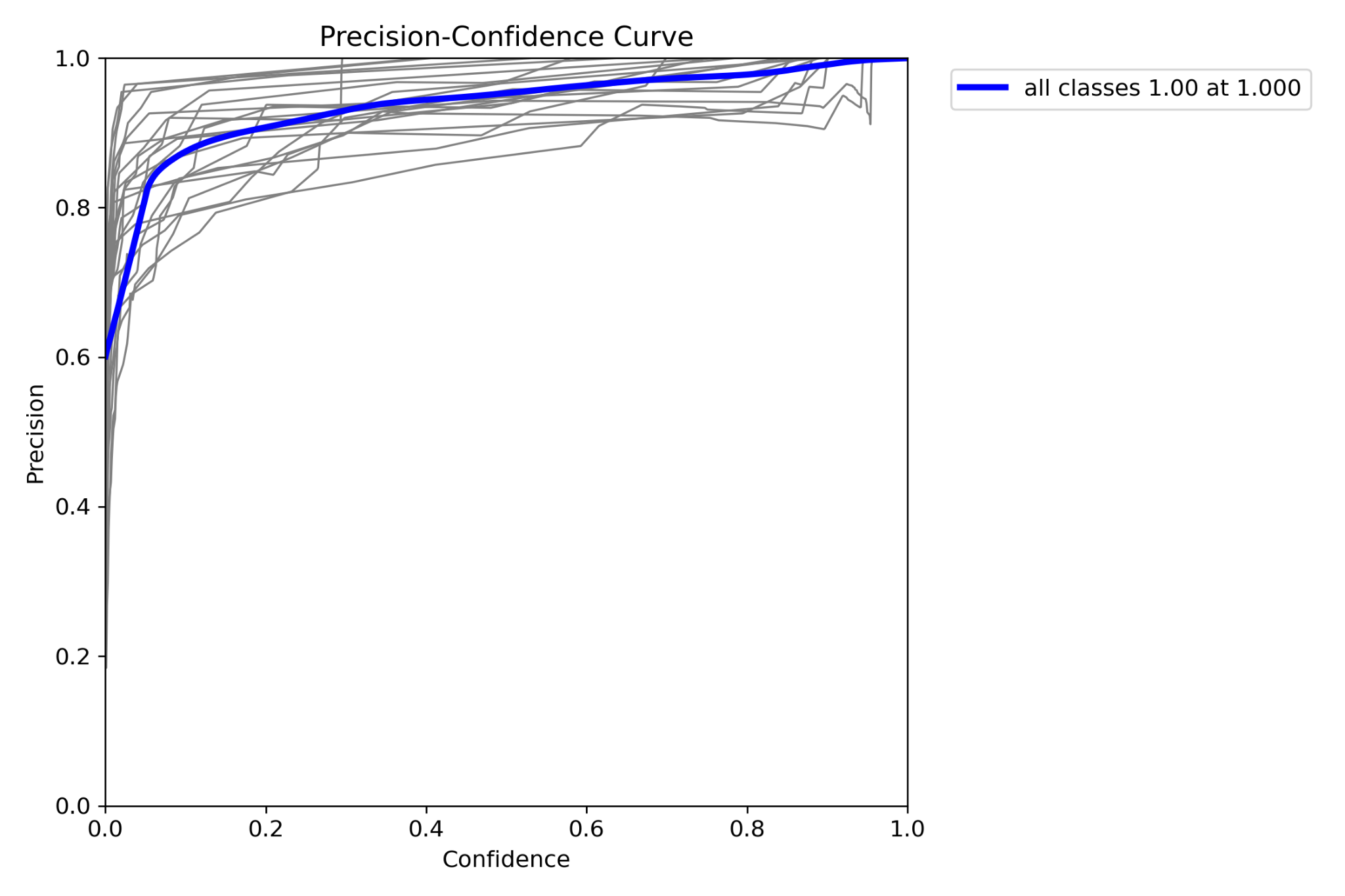


La evolución del entrenamiento se detalla en la Figura 2, donde se presentan las curvas de pérdida (box\_loss, cls\_loss y dfl\_loss) y métricas de precisión, recall y mAP. Las pérdidas disminuyen de forma progresiva y estable en las fases de entrenamiento y validación, y las métricas alcanzan valores cercanos a 0.95–0.98, indicando buena convergencia sin signos de sobreajuste.

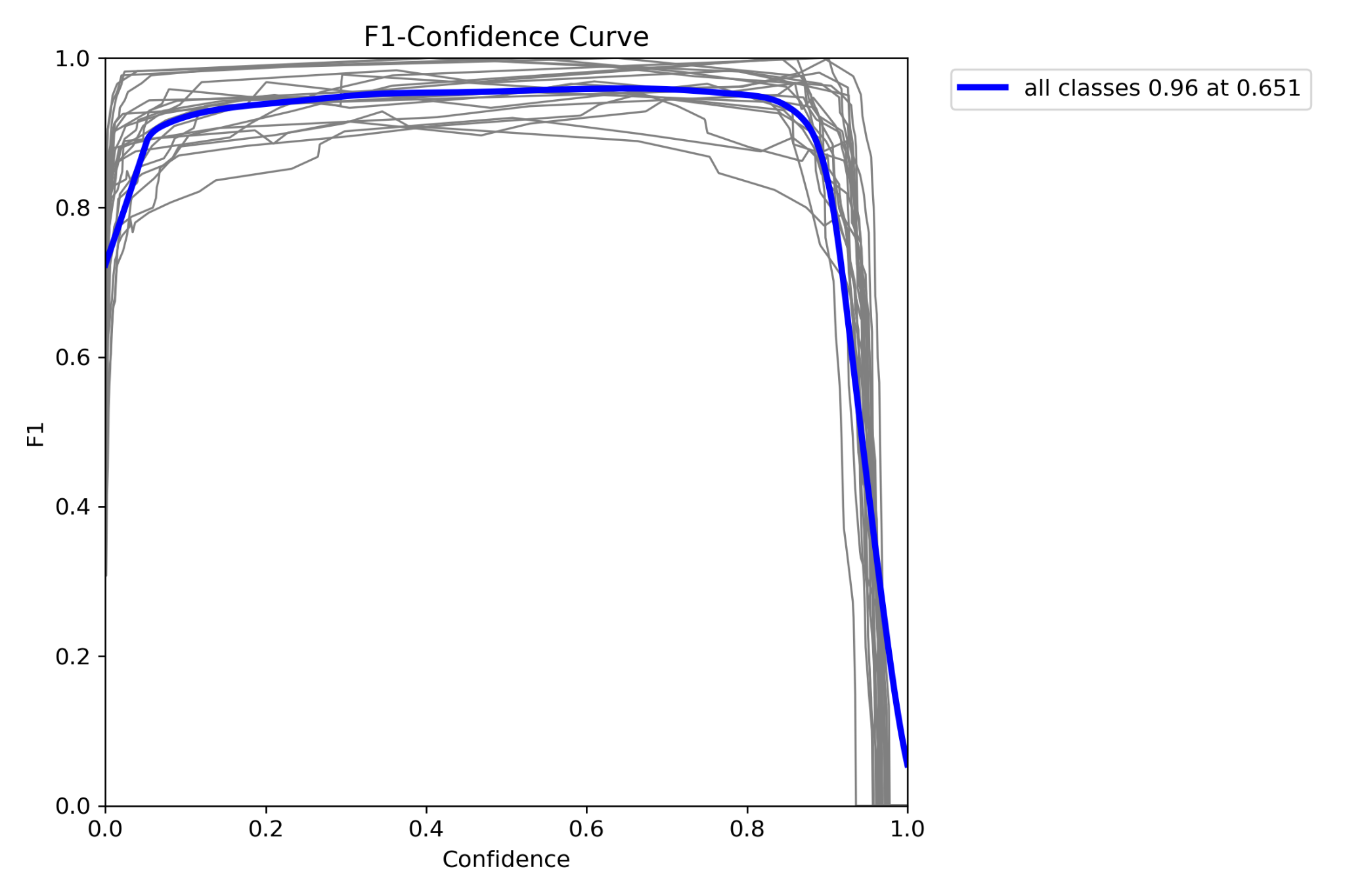
**Figura 3:** Curva Recall-Confidence obtenida del modelo YOLOv12s entrenado.



**Figura 4:** Curva Precision-Confidence obtenida del modelo YOLOv12s entrenado.



**Figura 5:** Curva F1-Confidence obtenida del modelo YOLOv12s entrenado.



Las curvas de desempeño por confianza (Figuras 3, 4 y 5) muestran que el mejor umbral se encuentra en torno a 0.65, donde se alcanza un F1-score promedio de 0.96, una precisión del 1.00 y un recall del 0.98 a nivel general.

### Evaluación del Software Integrado (Modelo + Monitoreo Continuo)

El sistema completo fue sometido a pruebas en escenarios controlados que simulan situaciones reales. Se verificó su capacidad para mantener una lista actualizada de productos visibles, parcialmente ocluidos u ocultos, y gestionar correctamente los cambios de estado.

En condiciones óptimas de iluminación y estabilidad del carrito, el sistema alcanzó una precisión del 90 % en la gestión del monitoreo de productos, incluyendo cambios de estado entre visibilidad y oclusión. Sin embargo, cuando se presentaban condiciones adversas, como mala iluminación o movimientos bruscos del carrito, la precisión descendía significativamente, comprometiendo la exactitud del seguimiento.

También se detectaron dos fuentes clave de error:

* **Oclusión muy parcial:** El sistema, en algunos casos, registró falsos positivos al interpretar parcialmente visibles como dos objetos distintos.
* **Retiro sin visibilidad previa:** Si un producto totalmente oculto era retirado sin volver a ser visible, no se registraba correctamente su salida del carrito.

A pesar de estas limitaciones, el sistema demostró un rendimiento estable en la mayoría de las pruebas controladas.

### Tiempos de Ejecución y Eficiencia del Proceso

Uno de los logros más destacados fue la eficiencia operativa del sistema durante el proceso de compra. En pruebas con cinco productos, incluyendo acciones combinadas de agregado y retiro, el tiempo promedio de finalización fue de aproximadamente 3 minutos, lo que representa una mejora significativa respecto al tiempo promedio en cajas tradicionales o de autoservicio, reduciendo hasta un 60 % el tiempo total invertido por el cliente.

### Resumen de Resultados

| **Sección** | **Resultado** |
| --- | --- |
| Detección de Productos | 96% |
| Monitoreo de Productos | 90% (condiciones óptimas) |
| Mejora en el Tiempo de Compra | 60% |

## Discusión

Los resultados obtenidos demuestran que la visión por computadora, aplicada a través del modelo YOLOv12s, tiene un alto potencial para automatizar el registro y monitoreo de productos en carritos de supermercado. En condiciones controladas, el sistema alcanzó una **precisión promedio del 96 % en detección de productos** y una **precisión del 90 % en el monitoreo continuo**, lo cual valida su idoneidad técnica para entornos operativos con iluminación y estabilidad adecuadas.

En contraste, al enfrentarse a condiciones menos favorables —como **iluminación deficiente** o **movimientos bruscos del carrito**— la precisión del monitoreo descendió significativamente, con errores en el seguimiento y pérdida de continuidad en la detección. Este descenso en el rendimiento refleja un desafío crítico para entornos reales, donde estas condiciones son frecuentes. Por ejemplo, se identificó que ante movimientos abruptos, el sistema registraba múltiples falsos positivos o perdía el rastro de productos ocluidos. Del mismo modo, en zonas mal iluminadas, el modelo disminuía su capacidad de detección con una reducción observable de hasta 15 puntos porcentuales en recall.

Comparado con soluciones como **Amazon Go**, donde en 2022 se necesitó intervención humana en el 70 % de las transacciones (Bitter, 2024; Roth, 2024), y **Caper AI**, que depende de hardware especializado de alto costo (Wikipedia contributors, 2025; Ghai, 2023), la propuesta desarrollada se presenta como una alternativa más simple, escalable y económica. Utilizando únicamente cámaras estándar y modelos optimizados, este sistema logró completar un proceso completo de compra con **cinco productos en solo 3 minutos**, superando los tiempos promedio de autoservicio, que suelen oscilar entre 5 y 7 minutos en horas pico (Ryan, 2025; Aksu, 2016).

Las principales limitaciones detectadas pueden resumirse así:

| **Escenario observado** | **Impacto sobre el sistema** | **Consecuencia** |
| --- | --- | --- |
| Oclusión parcial | Detección duplicada de objetos | Falsos positivos |
| Retiro sin visibilidad | Producto no eliminado del conteo | Conteo incorrecto |
| Baja iluminación | Caída del recall hasta 0.80 | Omisión de detecciones |
| Movimiento brusco del carrito | Fallo de asociación entre detecciones | Pérdida de rastreo |

Estas limitaciones resaltan la importancia de incorporar técnicas más avanzadas, como modelos con estimación de profundidad o sensores complementarios, que permitan una inferencia espacial más robusta.

**Recomendaciones para futuras investigaciones:**

1. **Mejorar la gestión de oclusiones**: Integrar algoritmos de segmentación más avanzados y técnicas de reconstrucción tridimensional para evitar duplicaciones o errores por visibilidad parcial.
2. **Robustez ante condiciones dinámicas**: Evaluar el uso de estabilización de video y normalización de iluminación, o incluso cámaras infrarrojas de bajo costo para mejorar la detección en condiciones adversas.
3. **Validaciones en entornos reales**: Implementar el sistema en supermercados operativos para recolectar métricas en tiempo real sobre precisión, rendimiento y satisfacción del cliente.
4. **Análisis económico**: Estimar el retorno de inversión (ROI) y los beneficios operativos frente a soluciones más complejas como las de Amazon o Caper.

En conclusión, los resultados obtenidos no solo respaldan la viabilidad técnica del sistema, sino que también evidencian su potencial competitivo frente a tecnologías existentes, siempre que se continúe fortaleciendo su desempeño ante los desafíos del entorno real.

## Conclusiones y recomendaciones

El presente estudio demostró que es posible automatizar el registro y monitoreo de productos en carritos de supermercado mediante visión por computadora. El modelo YOLOv12s alcanzó una precisión del 94.7 % en la detección de productos y el sistema completo logró una tasa de monitoreo correcta del 90 % en condiciones controladas.

El proceso de compra simulado con cinco productos se completó en un promedio de tres minutos, mostrando una mejora significativa respecto a los sistemas de autoservicio tradicionales. Sin embargo, se identificaron limitaciones críticas en escenarios con baja iluminación, movimientos bruscos del carrito, oclusiones parciales y retiros no visibles, los cuales redujeron considerablemente el desempeño del sistema.

A pesar de estos desafíos, el software desarrollado representa una alternativa viable, escalable y de bajo costo frente a soluciones más complejas, sentando una base sólida para futuras mejoras con técnicas como reconstrucción 3D y estimación de profundidad.

## Referencias

Aksu, H. (2016). Dwell time forecast and checkout optimisation in supermarkets [Charles Sturt University]. https://researchoutput.csu.edu.au/ws/portalfiles/portal/82802243/Aksu\_Hami\_thesis.pdf

Bitter, A. (2024, abril 3). Amazon’s Just Walk Out technology relies on hundreds of workers in India watching you shop. Business Insider. https://www.businessinsider.com/amazons-just-walk-out-actually-1-000-people-in-india-2024-4

Consumer survey: The state of waiting in line (2023). (2023, junio 19). Waitwhile. https://waitwhile.com/blog/consumer-survey-waiting-in-line-2023

Ghai, N. (2023, junio 19). Electrifying the in-store experience with smart carts. Grocery Doppio. https://www.grocerydoppio.com/articles/electrifying-the-in-store-experience-with-smart-carts

Roth, E. (2024, abril 17). Amazon insists Just Walk Out isn’t secretly run by workers watching you shop. The Verge. https://www.theverge.com/2024/4/17/24133029/amazon-just-walk-out-cashierless-ai-india

Ryan, T. (2008, julio 8). Checkout time limit around four minutes. RetailWire. https://retailwire.com/discussion/checkout-time-limit-around-four-minutes

van Riel, A., Semeijn, J., C. R, A., Ribbink, & Peters. (2012, febrero 1). Waiting for service at the checkout: Negative emotional responses, store image and overall satisfaction. Journal of Service Management, 144–169.

Wikipedia contributors. (2025a, marzo 27). Caper AI. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Caper\_AI&oldid=1282596953

Wikipedia contributors. (2025b, junio 4). Amazon Go. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Amazon\_Go&oldid=1293876391