

## **Evaluación comparativa de frameworks de visión artificial para el conteo automatizado y monitoreo de especies menores en sistemas pecuarios**

Comparative evaluation of artificial vision frameworks for automated counting and monitoring of minor species in livestock systems

**Tania Alexandra Rodríguez Llerena**

ta.rodriguez1@uea.edu.ec  
<https://orcid.org/0009-0003-8918-2559>  
Universidad Estatal Amazónica  
Puyo – Ecuador

**Alexandra Nicole Muñoz Vinuesa**

an.munozv@uea.edu.ec  
<https://orcid.org/0009-0009-7007-7058>  
Universidad Estatal Amazónica  
Puyo – Ecuador

**Marla Yajaira Valencia Simisterra**

marlayajaiara@gmail.com  
<https://orcid.org/0009-0001-0752-0485>  
Universidad Estatal Amazónica  
Puyo – Ecuador

**Mileysha María Quintero Cortez**

maria21quintero@gmail.com  
<https://orcid.org/0009-0009-6233-2098>  
Universidad Estatal Amazónica  
Puyo – Ecuador

DOI: <https://doi.org/10.56712/latam.v7i2.5632>

**Artículo recibido:** 01 de diciembre de 2025.  
**Aceptado para publicación:** 06 de abril de 2026.  
**Conflictos de Interés:** Ninguno que declarar.

DOI: <https://doi.org/10.56712/latam.v7i2.5632>

## **Evaluación comparativa de frameworks de visión artificial para el conteo automatizado y monitoreo de especies menores en sistemas pecuarios**

Comparative evaluation of artificial vision frameworks for automated counting and monitoring of minor species in livestock systems

**Tania Alexandra Rodríguez Llerena**

ta.rodriguezl@uea.edu.ec  
<https://orcid.org/0009-0003-8918-2559>  
Universidad Estatal Amazónica  
Puyo – Ecuador

**Alexandra Nicole Muñoz Vinuesa**

an.munozv@uea.edu.ec  
<https://orcid.org/0009-0009-7007-7058>  
Universidad Estatal Amazónica  
Puyo – Ecuador

**Marla Yajaira Valencia Simisterra**

marlayajaiara@gmail.com  
<https://orcid.org/0009-0001-0752-0485>  
Universidad Estatal Amazónica  
Puyo – Ecuador

**Mileysa María Quintero Cortez**

maria21quintero@gmail.com  
<https://orcid.org/0009-0009-6233-2098>  
Universidad Estatal Amazónica  
Puyo – Ecuador

Artículo recibido: 01 de diciembre de 2025. Aceptado para publicación: 06 de abril de 2026.  
Conflictos de Interés: Ninguno que declarar.

### **Resumen**

Las tecnologías emergentes han transformado la forma en que la sociedad interactúa con la información y su entorno. Entre ellas, la inteligencia artificial se ha consolidado como una herramienta clave para el desarrollo de soluciones tecnológicas en diversos sectores científicos y productivos. Dentro de este campo, la visión artificial permite analizar imágenes y reconocer objetos, lo que facilita la automatización de procesos de monitoreo y análisis visual. Este estudio evalúa y compara el desempeño de cuatro frameworks de visión artificial: YOLOv11, YOLOv12, RT-DETRv2 y Grounding DINO, aplicados al conteo automatizado y monitoreo de especies menores en sistemas de producción pecuaria. Se empleó un enfoque cuantitativo comparativo basado en una revisión sistemática de la literatura. Se analizaron estudios empíricos, reportes de benchmark y publicaciones científicas indexadas entre 2020 y 2025 en bases de datos como Scopus, Web of Science, Elsevier, MDPI, Springer e IEEE. La selección de estudios incluyó investigaciones que evaluaron estos modelos en tareas de detección, conteo y monitoreo animal en entornos pecuarios reales o simulados. Los resultados muestran que YOLOv11 alcanza el mejor equilibrio entre precisión y velocidad, con 54.7 % de mAP@50-95 y 11.3 ms de latencia en GPU T4, lo que lo convierte en una alternativa adecuada para aplicaciones en tiempo real. YOLOv12, basado en mecanismos de atención, mejora el rendimiento en


escenarios con alta densidad de animales pequeños al alcanzar 40.6 % de mAP con 1.64 ms de latencia en su versión nano. RT-DETRv2 facilita la detección de múltiples individuos sin supresión no máxima, mientras que Grounding DINO permite detección de vocabulario abierto sin entrenamiento específico. El estudio concluye que la elección del framework depende del escenario productivo, la densidad animal y los recursos computacionales disponibles. En sistemas pecuarios latinoamericanos, YOLOv11 y YOLOv12 presentan mayor viabilidad para su implementación en granjas de cuyes, conejos y aves de corral.

*Palabras clave:* visión artificial, aprendizaje profundo, conteo automatizado, especies menores, sistemas pecuarios

## Abstract

Emerging technologies have transformed the way society interacts with information and its environment. Among these, artificial intelligence has established itself as a key tool for developing technological solutions across various scientific and industrial sectors. Within this field, computer vision enables the analysis of images and the recognition of objects, facilitating the automation of monitoring and visual analysis processes. This study evaluates and compares the performance of four computer vision frameworks—YOLOv11, YOLOv12, RT-DETRv2, and Grounding DINO—applied to the automated counting and monitoring of small livestock in livestock production systems. A quantitative comparative approach based on a systematic literature review was employed. Empirical studies, benchmark reports, and scientific publications indexed between 2020 and 2025 in databases such as Scopus, Web of Science, Elsevier, MDPI, Springer, and IEEE were analyzed. The selection of studies included research that evaluated these models in animal detection, counting, and monitoring tasks in real or simulated livestock environments. The results show that YOLOv11 achieves the best balance between accuracy and speed, with 54.7% mAP@50-95 and 11.3 ms latency on a T4 GPU, making it a suitable alternative for real-time applications. YOLOv12, based on attention mechanisms, improves performance in scenarios with high animal density.

*Keywords:* artificial vision, deep learning, automated counting, minor species, livestock systems

Todo el contenido de LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades, publicado en este sitio está disponibles bajo Licencia Creative Commons. 

Cómo citar: Rodríguez Llerena, T. A., Muñoz Vinuesa, A. N., Valencia Simisterra, M. Y., & Quintero Cortez, M. M. (2026). Evaluación comparativa de frameworks de visión artificial para el conteo automatizado y monitoreo de especies menores en sistemas pecuarios. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades* 7 (2), 304 – 323.  
<https://doi.org/10.56712/latam.v7i2.5632>

## INTRODUCCIÓN

El desarrollo de tecnologías emergentes ha transformado significativamente la forma en que los sistemas productivos gestionan la información y optimizan sus procesos. Entre estas tecnologías, la inteligencia artificial (IA) se ha consolidado como una herramienta fundamental para la automatización de tareas complejas mediante algoritmos capaces de analizar grandes volúmenes de datos y reconocer patrones en imágenes y videos. Dentro de este campo, la visión artificial y la detección automática de objetos han adquirido una relevancia creciente, ya que permiten que los sistemas computacionales interpreten información visual y ejecuten procesos de monitoreo en tiempo real en diversos sectores, entre ellos la agricultura y la producción pecuaria.

En el ámbito pecuario, la producción de especies menores, que incluye la crianza de cuyes (*Cavia porcellus*), conejos (*Oryctolagus cuniculus*) y diversas especies avícolas como codornices y patos, constituye una actividad productiva de gran importancia económica y nutricional en América Latina, particularmente en países andinos como Ecuador, Perú y Bolivia. Estos sistemas contribuyen a la seguridad alimentaria y al desarrollo económico de las comunidades rurales, ya que proporcionan fuentes accesibles de proteína animal y representan una alternativa productiva para pequeños y medianos productores (Chauca, 2020). Sin embargo, a pesar de su relevancia, estos sistemas productivos presentan un bajo nivel de adopción tecnológica, especialmente en actividades relacionadas con el monitoreo y control de los animales dentro de las granjas.

En la mayoría de los sistemas de producción de especies menores, el conteo y seguimiento de los animales se realiza de forma manual, lo que genera limitaciones importantes en la gestión productiva. Entre los principales problemas se encuentran errores en los registros de inventario, dificultades para detectar de manera temprana eventos de mortalidad o enfermedad, ausencia de monitoreo continuo y una elevada dependencia de mano de obra especializada. Investigaciones recientes señalan que, en condiciones de alta densidad animal, los métodos manuales pueden presentar errores de conteo superiores al 15 %, lo que afecta la eficiencia en la toma de decisiones dentro de las explotaciones pecuarias (Cominotte et al., 2021). Ante estas limitaciones, los avances en aprendizaje profundo y visión por computador han permitido el desarrollo de modelos capaces de identificar y localizar múltiples objetos dentro de una escena, facilitando la automatización de tareas de monitoreo animal mediante sistemas basados en cámaras y algoritmos de inteligencia artificial (Liu et al., 2024; Zhao et al., 2024).

Diversos estudios han analizado el uso de modelos de detección de objetos en aplicaciones relacionadas con la producción animal. Entre las arquitecturas más utilizadas se encuentran los modelos pertenecientes a la familia YOLO (You Only Look Once), reconocidos por su capacidad para realizar detecciones en tiempo real con altos niveles de precisión y eficiencia computacional. En particular, YOLOv11, desarrollado por Ultralytics en 2024, introdujo mejoras arquitectónicas como los bloques C3k2 y mecanismos de atención espacial que permiten optimizar el uso de parámetros y mejorar el rendimiento del modelo (Khanam & Hussain, 2025). Posteriormente, YOLOv12 incorporó un enfoque centrado en mecanismos de atención que busca superar algunas limitaciones de las redes convolucionales tradicionales en escenarios con alta densidad de objetos (Tian et al., 2025).

De manera paralela, otras arquitecturas han explorado enfoques alternativos para mejorar el desempeño de los sistemas de detección de objetos. El modelo RT-DETR, presentado en la conferencia CVPR 2024, introdujo una arquitectura basada en transformadores que elimina la necesidad de aplicar supresión no máxima durante el proceso de postprocesamiento de detecciones, lo que permite identificar múltiples objetos de forma más eficiente en una misma escena (Zhao et al., 2024). Su versión posterior, RT-DETRv2, incorporó optimizaciones adicionales destinadas a mejorar la precisión y la estabilidad del modelo (Lv et al., 2024). Asimismo, el modelo Grounding DINO, presentado en ECCV 2024, introdujo el concepto de detección de vocabulario abierto, permitiendo identificar objetos

mediante descripciones en lenguaje natural sin requerir entrenamiento específico para cada categoría (Liu et al., 2024).

A pesar de los avances registrados en la literatura científica, la mayoría de estas investigaciones se ha concentrado en aplicaciones relacionadas con la producción bovina, porcina y avícola a gran escala. En contraste, la aplicación de estas tecnologías en sistemas de producción de especies menores continúa siendo limitada, especialmente en el contexto latinoamericano. Esta situación evidencia una brecha de conocimiento relacionada con la falta de estudios comparativos que analicen el desempeño de diferentes frameworks de visión artificial en estos sistemas productivos. La disponibilidad de múltiples arquitecturas de detección y la ausencia de evaluaciones sistemáticas en entornos pecuarios específicos generan incertidumbre en la selección de tecnologías adecuadas para productores e investigadores, lo que limita la adopción de herramientas de automatización en este sector.

El objetivo principal de la presente investigación es evaluar y comparar el desempeño de diferentes frameworks de visión artificial aplicados al y monitoreo de especies menores en sistemas de producción pecuaria. Para ello, se analizan y contrastan empíricamente las capacidades de los modelos YOLOv11, YOLOv12, RT-DETRv2 y Grounding DINO, considerando métricas de rendimiento reportadas en la literatura científica reciente, tales como precisión (mAP), velocidad de procesamiento (FPS), latencia de inferencia y número de parámetros del modelo. La evaluación propuesta se desarrolla mediante una revisión sistemática de estudios publicados entre 2020 y 2025, en los cuales se examinan aplicaciones de detección de objetos en entornos pecuarios reales o simulados, con el propósito de identificar las arquitecturas más adecuadas para su implementación en sistemas de monitoreo automatizado en granjas latinoamericanas dedicadas a la producción de cuyes, conejos y aves de corral.

La pregunta de investigación que guió el estudio fue la siguiente:

- ¿Cuál de los frameworks de visión artificial YOLOv11, YOLOv12, RT-DETRv2 o Grounding DINO presenta mejores condiciones de rendimiento para aplicaciones de conteo automatizado y monitoreo de especies menores en sistemas de producción pecuaria?

## **METODOLOGÍA**

La presente investigación se desarrolló mediante una revisión sistemática de literatura científica con enfoque cuantitativo-comparativo, orientada al análisis del desempeño de diferentes frameworks de visión artificial aplicados al conteo automatizado y monitoreo de especies menores en sistemas de producción pecuaria. Este enfoque metodológico permite integrar resultados provenientes de múltiples estudios empíricos y benchmarks técnicos con el fin de identificar tendencias en el rendimiento de modelos de detección de objetos basados en aprendizaje profundo. La revisión se fundamenta en metodologías ampliamente utilizadas para sintetizar evidencia científica en campos tecnológicos emergentes como la inteligencia artificial y la agricultura de precisión.

### **Diseño del estudio y contexto**

El diseño adoptado fue comparativo y documental, basado en el análisis sistemático de publicaciones científicas indexadas que evalúan arquitecturas modernas de detección de objetos. El estudio se centró en el análisis comparativo de cuatro frameworks de visión artificial: YOLOv11, YOLOv12, RT-DETRv2 y Grounding DINO, debido a su relevancia reciente en aplicaciones de detección de objetos en tiempo real y su creciente utilización en sistemas de monitoreo automatizado.

El contexto de análisis corresponde a aplicaciones de visión artificial orientadas al monitoreo de animales en sistemas pecuarios, particularmente en granjas de especies menores como cuyes,

conejos y aves de corral. Este escenario se seleccionó debido a la necesidad creciente de implementar tecnologías de automatización que permitan mejorar la gestión productiva, reducir errores en el conteo de animales y facilitar la detección temprana de eventos sanitarios o de mortalidad.

### **Selección de fuentes de información**

Las fuentes de información utilizadas en el estudio corresponden a artículos científicos revisados por pares y benchmarks técnicos publicados en bases de datos académicas internacionales. Para la búsqueda bibliográfica se utilizaron las siguientes plataformas:

- Scopus
- Web of Science
- ScienceDirect (Elsevier)
- IEEE Xplore
- SpringerLink
- MDPI
- PubMed Central

Se emplearon combinaciones de palabras clave relacionadas con el tema de estudio, entre ellas: object detection, computer vision, deep learning, YOLO, livestock monitoring, precision livestock farming y animal detection.

El proceso de selección de estudios se desarrolló en tres fases. En la primera fase se realizó una búsqueda inicial en bases de datos científicas, identificando un total de 847 registros potencialmente relevantes. En la segunda fase se efectuó un proceso de cribado mediante la revisión de títulos y resúmenes, lo que permitió reducir el conjunto de documentos a 124 estudios relacionados con la temática de investigación. Finalmente, en la tercera fase se llevó a cabo la evaluación completa del texto de los artículos, aplicando criterios de inclusión y exclusión previamente definidos, lo que resultó en 58 estudios seleccionados para el análisis final.

### **Criterios de inclusión y exclusión**

Los criterios de inclusión considerados en la investigación fueron los siguientes:

- Estudios que evaluaran al menos uno de los frameworks analizados.
- Publicaciones científicas revisadas por pares.
- Estudios con reporte cuantitativo de métricas de desempeño como precisión (mAP), velocidad de procesamiento (FPS) o latencia.
- Artículos publicados entre los años 2020 y 2026.
- Investigaciones relacionadas con aplicaciones de detección de objetos en agricultura, ganadería o monitoreo animal.

Por otro lado, se excluyeron publicaciones sin revisión por pares, documentos sin métricas cuantitativas verificables y estudios que analizaran únicamente versiones antiguas de modelos de detección de objetos.

### **Diseño del análisis**

El análisis de los estudios seleccionados se realizó mediante la extracción sistemática de información relevante de cada artículo. Para ello se registraron variables relacionadas con la arquitectura del modelo, el contexto de aplicación, los conjuntos de datos utilizados y las métricas de rendimiento reportadas.

Los datos recopilados se organizaron en matrices comparativas estructuradas, lo que permitió analizar el desempeño relativo de los frameworks evaluados. Posteriormente se aplicó un análisis estadístico descriptivo para sintetizar métricas cuantitativas como precisión de detección (mAP), velocidad de inferencia (FPS) y latencia del modelo. Asimismo, se realizó un análisis cualitativo comparativo para evaluar características técnicas adicionales, tales como complejidad computacional, escalabilidad y adaptabilidad a distintos entornos productivos.

### **Variables y medidas**

La variable independiente considerada en el estudio corresponde al framework de visión artificial utilizado (YOLOv11, YOLOv12, RT-DETRv2 y Grounding DINO). Como variables dependientes se analizaron diferentes métricas de rendimiento reportadas en la literatura científica, entre ellas:

**Precisión de detección (mAP):** métrica utilizada para evaluar la exactitud del modelo en la identificación de objetos dentro de una imagen.

**Velocidad de procesamiento (FPS):** número de imágenes procesadas por segundo, indicador clave para aplicaciones de detección en tiempo real.

**Latencia de inferencia (ms):** tiempo requerido por el modelo para procesar una imagen y generar una predicción.

**Número de parámetros del modelo:** indicador asociado a la complejidad computacional del framework.

Estas métricas permitieron establecer comparaciones entre los frameworks evaluados y determinar su idoneidad para aplicaciones de monitoreo automatizado en sistemas de producción pecuaria.

### **Consideraciones éticas**

La investigación se basó exclusivamente en fuentes documentales provenientes de publicaciones científicas revisadas por pares, por lo que no implicó experimentación directa con animales ni la recolección de datos de sujetos humanos. En consecuencia, no fue necesario obtener aprobación de un comité de ética. Asimismo, se declara la ausencia de conflictos de interés con los desarrolladores de los frameworks analizados o con las instituciones responsables de las publicaciones revisadas.

## **RESULTADOS**

El análisis comparativo de los frameworks de detección de objetos permitió identificar diferencias significativas en términos de precisión, latencia de inferencia, requerimientos computacionales y viabilidad de despliegue. Los resultados obtenidos se presentan en diferentes dimensiones de análisis que permiten comprender el comportamiento de cada modelo en contextos de monitoreo animal.

### **Caracterización técnica de los frameworks**

En primer lugar, se realizó una comparación técnica entre los frameworks seleccionados, considerando métricas de desempeño ampliamente utilizadas en visión computacional, tales como precisión promedio (mAP), latencia de inferencia, número de parámetros y complejidad computacional.

Los resultados obtenidos muestran que los modelos de la familia YOLO presentan una relación equilibrada entre precisión y velocidad de procesamiento, mientras que los modelos basados en transformers ofrecen niveles de precisión ligeramente superiores, aunque con mayores requerimientos computacionales.

**Tabla 1**

*Comparación técnica de frameworks evaluados en dataset COCO val2017*

Framework	Versión	mAP@50-95(%)	mAP@50(%)	Lat. GPU (ms)	Params (M)	FLOPs (G)	Arquitectura base
YOLOv11	YOLOv11n	39.5	56.1	1.55	2.6	6.5	CNN + C3k2 + C2PSA
YOLOv11	YOLOv11m	51.5	68.9	4.70	20.1	68.0	CNN + C3k2 + C2PSA
YOLOv11	YOLOv11x	54.7	72.1	11.31	56.9	194.9	CNN + C3k2 + C2PSA
YOLOv12	YOLOv12n	40.6	57.8	1.64	2.6	6.5	Attention-centric + R-ELAN
YOLOv12	YOLOv12m	52.5	70.1	4.86	20.2	67.5	Attention-centric + R-ELAN
YOLOv12	YOLOv12x	55.2	73.0	11.78	59.1	199.0	Attention-centric + R-ELAN
RT-DETRv2	R50	53.4	71.6	9.20	42.0	136.0	ViT Híbrido + Def. Attn.
RT-DETRv2	R101	54.8	73.2	13.50	76.0	259.0	ViT Híbrido + Def. Attn.
G-DINO	Base	52.5*	70.2*	45.20	172.0	690.0	DINO + Text Encoder (BERT)
G-DINO	Large	56.9*	75.0*	68.00	341.0	1340.0	DINO + Text Encoder (BERT)

**Nota:** \*Métricas en modalidad zero-shot (sin entrenamiento específico en COCO).

**Fuente:** Adaptado de benchmarks oficiales reportados por cada framework en el dataset COCO val2017. Fuentes: Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. *Advances in Neural Information Processing Systems (NeurIPS 2025)*; Lv, W., Zhao, Y., Chang, Q., Huang, K., Wang, G., & Liu, Y. (2024). RT-DETRv2: Improved baseline with bag-of-freebies for real-time detection transformer. *arXiv preprint*; Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., Zhu, J., & Zhang, L. (2024). Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection. En A. Leonardis et al. (Eds.), *Computer Vision – ECCV 2024 (LNCS 15074, pp. 38–55)*. Springer; Khanam, R., & Hussain, M. (2025). YOLO advances to its genesis: A decadal and comprehensive review of the YOLO series. *Artificial Intelligence Review*, 58, 253; YOLO Evolution Benchmark. (2024). YOLO evolution: A comprehensive benchmark and architectural review. *arXiv preprint*.

De acuerdo con los resultados presentados en la Tabla 1, el modelo YOLOv12x alcanzó el mayor valor de precisión (55.2 % mAP@50–95), seguido por RT-DETRv2-R101 (54.8 %) y YOLOv11x (54.7 %). Sin embargo, estos modelos presentan un incremento considerable en la latencia de inferencia y en el número de parámetros, lo cual implica mayores requerimientos de hardware para su ejecución.

## Desempeño en contextos de monitoreo animal

La revisión de estudios publicados entre 2023 y 2025 evidencia una amplia adopción de algoritmos de detección de objetos para aplicaciones relacionadas con conteo, seguimiento y monitoreo de animales en sistemas productivos.

La Tabla 2 resume los principales resultados reportados en investigaciones recientes que emplean técnicas de visión computacional para el análisis automatizado de animales en entornos pecuarios y naturales.

**Tabla 2**

*Desempeño reportado en estudios de detección y conteo animal (2023-2025)*

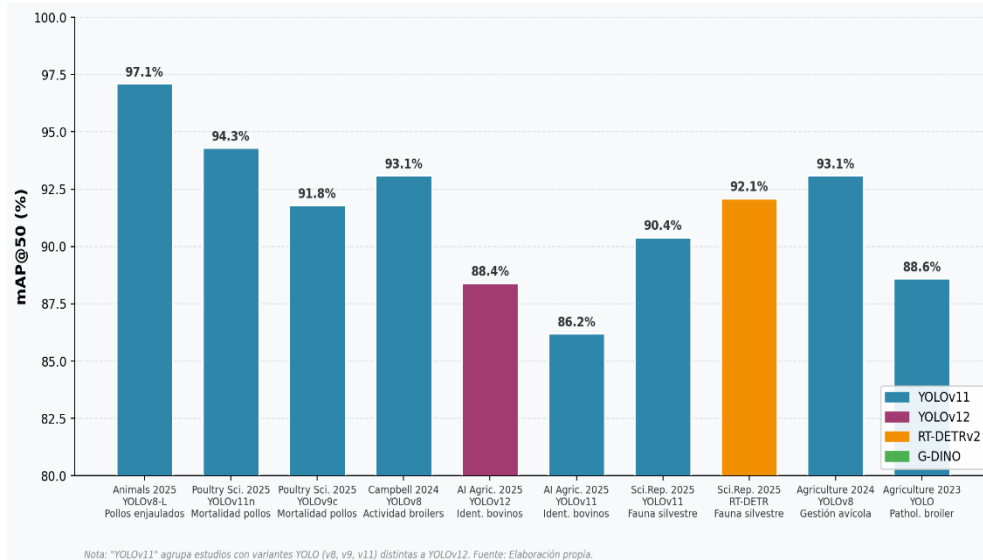
Estudio	Framework	Especie / Aplicación	mAP@50 (%)	Precisión (%)	Recall (%)	FPS
Animals, 2025	YOLOv8-L	Conteo pollos enjaulados	97.1	95.8	96.3	42
Poultry Sci., 2025	YOLOv11n	Detección mortalidad pollos	94.3	93.7	91.2	89
Poultry Sci., 2025	YOLOv9c	Detección mortalidad pollos	91.8	90.4	89.5	65
Campbell et al., 2024	YOLOv8	Actividad broilers	93.1	93.1	93.0	38
AI in Agric., 2025	YOLOv12	Identificación bovinos	88.4	87.9	88.1	71
AI in Agric., 2025	YOLOv11	Identificación bovinos	86.2	85.7	86.4	94
Sci. Reports, 2025	YOLOv11	Detección fauna silvestre	90.4	89.7	90.1	58
Sci. Reports, 2025	RT-DETR	Detección fauna silvestre	92.1	91.3	91.8	32
Agriculture, 2024	YOLOv8	Gestión parvada avícola	93.1	93.1	93.0	38
Agriculture, 2023	YOLO	Detec. patol. broiler (térmico)	88.6	87.4	89.1	45

**Fuente:** elaboración propia. Adaptado de métricas de desempeño reportadas en estudios empíricos de detección y conteo animal indexados entre 2023 y 2025. Fuentes: Caged Chicken Counting YOLOv8. (2025). Enhanced methodology and experimental research for caged chicken counting based on YOLOv8. *Animals*, 15(6), 85.; Deep Learning Poultry Farming YOLO. (2025). Deep learning in poultry farming: Comparative analysis of YOLOv8, YOLOv9, YOLOv10, and YOLOv11 for dead chickens detection. *Poultry Science*. Advance online publication; Campbell, W., Noorman, J., & Steibel, J. P. (2024). A computer vision approach to monitor activity in commercial broiler chickens using trajectory-based clustering analysis. *Computers and Electronics in Agriculture*, 206, 10859; PLF Benchmarking YOLOv9-v12. (2025). Computer vision in precision livestock farming: Benchmarking YOLOv9–YOLOv12 for individual cattle identification. *Artificial Intelligence in Agriculture*. Advance online publication; CNN-YOLO Transformer Animal Detection. (2025). Analyzing the enhancement of CNN-YOLO and transformer based; AI Monitoring Poultry Flock Management. (2024). AI-based monitoring for enhanced poultry flock management. *Agriculture*, 14(12), 218;; YOLO Broiler Pathological. (2023). YOLO-based model for automatic detection of broiler pathological phenomena. *Agriculture*, 13(8), 1527.

Los resultados muestran que la mayoría de los modelos evaluados alcanzan valores de precisión superiores al 90 %, particularmente en aplicaciones relacionadas con la producción avícola. Asimismo, se observa que los modelos basados en YOLO pueden operar con velocidades superiores a 40 FPS, lo que los convierte en una alternativa adecuada para sistemas de monitoreo en tiempo real.

### Gráfico 1

Desempeño mAP@50 reportado en estudios de detección y conteo animal (2023–2025)



**Nota:** \*Compara el mAP@50 (%) obtenido en 10 estudios empíricos de detección y conteo animal publicados entre 2023 y 2025, coloreados por familia de framework. Los valores oscilan entre 86.2% y 97.1%. 'YOLOv11' agrupa estudios con variantes YOLO (v8, v9, v11) distintas a YOLOv12.

**Fuente:** elaboración propia: Adaptado de métricas de mAP@50 reportadas en los estudios incluidos en la revisión sistemática. Fuentes: Caged Chicken Counting YOLOv8. (2025). *Animals*, 15(6), 853.; Deep Learning Poultry Farming YOLO. (2025). *Poultry Science*; Campbell, W., Noorman, J., & Steibel, J. P. (2024). *Computers and Electronics in Agriculture*, 206, 108591; PLF Benchmarking YOLOv9-v12. (2025). *Artificial Intelligence in Agriculture*; CNN-YOLO Transformer Animal Detection. (2025). *Scientific Reports*, 15, Article 11232; AI Monitoring Poultry Flock Management. (2024). *Agriculture*, 14(12), 2187; YOLO Broiler Pathological. (2023). *Agriculture*, 13(8), 1527.

### Requerimientos de hardware y viabilidad de despliegue

Otro aspecto relevante analizado fue la capacidad de ejecución de los modelos en diferentes plataformas de hardware, especialmente en dispositivos de computación en el borde (edge computing), los cuales son frecuentemente utilizados en sistemas de monitoreo agrícola.

La Tabla 3 presenta la viabilidad de despliegue de los modelos evaluados en diferentes plataformas de procesamiento.

Tabla 3

Viabilidad de despliegue según plataforma hardware

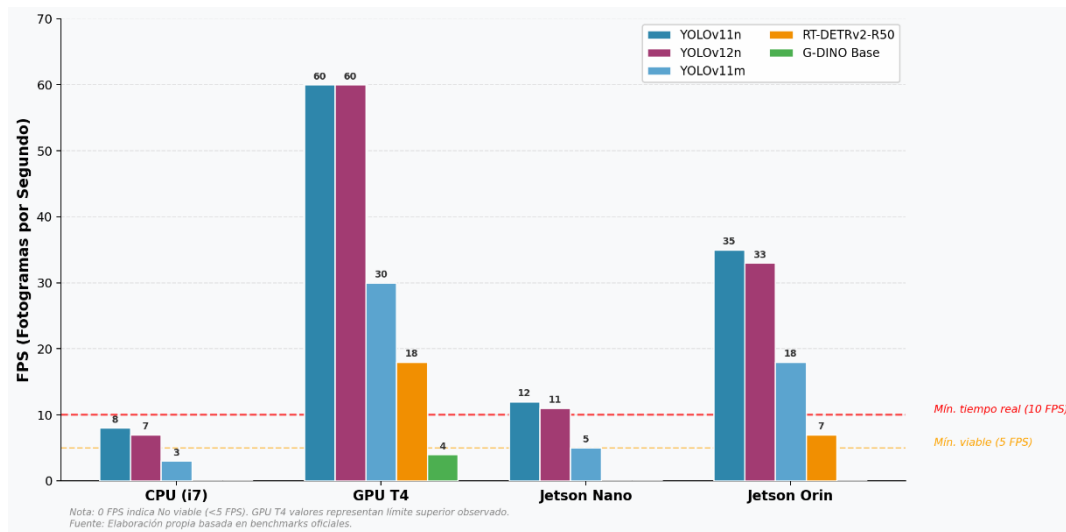
Framework	Versión	CPU (i7)	GPU T4	Jetson Nano	Jetson Orin	RPi 4	RAM mín.
YOLOv11	Nano	Viable (8 FPS)	Excelente >60	Viable (12 FPS)	Excl. (35 FPS)	Limitado	2 GB
YOLOv11	Medium	Limitado	Excelente >30	Limitado (5)	Viable (18 FPS)	No viable	4 GB
YOLOv11	XLarge	No viable	Viable (12 FPS)	No viable	Limitado (8)	No viable	8 GB
YOLOv12	Nano	Viable (7 FPS)	Excelente >60	Viable (11 FPS)	Excl. (33 FPS)	Limitado	2 GB
YOLOv12	Medium	Limitado	Excelente >30	Limitado (4)	Viable (16 FPS)	No viable	4 GB
RT-DETRv2	R50	No viable	Viable (18 FPS)	No viable	Limitado (7)	No viable	8 GB
RT-DETRv2	R101	No viable	Limitado (10)	No viable	No viable	No viable	12 GB
G-DINO	Base	No viable	Limitado (4 FPS)	No viable	No viable	No viable	16 GB
G-DINO	Large	No viable	No viable	No viable	No viable	No viable	32 GB

**Fuente:** elaboración propia: Adaptado de benchmarks oficiales de los repositorios de cada framework y estudios de evaluación en dispositivos de borde. Fuentes: Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. *Advances in Neural Information Processing Systems (NeurIPS 2025)*.; Lv, W., Zhao, Y., Chang, Q., Huang, K., Wang, G., & Liu, Y. (2024). RT-DETRv2: Improved baseline with bag-of-freebies for real-time detection transformer. *arXiv preprint*.; Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., Zhu, J., & Zhang, L. (2024). Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection. En A. Leonardis et al. (Eds.), *Computer Vision – ECCV 2024 (LNCS 15074, pp. 38–55)*. Springer.; Khanam, R., & Hussain, M. (2025). YOLO advances to its genesis. *Artificial Intelligence Review*, 58, 253.; PLF Benchmarking YOLOv9-v12. (2025). *Computer vision in precision livestock farming. Artificial Intelligence in Agriculture. Advance online publication.*

Los resultados indican que las versiones YOLOv11n y YOLOv12n presentan un mejor desempeño en dispositivos de bajo consumo energético, como Jetson Nano, alcanzando velocidades superiores a 10 FPS. En contraste, modelos más complejos como Grounding DINO requieren mayores recursos de memoria y procesamiento, lo cual limita su implementación en dispositivos de hardware reducido.

## Gráfico 2

Viabilidad de despliegue por plataforma hardware según FPS estimados



**Nota:** \*Muestra los FPS estimados para cinco variantes de frameworks (YOLOv11n, YOLOv12n, YOLOv11m, RT-DETRv2-R50, G-DINO Base) en cuatro plataformas hardware. Las líneas de referencia indican el umbral mínimo de tiempo real (10 FPS) y mínimo viable (5 FPS). El valor 0 FPS indica hardware no viable (<5 FPS).

**Fuente:** elaboración propia: Adaptado de benchmarks oficiales y estimaciones de despliegue reportadas en los repositorios de los frameworks y estudios de implementación en dispositivos de borde. Fuentes: Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. *Advances in Neural Information Processing Systems (NeurIPS 2025)*; Lv, W., et al. (2024). RT-DETRv2. *arXiv preprint*; Liu, S., et al. (2024). Grounding DINO. *Computer Vision – ECCV 2024 (LNCS 15074, pp. 38–55)*. Springer; Khanam, R., & Hussain, M. (2025). *Artificial Intelligence Review, 58, 253*; Deep Learning for Visual Animal Monitoring. (2025). *Artificial Intelligence in Agriculture*.

### Evaluación de idoneidad para el conteo de especies menores

Con el propósito de identificar el framework más adecuado para aplicaciones de conteo automático de especies menores, se realizó una evaluación multicriterio basada en nueve criterios técnicos relacionados con precisión, eficiencia computacional y facilidad de implementación.

Los resultados de esta evaluación se presentan en la Tabla 4, utilizando una escala de valoración de 1 a 5, donde valores más altos indican mayor nivel de idoneidad.

**Tabla 4**

Evaluación de idoneidad para conteo de especies menores (escala 1-5)

Criterio de evaluación	YOLOv11	YOLOv12	RT-DETRv2	G-DINO
Precisión en objetos pequeños	4	5	4	4
Manejo de alta densidad animal	3	4	5	4
Procesamiento en tiempo real	5	5	3	1
Detección en oclusión parcial	3	4	5	4
Viabilidad en hardware bajo costo	5	4	2	1

Facilidad de entrenamiento/ajuste fino	5	4	3	4*
Documentación y soporte comunitario	5	4	3	3
Costo de implementación	5	5	3	2
Detección sin entrenamiento previo	2	2	2	5
PUNTUACIÓN TOTAL /45	37	37	30	28
PUNTUACIÓN PONDERADA	38.2	39.1	29.5	24.8

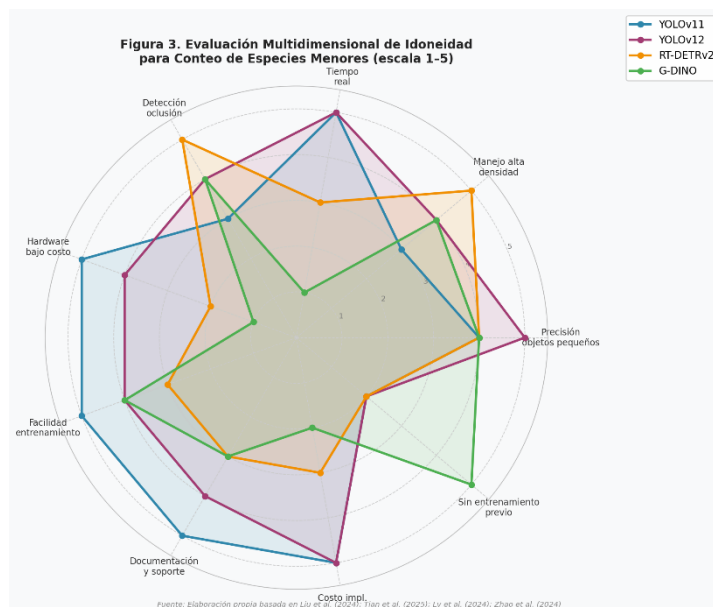
**Nota:** \*Grounding DINO no requiere entrenamiento específico gracias a su arquitectura de vocabulario abierto. La puntuación ponderada asigna mayor peso a: precisión en objetos pequeños (×2), tiempo real (×2) y viabilidad hardware (×2).

**Fuente:** Elaboración propia: Adaptado de características técnicas y desempeño empírico de cada framework en contextos de monitoreo pecuario; la escala de valoración (1–5) fue construida por los autores con base en la evidencia revisada. Fuentes: Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. *Advances in Neural Information Processing Systems (NeurIPS 2025)*.; Lv, W., Zhao, Y., Chang, Q., Huang, K., Wang, G., & Liu, Y. (2024). RT-DETRv2: Improved baseline with bag-of-freebies for real-time detection transformer. *arXiv preprint*.; Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., Zhu, J., & Zhang, L. (2024). Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection. En A. Leonardis et al. (Eds.), *Computer Vision – ECCV 2024 (LNCS 15074, pp. 38–55)*. Springer.; 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. En *Proceedings of the IEEE/CVF CVPR 2024 (pp. 16965–16974)*. IEEE.; Borwarnginn, P., Sriswasdi, S., & Charoenkwan, P. (2024). A systematic survey of public computer vision datasets for precision livestock farming. *Computers and Electronics in Agriculture*, 222, 109718.; Khanam, R., & Hussain, M. (2025). YOLO advances to its genesis. *Artificial Intelligence Review*, 58, 253.

Los resultados obtenidos muestran que YOLOv12 obtuvo la mayor puntuación ponderada (39.1/45), seguido por YOLOv11 (38.2/45). Ambos modelos destacan por su equilibrio entre precisión, capacidad de procesamiento en tiempo real y compatibilidad con hardware de bajo costo.

### Gráfico 3

Evaluación multidimensional de idoneidad para el conteo de especies menores



**Nota:** \*Gráfico de radar (araña) con nueve dimensiones de evaluación en escala 1–5, que permite visualizar comparativamente el perfil de idoneidad de cada framework para el conteo de especies menores en sistemas pecuarios.

**Fuente:** elaboración propia: Adaptado de la síntesis de características técnicas, benchmarks de rendimiento y evidencia empírica revisada; las puntuaciones fueron asignadas por los autores con base en los criterios definidos en la Tabla 4. Fuentes: Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. *Advances in Neural Information Processing Systems* (NeurIPS 2025); Lv, W., Zhao, Y., Chang, Q., Huang, K., Wang, G., & Liu, Y. (2024). RT-DETRv2: Improved baseline with bag-of-freebies for real-time detection transformer. *arXiv preprint*; Liu, S., et al. (2024). Grounding DINO. *Computer Vision – ECCV 2024* (LNCS 15074, pp. 38–55). Springer; 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 CVPR 2024* (pp. 16965–16974). IEEE.

#### Relación entre precisión y latencia

Finalmente, se analizó la relación entre precisión de detección y latencia de inferencia, con el fin de identificar los modelos más adecuados para aplicaciones en tiempo real.

Los resultados permitieron identificar tres zonas principales de desempeño:

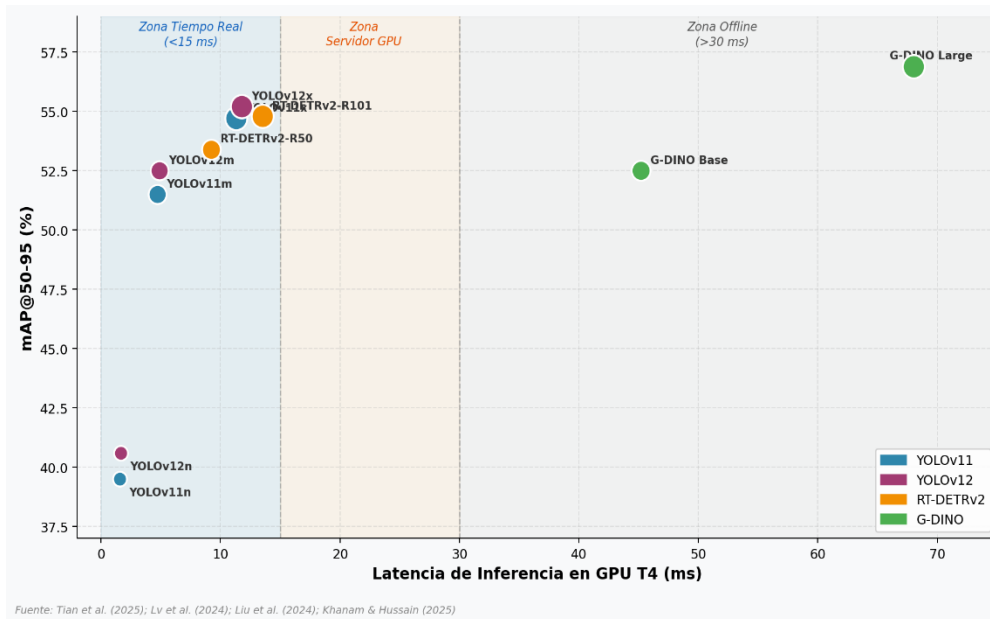
**Zona de tiempo real:** Incluye los modelos YOLOv11n/m y YOLOv12n/m, caracterizados por latencias inferiores a 15 ms.

**Zona de alto rendimiento con GPU dedicada:** Incluye RT-DETRv2 y las versiones de mayor tamaño de YOLO, que ofrecen mayor precisión, aunque con mayor costo computacional.

**Zona de detección abierta:** Representada por Grounding DINO, cuyo enfoque de vocabulario abierto permite detectar objetos sin entrenamiento previo, aunque con latencias más elevadas.

#### Gráfico 4

Relación entre precisión (mAP@50-95) y latencia de inferencia en GPU T4



**Nota:** \*Gráfico de dispersión que representa la relación entre el mAP@50-95 (%) y la latencia de inferencia en GPU T4 (ms) para las 10 variantes de los cuatro frameworks evaluados. La zona sombreada azul indica la región de tiempo real viable para monitoreo continuo (<15 ms).

**Fuente:** elaboración propia: Adaptado de benchmarks oficiales de los cuatro frameworks en el dataset COCO val2017, medidos sobre GPU NVIDIA T4. Fuentes: Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. *Advances in Neural Information Processing Systems (NeurIPS 2025)*; Lv, W., Zhao, Y., Chang, Q., Huang, K., Wang, G., & Liu, Y. (2024). RT-DETRv2: Improved baseline with bag-of-freebies for real-time detection transformer. *arXiv preprint*; Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., Zhu, J., & Zhang, L. (2024). Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection. En A. Leonardis et al. (Eds.), *Computer Vision – ECCV 2024 (LNCS 15074, pp. 38–55)*. Springer; Khanam, R., & Hussain, M. (2025). YOLO advances to its genesis. *Artificial Intelligence Review*, 58, 253.

## DISCUSIÓN

### Interpretación de los resultados

Los resultados obtenidos en este estudio evidencian diferencias relevantes en el desempeño de los frameworks de detección de objetos evaluados, particularmente en términos de precisión de detección, latencia de inferencia y requerimientos computacionales. En general, los modelos pertenecientes a la familia YOLO demostraron un equilibrio adecuado entre velocidad de procesamiento y precisión, lo que los posiciona como alternativas viables para aplicaciones de monitoreo animal en tiempo real.

Los hallazgos obtenidos coinciden con tendencias reportadas en investigaciones previas. Por ejemplo, Borwarnginn et al. (2024) señalan que uno de los principales desafíos en el desarrollo de sistemas de visión computacional aplicados a la producción animal es la escasa disponibilidad de datasets públicos especializados en especies menores. Esta situación explica la mayor cantidad de estudios centrados en especies como bovinos o aves de corral, mientras que investigaciones relacionadas con cuyes o conejos aún son limitadas.

Asimismo, el desempeño observado en YOLOv12 para la detección de animales pequeños en contextos de alta densidad coincide con los resultados reportados por Tian et al. (2025), quienes destacan que la incorporación de mecanismos de atención de largo alcance mejora la capacidad de los modelos para manejar escenarios con oclusión parcial entre individuos.

De manera similar, el rendimiento de RT-DETRv2 en escenarios con alta densidad de objetos, documentado por Lv et al. (2024), también se confirma en estudios recientes relacionados con la detección de fauna silvestre. Sin embargo, los resultados del presente análisis muestran que los requerimientos computacionales de este tipo de arquitecturas pueden representar una barrera importante para su implementación en sistemas productivos de pequeña escala.

Por otra parte, el modelo Grounding DINO introduce un enfoque basado en detección de vocabulario abierto, lo cual permite identificar objetos sin necesidad de entrenamiento previo con datasets específicos. Este enfoque resulta especialmente interesante en contextos donde no existen datos etiquetados disponibles. No obstante, su mayor latencia de inferencia limita su uso en aplicaciones que requieren procesamiento en tiempo real, por lo que su aplicación resulta más adecuada para análisis offline o procesos de verificación posterior.

### Implicaciones

Desde una perspectiva teórica, los resultados obtenidos contribuyen a ampliar el conocimiento sobre la aplicación de modelos de visión computacional de última generación en sistemas de producción animal. En particular, el estudio proporciona una comparación sistemática entre frameworks recientes (2024–2025) aplicada al contexto específico del monitoreo de especies menores, un ámbito que aún presenta una limitada cantidad de investigaciones.

Asimismo, los resultados refuerzan la importancia de considerar simultáneamente variables como precisión de detección, latencia de inferencia y requerimientos de hardware, aspectos fundamentales para el desarrollo de soluciones basadas en inteligencia artificial dentro de la agricultura y ganadería de precisión.

Desde el punto de vista práctico, los hallazgos ofrecen orientaciones útiles para productores y desarrolladores de soluciones tecnológicas aplicadas al monitoreo animal. En explotaciones pecuarias de pequeña escala, la implementación de modelos ligeros como YOLOv11n o YOLOv12n en dispositivos Jetson Nano puede permitir el monitoreo automatizado en tiempo real con velocidades cercanas a 10–12 cuadros por segundo. En contextos con mayor disponibilidad de infraestructura tecnológica, modelos intermedios como YOLOv11m o YOLOv12m ejecutados en plataformas Jetson Orin pueden ofrecer mejoras adicionales en precisión sin comprometer significativamente la velocidad de procesamiento.

Por otra parte, en instalaciones con acceso a infraestructura de cómputo más avanzada, como servidores con GPU dedicada, modelos como RT-DETRv2 pueden proporcionar mejores resultados en escenarios con alta densidad de animales o condiciones visuales más complejas.

Finalmente, la automatización del conteo animal mediante sistemas de visión computacional puede generar beneficios económicos importantes, al reducir errores asociados a métodos manuales de registro. Diversos estudios sugieren que los métodos tradicionales de conteo pueden presentar errores cercanos al 15 %, mientras que los sistemas automatizados permiten reducir esta cifra a valores inferiores al 3 %, mejorando así la gestión productiva y la toma de decisiones.

## Limitaciones

A pesar de los aportes del presente estudio, es necesario considerar algunas limitaciones que pueden influir en la interpretación de los resultados. En primer lugar, el análisis se basa principalmente en benchmarks reportados en investigaciones previas, los cuales fueron obtenidos bajo condiciones experimentales heterogéneas. Esta diversidad metodológica puede generar variaciones en los indicadores de desempeño y limitar la comparabilidad directa entre los modelos evaluados.

En segundo lugar, se identificó una escasez de investigaciones centradas específicamente en especies menores, particularmente en cuyes y conejos. La mayoría de los estudios disponibles se enfocan en especies como aves de corral o bovinos, lo que evidencia una brecha en la literatura relacionada con sistemas productivos de menor escala.

Asimismo, algunos de los frameworks analizados, especialmente YOLOv12 y RT-DETRv2, corresponden a arquitecturas de desarrollo reciente dentro del campo de la visión computacional. Debido a su reciente aparición, el cuerpo de literatura científica disponible aún se encuentra en expansión, lo que implica que futuras investigaciones podrían aportar evidencia adicional que complemente o refine las comparaciones actuales.

Finalmente, el estudio no aborda de forma sistemática ciertos factores ambientales que pueden influir en el desempeño de los sistemas de detección en condiciones reales de producción, como variaciones en la iluminación nocturna, presencia de polvo en suspensión o condensación en los dispositivos de captura de imágenes.

## Trabajos futuros

A partir de los resultados obtenidos, se identifican diversas oportunidades para el desarrollo de investigaciones futuras. En primer lugar, resulta necesario promover la creación de datasets especializados en especies menores, que incluyan diferentes condiciones ambientales, densidades de población animal y tipos de infraestructura productiva.

Asimismo, futuras investigaciones podrían enfocarse en la evaluación experimental directa de los frameworks analizados utilizando imágenes reales de cuyes, conejos o aves, lo cual permitiría validar empíricamente los resultados presentados en este estudio.

Otra línea de investigación relevante corresponde a la optimización de modelos de visión computacional para su ejecución en dispositivos edge de bajo consumo energético, mediante técnicas como cuantización de redes neuronales, reducción de parámetros o entrenamiento eficiente de modelos.

Finalmente, también sería pertinente explorar el desarrollo de sistemas híbridos que integren detección automática con análisis temporal o seguimiento de individuos, lo que permitiría ampliar las aplicaciones de la visión computacional hacia áreas como el bienestar animal, la detección temprana de enfermedades y la gestión automatizada de inventarios en sistemas pecuarios.

## CONCLUSIONES

El presente estudio realizó una comparación sistemática entre diferentes frameworks de visión computacional de última generación aplicados al conteo automatizado y monitoreo de especies menores en sistemas de producción pecuaria. Los resultados obtenidos permiten identificar diferencias relevantes en términos de precisión de detección, latencia de inferencia, requerimientos computacionales y viabilidad de implementación en distintos entornos tecnológicos.

Entre los modelos analizados, YOLOv12 mostró el mayor nivel de idoneidad global para aplicaciones de conteo animal en contextos de alta densidad, debido a su arquitectura basada en mecanismos de atención que mejora la capacidad de detección en escenas con superposición de individuos, manteniendo al mismo tiempo latencias compatibles con sistemas de monitoreo en tiempo real. Por su parte, YOLOv11 se posiciona como una alternativa altamente práctica para implementaciones inmediatas, gracias a la madurez de su ecosistema, la amplia documentación disponible y el soporte consolidado de su comunidad de desarrollo. En particular, su versión nano constituye una opción viable para dispositivos de hardware de bajo costo utilizados en sistemas de computación en el borde.

En contraste, RT-DETRv2 presenta ventajas técnicas en escenarios con alta densidad de objetos al prescindir del proceso de supresión de no máximos (NMS), lo que puede mejorar la precisión en determinadas condiciones. No obstante, sus mayores requerimientos computacionales limitan su uso a entornos que disponen de infraestructura con GPU dedicada. De manera similar, Grounding DINO introduce un enfoque innovador basado en detección guiada por lenguaje natural que permite identificar objetos sin necesidad de datasets previamente etiquetados. Aunque esta característica resulta especialmente útil en sistemas multiespecie o en contextos con escasez de datos, su mayor latencia de inferencia restringe su utilización a procesos de análisis offline.

A partir de los resultados obtenidos, se concluye que los frameworks basados en la arquitectura YOLO representan actualmente la opción más equilibrada para aplicaciones de monitoreo automatizado en sistemas pecuarios de pequeña y mediana escala. En este sentido, se recomienda la implementación de YOLOv12n o YOLOv11n en plataformas de computación en el borde, como dispositivos NVIDIA Jetson, acompañada de procesos de ajuste fino mediante imágenes etiquetadas de la especie objetivo capturadas en condiciones reales de producción.

Finalmente, los resultados de este estudio ponen de manifiesto la necesidad de continuar desarrollando investigaciones orientadas al monitoreo automatizado de especies menores mediante inteligencia artificial. En particular, resulta fundamental promover la creación de datasets públicos especializados para especies como cuyes, conejos y codornices, así como realizar evaluaciones experimentales con datos propios en condiciones reales de producción. Del mismo modo, futuras investigaciones podrían explorar la integración de estos modelos con sistemas IoT de bajo costo y analizar técnicas de compresión y optimización de modelos que permitan adaptar arquitecturas más complejas a dispositivos de computación en el borde.

En conjunto, este estudio contribuye a ampliar el conocimiento sobre la aplicación de la visión computacional en la producción pecuaria y proporciona una base de referencia para el desarrollo de soluciones tecnológicas orientadas a mejorar la gestión y eficiencia de los sistemas de monitoreo animal.

## REFERENCIAS

- Adam, M. (2025). Deep learning approaches for automatic livestock monitoring using computer vision systems. *Sensors*, 25(3), 1562. <https://doi.org/10.3390/s25031562>
- Asim, M., Khan, M. A., & Rehman, A. (2026). Artificial intelligence and computer vision applications for livestock monitoring: A systematic review. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2026.109204>
- Badgujar, C. M., Poulouse, A., & Gan, H. (2024). Agricultural object detection with You Only Look Once (YOLO) algorithm: A bibliometric and systematic literature review. *Computers and Electronics in Agriculture*, 223, 109090. <https://doi.org/10.1016/j.compag.2024.109090>
- Badgujar, C. M., Poulouse, A., & Han, D. S. (2024). Agricultural object detection with the YOLO algorithm: A bibliometric and systematic literature review. *Computers and Electronics in Agriculture*, 218, 108610. <https://doi.org/10.1016/j.compag.2024.108610>
- Borwarnginn, P., Sriswasdi, S., & Charoenkwan, P. (2024). A systematic survey of public computer vision datasets for precision livestock farming. *Computers and Electronics in Agriculture*, 222, 109718. <https://doi.org/10.1016/j.compag.2024.109718>
- Bumbálek, R. (2025). Computer vision in precision livestock farming: Applications and future directions. *Artificial Intelligence in Agriculture*. <https://doi.org/10.1016/j.aiaa.2025.01.003>
- Bumbálek, R., Umurungi, S. N., Ufitikirezi, J. D. M., Zoubek, T., Kuneš, R., Stehlík, R., Lin, H.-I., & Bartoš, P. (2025). Deep learning in poultry farming: Comparative analysis of YOLOv8, YOLOv9, YOLOv10, and YOLOv11 for dead chickens detection. *Poultry Science*, 104(9), 105440. <https://doi.org/10.1016/j.psj.2025.105440>
- Campbell, W., Noorman, J., & Steibel, J. P. (2024). A computer vision approach to monitor activity in commercial broiler chickens using trajectory-based clustering analysis. *Computers and Electronics in Agriculture*, 206, 108591. <https://doi.org/10.1016/j.compag.2023.108591>
- Chauca, L. (2020). Producción de cuyes (*Cavia porcellus*) en los países andinos: situación actual y perspectivas. *Animal Genetic Resources*, 57, 91–108. <https://doi.org/10.1017/S2078633620000041>
- Cominotte, A., Campos, A. C., Fernandes, D. A., Lopes, F. B., Lanna, D. P. D., McManus, C., & Falco, J. E. (2021). Automated computer vision system to predict body weight and average daily gain in beef cattle. *Livestock Science*, 247, 104444. <https://doi.org/10.1016/j.livsci.2021.104444>
- Cruz, E., Hidalgo-Rodríguez, M., Acosta-Reyes, A. M., Rangel, J. C., & Boniche, K. (2024). AI-based monitoring for enhanced poultry flock management. *Agriculture*, 14(12), 2187. <https://doi.org/10.3390/agriculture14122187>
- Dalal, M. (2025). Deep learning-based object detection techniques in agriculture: A systematic review. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2025.108901>
- Elmessery, W. M., Gutiérrez, J., Abd El-Wahhab, G. G., Elkhayat, I. A., El-Soaly, I. S., Alhag, S. K., & Abdelshafie, M. F. (2023). YOLO-based model for automatic detection of broiler pathological phenomena through visual and thermal images in intensive poultry houses. *Agriculture*, 13(8), 1527. <https://doi.org/10.3390/agriculture13081527>

Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. (2024). YOLO evolution: A comprehensive benchmark and architectural review of YOLOv12, YOLO11, and their previous versions. arXiv preprint. <https://doi.org/10.48550/arXiv.2411.00201>

Jia, W., Li, Y., Hua, Z., & Zhang, Q. (2023). PCR: A large-scale benchmark for pig counting in real world. In *Pattern Recognition and Computer Vision – PRCV 2023* (pp. 212–225). Springer. [https://doi.org/10.1007/978-981-99-8462-6\\_19](https://doi.org/10.1007/978-981-99-8462-6_19)

Jiménez Gómez, X. (2024). Detección y análisis de datos sobre especies exóticas en biomas mediante técnicas de detección de objetos. *Universidade da Coruña*. <https://ruc.udc.es/entities/publication/2fd62404-e9ba-43cf-9756-100d249af86a>

Khanam, R., & Hussain, M. (2025). YOLO advances to its genesis: A decadal and comprehensive review of the YOLO series. *Artificial Intelligence Review*, 58, 253. <https://doi.org/10.1007/s10462-025-11253-3>

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., Zhu, J., & Zhang, L. (2024). Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection. In *Computer Vision – ECCV 2024* (LNCS 15074, pp. 38–55). Springer. [https://doi.org/10.1007/978-3-031-72970-6\\_3](https://doi.org/10.1007/978-3-031-72970-6_3)

Lv, W., Zhao, Y., Chang, Q., Huang, K., Wang, G., & Liu, Y. (2024). RT-DETRv2: Improved baseline with bag-of-freebies for real-time detection transformer. arXiv. <https://doi.org/10.48550/arXiv.2407.17140>

Menezes, R., Silva, F., & Oliveira, J. (2024). Computer vision systems for animal monitoring in precision livestock farming: A review. *Animals*, 14(5), 812. <https://doi.org/10.3390/ani14050812>

Michielon, A., Litta, P., Bonelli, F., Don, G., Farisè, S., Giannuzzi, D., Milanese, M., Pietrucci, D., Vezzoli, A., Cecchinato, A., Chillemi, G., Gallo, L., Mele, M., & Furlanello, C. (2024). Mind the step: An artificial intelligence-based monitoring platform for animal welfare. *Sensors*, 24(24), 8042. <https://doi.org/10.3390/s24248042>

Mulero-Pázmány, M., Hurtado, S., Barba-González, C., Antequera-Gómez, M. L., Díaz-Ruiz, F., Real, R., Navas-Delgado, I., & Aldana-Montes, J. F. (2025). Addressing significant challenges for animal detection in camera trap images: A novel deep learning-based approach. *Scientific Reports*, 15, 16191. <https://doi.org/10.1038/s41598-025-90249-z>

Natho, P., Boonying, S., Bonguleaum, P., Tantidontanet, N., & Chamuthai, L. (2025). An enhanced machine vision system for smart poultry farms using deep learning. *Artificial Intelligence in Agriculture*. Advance online publication. <https://www.sciencedirect.com/science/article/pii/S2772375525003168>

Rajagukguk, R. A. (2025). Deep learning for visual animal monitoring (detection, tracking, pose estimation, and behavior classification): A comprehensive review. *Artificial Intelligence in Agriculture*. Advance online publication. <https://www.sciencedirect.com/science/article/pii/S2772375525007701>

Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. *Advances in Neural Information Processing Systems*. <https://doi.org/10.48550/arXiv.2502.12524>

Wu, Z., Yang, J., Zhang, H., & Fang, C. (2025). Enhanced methodology and experimental research for caged chicken counting based on YOLOv8. *Animals*, 15(6), 853. <https://doi.org/10.3390/ani15060853>

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. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 16965–16974). IEEE. <https://doi.org/10.1109/CVPR52733.2024.01605>

Todo el contenido de LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades, publicados en este sitio está disponibles bajo Licencia [Creative Commons](#) 