



Etiquetado Semi-Automático con SAM3 y Segmentación de Instancias de Pollos de Engorde mediante YOLO26n-seg

Semi-Automatic Labeling with SAM3 and Instance Segmentation of Broiler Chickens Using YOLO26n-seg

Autores

Jonathan Aguilar-Alvarado*
✉ jaguilar@utmachala.edu.ec

Karina Garcia-Galarza
✉ kgarcia@utmachala.edu.ec

John Patricio Orellana Preciado
✉ jporellana@utmachala.edu.ec

Universidad Técnica de Machala,
Facultad de Ingeniería Civil, Carrera de
Ciencia de Datos e Inteligencia Artificial,
Ecuador, Machala.

* Autor para correspondencia

Comó citar el artículo:

Aguilar-Alvarado, J., Garcia-Galarza, K. & Orellana Preciado, J.P. (2026). Etiquetado Semi-Automático con SAM3 y Segmentación de Instancias de Pollos de Engorde mediante YOLO26n-seg. *Informática y Sistemas*, 10(2), 184–196. <https://doi.org/10.33936/isrtic.v10i2.8599>

Enviado: 22/07/2026

Aceptado: 12/10/2026

Publicado: 13/10/2026

Resumen



La segmentación de instancias constituye una herramienta relevante para el monitoreo automatizado de pollos de engorde en escenas cenitales de alta densidad; sin embargo, la generación manual de máscaras demanda un esfuerzo considerable. El objetivo de este estudio fue desarrollar y evaluar un flujo de etiquetado semi-automático basado en Segment Anything Model 3 (SAM3), limpieza automática, revisión humana y entrenamiento con You Only Look Once (YOLO26n-seg). A partir de 14 videos se extrajeron 1.863 imágenes; 1.860 conformaron el conjunto final utilizado por YOLO26n-seg, distribuidas en 1.301 imágenes de entrenamiento, 307 de validación y 252 de prueba. SAM3 generó 210.762 anotaciones poligonales y la limpieza automática registrada eliminó 134 duplicados. El paquete de etiquetas disponible antes de la revisión humana contenía 210.069 anotaciones; después de la revisión en Roboflow, la exportación final registró 229.052 anotaciones y el cargador de Ultralytics utilizó 229.035 instancias tras eliminar 17 duplicados exactos. En validación, el modelo alcanzó una precisión de máscara de 0,78845, un recall de 0,90522, un mAP@0,50 de 0,89476 y un mAP@0,50:0,95 de 0,68705; en prueba obtuvo 0,78639, 0,87559, 0,88387 y 0,67593, respectivamente. Los resultados mostraron un desempeño estable entre ambos subconjuntos y una elevada capacidad para recuperar instancias. No obstante, la delimitación precisa de los contornos continuó siendo más exigente en regiones con alta densidad, contacto físico y oclusión parcial. El flujo propuesto constituye una alternativa reproducible para apoyar la construcción de conjuntos de datos de segmentación en entornos avícolas reales.

Palabras clave: pollos de engorde; segmentación de instancias; SAM3; YOLO26n-seg; visión por computador.

Abstract

Instance segmentation is a relevant tool for the automated monitoring of broiler chickens in high-density overhead scenes; however, manual mask generation requires considerable effort. This study aimed to develop and evaluate a semi-automatic labeling workflow based on Segment Anything Model 3 (SAM3), automated cleaning, human review, and You Only Look Once (YOLO26n-seg) training. A total of 1,863 images were extracted from 14 videos; 1,860 constituted the final dataset used by YOLO26n-seg, including 1,301 training images, 307 validation images, and 252 test images. SAM3 generated 210,762 polygon annotations, and the recorded automated cleaning stage removed 134 duplicates. The label package available before human review contained 210,069 annotations; after review in Roboflow, the final export contained 229,052 annotations, and the Ultralytics loader used 229,035 instances after removing 17 exact duplicates. On the validation set, the model achieved a mask precision of 0.78845, a recall of 0.90522, a mAP@0.50 of 0.89476, and a mAP@0.50:0.95 of 0.68705; on the test set, it obtained 0.78639, 0.87559, 0.88387, and 0.67593, respectively. The results showed stable performance across both subsets and a strong ability to recover instances. Nevertheless, accurate contour delineation remained more challenging in regions characterized by high density, physical contact, and partial occlusion. The proposed workflow represents a reproducible alternative for supporting the construction of instance-segmentation datasets in real poultry-farming environments.

Keywords: broiler chickens; instance segmentation; SAM3; YOLO26n-seg; computer vision.



1. Introducción

La avicultura de precisión ha incorporado progresivamente tecnologías de visión por computador para apoyar el monitoreo no invasivo de las aves, automatizar el análisis de imágenes y generar información útil sobre su distribución, comportamiento y estado dentro de los sistemas productivos. No obstante, el desarrollo de estos sistemas depende de la disponibilidad de conjuntos de datos representativos, obtenidos en condiciones reales de crianza y acompañados de anotaciones consistentes (Lamping et al., 2022). Una revisión de conjuntos de datos públicos relacionados con visión artificial en avicultura identificó limitaciones asociadas con el tamaño de las muestras, la diversidad de escenarios, la estandarización de las anotaciones y la disponibilidad de métricas comparables (Li et al., 2026). Frente a estas limitaciones, se han desarrollado conjuntos de datos como PIO, compuesto por 1.487 imágenes y 327.289 instancias de pollos de engorde anotadas manualmente, con variaciones de iluminación, densidad, morfología y etapa de crecimiento (Boniche et al., 2026).

En el contexto de la ganadería y la avicultura de precisión, la visión por computador ha sido utilizada para apoyar tareas de identificación, seguimiento y monitoreo de animales. (Reyes Zambrano & Moreira Santos, 2025), identifico a los modelos YOLO, las redes neuronales convolucionales, las técnicas de segmentación y los dispositivos de captura como herramientas relevantes para el control productivo, sanitario y de bienestar de los animales de granja. Estos antecedentes confirman la importancia de desarrollar procedimientos que permitan obtener datos visuales confiables y modelos capaces de analizar individualmente a los animales en condiciones productivas reales.

Los modelos de la familia You Only Look Once han sido ampliamente utilizados en aplicaciones avícolas debido a su capacidad para localizar múltiples individuos en una misma imagen (Paneru et al., 2026). Una investigación desarrolló un modelo YOLOv5-CBAM para la detección de pollos de engorde criados sobre cama, incorporando mecanismos de atención para mejorar la identificación de aves pequeñas u ocluidas (Guo et al., 2023). Asimismo, en otra investigación compararon distintas versiones de YOLO para la detección de aves muertas, evaluando su exactitud y velocidad de inferencia (Bumbálek et al., 2025). Estos estudios demuestran la utilidad de los detectores basados en YOLO; sin embargo, sus resultados se representan principalmente mediante cajas delimitadoras, las cuales permiten localizar cada ave, pero no describen con precisión su contorno ni la

superficie ocupada por cada individuo.

La segmentación de instancias amplía esta capacidad al identificar cada objeto y asignar una máscara a los píxeles que lo conforman. Esta representación resulta especialmente relevante en imágenes cenitales con alta densidad de aves, contacto físico y oclusiones parciales. (Sun et al., 2025) utilizaron segmentación de instancias para extraer regiones anatómicas de gallinas y combinar información visual y térmica en la identificación de animales enfermos. De manera similar, (Chen et al., 2025) propusieron una variante de Mask R-CNN para segmentar gallinas individualmente en imágenes infrarrojas obtenidas bajo condiciones de interferencia visual y oclusión. Estos antecedentes evidencian que las máscaras de instancia proporcionan información espacial más detallada que las cajas delimitadoras y pueden sustentar tareas posteriores de conteo, estimación de variables asociadas con el área corporal proyectada, seguimiento y análisis de distribución espacial (Saeidifar et al., 2024).

A pesar de estas ventajas, la construcción de conjuntos de datos para segmentación de instancias continúa siendo una etapa costosa, debido a que cada individuo debe delimitarse mediante máscaras o polígonos. La dificultad se incrementa en escenas avícolas densas, donde una sola imagen puede contener numerosas aves parcialmente superpuestas. Segment Anything Model introdujo un paradigma de segmentación guiada por prompts, diseñado para transferirse a nuevas imágenes y dominios sin requerir un entrenamiento específico para cada tarea (Kirillov et al., 2023). Posteriormente, SAM3 fue descrito como un modelo de segmentación de vocabulario abierto capaz de recuperar instancias asociadas con conceptos expresados mediante indicaciones nominales, lo que permite utilizarlo para generar máscaras iniciales que posteriormente pueden ser verificadas y corregidas mediante supervisión humana (Li et al., 2026).

Sin embargo, todavía existe incertidumbre respecto al desempeño de un flujo integrado que combine autoetiquetado con SAM3, depuración automática, revisión humana y entrenamiento de un modelo YOLO de segmentación de instancias en imágenes cenitales de pollos de engorde (Jiang et al., 2022). Los estudios revisados se han concentrado principalmente en detección mediante cajas delimitadoras, comparación de arquitecturas YOLO, construcción manual de conjuntos de datos o segmentación aplicada a tareas específicas. En consecuencia, permanece menos documentada la utilidad de un procedimiento completo que genere, refine y utilice máscaras individuales de pollos de engorde para entrenar un

modelo compacto de segmentación en escenas de alta densidad.

A partir de esta brecha se planteó la siguiente pregunta de investigación: ¿en qué medida un flujo de etiquetado semi-automático basado en SAM3, limpieza automática, revisión humana y entrenamiento con YOLO26n-seg permite segmentar instancias de pollos de engorde en imágenes cenitales?

Se formuló como hipótesis de trabajo que la combinación de máscaras iniciales generadas por SAM3, controles automáticos de calidad y revisión humana permitiría construir un conjunto de datos adecuado para entrenar YOLO26n-seg y mantener un desempeño estable en imágenes cenitales no utilizadas durante el ajuste (Çarklı Yavuz, 2026).

El objetivo general del estudio fue desarrollar y evaluar un procedimiento de etiquetado semi-automático con SAM3 para entrenar un modelo YOLO26n-seg orientado a la segmentación de instancias de pollos de engorde. Para alcanzar este objetivo, se extrajeron imágenes a partir de videos, se generaron máscaras iniciales mediante SAM3, se convirtieron las anotaciones al formato YOLO Segmentation, se aplicaron procesos de limpieza automática y revisión humana, y se entrenó y evaluó YOLO26n-seg mediante métricas de precisión, recall y mAP calculadas sobre las máscaras.

2. Materiales y Métodos

El procedimiento comprendió la adquisición de videos, la extracción de imágenes, el autoetiquetado mediante SAM3, la conversión de máscaras al formato YOLO Segmentation, la limpieza automática, la revisión humana de las anotaciones, el entrenamiento de YOLO26n-seg y la evaluación del modelo en subconjuntos independientes de validación y prueba. La Figura 1 resume el flujo metodológico.

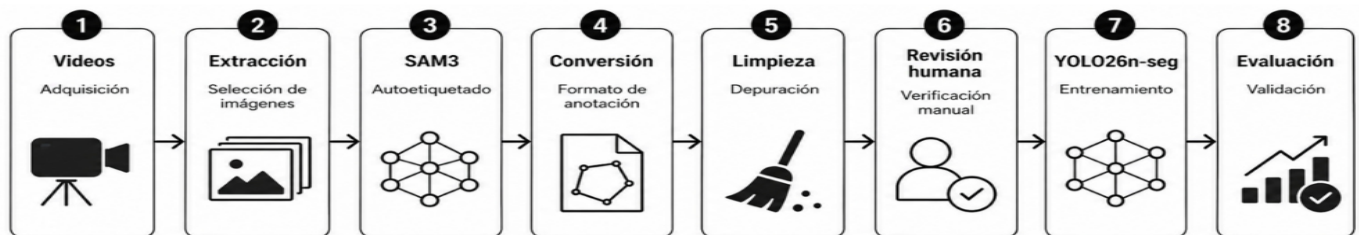


Figura 1. Pipeline metodológico del etiquetado semi-automático y la segmentación de instancias de pollos de engorde.

Fuente: Los autores.

Adquisición de Videos

La adquisición del material audiovisual se realizó en el área de crianza de pollos de engorde mediante una cámara instalada en posición superior, con el propósito de obtener una perspectiva cenital de las aves. Esta disposición permitió registrar escenas amplias con múltiples individuos, conservar información sobre su distribución espacial y documentar situaciones de proximidad, contacto y oclusión parcial. Los videos obtenidos

constituyeron la fuente primaria para la posterior extracción de imágenes y construcción del conjunto de datos (Doornweerd et al., 2024). Las condiciones técnicas bajo las cuales se obtuvo el material audiovisual se resumen en la Tabla 1, incluyendo el dispositivo utilizado, la resolución, la frecuencia de grabación y la perspectiva de captura.

Tabla 1. Características técnicas de la adquisición del material audiovisual.

Fuente: Los autores.

Parámetro	Especificación
Dispositivo de captura	Cámara Bullet IP 6 MP ColorVu 3.0
Fabricante	Hikvision
Tipo de registro	Video digital
Perspectiva de captura	Cenital
Ubicación de la cámara	Posición superior dentro del área de crianza
Número de videos	14
Resolución original	3200 × 1800 píxeles
Frecuencia de grabación	20 fotogramas por segundo
Población observada	Pollos de engorde
Condición de captura	Entorno de crianza

Nota. La información técnica corresponde a las condiciones documentadas durante la adquisición y procesamiento inicial del material audiovisual.

Como se observa en la Tabla 1, los videos fueron capturados mediante una cámara IP de 6 MP ubicada en posición superior, con una resolución original de 3200 × 1800 píxeles y una frecuencia de 20 fotogramas por segundo. La perspectiva cenital

permitió registrar simultáneamente un número elevado de aves y conservar información sobre su distribución espacial, proximidad y oclusión. Estas condiciones de captura proporcionaron el material visual necesario para la posterior extracción de imágenes y segmentación individual de los pollos de engorde.

Extracción de imágenes

Los videos adquiridos fueron procesados para obtener imágenes individuales que sirvieran como unidades de análisis durante el

autoetiquetado y el entrenamiento del modelo. La extracción se realizó de forma periódica para reducir la similitud entre fotogramas consecutivos y evitar una representación excesivamente redundante de las escenas. Cada imagen fue almacenada con una nomenclatura estructurada y vinculada con su video de origen mediante un registro de metadatos. Posteriormente, la distribución del conjunto de datos se efectuó a nivel de video, asegurando que los fotogramas procedentes de una misma grabación permanecieran en un único subconjunto y reduciendo así el riesgo de fuga de información entre entrenamiento, validación y prueba (Lilja et al., 2024). La configuración técnica utilizada para extraer, organizar y distribuir las imágenes obtenidas de los videos se resume en la Tabla 2.

Tabla 2. Configuración técnica de la extracción y organización de imágenes.

Fuente: Los autores.

Parámetro	Especificación
Fuente de entrada	14 videos cenitales
Resolución original	3200 × 1800 píxeles
Resolución de procesamiento	1280 × 720 píxeles
Intervalo de extracción	Una imagen cada 3 segundos
Formato de salida	JPG
Nomenclatura	Identificación del video, tiempo de captura y fotograma
Estrategia de partición	Separación por video
Entrenamiento	Videos 1–9 y 14
Validación	Videos 10 y 11
Prueba	Videos 12 y 13
Imágenes obtenidas inicialmente	1.863
Imágenes de entrenamiento	1.304
Imágenes de validación	307
Imágenes de prueba	252

La Tabla 2 muestra que se extrajo una imagen cada tres segundos y que la resolución utilizada durante el procesamiento fue de 1280×720 píxeles. Este intervalo buscó disminuir la redundancia entre fotogramas consecutivos, conservando al mismo tiempo variaciones relevantes en la posición y distribución de las aves. Las 1.863 imágenes obtenidas se distribuyeron en 1.304 imágenes de entrenamiento, 307 de validación y 252 de prueba. La partición se efectuó por video, evitando que imágenes temporalmente próximas de una misma grabación aparecieran en diferentes subconjuntos y reduciendo así el riesgo de fuga de información.

Autoetiquetado con SAM3

Las imágenes seleccionadas fueron procesadas mediante SAM3 para generar automáticamente las máscaras iniciales correspondientes a los pollos de engorde presentes en cada escena. El modelo recibió una indicación textual asociada con la clase de interés y generó máscaras para las instancias vinculadas con dicho concepto. Esta etapa permitió obtener una primera aproximación de las anotaciones antes de aplicar los procesos posteriores de conversión, depuración y revisión humana (Li et al., 2026; Yang et al., 2024). Los parámetros empleados durante el autoetiquetado de las imágenes mediante SAM3 se presentan en la Tabla 3.

Tabla 3. Configuración técnica del autoetiquetado mediante SAM3.

Fuente: Los autores.

Parámetro	Especificación
Modelo utilizado	SAM3
Tipo de tarea	Segmentación de instancias
Clase de interés	Pollo
Prompt textual	chicken
Entrada	Imágenes extraídas de los videos
Resolución máxima de procesamiento	Ancho de hasta 1280 píxeles
Umbral mínimo de confianza	0,30
Umbral de binarización de máscara	0,50
Tipo de salida	Máscaras binarias por instancia
Aplicación del procedimiento	Entrenamiento, validación y prueba

Nota. Las máscaras generadas por SAM3 se consideraron anotaciones iniciales. Estas no fueron utilizadas directamente para entrenar el modelo, sino que posteriormente se transformaron al formato de segmentación requerido, se depuraron y se sometieron a revisión humana.

Como se aprecia en la Tabla 3, SAM3 fue configurado para realizar segmentación de instancias utilizando el prompt textual chicken y un umbral mínimo de confianza de 0,30. Las máscaras generadas fueron binarizadas mediante un umbral de 0,50 y posteriormente convertidas en anotaciones poligonales. Esta configuración se aplicó a las imágenes de entrenamiento, validación y prueba. Las máscaras obtenidas en esta etapa se consideraron anotaciones iniciales y no fueron utilizadas directamente para entrenar YOLO26n-seg, debido a que primero fueron sometidas a procesos de conversión, depuración automática y revisión humana.

Conversión a formato de anotación

Las máscaras producidas por SAM3 fueron convertidas en anotaciones poligonales compatibles con YOLO Segmentation. Cada máscara se binarizó, se ajustó a las dimensiones de la imagen procesada cuando fue necesario y se transformó en un contorno mediante OpenCV. Posteriormente, se conservó el contorno de mayor área, se aproximó como polígono y se normalizaron sus coordenadas respecto al ancho y alto de la imagen. Cada instancia válida se almacenó en una línea de un archivo de texto con el identificador de clase seguido por los pares de coordenadas del polígono.

Limpeza automática de las anotaciones

Las anotaciones en formato YOLO Segmentation fueron sometidas a un proceso automático de validación y eliminación de duplicados (Jiang et al., 2022). El procedimiento conservó las etiquetas originales en una carpeta independiente y generó una nueva versión depurada para cada subconjunto. Primero se verificó la estructura de cada línea de anotación y se excluyeron aquellas con errores de clase, cantidad de coordenadas, valores no numéricos, valores no finitos o coordenadas fuera del intervalo normalizado.

Posteriormente, se aplicó un procedimiento de supresión no máxima basado en máscaras para comparar polígonos superpuestos y conservar, cuando se identificó una duplicación, el polígono de mayor área. El proceso se ejecutó sobre entrenamiento, validación y prueba (Jena et al., 2023). La configuración implementada para validar las anotaciones y eliminar posibles duplicados se detalla en la Tabla 4.

Tabla 4. Configuración implementada para la limpieza automática de anotaciones.

Fuente: Los autores.

Parámetro	Implementación documentada
Subconjuntos	train, val y test
Clase admitida	0
Número mínimo de elementos por línea	7
Número mínimo de puntos	3
Intervalo permitido	Coordenadas entre 0 y 1
Método de eliminación de duplicados	Supresión no máxima basada en máscaras
Umbral de IoU entre máscaras	0,45
Umbral de contención	0,70
Distancia máxima entre centroides	0,08
Umbral de IoU entre cajas	0,35
Resolución interna para comparar máscaras	640 × 360 píxeles
Criterio de conservación	Polígono de mayor área

Nota. El algoritmo consideró una anotación como duplicada cuando cumplía alguno de los criterios programados: IoU de máscara igual o superior a 0,45; contención igual o superior a 0,70 con centroides cercanos; IoU de cajas igual o superior a 0,35 con centroides cercanos; o IoU de máscara moderado con una distancia entre centroides especialmente reducida. Para cada subconjunto se registraron el número de objetos originales, válidos y depurados, las anotaciones duplicadas eliminadas, las líneas inválidas y el porcentaje de reducción.

La Tabla 4 evidencia que la limpieza automática combinó criterios de superposición, contención, proximidad entre centroides e intersección entre cajas delimitadoras. Una anotación podía considerarse duplicada cuando representaba una región ampliamente coincidente con otra máscara y ambas se encontraban espacialmente próximas. Ante la identificación de un duplicado, se conservó el polígono de mayor área. Esta configuración permitió aplicar un procedimiento reproducible de depuración antes de la revisión humana, aunque la superación de estos controles automáticos no garantizó por sí sola la precisión visual de los contornos.

Revisión y corrección humana de las anotaciones

Después de la limpieza automática, las anotaciones fueron importadas en Roboflow para realizar una inspección visual individual de las imágenes y sus polígonos asociados (Lin et al., 2022). Durante la revisión humana se conservaron las máscaras correctamente delimitadas, se eliminaron falsos positivos y duplicados, se corrigieron los polígonos incompletos o imprecisos y se incorporaron manualmente las aves omitidas. También se verificó la correspondencia con la clase 0 y la consistencia del formato YOLO Segmentation. Para cuantificar retrospectivamente el grado de intervención humana, se compararon los archivos de etiquetas disponibles inmediatamente antes de la revisión con la exportación final de Roboflow y, adicionalmente, con las instancias efectivamente cargadas por Ultralytics. La comparación se realizó por imagen y permitió calcular la variación neta del número de instancias y la proporción de imágenes cuyo conteo cambió. Este procedimiento no permite estimar el tiempo empleado ni contabilizar correcciones geométricas que mantuvieron sin cambio el número de instancias.

Entrenamiento del modelo YOLO26n-seg

El conjunto de datos depurado se utilizó para entrenar YOLO26n-seg, correspondiente a la variante nano (n) de YOLO26 para segmentación de instancias (seg), inicializada a partir del checkpoint yolo26n-seg.pt y orientada a la segmentación de pollos de engorde. La configuración incorporó una semilla fija y ejecución determinista para reducir la variabilidad entre ejecuciones, además de detención temprana para finalizar el entrenamiento cuando el desempeño de validación dejó de mejorar (Åkesson et al., 2024). Los hiperparámetros y las condiciones experimentales utilizadas para entrenar YOLO26n-seg se presentan en la Tabla 5.



Tabla 5. Configuración utilizada para el entrenamiento de YOLO26n-seg.

Fuente: Los autores.

Parámetro	Configuración
Arquitectura	YOLO26n-seg
Checkpoint inicial	yolo26n-seg.pt
Número máximo de épocas	150
Resolución de entrada	960 × 960 píxeles
Tamaño del lote	16
Optimizador	AdamW
Tasa de aprendizaje inicial	0,002
Regularización de pesos	0,0005
Detención temprana	30 épocas sin mejora
Programación de aprendizaje	Cosenoidal
Semilla experimental	42
Factor de reducción de máscaras	4
Aumento de datos	Rotación, traslación, escala, volteos, variación cromática y mosaico
Desactivación de mosaico	Últimas 15 épocas
GPU	NVIDIA A100-SXM4-80GB
Memoria GPU	81.153 MiB
CPU disponible	12
Memoria RAM	167,1 GB

Nota. El mejor peso obtenido durante el entrenamiento se seleccionó para la evaluación posterior en los subconjuntos de validación y prueba.

La configuración de la Tabla 5 estableció un máximo de 150 épocas, una resolución de entrada de 960 × 960 píxeles y un tamaño de lote de 16 imágenes. La resolución seleccionada buscó conservar suficiente detalle espacial para delimitar los contornos de las aves. La detención temprana de 30 épocas sin mejora permitió finalizar el entrenamiento cuando el desempeño de validación dejó de progresar. Asimismo, la utilización de una semilla experimental fija y una ejecución determinista favoreció la reproducibilidad del procedimiento.

Entorno computacional. Los experimentos de entrenamiento de YOLO26n-seg se ejecutaron utilizando Ultralytics 8.4.98 sobre Python 3.12.13 y PyTorch 2.11.0+cu128. La aceleración mediante GPU se realizó con una NVIDIA A100-SXM4-80GB,

con 81.153 MiB de memoria de GPU disponible. El entorno de ejecución reportó 12 CPU y 167,1 GB de memoria RAM. Esta configuración fue utilizada para el entrenamiento y la evaluación del modelo.

Evaluación del modelo

El mejor peso obtenido durante el entrenamiento fue evaluado de manera independiente en los subconjuntos de validación y prueba. El desempeño se determinó tanto para las cajas delimitadoras como para las máscaras de instancia mediante precisión, recall, mAP@0,50 y mAP@0,50:0,95. El conjunto de validación permitió comprobar el comportamiento del modelo durante el proceso experimental, mientras que el conjunto de prueba se mantuvo separado para estimar su capacidad de generalización sobre imágenes no utilizadas en el ajuste. Además de las métricas cuantitativas, se revisaron las predicciones generadas sobre imágenes de prueba para observar cualitativamente la delimitación individual de las aves (Cheng et al., 2021). La configuración utilizada para evaluar el desempeño de YOLO26n-seg en los subconjuntos de validación y prueba se resume en la Tabla 6.

Tabla 6. Configuración de la evaluación de YOLO26n-seg.

Fuente: Los autores.

Parámetro	Configuración
Subconjuntos evaluados	Validación y prueba
Tipo de tarea	Segmentación de instancias
Resolución de entrada	960 × 960 píxeles
Tamaño del lote	16
Número máximo de predicciones por imagen	500
Métricas principales	Precisión, recall, mAP@0,50 y mAP@0,50:0,95
Tipos de predicción evaluados	Cajas delimitadoras y máscaras
Criterio principal del estudio	Desempeño de las máscaras de instancia
Evaluación de generalización	Conjunto de prueba independiente

Como se indica en la Tabla 6, la evaluación se realizó de manera independiente sobre 307 imágenes de validación y 252 imágenes de prueba. Se calcularon métricas para cajas delimitadoras y máscaras de instancia; sin embargo, las métricas correspondientes a las máscaras constituyeron el principal criterio del estudio. Esta diferenciación permitió evaluar, por una parte, la capacidad del modelo para localizar las aves y, por otra, su precisión para

delimitar la superficie y el contorno de cada individuo.

4. Resultados

Conformación del conjunto de datos

La extracción periódica de fotogramas produjo 1.863 imágenes, cuya distribución inicial por subconjunto se presenta en la Tabla 7. La partición se realizó a nivel de video para evitar que imágenes temporalmente próximas de una misma grabación aparecieran simultáneamente en entrenamiento, validación y prueba. De las 1.863 imágenes inicialmente extraídas, el conjunto final utilizado por YOLO26n-seg quedó conformado por 1.860 imágenes: 1.301 de entrenamiento, 307 de validación y 252 de prueba. Tres imágenes del subconjunto de entrenamiento presentes en la extracción inicial no formaron parte del paquete final utilizado para la revisión y el entrenamiento.

Tabla 7. Distribución del conjunto de datos por subconjunto.
Fuente: Los autores.

Subconjunto	Videos	Imágenes
Entrenamiento	10	1.304
Validación	2	307
Prueba	2	252
Imágenes totales	14	1.863

Nota. El conjunto de entrenamiento se conformó con los videos 1 al 9 y el video 14; validación con los videos 10 y 11; y prueba con los videos 12 y 13. La partición se realizó a nivel de video.

La Tabla 7 describe la distribución inicial de las 1.863 imágenes extraídas: 1.304 imágenes de entrenamiento, equivalentes aproximadamente al 70,0 % del total, 307 de validación (16,5 %) y 252 de prueba (13,5 %). Diez videos se destinaron al entrenamiento y cuatro videos completos se reservaron para la evaluación. Posteriormente, el conjunto final utilizado por YOLO26n-seg quedó conformado por 1.860 imágenes, debido a que tres imágenes del subconjunto de entrenamiento no formaron parte del paquete final de revisión y entrenamiento. La separación por video se mantuvo en todas las etapas.

Autoetiquetado y limpieza automática

El autoetiquetado mediante SAM3 procesó las 1.863 imágenes y generó 210.762 anotaciones poligonales, sin errores registrados. La limpieza automática identificó duplicados en 126 imágenes y eliminó 134 anotaciones: 99 en entrenamiento, 20 en validación y 15 en prueba. Como resultado, el conjunto pasó de 210.762 a 210.628 anotaciones, equivalente a una reducción global aproximada del 0,064 %. No se detectaron líneas inválidas, etiquetas vacías, archivos faltantes ni imágenes sin polígonos válidos. Los resultados obtenidos del proceso se presentan en la Tabla 8.

Tabla 8. Resultados del autoetiquetado y la limpieza automática de anotaciones.

Fuente: Los autores.

Sub-conjunto	Imágenes	Anotaciones SAM3	Duplicados eliminados	Anotaciones después de la limpieza	Reducción
Entrenamiento	1.304	146.022	99	145.923	0,0678 %
Validación	307	35.616	20	35.596	0,0562 %
Prueba	252	29.124	15	29.109	0,0515 %
Total	1.863	210.762	134	210.628	0,0636 %

Como se aprecia en la Tabla 8, la limpieza automática eliminó 134 anotaciones duplicadas, por lo que el conjunto pasó de 210.762 a 210.628 polígonos. La reducción global fue del 0,0636 %. El subconjunto de entrenamiento concentró la mayor cantidad de duplicados eliminados, con 99 casos, seguido de validación con 20 y prueba con 15. En ninguno de los subconjuntos la reducción superó el 0,07 %.

Después de la limpieza automática, la densidad media fue de aproximadamente 111,90 anotaciones por imagen en entrenamiento, 115,95 en validación y 115,51 en prueba. Estos valores muestran que las escenas contenían simultáneamente una cantidad elevada de aves, lo que incrementó la dificultad de separar individuos próximos o parcialmente superpuestos. La baja proporción de duplicados no debe interpretarse como evidencia de que todas las máscaras generadas por SAM3 fueran visualmente correctas, debido a que la exactitud de los contornos y la correspondencia entre máscara y ave fueron verificadas posteriormente mediante revisión humana.

Después de la limpieza automática, las anotaciones fueron revisadas visualmente en Roboflow. El paquete de etiquetas disponible inmediatamente antes de esta revisión contenía 1.860 archivos y 210.069 anotaciones: 145.416 en entrenamiento, 35.557 en validación y 29.096 en prueba. La exportación final de Roboflow registró 229.052 anotaciones, lo que representa un incremento neto de 18.983 anotaciones (9,04 %) respecto del paquete previo a la revisión. Durante la generación de la caché de Ultralytics se eliminaron 17 duplicados exactos —4 en entrenamiento, 6 en validación y 7 en prueba—, por lo que el modelo utilizó efectivamente 229.035 instancias: 164.308 en entrenamiento, 35.610 en validación y 29.117 en prueba. Al comparar por imagen el paquete previo a la revisión con el conjunto efectivamente utilizado por el modelo, 1.090 de 1.860 imágenes (58,60 %) presentaron cambios en el número de instancias: 982 de 1.301 en entrenamiento, 65 de 307 en validación y 43 de 252 en prueba. Estas cifras cuantifican el efecto neto de la intervención humana sobre el conjunto de anotaciones, pero no equivalen al número de operaciones manuales realizadas, ya que las correcciones geométricas pueden modificar un polígono sin alterar el conteo de instancias.





Entrenamiento

El modelo YOLO26n-seg fue entrenado utilizando el conjunto final conformado por 1.860 imágenes y 229.035 instancias efectivamente cargadas por Ultralytics: 1.301 imágenes y 164.308 instancias en entrenamiento, 307 imágenes y 35.610 instancias en validación, y 252 imágenes y 29.117 instancias en prueba. Se mantuvo la separación por video entre los subconjuntos de entrenamiento, validación y prueba. La evaluación consideró métricas correspondientes a cajas delimitadoras y máscaras de instancia.

La pérdida de caja (Figura 2(a)) disminuyó de forma sostenida tanto en entrenamiento como en validación, lo que refleja una mejora progresiva en la localización espacial de las aves. De manera similar, la pérdida de segmentación (Figura 2(b)) presentó una reducción continua, evidenciando un mejor ajuste de las máscaras a los contornos de los pollos de engorde. En ambos casos, la curva de validación se mantuvo ligeramente por debajo de la de entrenamiento durante la mayor parte del proceso, sin observarse una divergencia creciente entre ellas.

La pérdida de clasificación (Figura 2(c)) descendió rápidamente

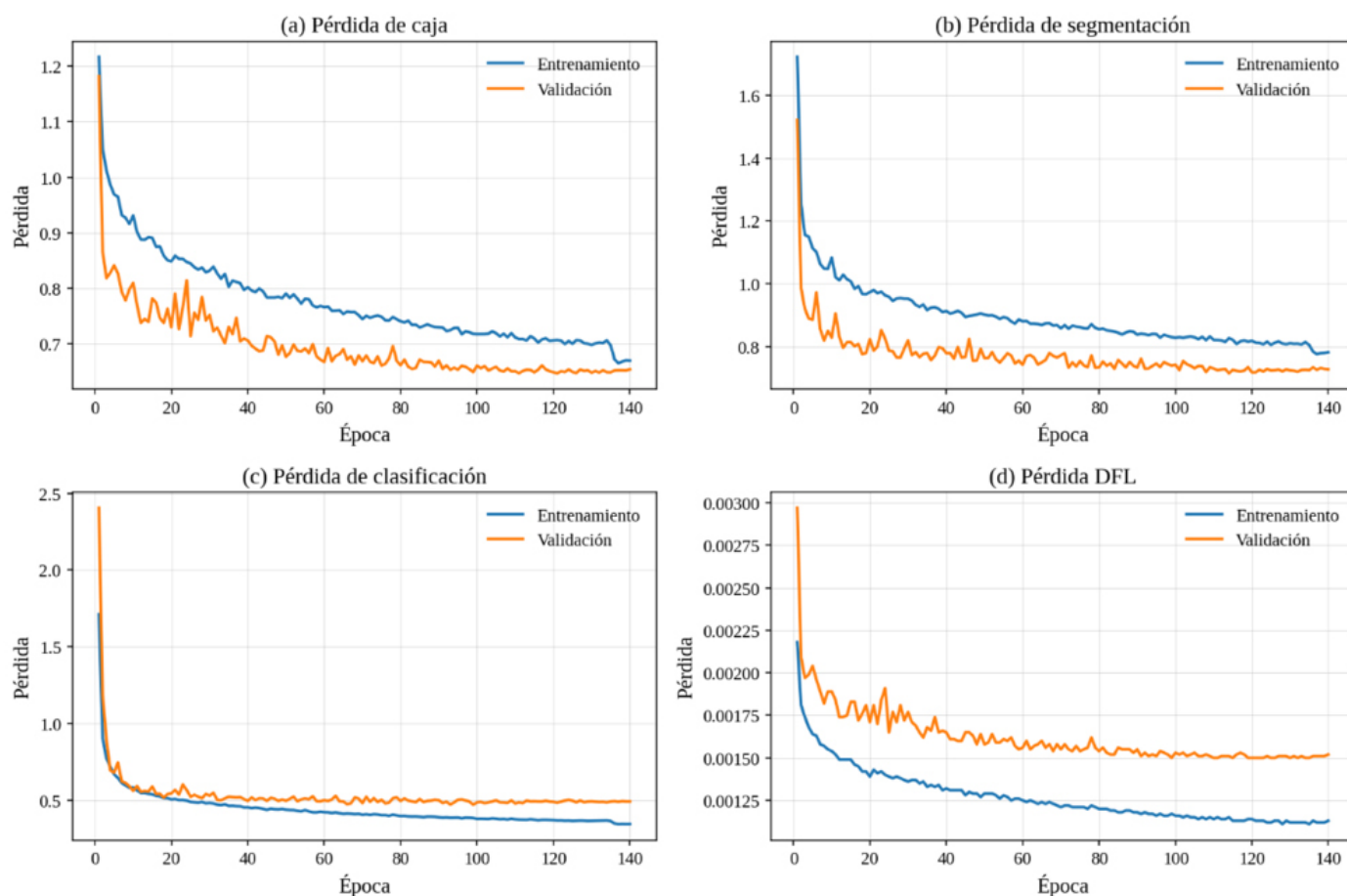


Figura 2. Evolución de las pérdidas de caja, segmentación, clasificación y DFL durante el entrenamiento y la validación de YOLO26n-seg.

Fuente: Los autores.

La Figura 2 muestra una reducción pronunciada de las cuatro funciones de pérdida durante las primeras épocas, seguida de un descenso más gradual y una posterior estabilización. Este comportamiento indica que el modelo adquirió rápidamente los patrones principales de las imágenes y continuó refinando sus predicciones conforme avanzó el entrenamiento.

durante las primeras épocas y posteriormente se estabilizó. La curva de entrenamiento alcanzó valores inferiores a la de validación, aunque ambas conservaron una tendencia estable hacia el final. Por su parte, la pérdida DFL (Figura 2(d)) mostró una disminución progresiva en entrenamiento, mientras que en validación presentó pequeñas oscilaciones iniciales antes de

estabilizarse. La separación moderada entre ambas curvas sugiere que la precisión espacial continuó siendo más exigente en las imágenes de validación.

En conjunto, las curvas evidencian un proceso de aprendizaje estable, sin observarse una divergencia sostenida entre las curvas de entrenamiento y validación. La disminución adicional observada en las últimas épocas, especialmente en las pérdidas de caja y segmentación, coincide con la fase final de ajuste del modelo.

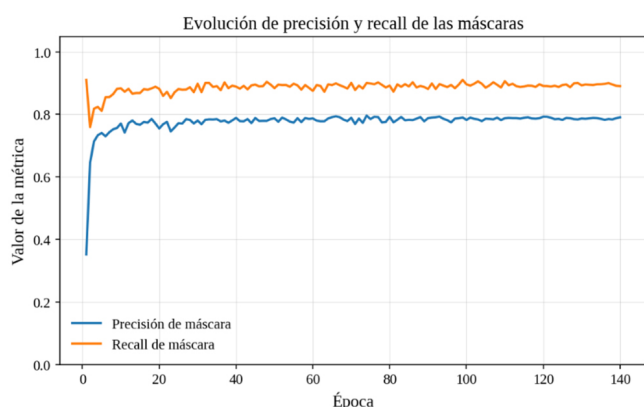


Figura 3. Evolución de la precisión y el recall de las máscaras durante el entrenamiento de YOLO26n-seg.

Fuente: Los autores.

La Figura 3 muestra que la precisión y el recall de las máscaras aumentaron rápidamente durante las primeras épocas y posteriormente alcanzaron una fase de estabilización. La precisión pasó de valores iniciales bajos a niveles cercanos a 0,78, mientras que el recall se mantuvo alrededor de 0,88 - 0,90 durante la mayor parte del entrenamiento. Este comportamiento evidencia que el modelo aprendió tempranamente a identificar una proporción elevada de las instancias presentes en las imágenes.

El recall se mantuvo por encima de la precisión, lo que indica una mayor capacidad para recuperar instancias que para evitar completamente predicciones adicionales. En términos prácticos, el modelo mostró una elevada sensibilidad para detectar y segmentar las aves, pero con una precisión relativamente menor al diferenciar algunos contornos en escenas de alta densidad.

Las pequeñas oscilaciones observadas después de las primeras épocas no modificaron la tendencia general y ambas métricas conservaron valores estables hasta el final del entrenamiento. La ausencia de caídas sostenidas sugiere un proceso de convergencia estable. La diferencia entre precisión y recall puede relacionarse con la proximidad, el contacto físico y la oclusión parcial entre los pollos, condiciones que dificultan la delimitación exacta de cada instancia.

La Figura 4 muestra un incremento pronunciado de ambas métricas durante las primeras épocas, seguido de una fase de crecimiento más gradual y posterior estabilización. El $mAP@0,50$

de las máscaras aumentó rápidamente desde valores iniciales cercanos a 0,67 hasta situarse alrededor de 0,88–0,89, mientras que el $mAP@0,50:0,95$ evolucionó desde aproximadamente 0,47 hasta valores próximos a 0,68–0,69.

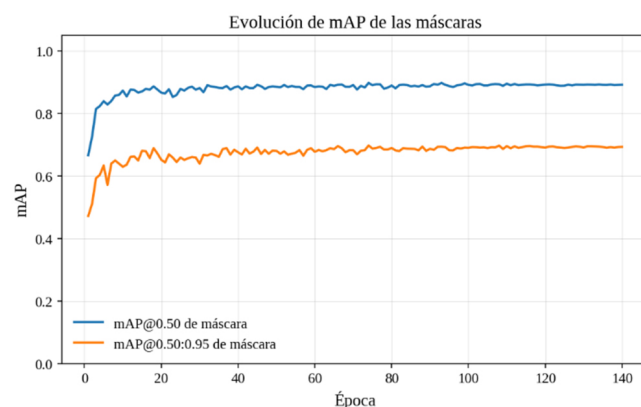


Figura 4. Evolución de $mAP@0,50$ y $mAP@0,50:0,95$ de las máscaras durante el entrenamiento de YOLO26n-seg.

Fuente: Los autores.

El $mAP@0,50$ se mantuvo por encima del $mAP@0,50:0,95$ durante todo el entrenamiento. Esta diferencia es esperable, debido a que el segundo indicador promedia el rendimiento en un rango más estricto de umbrales de intersección sobre unión y, por tanto, exige una mayor precisión en la delimitación de las máscaras. En consecuencia, el modelo mostró una alta capacidad para identificar correctamente las instancias bajo un umbral de IoU de 0,50, aunque la coincidencia exacta de los contornos resultó más exigente cuando se consideraron umbrales superiores.

Después de las primeras 20 épocas, ambas curvas presentaron oscilaciones de baja magnitud y una tendencia estable hasta el final del entrenamiento. Este comportamiento evidencia una convergencia progresiva del modelo y una mejora limitada durante las épocas posteriores. La separación persistente entre ambas métricas sugiere que la principal dificultad no radicó en reconocer la presencia de los pollos, sino en alcanzar una delimitación espacial más precisa en escenas caracterizadas por alta densidad, contacto físico y oclusión parcial entre individuos.

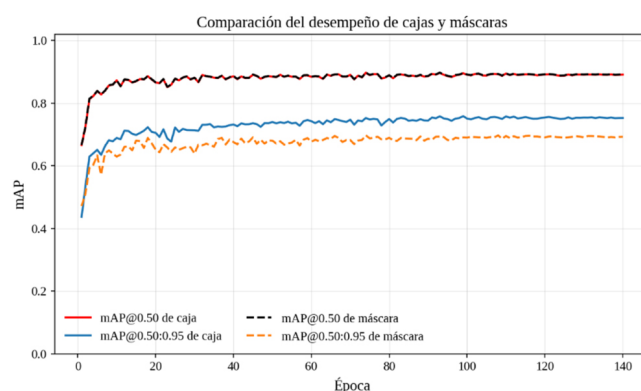


Figura 5. Comparación del desempeño de cajas delimitadoras y máscaras durante el entrenamiento de YOLO26n-seg.

Fuente: Los autores.

La Figura 5 compara la evolución de $mAP@0,50$ y $mAP@0,50:0,95$ para cajas delimitadoras y máscaras de instancia. En ambos tipos de predicción se observa un incremento pronunciado durante las primeras épocas, seguido de una fase de estabilización. Las curvas de $mAP@0,50$ para cajas y máscaras permanecen prácticamente superpuestas durante gran parte del entrenamiento, con valores cercanos a 0,88–0,89, lo que indica un desempeño similar en la identificación general de las instancias bajo un umbral de IoU de 0,50. La diferencia se hace más evidente en $mAP@0,50:0,95$.

Las cajas delimitadoras alcanzan valores próximos a 0,75, mientras que las máscaras se estabilizan alrededor de 0,69. Esta separación refleja la mayor exigencia de la segmentación de instancias, ya que no basta con localizar correctamente cada pollo, sino que también se requiere delimitar con precisión su contorno. Por esta razón, el desempeño de las máscaras disminuye al evaluarse bajo umbrales de IoU más estrictos. Las métricas finales de segmentación obtenidas en los subconjuntos de validación y prueba se presentan en la Tabla 9.

En validación, el modelo alcanzó una precisión de máscara de 0,78845, un recall de 0,90522, un $mAP@0,50$ de 0,89476 y un $mAP@0,50:0,95$ de 0,68705. En prueba, obtuvo una precisión de 0,78639, un recall de 0,87559, un $mAP@0,50$ de 0,88387 y un $mAP@0,50:0,95$ de 0,67593.

Entre validación y prueba, la precisión disminuyó solamente 0,00206; el $mAP@0,50$, 0,01089; y el $mAP@0,50:0,95$, 0,01112. Estas diferencias reducidas indican que el desempeño general se mantuvo estable en imágenes provenientes de videos no utilizados durante el entrenamiento. La mayor variación se presentó en el recall, que disminuyó 0,02963, lo que sugiere una cantidad relativamente mayor de instancias no recuperadas en el subconjunto de prueba.

En ambos subconjuntos, el recall fue superior a la precisión, lo que evidencia que el modelo priorizó la recuperación de una proporción elevada de aves, aunque mantuvo algunas predicciones adicionales o delimitaciones ambiguas. La diferencia entre $mAP@0,50$ y $mAP@0,50:0,95$ confirma que la

Tabla 9. Métricas de segmentación del modelo YOLO26n-seg en los conjuntos de validación y prueba.

Fuente: Los autores.

Conjunto	Imágenes	Instancias	Precisión de máscara	Recall de máscara	$mAP@0,50$	$mAP@0,50:0,95$
Validación	307	35.610	0,78845	0,90522	0,89476	0,68705
Prueba	252	29.117	0,78639	0,87559	0,88387	0,67593

Nota. Las cantidades de instancias presentadas en esta tabla corresponden a las anotaciones efectivamente cargadas por Ultralytics después de la revisión humana y de la generación de la caché del conjunto de datos. Durante este proceso se eliminaron 6 duplicados exactos en validación y 7 en prueba; por ello, los conteos utilizados para la evaluación fueron 35.610 y 29.117 instancias, respectivamente.

La Tabla 9 presenta las métricas de segmentación obtenidas por YOLO26n-seg en los subconjuntos de validación y prueba.

principal dificultad no estuvo en reconocer la presencia de los pollos, sino en delimitar con elevada precisión sus contornos



Figura 6. Ejemplo cualitativo de segmentación de instancias realizada por YOLO26n-seg en una imagen del conjunto de prueba.

Fuente: Los autores.

bajo umbrales de IoU más estrictos.

La Figura 6 presenta una comparación entre la imagen original y la predicción generada por YOLO26n-seg sobre una escena del conjunto de prueba. El modelo identificó 138 instancias de pollos de engorde y asignó una máscara individual a cada ave detectada. La visualización evidencia la capacidad del modelo para procesar escenas con alta densidad, múltiples individuos simultáneos y variaciones en la posición de las aves.

La mayor parte de las máscaras se concentró sobre los cuerpos de los pollos y permitió diferenciar individuos próximos entre sí. Sin embargo, en algunas regiones se observaron máscaras parciales o delimitaciones visualmente ambiguas, principalmente en zonas con contacto entre aves, oclusión parcial y límites poco definidos. Estas dificultades pueden relacionarse con la similitud visual entre el plumaje, el suelo y las aves adyacentes.

En conjunto, la predicción muestra que YOLO26n-seg puede segmentar múltiples aves en una sola imagen cenital y mantener una representación individual de gran parte de las instancias presentes. No obstante, la inspección cualitativa confirma que la precisión de los contornos continúa siendo más exigente en regiones densas o visualmente ambiguas.

5. Discusión

Los resultados respaldan la viabilidad de integrar SAM3, limpieza automática, revisión humana y YOLO26n-seg para segmentar pollos de engorde en imágenes cenitales de alta densidad. El desempeño fue estable entre validación y prueba: el $mAP@0,50$ disminuyó de 0,89476 a 0,88387 y el $mAP@0,50:0,95$ de 0,68705 a 0,67593. Estas diferencias reducidas sugieren que la partición por video permitió evaluar el modelo en grabaciones no utilizadas durante el entrenamiento y limitar la fuga de información entre subconjuntos. No obstante, esta estabilidad corresponde al entorno de captura estudiado y no demuestra todavía generalización hacia otras granjas o dispositivos.

En comparación con (Guo et al., 2023), quienes emplearon YOLOv5-CBAM para detectar pollos mediante cajas delimitadoras, el presente estudio proporciona máscaras individuales que representan con mayor detalle la superficie y el contorno aproximado de cada ave. De manera similar, (Bumbálek et al., 2025) compararon diferentes versiones de YOLO para detectar aves muertas, mientras que este trabajo aborda aves vivas en escenas cenitales con más de cien individuos por imagen y evalúa segmentación de instancias. Por tanto, la principal diferencia no radica únicamente en la arquitectura, sino en el nivel de representación: la detección localiza al individuo, mientras que la segmentación exige delimitar su forma espacial. Las métricas no deben compararse directamente debido a las diferencias en clases, conjuntos de datos y condiciones experimentales.

La diferencia entre el $mAP@0,50$ y el $mAP@0,50:0,95$ muestra que el modelo reconoció una proporción elevada de las aves, pero presentó mayores dificultades para reproducir sus contornos

bajo umbrales estrictos de IoU.

Este comportamiento coincide con (Chen et al., 2025), quienes reportaron que la oclusión y la interferencia visual dificultan la segmentación individual de gallinas en imágenes infrarrojas. También guarda relación con (Sun et al., 2025), cuyo método multimodal utilizó segmentación de instancias para analizar gallinas bajo condiciones visuales complejas. En el presente estudio, el contacto físico, la oclusión parcial y la similitud entre el plumaje, el suelo y las aves adyacentes explican razonablemente la menor precisión geométrica de las máscaras.

En relación con los modelos generalistas de segmentación, (Yang et al., 2024) demostraron la utilidad de Segment Anything Model para el monitoreo avícola, mientras que (Saeidifar et al., 2024) aplicaron segmentación zero-shot al análisis térmico individual de gallinas. A diferencia de esos enfoques, SAM3 no se utilizó aquí como modelo final de inferencia, sino como generador de anotaciones iniciales. Las máscaras fueron convertidas, depuradas y revisadas antes de entrenar YOLO26n-seg. En consecuencia, el aporte metodológico consiste en transferir la capacidad de un modelo generalista hacia un modelo compacto y especializado mediante un flujo con supervisión humana.

La limpieza automática eliminó 134 duplicados de las 210.762 anotaciones iniciales, equivalente al 0,0636 %. Esta reducción indica que se detectaron pocas máscaras altamente coincidentes, pero no constituye por sí sola una medida de la calidad global del autoetiquetado, debido a que el procedimiento no evaluó falsos positivos, contornos imprecisos ni aves omitidas. Por esta razón, la revisión humana continuó siendo indispensable para corregir errores semánticos y geométricos antes del entrenamiento.

Desde una perspectiva más amplia, (Reyes Zambrano & Moreira Santos, 2025) identificaron los modelos YOLO y las técnicas de segmentación como herramientas relevantes para la identificación, el seguimiento y el monitoreo de animales de granja. Mientras su estudio ofrece una revisión general de estas tecnologías y sus aplicaciones productivas y sanitarias, la presente investigación aporta una implementación experimental específica en pollos de engorde. En particular, integra la generación inicial de máscaras con SAM3, la depuración automática, la revisión humana y el entrenamiento de YOLO26n-seg dentro de un único flujo metodológico.

Entre las principales limitaciones se encuentran el uso de una sola granja, una única perspectiva cenital, una sola clase y la ausencia de un registro cuantitativo del tiempo y de las modificaciones realizadas durante la revisión humana. Por tanto, todavía no puede afirmarse una reducción demostrada del esfuerzo frente al etiquetado completamente manual. Futuras investigaciones deberían validar el flujo en otras condiciones productivas, comparar sus anotaciones con una muestra elaborada manualmente por especialistas y analizar el desempeño según los niveles de densidad y oclusión.



6. Conclusiones

El flujo metodológico integrado por autoetiquetado con SAM3, conversión de máscaras, limpieza automática, revisión humana y entrenamiento con YOLO26n-seg partió de 1.863 imágenes cenitales extraídas, de las cuales 1.860 conformaron el conjunto final utilizado para la segmentación de instancias de pollos de engorde. SAM3 generó 210.762 anotaciones iniciales y la depuración automática eliminó 134 duplicados, lo que contribuyó al control de calidad previo al entrenamiento. Los resultados del modelo fueron estables entre validación y prueba, con valores de $mAP@0,50$ de 0,89476 y 0,88387, y de $mAP@0,50:0,95$ de 0,68705 y 0,67593, respectivamente. Estos hallazgos indican que el modelo logró identificar y segmentar una proporción elevada de aves en escenas densas, aunque la precisión geométrica de los contornos disminuyó cuando se aplicaron umbrales de IoU más estrictos.

La comparación entre cajas y máscaras mostró que la principal dificultad se concentró en la delimitación precisa de individuos en contacto, parcialmente ocluidos o con límites visuales poco definidos. En consecuencia, la contribución del estudio radica tanto en el desempeño alcanzado como en la integración de un procedimiento completo para generar, depurar y aprovechar anotaciones de segmentación en un entorno avícola real. No obstante, los resultados se limitan a una única granja, una perspectiva cenital y una sola clase, y no se cuantificó el tiempo requerido para la revisión humana. Se recomienda validar el flujo en otras condiciones productivas, comparar su eficiencia con la anotación manual completa y evaluar estrategias orientadas a mejorar la precisión de las máscaras en escenas con mayor densidad y oclusión.

Agradecimientos

Agradecemos a la Universidad Técnica de Machala por el apoyo institucional brindado para el desarrollo de esta investigación, así como por facilitar los recursos académicos y tecnológicos necesarios para la adquisición, procesamiento y análisis de las imágenes. También se reconoce la colaboración de las personas responsables del entorno de crianza avícola, cuya participación permitió obtener el material audiovisual utilizado en el estudio.

Contribución de los autores

Jonathan Aguilar-Alvarado: Investigación, metodología, software, curación de datos y redacción del borrador original. **Karina García-Galarza:** Conceptualización, metodología, supervisión, revisión y edición del manuscrito. **John Patricio Orellana Preciado:** Conceptualización, metodología, supervisión, validación y edición del manuscrito.

Conflictos de interés

Los autores declaran no tener conflictos de interés financieros, profesionales, institucionales o personales que pudieran haber influido en el desarrollo, análisis o presentación de los resultados de esta investigación.

Referencias bibliográficas

- Åkesson, J., Töger, J., & Heiberg, E. (2024). Random effects during training: Implications for deep learning-based medical image segmentation. *Computers in Biology and Medicine*, 180, Article 108944. <https://doi.org/10.1016/j.compbiomed.2024.108944>
- Boniche, K., Cruz, E., Rangel, J. C., Hidalgo-Rodríguez, M., & Gómez-Donoso, F. (2026). PIO, A Large-Scale Dataset for Broiler Chicken Detection under Real Poultry Farming Conditions. *Scientific Data*, 13(1), Article 801. <https://doi.org/10.1038/s41597-026-07114-5>
- Bumbálek, R., Umurungi, S. N., de Dieu Marcel Ufitikirezi, J., Zoubek, T., Kuneš, R., Stehlik, 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), Article 105440. <https://doi.org/10.1016/j.psj.2025.105440>
- Çarklı Yavuz, B. (2026). Scale-Dependent Performance Analysis of YOLO26 and YOLOv11 for PPE Detection. *Electronics*, 15(6), Article 1146. <https://doi.org/10.3390/electronics15061146>
- Chen, Y., Liu, H., Wang, L., Chen, C., Li, S., Zhong, B., Qiao, J., Ye, R., & Li, T. (2025). Research on Instance Segmentation Algorithm for Caged Chickens in Infrared Images Based on Improved Mask R-CNN. *Sensors*, 25(19), Article 6237. <https://doi.org/10.3390/s25196237>
- Cheng, B., Girshick, R., Dollar, P., Berg, A. C., & Kirillov, A. (2021). Boundary IoU: Improving Object-Centric Image Segmentation Evaluation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15334–15342).
- Doornweerd, J. E., Veerkamp, R. F., de Klerk, B., van der Sluis, M., Bouwman, A. C., Ellen, E. D., & Kootstra, G. (2024). Tracking individual broilers on video in terms of time and distance. *Poultry Science*, 103(1), Article 103185. <https://doi.org/10.1016/j.psj.2023.103185>

- Guo, Y., Aggrey, S. E., Yang, X., Oladeinde, A., Qiao, Y., & Chai, L. (2023). Detecting broiler chickens on litter floor with the YOLOv5-CBAM deep learning model. *Artificial Intelligence in Agriculture*, 9, 36–45. <https://doi.org/10.1016/j.aiia.2023.08.002>
- Jena, R., Zhorniyak, L., Doiphode, N., Chaudhari, P., Buch, V., Gee, J., & Shi, J. (2023). Beyond mAP: Towards Better Evaluation of Instance Segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11309–11318.
- Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo Algorithm Developments. In *Procedia Computer Science*, 199, 1066–1073. <https://doi.org/10.1016/j.procs.2022.01.135>
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., & others. (2023). Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, (pp 4015–4026).
- Lamping, C., Derks, M., Groot Koerkamp, P., & Kootstra, G. (2022). ChickenNet - an end-to-end approach for plumage condition assessment of laying hens in commercial farms using computer vision. *Computers and Electronics in Agriculture*, 194, Article 106695. <https://doi.org/10.1016/j.compag.2022.106695>
- Li, J., Feng, Y., Guo, Y., Huang, J., Ji, W., Bi, Q., Piao, Y., Zhang, M., Zhao, X., Chen, Q., Zou, S., Lu, H., & Cheng, L. (2026). SAM3-I: Segment Anything with Instructions. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 27232–27249). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2026.acl-long.1255>
- Lilja, A., Fu, J., Stenborg, E., & Hammarstrand, L. (2024). Localization Is All You Evaluate: Data Leakage in Online Mapping Datasets and How to Fix It. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 22150–22159).
- Lin, Q., Ye, G., Wang, J., & Liu, H. (2022). RoboFlow: a Data-centric Workflow Management System for Developing AI-enhanced Robots. En A. Faust, D. Hsu, & G. Neumann (Eds.), *Proceedings of the 5th Conference on Robot Learning* (Vol. 164, pp. 1789–1794). PMLR. <https://proceedings.mlr.press/v164/lin22c.html>
- Paneru, B., Dhungana, A., Dahal, S., Ritz, C. W., Kim, W., Liu, T., & Chai, L. (2026). Computer vision models for precision poultry farming: A narrative review of behavioral and welfare monitoring studies. *Poultry Science*, 105(7), 106887. <https://doi.org/10.1016/j.psj.2026.106887>
- Reyes Zambrano, J. K., & Moreira Santos, M. G. (2025). Visión por Computadora en el ámbito de control productivo y sanidad en Animales de Granja. *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*, 9(1), 16–29. <https://doi.org/10.33936/isrtic.v9i1.7156>
- Saeidifar, M., Li, G., Chai, L., Bist, R., Rasheed, K. M., Lu, J., Banakar, A., Liu, T., & Yang, X. (2024). Zero-shot image segmentation for monitoring thermal conditions of individual cage-free laying hens. *Computers and Electronics in Agriculture*, 226, Article 109436. <https://doi.org/10.1016/j.compag.2024.109436>
- Sun, R., Wang, Q., Yu, C., Yang, Z., Wu, J., & Fan, W. (2025). A multimodal detection method for caged diseased hens integrating behavioral and thermal features via instance segmentation. *Computers and Electronics in Agriculture*, 239, Article 110926. <https://doi.org/10.1016/j.compag.2025.110926>
- Yang, X., Dai, H., Wu, Z., Bahadur Bist, R., Subedi, S., Sun, J., Lu, G., Li, C., Liu, T., & Chai, L. (2024). An innovative segmentanything model for precision poultry monitoring. *Computers and Electronics in Agriculture*, 222, Article 109045. <https://doi.org/10.1016/j.compag.2024.109045>