



DETECCIÓN Y CONTEO DE ESPIGAS DE MAÍZ CON INTELIGENCIA ARTIFICIAL EN IMÁGENES AÉREAS DE DRON

DETECTION AND COUNTING OF MAIZE TASSELS USING ARTIFICIAL INTELLIGENCE IN AERIAL DRONE IMAGES

Rodrigo Santacruz-Martínez, Héctor Flores-Magdaleno*, María Guadalupe López-Pérez, Nora Meraz-Maldonado y Roberto Ascencio-Hernández

Colegio de Postgraduados, Campus Montecillo, Hidrociencias, Montecillo, Texcoco, Estado de México, México.

*Autor de correspondencia (mhector@colpos.mx)

RESUMEN

La producción de semilla híbrida de maíz (*Zea mays* L.) es esencial a nivel global, pero su pureza genética depende del desespigamiento riguroso de los progenitores femeninos, con un umbral de tolerancia inferior al 1 % de liberación de polen. Los métodos de desespigamiento actuales son costosos, intensivos en mano de obra o ineficaces ante variaciones de altura y densidad. Como una alternativa para evaluar el desespigamiento, el objetivo de esta investigación fue desarrollar dos modelos de visión por computadora basados en la arquitectura YOLOv12m para la detección y conteo automatizado de espigas de maíz a partir de imágenes aéreas tomadas con dron, analizando el impacto del ángulo de captura (Nadir u Oblicuo) en el desempeño de la detección. Se generó una base de datos con información colectada en múltiples regiones, con varias variedades de maíz y condiciones de siembra, lo cual se considera crucial para la capacidad de generalización del modelo. Los datos fueron preprocesados y aumentados, resultando en 226 imágenes Nadir y 4900 imágenes Oblicuo. Se entrenaron dos modelos basados en YOLOv12m mediante *fine-tuning*. El Modelo Oblicuo superó sistemáticamente al Modelo Nadir en todas las métricas de desempeño. El Modelo Oblicuo alcanzó una precisión media promedio (mAP50) de 0.995 y un F1 Score de 0.983, mientras que el Modelo Nadir obtuvo un mAP50 de 0.886. Se concluye que el Modelo Oblicuo establece un nuevo punto de referencia en rendimiento para la detección de espigas; no obstante, el Modelo Nadir sigue siendo la mejor aproximación metodológica para el conteo preciso y la estimación de densidad en cultivos densos, ya que la vista cenital minimiza la oclusión por el follaje superior.

Palabras clave: *Zea mays* L., agricultura de precisión, desespigamiento, detección de objetos, VANT.

SUMMARY

Production of hybrid maize (*Zea mays* L.) seed is essential globally, but its genetic purity depends on the rigorous detasseling of female parents, with a tolerance threshold of less than 1 % pollen shedding. Current detasseling methods are expensive, labor-intensive, or inefficient at variations in plant height and plant density. As an alternative to assess detasseling, the objective of this research was to develop two computer vision models based on the YOLOv12m architecture for the automated detection and counting of maize tassels from drone aerial imagery, analyzing the impact of the capture angle (Nadir or Oblique) on detection performance. A dataset was constructed from information collected across multiple regions, with various maize varieties and

planting conditions, which is considered crucial for the models capacity for generalization. The data were preprocessed and augmented, resulting in 226 Nadir images and 4900 Oblique images. Two models based on YOLOv12m were trained by fine-tuning. The Oblique model systematically outperformed the Nadir model across all performance metrics. The Oblique Model achieved a mean average precision (mAP50) of 0.995 and an F1 Score of 0.983, while the Nadir Model achieved an mAP50 of 0.886. It is concluded that the Oblique Model sets a new performance benchmark for tassel detection. Nevertheless, the Nadir Model remains the preferred methodological approach for precise counting and density estimation in dense crops, as the Nadir view minimizes occlusion by the upper foliage.

Index words: *Zea mays* L., detasseling, object detection, precision agriculture, UAV.

INTRODUCCIÓN

A nivel global, el maíz (*Zea mays* L.) se posiciona como un cultivo esencial, con alrededor del 40 % de la producción mundial de los principales granos (FAO, 2024). La relevancia de este cultivo trasciende la alimentación humana, convirtiéndose en un recurso importante para la alimentación animal, producción de biocombustibles y obtención de materias primas industriales. La adopción masiva del maíz híbrido, que se caracteriza por su notable rendimiento y adaptabilidad, ha consolidado su posición, dominando más de dos tercios de la superficie sembrada de maíz en todo el planeta (Erenstein et al., 2022).

En México, durante los ciclos otoño-invierno 2023-2024 y primavera-verano 2024 se registraron 25,347.12 ha para producción de semilla híbrida, y en dicha superficie se produjeron 110,595.75 t de semilla (SNICS, 2024), lo que resalta la importancia de monitorear y hacer más eficiente el proceso de desespigamiento.

La pureza genética es el indicador de calidad más crítico en la producción anual de semilla híbrida de maíz.

Para salvaguardarla, el desespigamiento debe ejecutarse con absoluta precisión en un periodo crítico de apenas 7 a 10 días. En la industria contemporánea, se emplea una combinación de desespigamiento mecánico –utilizando máquinas como la Hagie 204SP o la Bourgoin BD864– y manual. No obstante, debido a variaciones en la altura de las plantas, la maquinaria suele alcanzar una tasa de remoción de solo el 80 al 90 %, lo que hace indispensables múltiples rondas de desespigamiento manual para asegurar que la liberación acumulada de polen no supere el umbral del 1 %, cumpliendo así con los estándares internacionales de certificación (Yang *et al.*, 2025; Zhang *et al.*, 2024).

En los últimos años, los vehículos aéreos no tripulados (VANTs), también conocidos como drones, se han convertido en herramientas valiosas para la agricultura, especialmente para la teledetección de cultivos. Su creciente aceptación en el área agrícola se debe a una combinación de factores clave: la rapidez con que pueden cubrir grandes extensiones de terreno, su capacidad de maniobra que les permite adaptarse a diversos tipos de topografías, la nula compactación del suelo donde operan, y el imperceptible daño que infligen a las plantas en comparación con métodos tradicionales (Wang *et al.*, 2025). Estas ventajas han posicionado a los drones como aliados estratégicos para optimizar la gestión de recursos y la toma de decisiones informadas en el sector agrícola.

El modelo YOLO (You Only Look Once), presentado por Redmond *et al.* (2016) constituye un enfoque unificado para la detección de objetos que replantea esta tarea como un único problema de regresión, el cual mapea directamente los píxeles de la imagen hacia las coordenadas de cajas delimitadoras y sus respectivas probabilidades de clase. Su relevancia en la visión por computadora se debe a su extrema rapidez, alcanzando procesamientos en tiempo real de 45 cuadros por segundo en su versión base. Sus ventajas principales incluyen una reducción drástica de errores de fondo (falsos positivos) gracias a su razonamiento contextual global al ver la imagen completa, además de poseer una alta capacidad de generalización.

A partir de su presentación en 2016, diferentes versiones de YOLO se han liberado, cada una optimizando la precisión, velocidad y eficiencia computacional. Las versiones más recientes como YOLOv8 (Varghese y Sambath, 2024), YOLOv11 (Khanam y Hussain, 2024) y el recientemente publicado YOLOv12 (Khanam y Hussain, 2025) han establecido nuevos estándares en la detección de objetos, especialmente en entornos complejos. La arquitectura de YOLO facilita su personalización, convirtiéndolo en una opción popular en aplicaciones agrícolas como estimación de rendimiento en frutales (Mirhaji *et al.*, 2021), detección

temprana de enfermedades (Kapetas *et al.*, 2025), control de calidad de granos de arroz (Liang *et al.*, 2025), detección de variedades y defectos en semilla de maíz (Che *et al.*, 2025), detección de malezas, detección de cultivos, seguimiento de animales y agricultura de precisión (Alif *et al.*, 2024).

En cuanto a la detección de espigas de maíz con imágenes tomadas desde un dron, Pu *et al.* (2023) modificaron YOLOv7 y propusieron un modelo llamado Tassel-YOLO donde se modificó la arquitectura del modelo original, con la cual se logró una precisión mAP@050 de 94.71 % y con dicha modificación fue mejorada a 96.14 %. Paralelamente, Jia *et al.* (2024) propusieron un modelo mejorado de YOLOv5 al que llamaron CA-YOLO, donde se modificó el mecanismo de atención (CA) con el propósito de mejorar la capacidad de regresión y localización del modelo, consiguiendo una precisión mAP@050 del 96 %, mejorando en 1 %, en contraste con el modelo sin modificar (96 vs. 95 %), demostrando así el potencial de YOLO para automatizar la detección y conteo de espigas de maíz.

Ambos estudios mencionados en el párrafo anterior basaron el entrenamiento y evaluación de sus modelos en datos obtenidos de una sola ubicación específica, aunque intentaron incluir diversidad dentro de esa ubicación como condiciones de luz y altura de captura de imagen, en ambos estudios los autores reconocen esta limitación y sugieren la recolección de datos de diferentes variedades de maíz y múltiples regiones para mejorar la generalidad de los modelos; es decir, la capacidad de un modelo para desempeñarse de manera efectiva y precisa en datos nuevos, no vistos durante el proceso de entrenamiento. Por lo anterior, en este estudio se proponen dos modelos, que en lo sucesivo se referirán como Modelo Nadir y Modelo Oblicuo, basados en YOLO para la detección y conteo automático de espigas de maíz a partir de un conjunto de datos obtenido de diferentes regiones, ángulos de captura de imagen, variedades de maíz, densidades de siembra, condiciones de luz y altura de captura de imagen para que el modelo resultante tenga la capacidad de generalizar.

MATERIALES Y MÉTODOS

Captura de imágenes

La fase inicial de la investigación consistió en la captura de imágenes aéreas para generar una base de datos representativa. Para la adquisición de los datos se empleó un dron (Modelo DJI Mini 2 SE, DJI, Shenzhen, China), el cual está equipado con un sensor de cámara CMOS de 1/2.3" con 12 millones de píxeles efectivos. Las imágenes aéreas se capturaron utilizando dos ángulos de visión distintos: Nadir (vista cenital) y Oblicuo (con la cámara posicionada

a 45 ° respecto al dron). Las tomas en Nadir ofrecen una vista cenital que minimiza la oclusión por el follaje superior, facilitando la identificación individual y siendo la mejor aproximación metodológica para el conteo preciso y la estimación de densidad. En contraste, las tomas oblicuas brindan una vista lateral valiosa para capturar la forma y estructura tridimensional de las espigas, lo cual es crucial para distinguir las espigas que podrían estar ocultas desde una vista cenital.

Para las imágenes capturadas en ángulo oblicuo, la altura de vuelo desde el nivel del suelo se mantuvo constante a 5 m. Para las imágenes capturadas en ángulo Nadir (vista cenital), la adquisición de datos se realizó a alturas variables de 7, 10, 12 y 15 m desde el nivel del suelo, buscando enriquecer la información cubierta y mejorar la robustez del modelo al exponerlo a diversas escalas de espigas. Debido a las fortalezas inherentes de cada perspectiva, se optó por entrenar los dos modelos específicos de detección (Modelo Nadir y Modelo Oblicuo) a fin de explotar las características visuales únicas de cada tipo de toma.

Recolección de datos

La fase inicial para el desarrollo de un modelo de detección y conteo de objetos, en este caso espigas de maíz, basado en la arquitectura YOLO es la creación de una base de datos representativa. Con el fin de asegurar que el modelo resultante tenga una capacidad efectiva de generalizar; es decir, que pueda desempeñarse con precisión en datos no vistos antes, se colectaron datos procurando incluir diversidad en las condiciones de captura. Esto abarcó imágenes tomadas en diferentes localidades y regiones, con distintas variedades de maíz y en diferentes densidades de siembra. La colecta de datos se realizó durante el periodo de espigamiento del maíz, la etapa de interés para el desarrollo de un modelo de detección de espigas de maíz. En el Cuadro 1 se detallan los sitios donde se colectaron imágenes aéreas, sus coordenadas geográficas, variedad de maíz y densidad de siembra en cada sitio al momento de la colecta.

Conjunto de datos

Una vez completada la fase de colecta de datos, el siguiente paso para el desarrollo del modelo de detección de objetos fue la preparación de un conjunto de datos etiquetados como el insumo para el entrenamiento del modelo. El etiquetado de imágenes es el proceso de definir y localizar el objeto de interés en las imágenes para que el algoritmo de detección de objetos utilice esta información como el conocimiento de referencia durante el entrenamiento del modelo.

Del total de imágenes recolectadas para la conformación del conjunto de datos, se seleccionaron las imágenes de mayor calidad y pertinencia con base en tres criterios técnicos principales: 1) nitidez visual, descartando aquellas tomas con desenfoque (*motion blur*) provocado por el movimiento del dron o la vibración por viento; 2) condiciones de iluminación, seleccionando imágenes con una exposición equilibrada que evitara zonas sobreexpuestas o sombras densas que ocultaran la textura de la espiga; y 3) integridad representativa, priorizando imágenes donde los objetos de interés fueran claramente discernibles frente al dosel vegetal bajo las diferentes escalas. Se etiquetó manualmente cada espiga en cada imagen del conjunto de datos. En el Cuadro 2 se presentan las características del conjunto de datos tomadas en ángulo Nadir y oblicuo.

El redimensionamiento de las imágenes es un requisito para la mayoría de las arquitecturas para la detección de objetos. Al transformar todas las imágenes a un tamaño fijo de píxeles se garantiza que el modelo reciba lotes de datos con dimensiones consistentes. Las imágenes del conjunto de datos capturadas en ángulo de captura oblicuo se redimensionaron a un tamaño de 640 × 640 píxeles, mientras que las imágenes del conjunto de datos capturadas en ángulo Nadir se redimensionaron a un tamaño de 1152 × 2048 píxeles.

Aumento de datos

El aumento de datos es una técnica ampliamente usada en el área de visión por computadora, como la detección de objetos. Consiste en generar nuevas imágenes para el entrenamiento a partir de las imágenes existentes a través de transformaciones aleatorias, estas transformaciones crean versiones modificadas de las imágenes originales que el modelo interpreta como datos nuevos (Chollet, 2021).

El propósito principal de la aumentación es incrementar la diversidad de la base de datos. Al exponer al modelo a una mayor variabilidad en la apariencia de los objetos de interés, en este caso espigas de maíz, se obtendrá un modelo más robusto y con una mayor capacidad de generalización. Esta técnica es de gran utilidad para prevenir el sobreajuste (*overfitting*) en el área de visión por computadora, que ocurre cuando el modelo aprende demasiado bien los detalles específicos de la base de datos, pero falla al enfrentarse a datos nuevos o ligeramente diferentes.

En el contexto de la detección de espigas de maíz en entornos de campo complejos, como se aborda en este estudio, la variabilidad de las imágenes puede ser

Cuadro 1. Sitios de colecta de imágenes.

ID	Sitio	Localidad	Ubicación	Variedad (Empresa)	Densidad (plantas ha ⁻¹)
1	Módulo Producción de Leche en Pastoreo	Chapingo [†]	19° 29' 17.0" N, 98° 52' 46.0" O	NK858 (Syngenta)	85,000
2	Campo Experimental San Juan	Chapingo [†]	19° 29' 27.0" N, 98° 51' 09.0" O	Antílope (Asgrow)	75,000
3	Campo Experimental San Juan	Chapingo [†]	19° 29' 27.0" N, 98° 51' 10.0" O	Comanche (Aspros)	90,000
4	Campo Experimental San Juan	Chapingo [†]	19° 29' 32.0" N, 98° 51' 33.0" O	Tigrillo (Asgrow)	85,000
5	San Nicolas de los Agustinos	Guanajuato ^{††}	20° 15' 09.6" N, 100° 57' 05.7" O	Berrendo (Asgrow)	100,000
6	San Nicolas de los Agustinos	Guanajuato ^{††}	20° 15' 06.6" N, 100° 57' 41.5" O	Antílope (Asgrow)	90,000
7	San Nicolas de los Agustinos	Guanajuato ^{††}	20° 15' 35.4"N, 100° 57' 48.9" O	NK925W (Syngenta)	100,000

[†]: Universidad Autónoma Chapingo, Chapingo, Estado de México, México, ^{††}: Ejido San Nicolás de los Agustinos, Salvatierra, Guanajuato, México.

Cuadro 2. Conjunto de datos etiquetados.

Ángulo de captura	Número de imágenes	Número de etiquetas	Promedio de etiquetas por imagen
Nadir	92	17,939	195
Oblicuo	511	9612	18.8

considerable debido a factores como las diferentes etapas de crecimiento de las espigas, condiciones de iluminación variables, espigas obstruidas por hojas y las posibles alturas de captura de imagen (Jia *et al.*, 2024). La aumentación de datos simula muchas de estas variaciones, mejorando el desempeño del modelo en situaciones del mundo real.

En este estudio, la aumentación de datos se aplicó sobre las imágenes originales mediante la implementación de transformaciones aleatorias. En el Cuadro 3 se definen las técnicas de aumentación y su rango de valores, las transformaciones se realizaron aleatoriamente dentro de ese rango, lo cual simula la complejidad inherente de los entornos en campo.

Al aplicar las técnicas de aumento de datos del Cuadro 3, el conjunto de datos para el modelo Nadir aumentó de 92 a 226 imágenes, mientras que el conjunto de datos para el modelo Oblicuo aumentó de 511 a 4900 imágenes.

Entorno de desarrollo

El entorno de desarrollo usado para la construcción de los modelos propuestos en este estudio constó de una computadora con sistema operativo Windows 10, con una GPU Nvidia Titan Xp 12Gb, procesador Intel Xeon @ Ghz,

128Gb RAM DDR4, usando Python 3.12.6, Pytorch 2.7.1 en el IDE PyCharm 2024.2.1.

Conjuntos de entrenamiento, prueba y validación

El diseño metodológico de cualquier estudio de *deep learning* que aspire a una validez científica impone la segmentación de la base de datos. Esta práctica es fundamental para evaluar con precisión la capacidad de generalización de un modelo; es decir, su habilidad para desempeñarse de manera efectiva y precisa en datos nuevos, no vistos durante su fase de entrenamiento. Sin esta división, se corre el riesgo de un sobreajuste, donde el modelo memoriza el conjunto de datos de entrenamiento en lugar de aprender patrones subyacentes, resultando en un modelo deficiente.

Las tres secciones en las que se divide una base de datos para un proyecto de esta naturaleza son:

- 1. **Conjunto de entrenamiento.** Esta es la porción más extensa de la base de datos y su función primordial es la de nutrir el proceso de aprendizaje del modelo. El algoritmo de *deep learning* (en este caso, un modelo YOLO para detección de objetos) examina los datos de entrada junto con sus etiquetas correspondientes

Cuadro 3. Técnicas de aumento de datos y rango de valores.

Ángulo de captura	Acercamientos	Rotaciones	Cortes	Brillo	Tono	Saturación	Exposición	Ruido (% max.)	Recortes (%)
Oblicuo	0 – 15 %	± 15°	± 10°	± 10°	NA	± 15°	± 5°	0.1	2-10
Nadir	NA	± 15°	± 10°	± 15°	± 15°	± 22°	± 8°	0.1	3-10

(las cajas delimitadoras que definen la ubicación de las espigas de maíz). Durante esta fase, la red neuronal ajusta iterativamente sus parámetros internos para identificar las características visuales que definen una espiga (forma, textura, color, contexto), aprendiendo a predecir con exactitud dónde dibujar las cajas delimitadoras. En síntesis, el conjunto de entrenamiento es el manual que el modelo utiliza para aprender los patrones inherentes a las espigas de maíz.

2. **Conjunto de validación.** Este subconjunto de datos se utiliza para monitorizar el rendimiento del modelo durante el entrenamiento y, crucialmente, para ajustar sus hiperparámetros, que son configuraciones externas que no se aprenden directamente de los datos, como la tasa de aprendizaje o el número de épocas (Chollet, 2021). El modelo no aprende de este conjunto de datos; en cambio, sus métricas de rendimiento (como la precisión o la pérdida) se calculan periódicamente sobre el mismo para proporcionar una evaluación imparcial de cómo el modelo está generalizando a medida que se entrena. Esto permite detectar señales de sobreajuste de manera temprana, evitando que el modelo se especialice demasiado en los datos de entrenamiento y pierda su capacidad de desempeño en datos no vistos.
3. **Conjunto de prueba.** Es la porción final e independiente de la base de datos, utilizada para realizar una evaluación definitiva y sin sesgos del rendimiento del modelo, una vez que el entrenamiento y el ajuste de hiperparámetros han concluido. El modelo nunca interactúa con estos datos durante el entrenamiento o la validación, lo que garantiza que la evaluación de su desempeño sea una estimación lo más fidedigna posible de la forma en que se comportará en un entorno de producción con datos completamente nuevos. Su propósito es simular la aplicación del modelo en el mundo real, ofreciendo una medida robusta de su capacidad de generalización.

Para el conjunto de datos Nadir, el 88.9 % (201/226) de las imágenes se destinaron a entrenamiento, el 7.5 % (17/226) a prueba y el 3.5 % (8/226) a validación. De

manera similar, para el conjunto de datos Oblicuo, el 87.4 % (4284/4900) se utilizó para entrenamiento, el 8.4 % (412/4900) para prueba y el 4.2 % (204/4900) para validación. La partición de los conjuntos de datos se fundamentó en la necesidad de proporcionar una densidad de muestreo suficiente para que la arquitectura YOLOv12m logre aproximar una superficie continua que representa visualmente a las espigas de maíz en entornos complejos (Chollet, 2021). Debido a la alta capacidad de representación del modelo, se asignó una proporción mayor al entrenamiento (aproximadamente 88 %) para asegurar que la red aprenda características robustas y logre un ajuste óptimo antes de que ocurra el sobreajuste. Por otro lado, la reserva de subconjuntos independientes de validación y prueba responde al protocolo estándar de evaluación para evitar fugas de información, asegurando que el rendimiento reportado sea una medida fidedigna de la capacidad del modelo para interpolar en datos nunca vistos (Chollet, 2021).

Hiperparámetros

Los hiperparámetros son variables de configuración externa al proceso de entrenamiento del modelo con el conjunto de entrenamiento, los hiperparámetros deben ser definidos antes del entrenamiento. La selección adecuada de los hiperparámetros es crucial, ya que influyen directamente en la velocidad de convergencia, la estabilidad y el rendimiento final del modelo, incluida su capacidad de generalización a datos no vistos. El Cuadro 4 resume los hiperparámetros utilizados para el entrenamiento de los modelos Nadir y Oblicuo.

Métricas de evaluación

Para cuantificar el rendimiento de los modelos, se emplearon métricas estandarizadas en el campo de la visión por computadora y la detección de objetos, mismas que se describen a continuación:

Precisión (P). Esta métrica cuantifica la proporción de detecciones positivas que son realmente correctas (Pu *et al.*, 2023). En el contexto de la detección de espigas, la precisión mide la probabilidad de que un objeto identificado como espiga sea una espiga verdadera. Se calcula como:

Cuadro 4. Hiperparámetros utilizados para el entrenamiento de los modelos.

Base de datos	Modelo base	Épocas	Tamaño imagen	Tamaño de lote	Tasa de aprendizaje	Optimizador	Regularización L2
Nadir	YOLO12m	600	1152	1	0.02	AdamW	0.0005
Oblicuo	YOLO12m	600	640	7	0.01	SGD	0.0005

$$P\ (\%) = \frac{TP}{(TP + FP)} \times 100$$

Donde TP son los verdaderos positivos (detecciones correctas) y FP son los falsos positivos (detecciones incorrectas).

Recall (R). Mide la capacidad del modelo para identificar todas las espigas de maíz presentes en una imagen. Se calcula como:

$$R\ (\%) = \frac{TP}{(TP + FN)} \times 100$$

Donde FN son los falsos negativos (espigas no detectadas).

F1-Score: Esta métrica es la media armónica de la precisión y el *recall*, ofreciendo un balance entre ambas (Pu et al., 2023). Es particularmente útil cuando se busca un equilibrio entre la cantidad de detecciones correctas y la capacidad de encontrar todas las instancias relevantes, siendo su fórmula:

$$F1\ (\%) = 2 \frac{PR}{P + R} \times 100$$

Intersección sobre Unión (IoU - Intersection over Union)
El IoU es una medida de cuánto se superpone la caja delimitadora predicha por el modelo y la caja delimitadora real. Su valor oscila entre 0 y 1, donde 0 indica que no hay solapamiento y 1 representa una coincidencia perfecta. En la detección de objetos, el IoU se utiliza como umbral para determinar si una detección es un verdadero positivo (IoU > umbral) o un falso positivo (IoU ≤ umbral) (Kanna et al., 2024).

Promedio de precisión media (mAP - Mean Average Precision). El mAP es una métrica global que promedia los valores de la precisión media (AP) para todas las clases detectadas, dado que en este estudio el modelo se entrenó para detectar una única clase ('espiga'), el mAP puede interpretarse directamente como la precisión general del modelo.

mAP50. Se refiere al *mean Average Precision* (promedio de la precisión media) calculado con un umbral de IoU de

0.5. El mAP50 es el estándar más común y ampliamente aceptado para medir la precisión en los modelos de detección de objetos, reflejando un rendimiento robusto en la localización de objetos.

RESULTADOS Y DISCUSIÓN

En esta sección se presentan los resultados obtenidos de la evaluación de los modelos de detección de espigas de maíz basados en la arquitectura YOLO, entrenados con imágenes aéreas capturadas con dron. Se exploró el impacto de los diferentes ángulos de captura (Nadir y oblicuo) en el desempeño de la detección y se analizaron los hallazgos a la luz de la literatura existente.

Rendimiento general y comparación entre los modelos Nadir y Oblicuo

Esta sección se centra en la presentación y análisis de los resultados cuantitativos obtenidos del entrenamiento y evaluación de los dos modelos de detección y conteo de espigas de maíz desarrollados en este estudio, ambos basados en la arquitectura YOLO. Se evaluó el impacto del ángulo de captura de las imágenes aéreas (Nadir y oblicuo) en el desempeño de la detección, dado que las diferentes perspectivas presentan patrones visuales diferentes al capturar espigas de maíz.

El modelo Nadir fue entrenado con imágenes de vista cenital, una perspectiva que minimiza la oclusión de las espigas por el follaje superior, lo que favorece una identificación individual más clara y una estimación precisa de la densidad. Por su parte, el modelo Oblicuo fue desarrollado utilizando tomas anguladas (45°), buscando capturar la forma y estructura tridimensional de las espigas, lo cual es crucial para distinguir aquellas que podrían estar ocultas desde una vista Nadir o situadas en diferentes planos de profundidad dentro del dosel vegetal.

Tras el entrenamiento y validación de los modelos Nadir y Oblicuo, se obtuvieron los desempeños cuantitativos que se detallan en el Cuadro 5. La evaluación se centró en la capacidad de localización y clasificación de las espigas mediante el análisis de la Precisión (P), el Recall (R), el F1 Score y el Promedio de Precisión Media (mAP50).

El análisis comparativo de las métricas de rendimiento expuestas en el Cuadro 5 revela una diferencia sustancial y sistemática en el desempeño entre los dos modelos de detección desarrollados, contrario a la expectativa de que la vista cenital (Nadir) pudiera ofrecer una ventaja clara al minimizar la oclusión por el follaje superior, el modelo Oblicuo superó consistentemente al modelo Nadir en todas las métricas de evaluación.

El modelo Oblicuo demostró una precisión excepcionalmente alta, de 0.994 (una mejora de 12% respecto al Nadir), lo que significa que casi todas las detecciones positivas fueron correctas, minimizando drásticamente los falsos positivos. Su valor de recall de 0.973 (una mejora de 17 %) indica una capacidad superior para identificar la gran mayoría de las espigas presentes en el campo, reduciendo la omisión de objetivos (falsos negativos). El F1 Score del modelo Oblicuo alcanzó 0.983, reflejando un equilibrio superior entre precisión y exhaustividad. En cuanto al estándar de la industria, el mAP50 (calculado con un umbral IoU de 0.5), el modelo Oblicuo obtuvo 0.995, superando al Nadir (0.886) en un 12 %.

La vista oblicua (45 °) es valiosa porque captura la forma y la estructura tridimensional de las espigas, lo que ayuda a distinguir aquellas espigas que no están completamente visibles desde una vista cenital porque se encuentran en diferentes planos de profundidad dentro del dosel vegetal.

Para situar los modelos desarrollados (Nadir y Oblicuo) en el contexto de la investigación actual sobre la detección de espigas de maíz mediante visión por computadora,

se procedió a compararlos con modelos de referencia del estado del arte que utilizan la arquitectura YOLO en imágenes aéreas tomadas con dron. Específicamente, se seleccionaron Tassel-YOLO (Pu *et al.*, 2023), un modelo optimizado a partir de YOLOv7 y CA-YOLO (Jia *et al.*, 2024), una mejora del modelo YOLOv5. Esta comparación se justifica por el uso común de la arquitectura YOLO en este campo, lo que permite un contraste justo de las métricas de rendimiento clave: precisión, recall, mAP50 y F1 Score. El Cuadro 6 presenta las métricas de desempeño comparadas entre los modelos internos de este estudio y modelos de referencia externos.

La comparación demuestra que el modelo Oblicuo establece un nuevo punto de referencia en el rendimiento (mAP50 de 0.995) para la detección de espigas de maíz a partir de imágenes aéreas, lo que sugiere que la combinación de un ángulo de captura que favorece la localización estructural (oblicuo) y una base de datos diversa proporciona una ventaja crucial sobre los modelos previamente publicados en la literatura.

Ahora bien, al contrastar el modelo Nadir (mAP50 de 0.886) con los modelos del estado del arte, como Tassel-YOLO (mAP50 de 0.961) y CA-YOLO (mAP50 de 0.960), se puede argumentar que estos últimos superan ligeramente al modelo Nadir en términos de métricas de precisión absolutas. Esta aparente superioridad del estado del arte debe interpretarse críticamente considerando las metodologías de recolección y validación de los datos utilizados en cada estudio. Tanto el modelo Tassel-YOLO como CA-YOLO basaron su entrenamiento y validación

Cuadro 5. Rendimiento y comparación de los modelos Nadir y Oblicuo.

Métrica	Modelo Nadir	Modelo Oblicuo	Diferencia porcentual
Precisión (P)	0.889	0.994	12
Recall (R)	0.831	0.973	17
mAP50	0.886	0.995	12
F1 Score	0.859	0.983	14

Cuadro 6. Comparación de los modelos Nadir y Oblicuo con modelos del estado del arte.

Modelo	Precisión (P)	Recall (R)	mAP50	F1 Score
Modelo Nadir	0.889	0.831	0.886	0.859
Tassel-YOLO	0.931	0.932	0.961	0.932
CA-YOLO	0.917	0.92	0.96	0.918
Modelo Oblicuo	0.994	0.973	0.995	0.983

en datos obtenidos de una sola ubicación específica. El conjunto de datos de CA-YOLO fue recolectado de la Base Xiangyang de la Universidad Agrícola del Noreste en Harbin, China y el conjunto de datos de Tassel-YOLO se obtuvo de la Base de Investigación y Desarrollo Agrícola Moderna de la Universidad Agrícola de Sichuan en Chengdu, China. Aunque estos estudios procuraron incluir diversidad en las condiciones de luz y altura de captura dentro de su sitio de estudio, sus autores reconocieron esta limitación y sugirieron la recolección de datos de diferentes variedades de maíz y múltiples regiones para mejorar la capacidad de generalización de sus modelos. En contraste, el modelo Nadir de este estudio fue entrenado y evaluado con el propósito explícito de obtener una alta capacidad de generalización. La base de datos del modelo Nadir se compone de imágenes capturadas de múltiples sitios (incluyendo Salvatierra, Guanajuato, y varios sitios de la Universidad Autónoma Chapingo, México), variedades de maíz (NK858, Antilope, Comanche, Tigrillo y NK925W), bajo densidades de siembra variadas, y diversas condiciones de tiempo atmosférico e iluminación.

Es crucial destacar que el conjunto de validación del modelo Nadir se seleccionó de manera aleatoria a partir de la base de datos diversa, asegurando que la evaluación final del rendimiento (sus métricas P, R, mAP50 y F1 Score) fuera una estimación robusta de su desempeño en un entorno de producción con datos completamente nuevos. Por lo tanto, mientras que las métricas del modelo Nadir puedan ser numéricamente inferiores a las de Tassel-YOLO o CA-YOLO en mAP50, su desempeño demuestra la capacidad de generalizar de manera efectiva en un entorno agronómico y geográfico significativamente más complejo y variado que el utilizado para validar los modelos del estado del arte. La heterogeneidad de la base de datos expone al Modelo Nadir a un desafío considerablemente mayor.

En la Figura 1 se presentan ejemplos representativos de las detecciones; en la Figura 1A se ilustra la capacidad del Modelo Nadir para aislar individuos al minimizar la oclusión por el follaje superior, mientras que en la Figura 1B se ilustran las detecciones por el Modelo Oblicuo, el cual fue entrenado para delimitar espigas respecto a su estructura tridimensional captada en cámara.

Análisis de entrenamiento modelo Oblicuo

El modelo Oblicuo se entrenó utilizando la arquitectura YOLOv12m, un modelo preentrenado que fue ajustado mediante *fine-tuning* durante 600 épocas. El mejor resultado se obtuvo en la época 595, momento en el que se guardó el mejor modelo. La validación final de este

modelo consiguió una Precisión (P) de 0.994, Recall (R) de 0.973, mAP50 de 0.995 y F1 Score de 0.983.

El entrenamiento de un modelo de detección de objetos implica la minimización de una función de pérdida que cuantifica la discrepancia entre las predicciones del modelo y las etiquetas creadas manualmente durante el proceso de etiquetado, las cuáles usarán el modelo como referencia para ajustar sus parámetros. La memoria del entrenamiento registra los siguientes componentes de pérdida: 1) Pérdida de localización, mide el error de localización y el ajuste de las cajas delimitadoras, 2) Pérdida de clasificación, mide el error de clasificación, en este caso identificar correctamente la clase 'espiga', y 3) Pérdida focal de distribución, esta pérdida tiene el propósito de refinar la regresión de las cajas delimitadoras.

En la Figura 2 se observa una disminución constante y significativa en los tres componentes de pérdida, desde la época 1 hasta el término del entrenamiento en la época 600. Esta tendencia a la baja es resultado del ajuste iterativo de la red neuronal. La reducción gradual en la pérdida de localización y en la pérdida de clasificación indica que el modelo mejoró progresivamente en la localización precisa de las espigas, mientras que la disminución de la pérdida focal de distribución muestra una mayor confianza en la clasificación de las espigas detectadas. La convergencia de las pérdidas hacia valores muy bajos en las últimas épocas sugiere que el modelo ha aprendido de manera efectiva los patrones visuales de la base de datos oblicua. En cuanto a las métricas de desempeño (precisión, recall y mAP50), estas muestran un crecimiento acelerado durante las primeras 100 épocas y continúan mejorando hasta estabilizarse en un rendimiento excepcional cerca del final del entrenamiento.

Análisis de entrenamiento modelo Nadir

El modelo Nadir se entrenó utilizando la arquitectura YOLOv12m, un modelo preentrenado que fue ajustado mediante *fine-tuning* durante 397 épocas, aunque el mejor resultado de observó en la época 297, el proceso se ejecutó con la instrucción de 600 épocas con la técnica de suspensión temprana (*Early stopping*), que consiste en detener automáticamente el entrenamiento si la métrica de validación no muestra mejora durante 100 épocas consecutivas; con esta técnica, se evita el sobreajuste del modelo, mencionado antes. El sobreajuste ocurre cuando el modelo aprende los detalles específicos y el ruido de los datos de entrenamiento a tal grado que pierde su capacidad de generalizar a datos nuevos y no vistos (Ultralytics, 2024).

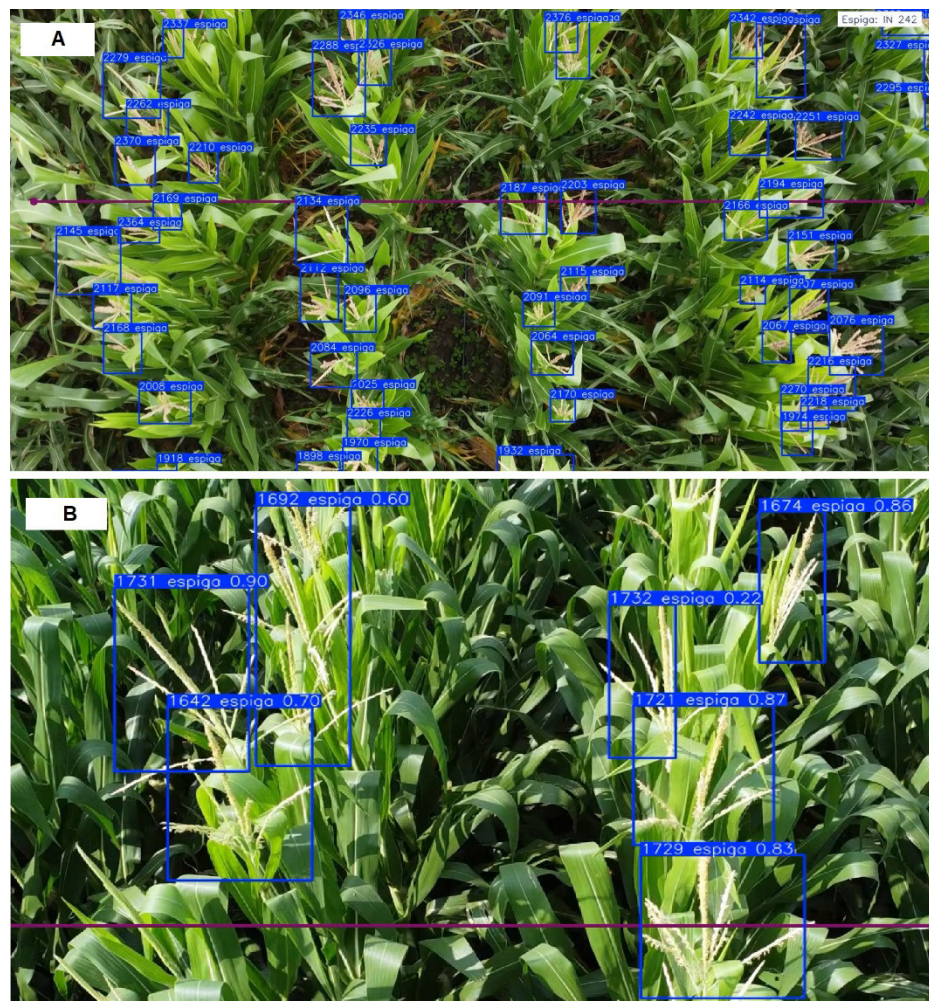


Figura 1. Detecciones de espigas en imágenes aéreas tomadas con dron. A) Detección con modelo Nadir, B) Detección con modelo Oblicuo.

Los resultados más destacados del modelo Nadir se obtuvieron en la época 297, con una Precisión (P) de 0.889, Recall (R) de 0.831, mAP50 de 0.886 y F1 Score de 0.859. En la Figura 3 se observan las métricas a través del entrenamiento del modelo, éstas reflejan el proceso de optimización iterativa de la red neuronal convolucional sobre el conjunto de entrenamiento. Todas las curvas de pérdida del entrenamiento exhiben una marcada tendencia decreciente a lo largo de las épocas, especialmente en las primeras 100 épocas, seguida de una convergencia más gradual. La disminución constante de la pérdida de localización indica que el modelo mejora progresivamente su capacidad para localizar con precisión las espigas en las imágenes. La rápida caída y la subsiguiente estabilización en valores bajos de la pérdida de clasificación sugieren que el modelo aprendió eficientemente a distinguir la clase 'espiga' del fondo y de otros elementos del dosel vegetal. En cuanto a la pérdida focal de distribución, que

optimiza la distribución de los cuadros delimitadores, su reducción, corrobora la mejora general en la calidad de las predicciones del modelo. La tendencia a la baja en las pérdidas de entrenamiento indica que el modelo está aprendiendo de los datos de manera efectiva, ajustando iterativamente sus pesos y sesgos para minimizar el error entre sus predicciones y las etiquetas verdaderas.

Las métricas de desempeño (precisión, recall y mAP50) muestran un rápido ascenso inicial y una posterior estabilización que se mantiene hasta el final del entrenamiento. La rápida estabilización de estas métricas en valores altos es un indicador de buena generalización. Esto significa que, si bien el modelo continuó refinando las pérdidas en el conjunto de entrenamiento, ya había alcanzado su capacidad máxima para desempeñarse en datos no vistos, evitando el sobreajuste.

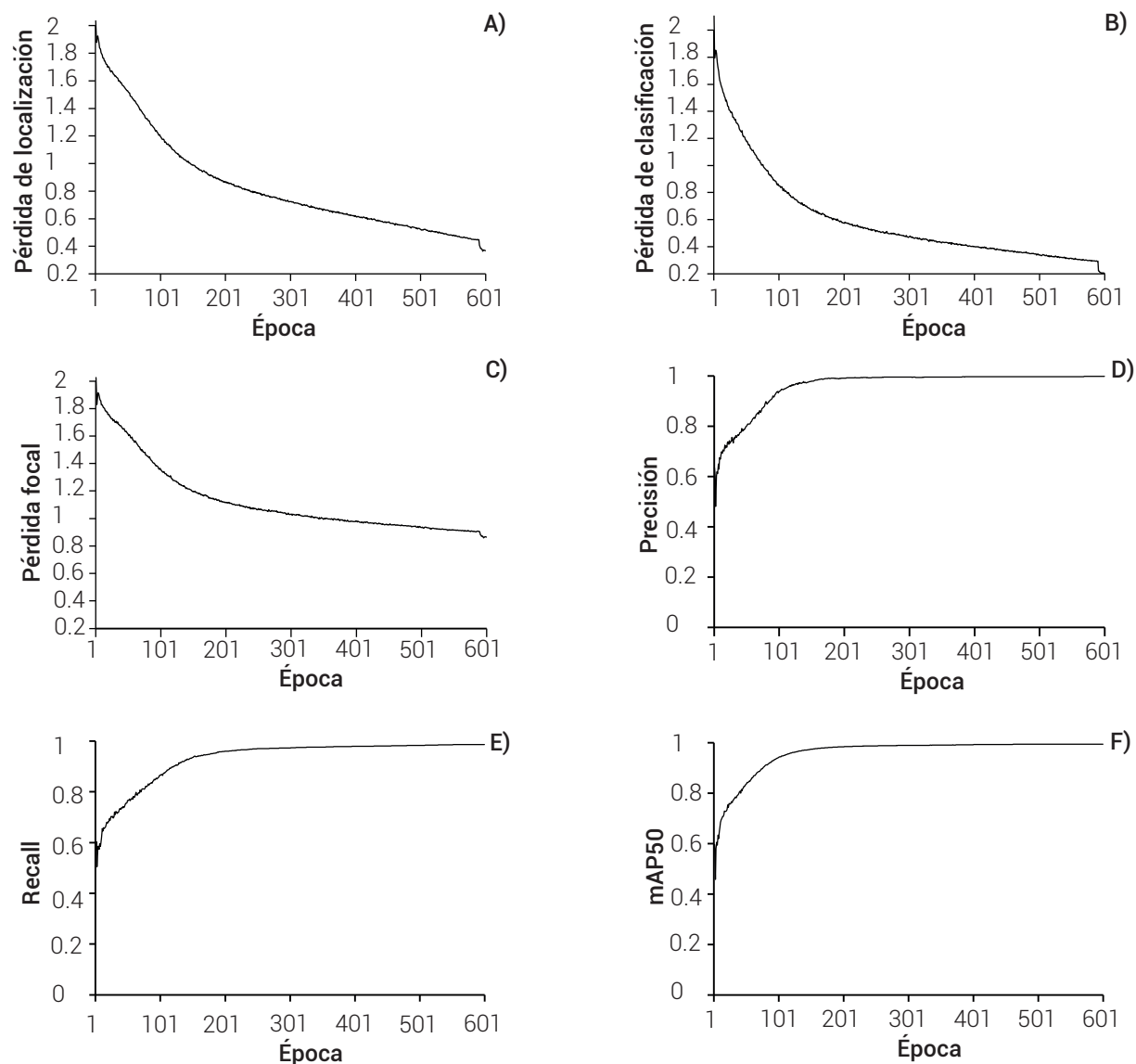


Figura 2. Análisis de entrenamiento modelo Oblicuo. A) Pérdida de localización durante entrenamiento, B) Pérdida de clasificación durante entrenamiento, C) Pérdida focal de distribución durante entrenamiento, D) Precisión durante entrenamiento, E) Recall durante entrenamiento, F) mAP50 durante entrenamiento.

Si se comparan las métricas de desempeño a través del entrenamiento de los modelos Nadir y Oblicuo, se observan resultados más variables o ruidosos en el modelo Nadir, esto se debe a que fue entrenado con un tamaño de lote de 1, lo que significa que el modelo actualiza sus pesos y sesgos después de procesar una sola imagen. La información contenida en una sola imagen rara vez es representativa del conjunto de datos completo, lo que resulta en un cálculo de gradiente altamente ruidoso y volátil, ya que la dirección de optimización se basa en un punto de dato individual y no en un promedio estadístico; por el contrario, el modelo

Oblicuo, al promediar los errores sobre un tamaño de lote de siete imágenes, tiene actualizaciones de peso más estables y una curva de pérdida significativamente más suave.

CONCLUSIONES

Los resultados de esta investigación demuestran que la implementación de la arquitectura YOLOv12m es apropiada para detectar y contar de forma automatizada espigas de maíz en imágenes aéreas. El uso de una base de datos diversa y multirregional dotó a ambos

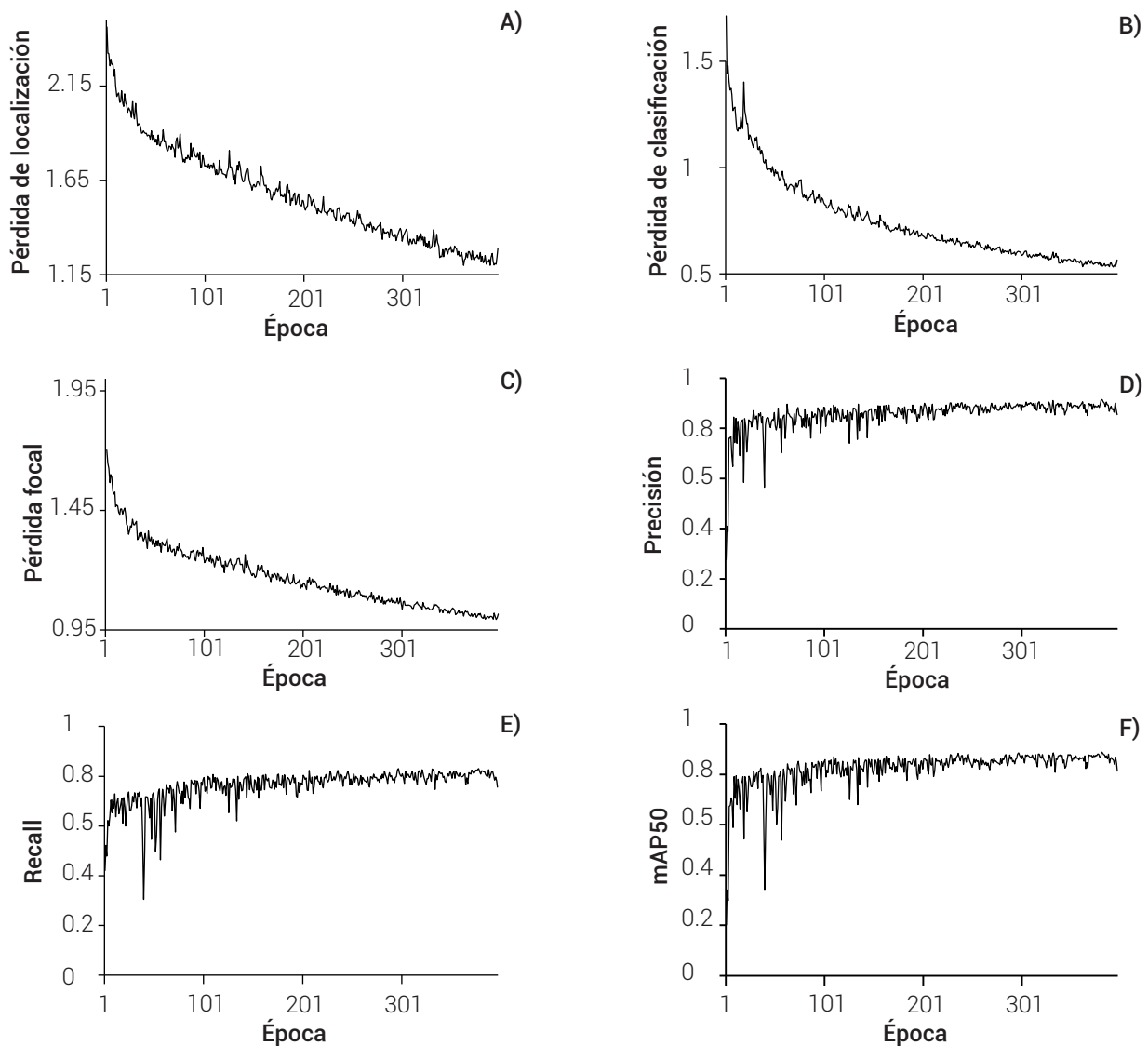


Figura 3. Análisis de entrenamiento modelo Nadir. A) Pérdida de localización durante entrenamiento, B) Pérdida de clasificación durante entrenamiento, C) Pérdida focal de distribución durante entrenamiento, D) Precisión durante entrenamiento, E) Recall durante entrenamiento, F) mAP50 durante entrenamiento.

modelos con la capacidad de generalización, validando su aplicabilidad en entornos agronómicos reales.

AGRADECIMIENTOS

Se extiende un agradecimiento a la Secretaría de Ciencias Humanidades Tecnología e Innovación por la beca otorgada al primer autor para la realización de este estudio; asimismo, se agradece a Secretaría de Agricultura y Desarrollo Rural que a través del Colegio de Posgraduados facilitó las instalaciones y equipo necesario para la ejecución de este estudio.

BIBLIOGRAFÍA

- Alif M. A. R. and M. Hussain (2024) YOLOv1 to YOLOv10: a comprehensive review of YOLO variants and their application in the agricultural domain. *Computer Vision and Pattern Recognition* arXiv:2406.10139, <https://doi.org/10.48550/arXiv.2406.10139>
- Che Y., H. Bai, L. Sun, Y. Fang, X. Guo and S. Yin (2025) Real-time detection of varieties and defects in moving corn seeds based on YOLO-SBWL. *Agriculture* 15:685, <https://doi.org/10.3390/agriculture15070685>
- Chollet F. (2021) Introduction to deep learning for computer vision. In: *Deep Learning with Python*. 2nd edition. F. Chollet (ed). Manning Publications. Shelter Island, New York, USA. pp:201-237.
- Erenstein O., M. Jaleta, K. Sonder, K. Mottaleb and B. M. Prasanna (2022) Global maize production, consumption and trade: trends and R&D implications. *Food Security* 14:1295-1319, <https://doi.org/10.1007/s12571-022-01288-7>

- FAO, Food and Agriculture Organization of the United Nations (2024) FAOSTAT data. Food and Agriculture Organization of the United Nations. Rome, Italy. <https://www.fao.org/faostat/en/#data> (October 2025).
- Jia Y., K. Fu, H. Lan, X. Wang and Z. Su (2024) Maize tassel detection with CA-YOLO for UAV images in complex field environments. *Computers and Electronics in Agriculture* 217:108562, <https://doi.org/10.1016/j.compag.2023.108562>
- Kanna K. S., R. Kumaraperumal, P. Pazhanivelan, R. Jagadeeswaran and P. C. Prabu (2024) YOLO deep learning algorithm for object detection in agriculture: a review. *Journal of Agricultural Engineering* 55:1641, <https://doi.org/10.4081/jae.2024.1641>
- Kapetas D., E. Kalogeropoulou, P. Christakakis, C. Klaridopoulos and E. M. Pechlivani (2025) Comparative evaluation of AI-based multi-spectral imaging and PCR-based assays for early detection of *Botrytis cinerea* infection on pepper plants. *Agriculture* 15:164, <https://doi.org/10.3390/agriculture15020164>
- Khanam R. and M. Hussain (2024) YOLOv11: an overview of the key architectural enhancements. *Computer Vision and Pattern Recognition*. arXiv:2410.17725, <https://doi.org/10.48550/arXiv.2410.17725>
- Khanam R. and M. Hussain (2025) A review of YOLOv12: attention-based enhancements vs. previous versions. *Computer Vision and Pattern Recognition* arXiv:2504.11995, <https://doi.org/10.48550/arXiv.2504.11995>
- Liang Z., X. Xu, D. Yang and Y. Liu (2025) The development of a lightweight DE-YOLO model for detecting impurities and broken rice grains. *Agriculture* 15:848, <https://doi.org/10.3390/agriculture15080848>
- Mirhaji H., M. Soleymani, A. Asakereh and S. A. Mehdizadeh (2021) Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions. *Computers and Electronics in Agriculture* 191:106533, <https://doi.org/10.1016/j.compag.2021.106533>
- Pu H., X. Chen, Y. Yang, R. Tang, J. Luo, Y. Wang and J. Mu (2023) Tassel-YOLO: a new high-precision and real-time method for maize tassel detection and counting based on UAV aerial images. *Drones* 7:492, <https://doi.org/10.3390/drones7080492>
- Redmon J., S. Divvala, R. Girshick and A. Farhadi (2016) You only look once: unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada, USA. pp:779-788, <https://doi.org/10.1109/CVPR.2016.91>
- SNICS, Servicio Nacional de Inspección y Certificación de Semillas (2024) Boletín informativo de producción de semilla calificada por el SNICS, ciclos agrícolas OI 2023-2024 y PV 2024. Año 6, No. 2. Servicio Nacional de Inspección y Certificación de Semillas. Ciudad de México. México. https://www.gob.mx/cms/uploads/attachment/file/958931/BOLETIN-NOVIEMBRE-2024_READY.pdf (Octubre 2024).
- Ultralytics (2024) Entrenamiento de modelos con Ultralytics YOLO. Ultralytics. Los Angeles, California, USA. <https://docs.ultralytics.com/es/modes/train/> (Octubre 2025).
- Varghese R. and M. Sambath (2024) YOLOv8: a novel object detection algorithm with enhanced performance and robustness. 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems. Institute of Electrical and Electronics Engineers. Chennai, India. pp:1-6, <https://doi.org/10.1109/ADICS58448.2024.10533619>
- Wang D., M. Zhao, Z. Li, S. Xu, X. Wu, X. Ma and X. Liu (2025) A survey of unmanned aerial vehicles and deep learning in precision agriculture. *European Journal of Agronomy* 164:127477, <https://doi.org/10.1016/j.eja.2024.127477>
- Yang J., R. Zhang, C. Ding, L. Chen, Y. Xie and H. Ou (2025) YOLO-detassel: efficient object detection for omitted pre-tassel in detasseling operation for maize seed production. *Computers and Electronics in Agriculture* 231:109951, <https://doi.org/10.1016/j.compag.2025.109951>
- Zhang R., J. Yang, L. Chen, C. Ding, L. L. Li and L. H. Zhang (2024) Review of the detasseling techniques for maize (*Zea mays* L.) hybrid seed production. *International Journal of Agricultural and Biological Engineering* 17:1-11, <https://doi.org/10.25165/ijabe.20241703.8423>