



UNIVERSIDAD AUTÓNOMA CHAPINGO

POSGRADO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

**DETECCIÓN DE FRESAS CON VISIÓN ARTIFICIAL Y ALGORITMOS DE
APRENDIZAJE PROFUNDO**

TESIS

Que como requisito parcial para obtener el título de:

MAESTRO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

Presenta:

GONZALEZ JESUS ELEUTERIO

Bajo la dirección de:

EFRÉN FITZ RODRÍGUEZ, Ph.D.



Chapingo, Estado de México, noviembre de 2024

**DETECCIÓN DE FRESAS CON VISIÓN ARTIFICIAL Y ALGORITMOS DE
APRENDIZAJE PROFUNDO**

Tesis realizada por **C. ELEUTERIO GONZALEZ JESUS** bajo la supervisión del
Comité Asesor indicado, aprobada por el mismo y aceptada como requisito
parcial para obtener el grado de:

MAESTRO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

DIRECTOR: _____

Ph.D. EFRÉN FITZ RODRÍGUEZ

ASESOR: _____

Ph.D. RAQUEL SALAZAR MORENO

ASESOR: _____

Ph.D. JOSÉ ARMANDO RAMÍREZ ARIAS

ASESOR: _____

DR. LUIS ARTURO SORIANO AVENDAÑO

CONTENIDO

CONTENIDO	iii
LISTA DE CUADROS	v
LISTA DE FIGURAS	vi
AGRADECIMIENTOS.....	viii
DATOS BIOGRÁFICOS.....	ix
RESUMEN GENERAL.....	x
GENERAL ABSTRACT	xi
1 INTRODUCCIÓN GENERAL	12
1.1 Objetivos.....	14
1.1.1 Objetivo general	14
1.1.2 Objetivos específicos	14
1.1.3 Hipótesis.....	14
1.1.4 Estructura de la tesis	14
2 REVISIÓN DE LITERATURA.....	15
2.1 El cultivo de la fresa.....	15
2.1.1 Panorama global y nacional de la producción de Fresas	15
2.2 Visión artificial en la detección de frutos.....	19
2.2.1 Métricas de la detección de objetos	23
2.2.2 Detección de fresas.....	25
2.2.3 segmentación de fresas	26
2.2.4 Detección de la orientación de fresas.....	28
2.3 Referencias	29
3. DETECCIÓN DE FRESAS CON VISION ARTIFICIAL Y ALGORITMOS DE APRENDIZAJE PROFUNDO	35
3.1 RESUMEN.....	35
3.2 INTRODUCCIÓN.....	36
3.3 MATERIALES Y MÉTODOS	37
3.3.1 Construcción del conjunto de datos (imágenes anotadas)	38

3.3.2 Selección de algoritmos de aprendizaje profundo	42
3.3.3 Proceso de entrenamiento	43
3.3.4 Métricas de evaluación.....	44
3.4 RESULTADOS Y DISCUSIÓN	45
3.4.1 Entrenamiento	45
3.4.2 Evaluación.....	47
3.5 CONCLUSIÓN.....	53
3.6 LITERATURA CITADA	54
4. IDENTIFICACIÓN DE PUNTOS CLAVE EN FRESAS CON ALGORITMOS DE APRENDIZAJE PROFUNDO	57
4.1 RESUMEN.....	57
4.1 ABSTRACT	58
4.2. INTRODUCCION.....	59
4.3 MATERIALES Y MÉTODOS	60
4.3.1 Construcción del conjunto de datos	60
4.3.2 Algoritmo de estimación de pose	61
4.3.3 Proceso de entrenamiento	61
4.3.4 Métricas de evaluación.....	61
4.4 RESULTADOS Y DISCUSIONES	62
4.4.1 Entrenamiento	62
4.4.2 Evaluación.....	64
4.5 CONCLUSION.....	71
4.6 LITERATURA CITADA	72

LISTA DE CUADROS

Cuadro 1. Investigaciones realizadas en la detección de fresas con algoritmos de aprendizaje profundo	25
Cuadro 2. Investigaciones realizadas en la segmentación de instancias de fresas con algoritmos de aprendizaje profundo	27
Cuadro 3. Comparativa de arquitecturas de los modelos resultantes.....	46
Cuadro 4. Resultado de las métricas de detección obtenidas durante el entrenamiento	47
Cuadro 5. Métricas generales de cada modelo obtenidas en la evaluación	48
Cuadro 6. Resultado de las métricas de detecciones evaluadas en el conjunto de datos de prueba	50
Cuadro 7. Velocidad de predicción en cada uno de los modelos	51
Cuadro 8. Balance de etiquetas para las siete clases de grado madurez	60
Cuadro 9. Comparativa de arquitecturas de los modelos resultantes.....	62
Cuadro 10. Rendimiento alcanzado durante el entrenamiento de los modelos	63
Cuadro 11. Desempeño de las métricas de detección de puntos clave en fresas con umbral de confianza y IoU de 0.8.....	68
Cuadro 12. Predicción en el conjunto de datos de prueba con umbral de <i>IoU</i> y confianza	68

LISTA DE FIGURAS

Figura 1: Producción/rendimiento de fresas en el mundo del año 2000 al 2022, <i>datos de FAOSTAT (2024)</i>	15
Figura 2: Producción de fresas en México del año 2000 al 2023, <i>datos del SIAP (2024)</i>	16
Figura 3: Producción total de fresas en los distintos sistemas de producción en México de 2016 a 2023, <i>datos del SIAP (2024)</i>	17
Figura 4. Rendimiento en kg/m ² de los distintos sistemas de producción en México, <i>datos del SIAP (2024)</i>	18
Figura 5. Estructura de los métodos de una etapa y los métodos de dos etapas. Modificado de (Huang et al., 2023).	20
Figura 6. Línea de tiempo de las versiones YOLO. <i>Adaptado de Terven et al. (2023)</i>	21
Figura 7. Arquitectura principal de los detectores de objetos modernos. <i>Modificado de Terven et al. (2023)</i>	21
Figura 8. Intersección sobre Unión (<i>IoU</i>). (a) cálculo; (b) ejemplo de tres valores diferentes. <i>Modificado de Terven et al. (2023)</i>	24
Figura 9. Funcionamiento del algoritmo de supresión de no máximos. <i>Modificado de Terven et al. (2023)</i>	24
Figura 10. Salida de una segmentación de instancias de fresas	27
Figura 11. Métodos basados en aprendizaje profundo para la estimación de la pose del objeto. <i>Modificado de Liu et al. (2024)</i>	28
Figura 12. Diagrama de flujo del entrenamiento de la red para la detección de frutos de fresas	38
Figura 13. Imágenes muestra con diferentes grados de iluminación, a) luz a través del túnel, b) luz directa y c) exterior nublado	39
Figura 14. Imágenes de muestra con diferentes grados de oclusión, a) alta b) moderada y c) sin oclusión	40

Figura 15. Grados de madurez en fresas, de acuerdo con la Norma Mexicana NMX-FF-062-SCFI-2002.....	40
Figura 16. Captura de imágenes, (a) al borde del camellón y (b) centrado	41
Figura 17. Enmascaramiento en color negro de frutos sobrerrepresentados, a) original y b) enmascarado.....	42
Figura 18. Cambios en las curvas $mAP@0.50$ durante el entrenamiento de varios modelos	46
Figura 19. Impacto del umbral de confianza en el $mAP@0.50$	49
Figura 20. Impacto del umbral de confianza en el $mAP@0.50-0.95$	49
Figura 21. Evaluación del puntaje $F1$ en el conjunto de datos de prueba con confianza de 0.05 a 0.95.....	50
Figura 22. Matriz de confusión normalizada del modelo YOLOv8n evaluada en el conjunto de datos de prueba	52
Figura 23. Salidas de las predicciones del modelo YOLOv8n	53
Figura 24. Etiquetado de imágenes con esqueleto de orientación en las fresas	60
Figura 25. Rendimiento del $mAP@0.50-0.95$ para los puntos clave durante el entrenamiento	64
Figura 26. Desempeño del $mAP@0.50$ según el Umbral de Confianza.....	65
Figura 27. Desempeño del $mAP@0.50-0.95$ según el Umbral de Confianza.....	66
Figura 28. Puntuaciones $F1$ para puntos claves en umbrales de confianza de 0.05 a 0.95.....	67
Figura 29. Matriz de confusión normalizada de YOLOv8n y YOLOv11l con un umbral de IoU y confianza de 0.8.....	69
Figura 30. Comparativo de la predicción del modelo YOLOv8n y las etiquetas reales	70
Figura 31. Comparativo de la predicción del modelo YOLOv11l y las etiquetas reales	71

AGRADECIMIENTOS

A todas las personas que han sido fundamentales en la realización de este proyecto.

A mis asesores académicos, por su invaluable guía, paciencia y apoyo constante.

A mis padres, por su incondicional apoyo y comprensión en cada paso de mi formación académica. Su confianza en mí ha sido una fuente de motivación constante.

A la Universidad Autónoma Chapingo, por brindarme las herramientas y el entorno necesario para crecer tanto personal como profesionalmente.

Al posgrado en Ingeniería Agrícola y Uso Integral del Agua, por brindarme extraordinarios profesores y personal administrativo que han contribuido a mi educación.

Al CONAHCYT por su apoyo financiero que ha facilitado la realización de este proyecto.

DATOS BIOGRÁFICOS



Nombre: ELEUTERIO GONZALEZ JESUS
Fecha de nacimiento: 2 de enero de 1997
Lugar de nacimiento: Santa María Tlahuitoltepec, Oaxaca
CURP: GOJE970102HOCNSL00
Profesión: Ingeniero Mecánico Agrícola
Cédula profesional Licenciatura: 12920185

DESARROLLO ACADÉMICO:

Se tituló como Ingeniero Mecánico Agrícola en 2019 en la Universidad Autónoma Chapingo.

RESUMEN GENERAL

DETECCIÓN DE FRESAS CON VISIÓN ARTIFICIAL Y ALGORITMOS DE APRENDIZAJE PROFUNDO

Dada la importancia comercial de las fresas a nivel mundial, resulta fundamental tener información útil y oportuna sobre el desarrollo y crecimiento del cultivo. Para ello, los sistemas de visión artificial, en conjunto con algoritmos de aprendizaje profundo, han comenzado a dar resultados prometedores en el monitoreo de los cultivos. Al obtener información previa sobre el rendimiento y calidad de las cosechas, a través de la detección y caracterización por grado de madurez de los frutos, se obtiene información útil para la logística de la cosecha y los procesos de postcosecha. En esta investigación se evaluaron algoritmos de aprendizaje profundo, incluyendo las variantes s, n, y en ciertos casos, la variante t de YOLO v8 a v11 para detectar y clasificar según el grado de madurez; y las versiones v8 y v11 para determinar dos puntos clave en los frutos de fresas en cultivos establecidos en camellones bajo macrotúneles. La red YOLOv8n tuvo la mayor velocidad de procesamiento de cuadros por segundo, aunque en las métricas de detección ocupó el tercer lugar, con valores que difieren del primer lugar en menos del 1 %. También se identificaron dos puntos clave en las fresas para determinar su eje, información que resulta necesaria para orientar a un robot en el acercamiento y desprendimiento del fruto, así como permitir el cálculo de su tamaño mediante la información de su longitud y diámetro. La mejor estimación de los puntos clave en un espacio 2D se logró con YOLO11l, mientras que YOLOv8n obtuvo el tercer lugar con valores menores al 2 %, pero con una velocidad superior en un 227.28 %. La red YOLOv8n presentó el mejor desempeño en general, considerando velocidad, precisión, tamaño y la estimación de los puntos clave al detectar y clasificar las fresas

Palabras clave: Sistema de monitoreo de cultivos, Detección y clasificación de frutos, Visión por computadora, Redes neuronales convolucionales, Aprendizaje de máquina.

GENERAL ABSTRACT

STRAWBERRY DETECTION WITH COMPUTER VISION AND DEEP LEARNING ALGORITHMS

Given the global commercial importance of strawberries, it is essential to have useful and timely information about the development and growth of the crop. To achieve this, computer vision systems, combined with deep learning algorithms, have started to show promising results in crop monitoring. By obtaining prior information on the performance and quality of the harvest through the detection and characterization of fruits based on their maturity level, useful data is obtained for harvest logistics and post-harvest processes. In this research, deep learning algorithms were evaluated, including the s, n variants, and in some cases, the t variant of YOLO v8 to v11, for detecting and classifying based on maturity level; and versions v8 and v11 were used to determine two key points on strawberry fruits in fields established in ridges under macro-tunnels. The YOLOv8n model had the highest frames-per-second processing speed, although in detection metrics, it ranked third, with values differing from the first place by less than 1%. Two key points on the strawberries were also identified to determine their axis, information that is necessary for guiding a robot in approaching and detaching the fruit, as well as allowing the calculation of its size through its length and diameter. The best estimation of key points in a 2D space was achieved with YOLO11l, while YOLOv8n placed third with values lower than 2%, but with a processing speed 227.28% faster. Overall, the YOLOv8n model showed the best performance, considering speed, precision, size, and the estimation of key points in detecting and classifying the strawberries.

Keywords: Crop monitoring system, Fruit detection and classification, Computer vision, Convolutional neural networks, Machine learning

1 INTRODUCCIÓN GENERAL

La fresa (*Fragaria × ananassa* Duch) es un cultivo de alto valor comercial global y sus frutos son una fuente de minerales, vitamina C, folatos y compuestos fenólicos antioxidantes (Martínez-Ferri et al., 2021; Giampieri et al., 2021).

En el proceso de producción, la detección y estimación de la madurez de los frutos de fresas, permite obtener información del estado de crecimiento, salud y estado de producción de la fruta (Mejia et al., 2023). Su conteo preciso y en tiempo real es esencial por su corta vida útil, aunque resulta muy complejo por el patrón de crecimiento en racimos (Zhou et al., 2024).

En el ámbito de la visión por computadora, el aprendizaje profundo está a la vanguardia del desarrollo del aprendizaje automático y ha logrado excelentes resultados. Las redes neuronales convolucionales profundas (DCNNs), por sus siglas en inglés de Deep Convolutional Neural Networks, han mejorado significativamente el rendimiento en la clasificación, segmentación y detección de imágenes (Zhang et al., 2022).

Dentro del aprendizaje profundo, existen dos tipos principales de algoritmos para la detección de objetos: los de dos etapas, que ofrecen alta precisión con baja velocidad debido a que generan regiones de interés y luego clasifican los objetos, y los de una etapa, que son más rápidos, pero menos precisos porque predicen directamente la ubicación y clase de los objetos (Huang et al., 2023).

El modelo YOLO (You Only Look Once) es una opción prometedora para la detección rápida de objetos. Este algoritmo de una sola etapa utiliza una red de retroalimentación directa para generar cajas delimitadoras y probabilidades de clase, lo que lo hace más eficiente computacionalmente que los modelos de dos etapas, como los R-CNN (Tituaña et al., 2024).

Algunas investigaciones han implementado el modelo YOLO en sus diferentes versiones, para la detección de grados de madurez en fresas, incluyendo: YOLOv3 (Xiaogang et al., 2020), YOLOv4-Tiny (Zhang et al., 2022; Tituaña et al., 2024), YOLOv5 (Zhou et al., 2024), YOLOv7 (Bai et al., 2024; Du et al., 2023), YOLOv8 (Yang et al., 2023; C. Wang, Wang, et al., 2024; C. Wang, Han, et al.,

2024), RTF-YOLO (Shen et al., 2024) y YOLACT (C. Wang, Ding, et al., 2024). La diferencia entre estas investigaciones se centra en la cantidad de clases de madurez consideradas para la detección y clasificación, el sistema de producción de las fresas y el hardware para el despliegue del modelo. Estas consideraciones determinan el tipo de ajuste o hibridación de los modelos usados.

El desafío de la recolección robótica de frutas, debido a las obstrucciones causadas por ramas, hojas, frutos adyacentes, y la variabilidad en tamaños, formas y ubicaciones aleatorias, plantea dos problemas principales para la visión artificial: 1) el reconocimiento y la detección precisa de las frutas, y 2) la determinación exacta del punto de cosecha (Xiong et al., 2018; Zhang et al., 2024).

Ante estos escenarios de visibilidad parcial de frutas, diversas investigaciones han centrado sus esfuerzos en determinar la orientación (pose) de las frutas, del tallo o racimo según el caso, con la finalidad de corte o estimación de tamaño. Estos enfoques ya se han utilizado en pimientos verdes japoneses (Eizentals & Oka, 2016), racimos de jitomate (Rong, Dai, et al., 2022, Rong, Wang, et al., 2022) y Fresas (G. Lin et al., 2019; Yu et al., 2019;2020) por citar algunos autores. Recientemente, la determinación de orientación o pose de frutos en coordenadas 2D o 3D, se ha abordado utilizando algoritmos que identifican y localizan partes específicas de un objeto (puntos clave). Algunos ejemplos de algoritmos incluyen: HRNet utilizado para medir ancho y altura de fresas (Jeong et al., 2024) y OpenPose para la orientación 2D de tomates (T. Kim et al., 2023). A estos tipos de algoritmos se ha sumado YOLO, ofreciendo por primera vez en su versión 8 y después 11 la detección de puntos clave, generando altas expectativas en velocidad de ejecución.

1.1 Objetivos

1.1.1 Objetivo general

Evaluar los algoritmos YOLOv8 a v11 en la detección y clasificación de fresas en sistemas de producción bajo macrotúneles.

1.1.2 Objetivos específicos

- Conformar una base datos con imágenes de fresas cultivadas en macrotúneles, considerando variaciones de iluminación y oclusión.
- Detectar y clasificar las fresas según su grado de madurez (A-G)
- Identificar dos puntos clave para determinar el eje de la fresa.

1.1.3 Hipótesis

En la detección de objetos, la capacidad de los algoritmos de aprendizaje profundo para aprender patrones complejos, a partir de grandes y diversos datos, les permite mejorar y superar los límites de los métodos basados solo en el procesamiento de imágenes.

1.1.4 Estructura de la tesis

Este documento está dividido en cuatro capítulos. 1) Introducción general, 2) revisión de literatura, 3) Detección de fresas con visión artificial y algoritmos de aprendizaje profundo, y 4) Identificación de puntos clave en fresas con algoritmos de aprendizaje profundo.

El capítulo 2 presenta una revisión de la literatura sobre el cultivo de fresa, destacando la relevancia de su producción a nivel mundial y nacional, así como la aplicación del aprendizaje profundo en visión por computadora, especialmente en la detección de fresas

El capítulo 3, es un artículo científico, que cumple los objetivos de detección y clasificación de fresas según la Norma Mexicana NMX-FF-062-SCFI-2002.

El capítulo 4 es un artículo científico que, con el objetivo de cumplir con los fines planteados, no solo se enfoca en la detección y clasificación, sino que también identifica dos puntos clave en las fresas.

2 REVISIÓN DE LITERATURA

2.1 El cultivo de la fresa

La variedad más cultivada a nivel mundial de la fresa es *Fragaria × ananassa*, perteneciente a la familia Rosaceae. Es una planta perenne que surge de una corona de tejido meristemático, de la cual se desarrollan hojas, tallos, estolones, inflorescencias y raíces. La fresa es una 'fruta accesoria', cuya pulpa proviene del receptáculo agrandado de la flor, que alberga los aquenios (Husaini & Neri, 2016).

2.1.1 Panorama global y nacional de la producción de Fresas

La producción mundial de fresas ha mostrado una tendencia al alza, reflejando el incremento en su consumo (Figura 1).

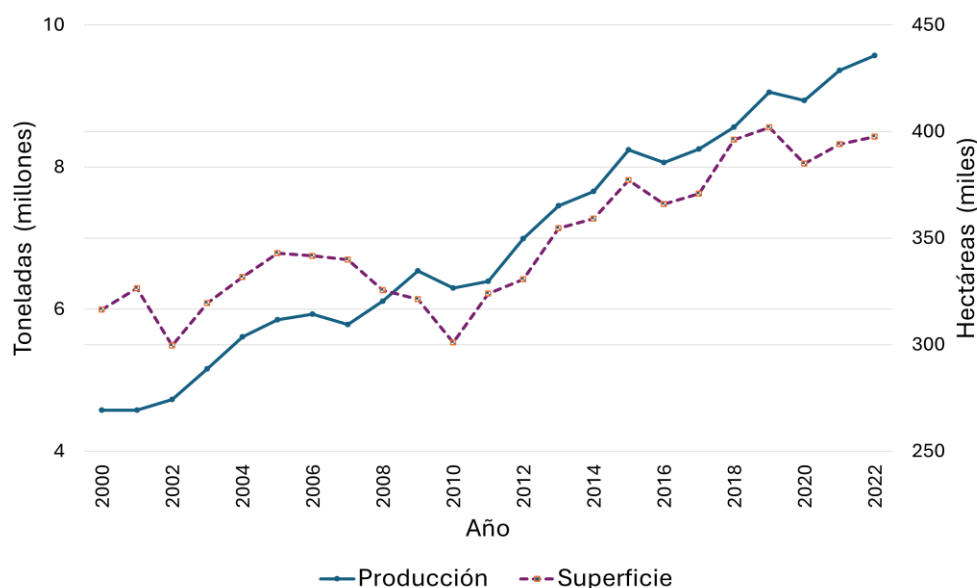


Figura 1: Producción/rendimiento de fresas en el mundo del año 2000 al 2022, *datos de FAOSTAT (2024)*.

En 2022, los diez principales países productores de fresas fueron China continental, Estados Unidos, Turquía, Egipto, México, España, Rusia, Polonia, Corea del Sur y Brasil, alcanzando una producción conjunta de 7.7 millones de toneladas (FAOSTAT, 2024).

En México, considerando los diversos ciclos y modalidades del cultivo de fresa, la producción total muestra una tendencia al alza tras la caída de 2020 (Figura 2).

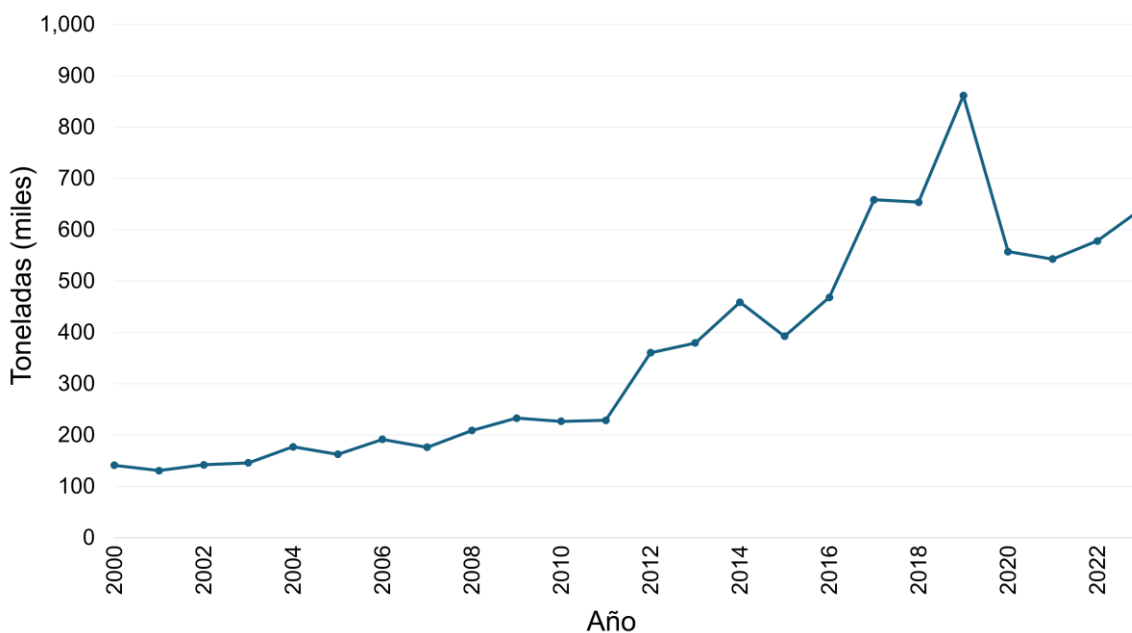


Figura 2: Producción de fresas en México del año 2000 al 2023, *datos del SIAP (2024)*

Aunque la mayor parte de la producción mundial de fresas es al aire libre, el cultivo en túneles e invernaderos ha crecido notablemente en la última década, permitiendo la producción fuera de temporada (Hancock, 2020).

La producción anual de fresas en México, considerando sus cuatro principales métodos de cultivo (a cielo abierto, macro túnel, malla sombra e invernadero) y los diversos ciclos y modalidades, se presenta en la Figura 3. Donde el sistema de producción bajo macro túnel se posiciona como la segunda forma de producción más utilizada.

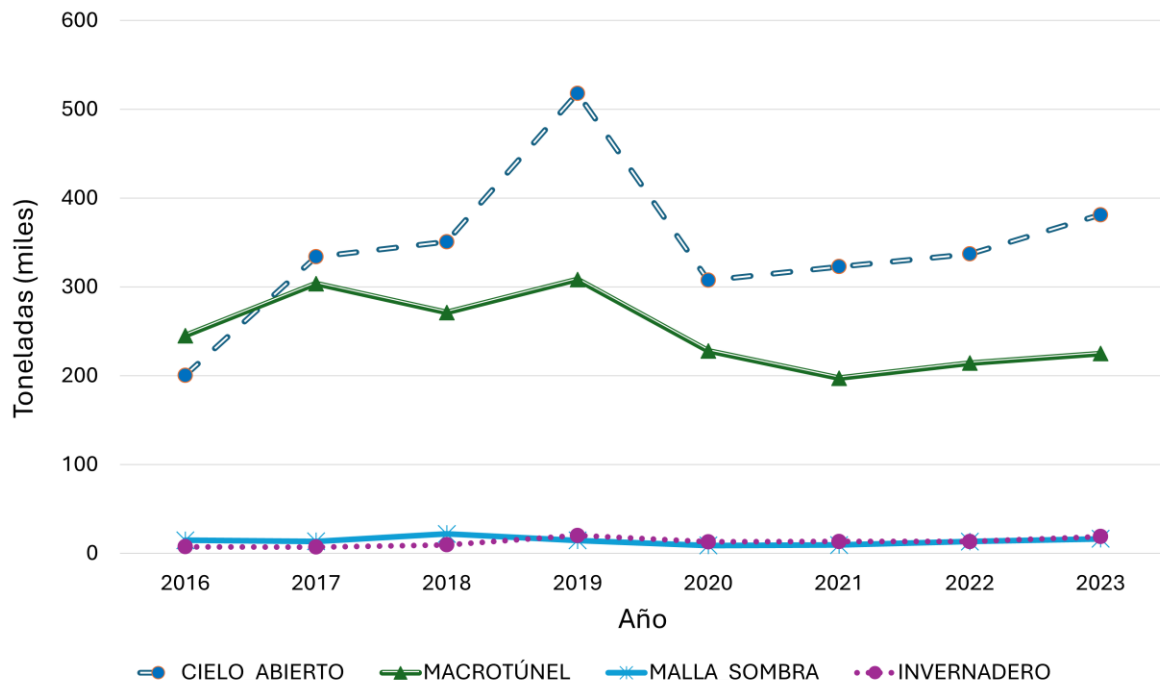


Figura 3: Producción total de fresas en los distintos sistemas de producción en México de 2016 a 2023, *datos del SIAP (2024)*

La producción en macro túneles es una técnica intermedia entre la producción a campo abierto y en invernadero. Consiste en una estructura de acero arqueada cubierta de plástico, con paredes laterales que pueden abrirse para regular la temperatura y la opción de añadir calefacción. A diferencia del invernadero, los macro túneles son móviles y no tienen cimientos de concreto, y están recubiertos de plástico en lugar de cristal, metacrilato o doble capa de polietileno (Both et al., 2019).

De acuerdo con los datos del SIAP (2024), el rendimiento en kg/m^2 de los distintos sistemas de producción en México es mayor en invernadero, seguido por el sistema en macrotúnel, malla sombra y a cielo abierto (Figura 4).

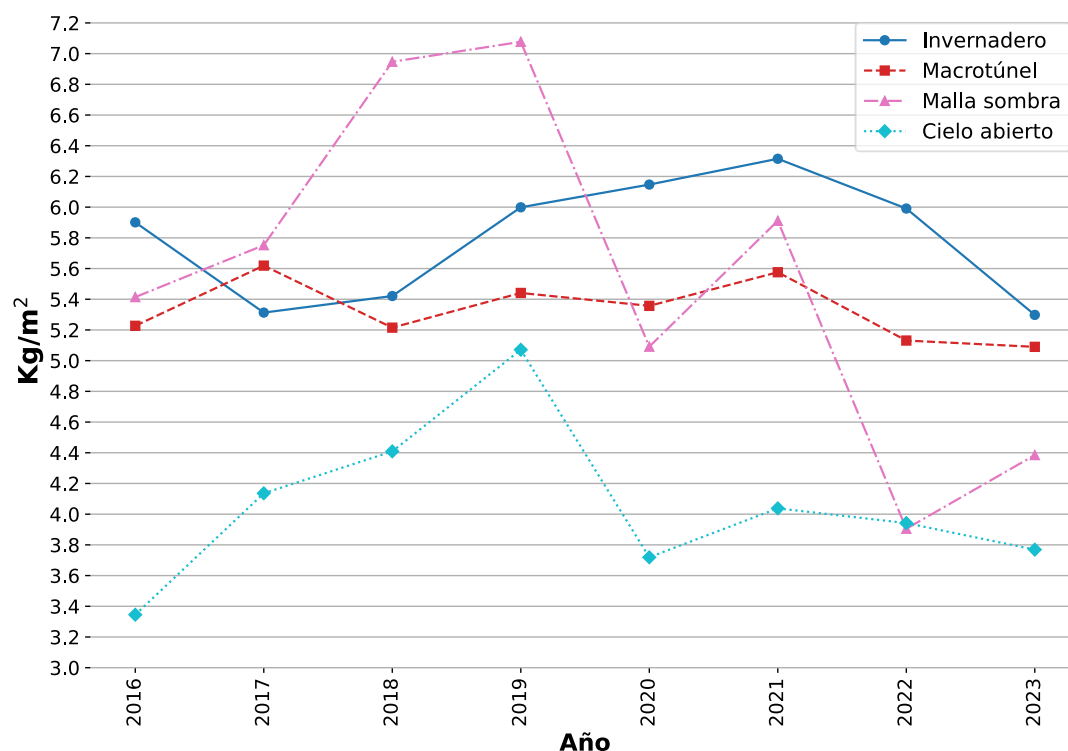


Figura 4. Rendimiento en kg/m² de los distintos sistemas de producción en México, *datos del SIAP (2024)*

La desviación estándar de los rendimientos por kg/m² es de 0.4, 0.2, 1.12 y 0.52 kg para los sistemas de producción en invernadero, macrotúnel, malla sombra y cielo abierto, respectivamente. El sistema de producción en macrotúnel presenta una mayor estabilidad en sus rendimientos, lo que podría explicar su creciente adopción.

De acuerdo con las Figura 3 y Figura 4, el sistema de producción en macrotúnel no solo ocupa el segundo lugar en cuanto a la contribución de toneladas de fresa a la producción nacional total, sino que también presenta un rendimiento elevado y estable en kg/m², con una diferencia promedio de 0.47 kg menos que el sistema de invernaderos. Además, según Both et al. (2019), su adopción entre los productores de fresa en México está en aumento. Por esta razón resulta pertinente enfocar los esfuerzos en este sistema de producción para implementar tecnologías de visión artificial que proporcionen información relevante.

2.2 Visión artificial en la detección de frutos

Los métodos tradicionales de procesamiento de imágenes para identificar frutas se basan en umbrales de color y características geométricas y texturales, pero su especificidad limita su aplicación general. Por otro lado, los algoritmos de aprendizaje profundo permiten a redes neuronales “aprender” automáticamente las características de las frutas, logrando mayor precisión en identificación y conteo; por lo que se ha vuelto ampliamente adoptado en este campo (Shi et al., 2024). Especialmente las redes neuronales convolucionales (CNN, por sus siglas en inglés) han ganado popularidad por su capacidad para extraer características complejas en el análisis de imágenes. Las CNN progresan desde el aprendizaje de detalles básicos como bordes y texturas hasta representaciones abstractas de formas y semántica de los objetos, capturando patrones detallados e información contextual (Kiran Sree Pokkuluri et al., 2024)

Los algoritmos de detección de objetos se clasifican principalmente en dos tipos (Figura 5): los detectores de dos etapas, que priorizan precisión mediante propuestas de regiones, y los de una etapa, que optimizan la velocidad y son ideales para aplicaciones en tiempo real al predecir directamente la ubicación y clase de los objetos (Rashmi & Chaudhry, 2024). Los detectores de una etapa tratan la posición del cuadro delimitador como un problema de regresión directa (Fan et al., 2025).

Entre los detectores de una etapa, los algoritmos YOLO (You Only Look Once) son los más populares debido a su facilidad de uso y alta capacidad de detección. Durante el entrenamiento, YOLO comprime o descarta la información irrelevante en cada capa durante la extracción de características (Wang et al., 2025).

YOLO fue desarrollado inicialmente por Redmon et al. (2016), con una arquitectura de 24 capas convolucionales que extraen características visuales y espaciales de los objetos, seguidas de 2 capas completamente conectadas que interpretan estas características para generar las predicciones finales. Este modelo redefinió la detección de objetos como un problema de regresión,

prediciendo directamente las cajas delimitadoras y probabilidades de clase desde imágenes completas en una sola evaluación.

En sus inicios, YOLO presentó dificultades para detectar objetos pequeños, ya que su división en una cuadrícula fija limitaba su capacidad para identificar objetos de diferentes escalas y tamaños de manera eficiente.

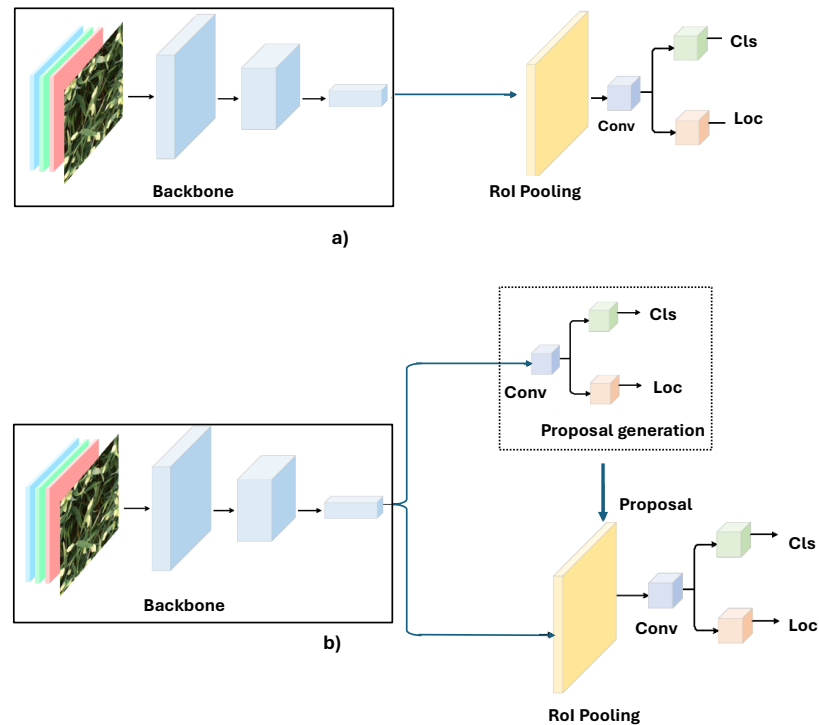


Figura 5. Estructura de los métodos de una etapa y los métodos de dos etapas.
Modificado de (Huang et al., 2023).

Desde su creación, la serie YOLO ha avanzado mediante diversas versiones, cada una corrigiendo las limitaciones de las anteriores y mejorando su rendimiento (Figura 6).

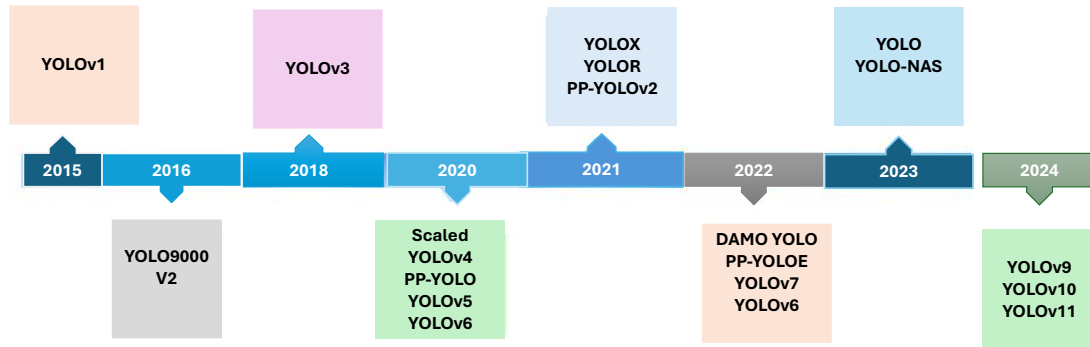


Figura 6. Línea de tiempo de las versiones YOLO. *Adaptado de Terven et al. (2023)*

La arquitectura de los detectores de objetos modernos (Figura 7) se divide en tres partes: la columna vertebral, que extrae características esenciales; el cuello, que refina estas características; y la cabeza, que usa la información refinada para hacer predicciones de detección de objetos (Terven et al., 2023).

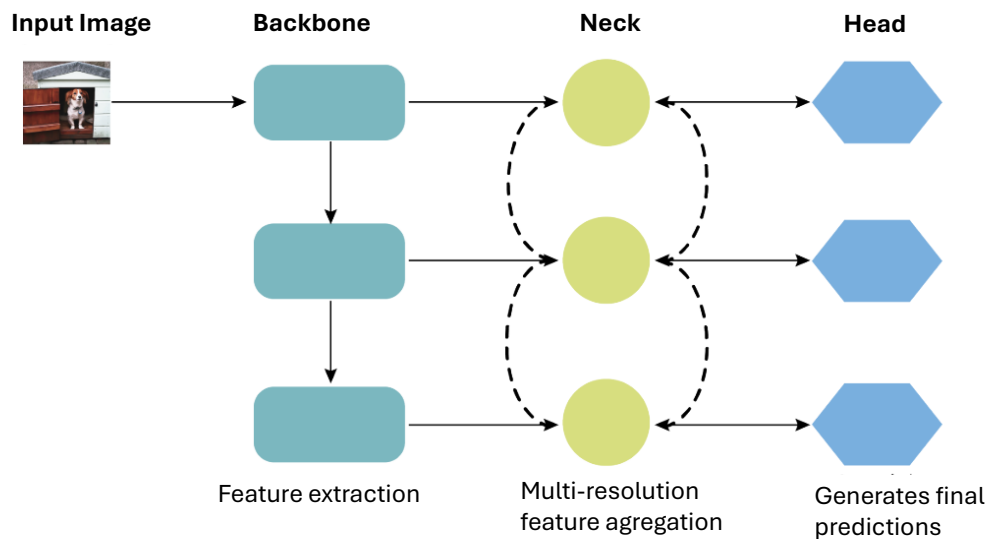


Figura 7. Arquitectura principal de los detectores de objetos modernos. *Modificado de Terven et al. (2023).*

Lanzado por Ultralytics en 2023, YOLOv8 introdujo mejoras significativas, incluyendo una arquitectura optimizada para una mejor extracción de características, un cabezal sin anclajes, y un balance ideal entre precisión y

velocidad para aplicaciones en tiempo real. Además, su facilidad de uso se ve reflejada en una implementación simplificada. Este modelo, empaquetado como una biblioteca de Python, es el primero en ofrecer funcionalidad para cinco tareas diferentes: detección de objetos, segmentación de instancias, detección de puntos clave (pose), detección orientada y clasificación.

YOLOv9 se distingue por su precisión y eficiencia en la detección de objetos, gracias a dos innovaciones clave: GELAN (Generalized Efficient Layer Aggregation Network) y PGI (Programmable Gradient Information). Con GELAN, el modelo logra gestionar datos complejos de mayor profundidad sin perder detalles esenciales, facilitando un flujo de información optimizado a través de sus capas más profundas para un procesamiento más eficaz. Por su parte, PGI ajusta dinámicamente la retropropagación de gradientes, incrementando la precisión sin elevar la demanda de recursos computacionales. Estas mejoras convierten a YOLOv9 en una herramienta ideal para aplicaciones en tiempo real, especialmente en dispositivos con recursos limitados (C.-Y. Wang et al., 2024).

YOLOv10 (A. Wang et al., 2024) optimiza la arquitectura de modelos YOLO anteriores mediante innovaciones clave como una versión mejorada de CSPNet en la red troncal, que facilita el flujo de gradientes y reduce la redundancia computacional. En el cuello, emplea Path Aggregation Network (PAN) para lograr una fusión eficiente de características multiescala. Su cabezal de uno a muchos genera múltiples predicciones por objeto durante el entrenamiento, enriqueciendo las señales de supervisión y mejorando la precisión. Durante la inferencia, el cabezal uno a uno produce una única predicción por objeto, eliminando la necesidad de *NMS*, lo que reduce la latencia y aumenta la eficiencia. Entre sus características más destacadas se encuentran el entrenamiento sin *NMS*, que mejora la velocidad de inferencia, y un diseño holístico que optimiza tanto la eficiencia como la precisión mediante cabezales ligeros y un muestreo desacoplado. Además, incorpora convoluciones de gran núcleo y autoatención parcial, lo que mejora el rendimiento sin aumentar significativamente los costos computacionales (Ultralytics, 2024b).

YOLOv11 a la fecha actual, representa el estado del arte en los modelos de detección de objetos, desarrollado por Ultralytics (2024b), no existe un artículo de investigación formal. Sin embargo, alguna de las mejoras mencionadas en su página incluye: Mayor precisión de detección debido a una red de extracción de características mejorada. Con un 22% menos de parámetros que YOLOv8m, YOLO11m alcanza una mAP superior en el conjunto de imágenes COCO, indicando mayor eficiencia en recursos computacionales. Puede ser implementado en dispositivos periféricos, plataformas en la nube y sistemas con GPU NVIDIA. Finalmente, es importante resaltar su capacidad para soportar tareas que incluye: detección de objetos, segmentación, clasificación, estimación de poses y detección orientada.

2.2.1 Métricas de la detección de objetos

La precisión promedio (AP), también conocida como precisión promedio media (mAP), es la métrica estándar para evaluar el rendimiento de los modelos de detección de objetos. Esta métrica mide tanto la precisión como la exhaustividad ($recall$), reflejando la capacidad del modelo para identificar correctamente objetos y detectar el mayor número de objetos reales posible. Utiliza una curva de precisión-exhaustividad, calculando el área bajo la curva para equilibrar el aumento en objetos detectados sin sacrificar precisión. Además, para múltiples categorías, la AP calcula la precisión promedio en cada clase y promedia estos valores, ofreciendo así una evaluación completa del rendimiento del modelo en cada categoría y en general (Terven et al., 2023).

La métrica AP o mAP utiliza la métrica Intersección sobre Unión (IoU) que, calcula la superposición entre el cuadro predicho y el real (Figura 8).

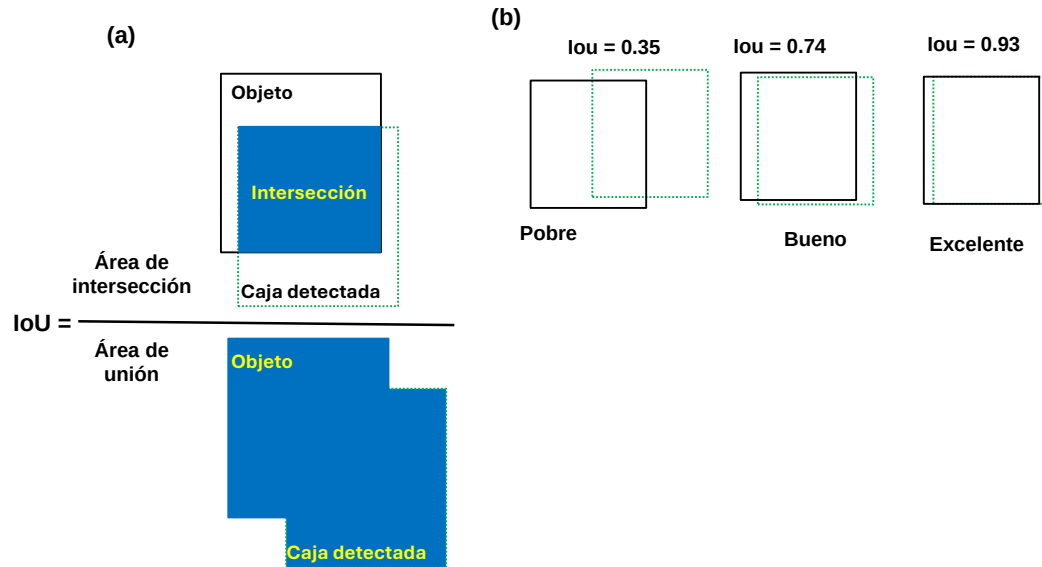


Figura 8. Intersección sobre Unión (*IoU*). (a) cálculo; (b) ejemplo de tres valores diferentes. *Modificado de Terven et al. (2023)*

El Non-maximum suppression (*NMS*) es una técnica de posprocesamiento que elimina cuadros delimitadores superpuestos en detección de objetos, reteniendo solo aquellos con puntuaciones de confianza más precisos (Figura 9).

Requiere: Conjunto de cajas delimitadoras predichas B , puntuaciones de confianza S , umbral de IoU τ , umbral de confianza T
Asegura: Conjunto de cajas delimitadoras filtradas F

- 1: $F \leftarrow \emptyset$
- 2: Filtrar las cajas: $B \leftarrow \{b \in B \mid S(b) \geq T\}$
- 3: Ordenar las cajas B según sus puntuaciones de confianza en orden descendente
- 4: **Mientras** $B \neq \emptyset$ **hacer**
- 5: Seleccionar la caja b con la puntuación de confianza más alta
- 6: Añadir b al conjunto de cajas finales F : $F \leftarrow F \cup \{b\}$
- 7: Remover b del conjunto de cajas B : $B \leftarrow B - \{b\}$
- 8: **Para todas** las cajas restantes r en B **hacer**
- 9: Calcular el IoU entre b y r : $iou \leftarrow IoU(b, r)$
- 10: **Si** $iou \geq \tau$ **entonces**
- 11: Remover r del conjunto de cajas B : $B \leftarrow B - \{r\}$
- 12: **Fin si**
- 13: **Fin para**
- 14: **Fin mientras**

Figura 9. Funcionamiento del algoritmo de supresión de no máximos. *Modificado de Terven et al. (2023)*

2.2.2 Detección de fresas

La detección de frutas en entornos no controlados presenta desafíos significativos, como la variabilidad en la iluminación, las oclusiones causadas por hojas y otros frutos, así como el tamaño reducido y el grado de madurez de las frutas. Las investigaciones recientes que han empleado algoritmos de aprendizaje profundo para la detección de fresas se detallan en el (Cuadro 1), en donde se aprecia repetidamente el uso de alguna versión o variante del modelo YOLO con algunas modificaciones, la elección en la mayoría de los casos se debe al enfoque de detección en tiempo real, tamaño reducido de parámetros que permite su despliegue en hardware de poca demanda computacional y al competitivo desempeño en la precisión de detección.

Cuadro 1. Investigaciones realizadas en la detección de fresas con algoritmos de aprendizaje profundo

Autor	Método	Escenario o Ambiente	Rendimiento
<i>Fresa madura e inmadura</i>			
(Habaragamuwa et al., 2018)	DCNN	Invernadero	$mAP = 82.62\%$
(Xiaogang et al., 2020)	YOLOv3 mejorado	Sombra, superposición	$mAP = 87.50\%$
(Gao et al., 2020)	AlexNet preentrena	Imágenes hiperespectrales	Precisión= 98.60%
(Yang et al., 2023)	YOLOv8s con Swin Transformer	Invernaderos con iluminación variada	Precisión = 94.4%
(Li et al., 2023).	Faster R-CNN	Campo abierto	$mAP = 87.33 \%$
(Shen et al., 2024)	RTF-YOLO	Iluminación y oclusión cambiante	$mAP = 90.24 \%$
(Wang, Wang, et al., 2024)	YOLOv8 mejorada	-	$F1 = 97.07\%$, Precisión = 97.81 %, Recall = 96.36 %
(Wang, Ding, et al., 2024)	YOLACT modificada	en camellones	$mAP 77.90 \%$
(Zhou et al., 2024)	YOLOv5s con Swin transformer	Agricultura vertical a gran escala, granja interior	$mAP = 90.40 \%$
<i>flor y fruto</i>			

(Bai et al., 2024)	YOLOv7 con Swin transformer y GS ELAN	Invernadero	$mAP = 92.1\%$
Fresa madura, inmadura y flor			
(Zhang et al., 2022)	RTSD-Net (basado en YOLOv4-tiny)	-	mAP ligeramente menor que YOLOv4-tiny
(Wang, Han, et al., 2024)	YOLOv8 con bloque de atención de mezcla y VoV-GSCSP	Imágenes de drones	$mAP = 84.2\%$
Fresa verde, semi madura y madura			
(Mejia et al., 2023)	MobileNetV2 + procesamiento de imágenes	Camellones dentro de túnel de polietileno	Precisión = 93.75 % para fresas maduras
Fresa madura con oclusión severa, moderada, leve y sin oclusión			
(Du et al., 2023)	DSW-YOLO (basado en YOLOv7)	-	$mAP = 86.7\%$
Fresa sobre madurada, madura, casi madura, inmadura y flor			
(Tituaña et al., 2024)	YOLOv4s	Plantaciones a campo abierto	$mAP = 80.20\%$
Del 0% al 110% de madurez (en intervalos de 10%)			
(Ge et al., 2024)	MiniNet	Fresas cultivadas sobre mesas en túneles de polietileno	$MAE = 4.8\%$

2.2.3 segmentación de fresas

En el análisis de imágenes agrícolas, es crucial identificar con precisión los objetos en una escena. La segmentación semántica agrupa píxeles por clases generales, pero no distingue entre instancias individuales, como los frutos en un cultivo. La segmentación de instancias, en cambio, asigna etiquetas a los píxeles y diferencia entre objetos individuales de la misma categoría (Figura 10), mejorando la precisión en tareas como el seguimiento, la recolección automatizada y el análisis de calidad, optimizando los procesos agrícolas. El Cuadro 2 es una muestra de investigaciones recientes en las que se emplearon algoritmos de aprendizaje profundo para la segmentación de instancias de fresas.

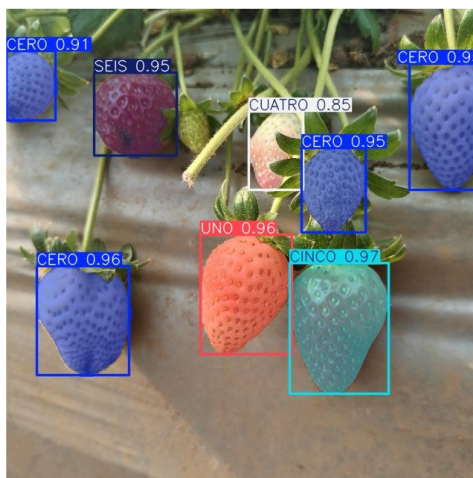


Figura 10. Salida de una segmentación de instancias de fresas

Cuadro 2. Investigaciones realizadas en la segmentación de instancias de fresas con algoritmos de aprendizaje profundo

Autor	Método	Escenario o Ambiente	Rendimiento
fresa madura, rosa, inmadura y deforme			
(Ge et al., 2019)	Mask-RCNN (compleción de oclusión)		F1 = 0.94 (maduras), 0.88 (resto)
Solo fresa			
(Pérez-Borrero et al., 2020)	Mask-RCNN mejorado	Base de datos pública StrawD1_Db1	AP = 43.85%
(Cao et al., 2023)	StrawSeg (MobileNetV2)		mAP = 79.7%
Fresa madura e inmadura			
(Yu et al., 2019)	Mask R-CNN	Entornos no estructurados, diferentes intensidades de iluminación, frutas superpuestas u ocultas	AP = 95.78%
(Wang et al., 2023)	ASFA-net con Swin-transformer	Cultivos elevados	mAP = 95.91 ± .49%
Fresa y fondo			
(Perez-Borrero et al., 2021)	Fully CNN		mAP = 52.61%
Fresa madura, casi madura y verde			
(Cai et al., 2022)	DeepLabV3+		AP = 90.9%

2.2.4 Detección de la orientación de fresas.

Los diversos enfoques de algoritmos basados en aprendizaje profundo que se utilizan actualmente para la detección de puntos clave y la estimación de la pose de objetos se detallan en la Figura 11.

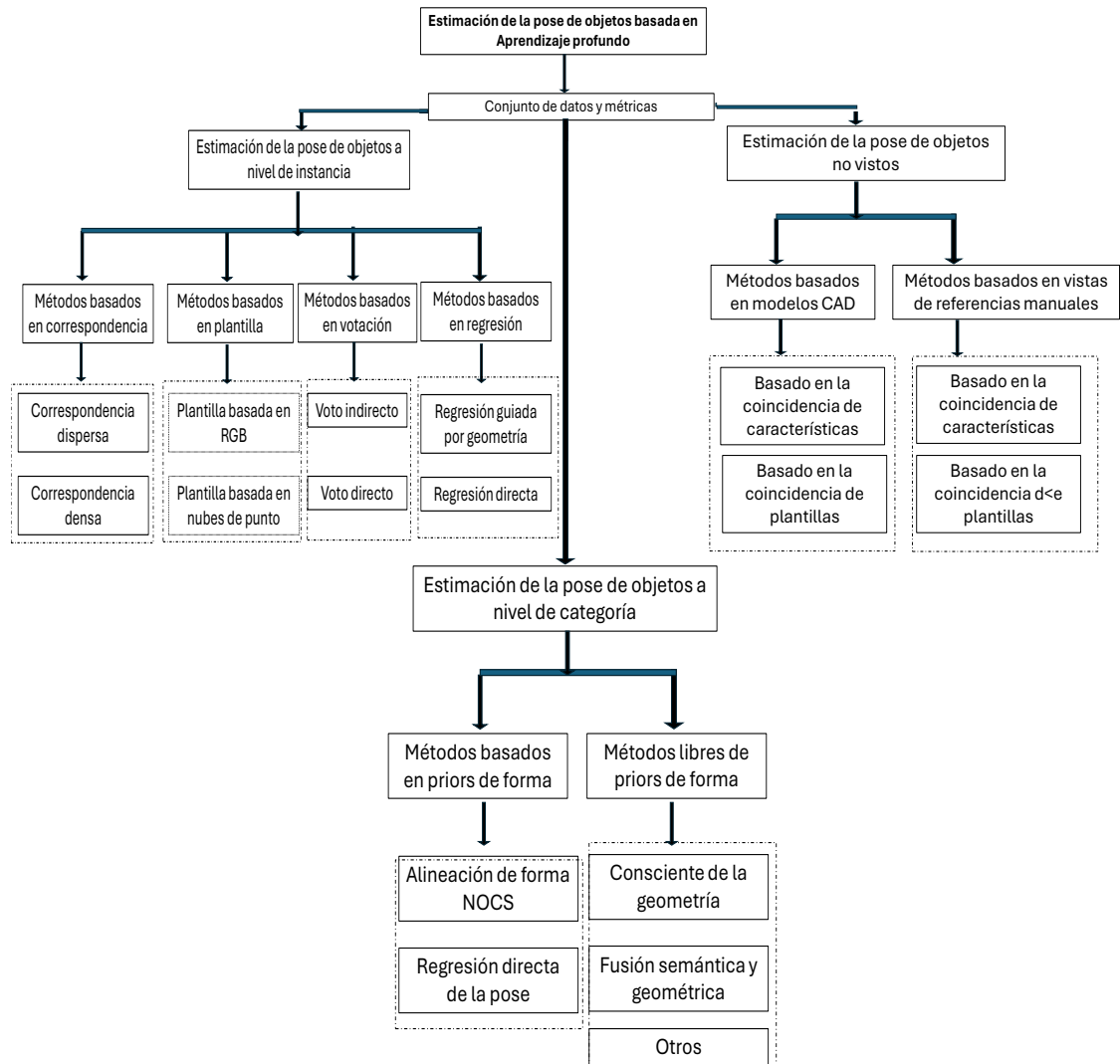


Figura 11. Métodos basados en aprendizaje profundo para la estimación de la pose del objeto. *Modificado de Liu et al. (2024)*

Los métodos a nivel de instancia se entrenan con datos específicos, lo que limita su generalización a otras instancias. La oclusión severa provoca texturas y características geométricas incompletas, dificultando su manejo eficaz Liu et al. (2024).

La orientación de las fresas se ha determinado mediante redes como Mask-RCNN, R-YOLO (Yu et al., 2019; 2020) y algoritmos de aprendizaje profundo para la detección de puntos clave como, HRNet combinado con sensores LiDAR (Jeong et al., 2024). La orientación 3D también se ha evaluado utilizando técnicas como la extracción de nube de puntos y construcción de características 3D (Ge et al., 2020a), la terminación de formas tridimensionales (Ge et al., 2020b), cámaras de visión estéreo en conjunto con YOLOv3 y Mask R-CNN (Hu et al., 2022) así como métodos basados en la agrupación en clústeres 3D, cuadros delimitadores 2D y valores de profundidad (Ge et al., 2023).

2.3 Referencias

- Bai, Y., Yu, J., Yang, S., & Ning, J. (2024). An improved YOLO algorithm for detecting flowers and fruits on strawberry seedlings. *Biosystems Engineering*, 237, 1–12. <https://doi.org/10.1016/j.biosystemseng.2023.11.008>
- Both, A.J., Demchak, Kathy., Hanson, Eric., Heidenreich, Cathy., Loeb, Greg., McDermott, Laura., Pritts, Marvin., & Weber, Courtney. (2019). *High tunnel production guide for raspberries & blackberries*.
- Cai, C., Tan, J., Zhang, P., Ye, Y., & Zhang, J. (2022). Determining Strawberries' Varying Maturity Levels by Utilizing Image Segmentation Methods of Improved DeepLabV3+. *Agronomy*, 12(8), 1875. <https://doi.org/10.3390/agronomy12081875>
- Cao, L., Chen, Y., & Jin, Q. (2023). Lightweight Strawberry Instance Segmentation on Low-Power Devices for Picking Robots. *Electronics*, 12(14), 3145. <https://doi.org/10.3390/electronics12143145>
- Du, X., Cheng, H., Ma, Z., Lu, W., Wang, M., Meng, Z., Jiang, C., & Hong, F. (2023). DSW-YOLO: A detection method for ground-planted strawberry fruits under different occlusion levels. *Computers and Electronics in Agriculture*, 214, 108304. <https://doi.org/10.1016/j.compag.2023.108304>
- Eizentals, P., & Oka, K. (2016). 3D pose estimation of green pepper fruit for automated harvesting. *Computers and Electronics in Agriculture*, 128, 127–140. <https://doi.org/10.1016/j.compag.2016.08.024>
- Fan, Q., Li, Y., Deveci, M., Zhong, K., & Kadry, S. (2025). LUD-YOLO: A novel lightweight object detection network for unmanned aerial vehicle. *Information Sciences*, 686, 121366. <https://doi.org/10.1016/j.ins.2024.121366>

- FAOSTAT. (2024). *Cultivos y productos ganaderos*. Organización de Las Naciones Unidas Para La Alimentación y La Agricultura. <https://www.fao.org/faostat/en/#data/QCL/visualize>
- Gao, Z., Shao, Y., Xuan, G., Wang, Y., Liu, Y., & Han, X. (2020). Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. *Artificial Intelligence in Agriculture*, 4, 31–38. <https://doi.org/10.1016/j.aiia.2020.04.003>
- Ge, Y., From, P. J., & Xiong, Y. (2024). Multi-view gripper internal sensing for the regression of strawberry ripeness using a mini-convolutional neural network for robotic harvesting. *Computers and Electronics in Agriculture*, 216, 108474. <https://doi.org/10.1016/j.compag.2023.108474>
- Ge, Y., Xiong, Y., & From, P. J. (2019). Instance Segmentation and Localization of Strawberries in Farm Conditions for Automatic Fruit Harvesting. *IFAC-PapersOnLine*, 52(30), 294–299. <https://doi.org/10.1016/j.ifacol.2019.12.537>
- Ge, Y., Xiong, Y., & From, P. J. (2020a). Classification of pickable and unpickable strawberries under farm conditions. *IEEE International Conference on Automation Science and Engineering, 2020-Augus*, 961–966. <https://doi.org/10.1109/CASE48305.2020.9217022>
- Ge, Y., Xiong, Y., & From, P. J. (2020b). Symmetry-based 3D shape completion for fruit localisation for harvesting robots. *Biosystems Engineering*, 197, 188–202. <https://doi.org/10.1016/j.biosystemseng.2020.07.003>
- Ge, Y., Xiong, Y., & From, P. J. (2023). Three-dimensional location methods for the vision system of strawberry-harvesting robots: development and comparison. *Precision Agriculture*, 24, 764–782. <https://doi.org/10.1007/s11119-022-09974-4>
- Giampieri, F., Cianciosi, D., Ansary, J., Regolo, L., & T.Y. Forbes-Hernandez, M. B. (2021). Antioxidant effects of strawberry extract on HepG2 stressed with an oxidant agent. *Acta Horticulturae*, 1309, 903–909.
- Habaragamuwa, H., Ogawa, Y., Suzuki, T., Shiigi, T., Ono, M., & Kondo, N. (2018). Detecting greenhouse strawberries (mature and immature), using deep convolutional neural network. *Engineering in Agriculture, Environment and Food*, 11(3), 127–138. <https://doi.org/10.1016/j.eaef.2018.03.001>
- Hancock, J. F. (2020). *Strawberries crop production science in horticulture* (2nd ed.). CAB International.
- Hu, H. M., Kaizu, Y., Zhang, H. D., Xu, Y. W., Imou, K., Li, M., Huang, J. J., & Dai, S. (2022). Recognition and localization of strawberries from 3D binocular cameras for a strawberry picking robot using coupled YOLO/Mask R-CNN.

- International Journal of Agricultural and Biological Engineering*, 15, 175–179.
<https://doi.org/10.25165/j.ijabe.20221506.7306>
- Huang, Y., Qian, Y., Wei, H., Lu, Y., Ling, B., & Qin, Y. (2023). A survey of deep learning-based object detection methods in crop counting. *Computers and Electronics in Agriculture*, 215, 108425.
<https://doi.org/10.1016/j.compag.2023.108425>
- Husaini, A. M., & Neri, Davide. (2016). *Strawberry Growth, Development and Diseases* (A. M. Husaini & Davide. Neri, Eds.; 1st ed.). CABI.
<https://cabidigitallibrary.org>
- Jeong, H., Moon, H., Jeong, Y., Kwon, H., Kim, C., Lee, Y., Yang, S., & Kim, S. (2024). Automated Technology for Strawberry Size Measurement and Weight Prediction Using AI. *IEEE Access*, 12, 14157–14167.
<https://doi.org/10.1109/ACCESS.2024.3356118>
- Kim, T., Lee, D.-H., Kim, K.-C., & Kim, Y.-J. (2023). 2D pose estimation of multiple tomato fruit-bearing systems for robotic harvesting. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2023.108004>
- Kiran Sree Pokkuluri, Usha Devi N. S.S.S.N, Martin Margala, & Prasun Chakrabarti. (2024). Enhancing Image Segmentation Accuracy using Deep Learning Techniques. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 49(1), 139–148.
<https://doi.org/10.37934/araset.49.1.139148>
- Li, J., Zhu, Z., Liu, H., Su, Y., & Deng, L. (2023). Strawberry R-CNN: Recognition and counting model of strawberry based on improved faster R-CNN. *Ecological Informatics*, 77, 102210.
<https://doi.org/10.1016/j.ecoinf.2023.102210>
- Lin, G., Tang, Y., Zou, X., Xiong, J., & Li, J. (2019). Guava Detection and Pose Estimation Using a Low-Cost RGB-D Sensor in the Field. *Sensors*, 19(2), 428.
<https://doi.org/10.3390/s19020428>
- Liu, J., Sun, W., Yang, H., Zeng, Z., Liu, C., Zheng, J., Liu, X., Rahmani, H., Sebe, N., & Mian, A. (2024). *Deep Learning-Based Object Pose Estimation: A Comprehensive Survey*. <http://arxiv.org/abs/2405.07801>
- Martínez-Ferri, E., Cervantes, L., Soria, C., Forbes-Hernández, T., Giampieri, F., P. Reboredo-Rodríguez, M. B., & Ariza, M. T. (2021). Health-promoting properties as a target for selecting strawberry cultivars in breeding programmes. *Acta Horticulturae*, 1309, 865–868.
- Mejia, G., Montes de Oca, A., & Flores, G. (2023). Strawberry localization in a ridge planting with an autonomous rover. *Engineering Applications of Artificial Intelligence*, 119, 105810. <https://doi.org/10.1016/j.engappai.2022.105810>

- Pérez-Borrero, I., Marín-Santos, D., Gegúndez-Arias, M. E., & Cortés-Ancos, E. (2020). A fast and accurate deep learning method for strawberry instance segmentation. *Computers and Electronics in Agriculture*, 178, 105736. <https://doi.org/10.1016/j.compag.2020.105736>
- Perez-Borrero, I., Marin-Santos, D., Vasallo-Vazquez, M. J., & Gegundez-Arias, M. E. (2021). A new deep-learning strawberry instance segmentation methodology based on a fully convolutional neural network. *Neural Computing and Applications*, 33(22), 15059–15071. <https://doi.org/10.1007/s00521-021-06131-2>
- Rashmi, & Chaudhry, R. (2024). SD-YOLO-AWDNet: A hybrid approach for smart object detection in challenging weather for self-driving cars. *Expert Systems with Applications*, 256, 124942. <https://doi.org/10.1016/j.eswa.2024.124942>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*.
- Rong, J., Dai, G., & Wang, P. (2022). A peduncle detection method of tomato for autonomous harvesting. *Complex and Intelligent Systems*, 8, 2955–2969. <https://doi.org/10.1007/s40747-021-00522-7>
- Rong, J., Wang, P., Wang, T., Hu, L., & Yuan, T. (2022). Fruit pose recognition and directional orderly grasping strategies for tomato harvesting robots. *Computers and Electronics in Agriculture*, 202, 14. <https://doi.org/10.1016/j.compag.2022.107430>
- Shen, S., Duan, F., Tian, Z., & Han, C. (2024). A Novel Deep Learning Method for Detecting Strawberry Fruit. *Applied Sciences*, 14(10), 4213. <https://doi.org/10.3390/app14104213>
- Shi, Y., Qing, S., Zhao, L., Wang, F., Yuwen, X., & Qu, M. (2024). YOLO-Peach: A High-Performance Lightweight YOLOv8s-Based Model for Accurate Recognition and Enumeration of Peach Seedling Fruits. *Agronomy*, 14(8), 1628. <https://doi.org/10.3390/agronomy14081628>
- SIAP. (2024). *Anuario Estadístico de la Producción Agrícola*. Gobierno de México. <https://nube.siap.gob.mx/cierreagricola/>
- Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, 5(4), 1680–1716. <https://doi.org/10.3390/make5040083>
- Tituaña, L., Gholami, A., He, Z., Xu, Y., Karkee, M., & Ehsani, R. (2024). A small autonomous field robot for strawberry harvesting. *Smart Agricultural Technology*, 8, 100454. <https://doi.org/10.1016/j.atech.2024.100454>

- Ultralytics. (2024). *Ultralytics*. Ultralytics YOLO Docs. <https://docs.ultralytics.com/models/>
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). *YOLOv10: Real-Time End-to-End Object Detection*.
- Wang, C., Ding, F., Wang, Y., Wu, R., Yao, X., Jiang, C., & Ling, L. (2024). Real-Time Detection and Instance Segmentation of Strawberry in Unstructured Environment. *Computers, Materials & Continua*, 78(1), 1481–1501. <https://doi.org/10.32604/cmc.2023.046876>
- Wang, C., Han, Q., Li, C., Li, J., Kong, D., Wang, F., & Zou, X. (2024). Assisting the Planning of Harvesting Plans for Large Strawberry Fields through Image-Processing Method Based on Deep Learning. *Agriculture*, 14(4), 560. <https://doi.org/10.3390/agriculture14040560>
- Wang, C., Wang, H., Han, Q., Zhang, Z., Kong, D., & Zou, X. (2024). Strawberry Detection and Ripeness Classification Using YOLOv8+ Model and Image Processing Method. *Agriculture*, 14(5), 751. <https://doi.org/10.3390/agriculture14050751>
- Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). *YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information*.
- Wang, D., Wang, X., Chen, Y., Wu, Y., & Zhang, X. (2023). Strawberry ripeness classification method in facility environment based on red color ratio of fruit rind. *Computers and Electronics in Agriculture*, 214, 108313. <https://doi.org/10.1016/j.compag.2023.108313>
- Wang, P., Luo, Y., & Zhu, Z. (2025). FDI-YOLO: Feature disentanglement and interaction network based on YOLO for SAR object detection. *Expert Systems with Applications*, 260, 125442. <https://doi.org/10.1016/j.eswa.2024.125442>
- Xiaogang, L., Cheng, F., Jianian, L., Yanli, G., Yuyang, Z., & Qiliang, Y. (2020). Identification method of strawberry based on convolutional neural network. *Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery*, 8. <https://doi.org/10.6041/j.issn.1000-1298.2020.02.026>
- Xiong, J., Lin, R., Liu, Z., He, Z., Tang, L., Yang, Z., & Zou, X. (2018). The recognition of litchi clusters and the calculation of picking point in a nocturnal natural environment. *Biosystems Engineering*, 166, 44–57. <https://doi.org/10.1016/j.biosystemseng.2017.11.005>
- Yang, S., Wang, W., Gao, S., & Deng, Z. (2023). Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-Swin Transformer. *Computers and Electronics in Agriculture*, 215, 108360. <https://doi.org/10.1016/j.compag.2023.108360>

- Yu, Y., Zhang, K., Liu, H., Yang, L., & Zhang, D. (2020). Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot. *IEEE Access*, 8, 13. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9119372>
- Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture*, 163, 104846. <https://doi.org/10.1016/j.compag.2019.06.001>
- Zhang, J., Kang, N., Qu, Q., Zhou, L., & Zhang, H. (2024). Automatic fruit picking technology: a comprehensive review of research advances. *Artificial Intelligence Review*, 57(3), 54. <https://doi.org/10.1007/s10462-023-10674-2>
- Zhang, Y., Yu, J., Chen, Y., Yang, W., Zhang, W., & He, Y. (2022). Real-time strawberry detection using deep neural networks on embedded system (rtsd-net): An edge AI application. *Computers and Electronics in Agriculture*, 192, 106586. <https://doi.org/10.1016/j.compag.2021.106586>
- Zhou, X., Zhang, Y., Jiang, X., Riaz, K., Rosenbaum, P., Lefsrud, M., & Sun, S. (2024). Advancing tracking-by-detection with MultiMap: Towards occlusion-resilient online multiclass strawberry counting. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2024.124587>

3. DETECCIÓN DE FRESAS CON VISION ARTIFICIAL Y ALGORITMOS DE APRENDIZAJE PROFUNDO

Eleuterio Gonzalez-Jesus¹; Efrén Fitz-Rodríguez^{1*}, Raquel Salazar Moreno¹; José Armando Ramírez Arias¹; Luis Arturo Soriano-Avenidaño¹;

¹Posgrado de Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo. Texcoco, Estado de México. *Correo-e: efitzr@chapingo.mx

3.1 RESUMEN

Dada la importancia comercial del cultivo de fresas a nivel mundial, el monitoreo continuo y preciso resulta crucial para maximizar rendimientos y optimizar el uso de recursos. Por otro lado, con la información recopilada con los sistemas de monitoreo, también se pueden mejorar las labores y procesos de postcosecha. Con los avances en las tecnologías y algoritmos de aprendizaje profundo aplicados en visión artificial, resulta de interés su implementación para monitorear los cultivos de fresas. En este trabajo se compararon 8 algoritmos de aprendizaje profundo, incluyendo: YOLOv8n, YOLOv8s, YOLOv9t, YOLOv9s, YOLOv10n, YOLOv10s, YOLOv11n y YOLOv11s; para detectar y clasificar por el grado madurez los frutos de fresas cultivadas en camellones en macro túneles. El modelo YOLOv8n mostró mayor capacidad de generalización superando al resto de los modelos en las métricas de detección. Con diferencias menores al 1 %, estuvo entre los tres mejores modelos. Evaluado en todo el conjunto de imágenes de prueba, con un umbral de IoU y una confianza de 0.7, alcanzó un $F1 = 0.755$, $mAP@0.50 = 0.76$ y $mAP@0.50-0.95 = 0.716$. En las inferencias realizadas en el conjunto de datos de prueba, incluyendo el tiempo de preprocesamiento, inferencia y post procesamiento, el modelo YOLOv8n alcanzó la mayor velocidad con hasta 144.93 FPS, seguida por YOLOv8s con 136.99 FPS.

Palabras clave: Sistema de monitoreo de cultivos, Visión por computadora, Redes neuronales convolucionales.

3.2 INTRODUCCIÓN

El conteo preciso y en tiempo real de frutos permite ajustar estrategias y planificar la introducción al mercado. En las fresas, es esencial por su corta vida útil, aunque complejo por su crecimiento en racimos (Zhou et al., 2024). Actualmente, los sistemas de visión artificial y algoritmos de aprendizaje profundo, específicamente las redes neuronales convolucionales, han mejorado la precisión y rapidez en la detección de frutos (Du et al., 2023). En particular, Huang et al. (2023) distinguen dos tipos principales de algoritmos para la detección de objetos: 1) los algoritmos de dos etapas, que generan regiones de interés y luego clasifican los objetos, ofreciendo alta precisión, pero con una velocidad relativamente baja, y 2) los algoritmos de una etapa, que predicen directamente la ubicación y clase de los objetos, proporcionando mayor rapidez a expensas de una menor precisión.

En investigaciones recientes, la detección de fresas maduras e inmaduras ha sido abordada mediante el uso de algoritmos de aprendizaje profundo. Ejemplos de ello incluyen: YOLOv3 mejorado para condiciones de sombra y superposición de frutos (Xiaogang et al., 2020), AlexNet en imágenes hiperespectrales (Gao et al., 2020), YOLOv8s con LW-Swin Transformer para invernaderos con iluminación variable (Yang et al., 2023), una versión modificada de Faster R-CNN para el campo abierto (Li et al., 2023), YOLOv8 mejorado (Wang, Wang, et al., 2024), RTF-YOLO para detección en tiempo real bajo iluminación cambiante y oclusión (Shen et al., 2024), YOLACT para detección en tiempo real en camellones (Wang, Ding, et al., 2024), y YOLOv5s con Swin Transformer en entornos de agricultura vertical (Zhou et al., 2024).

Para la detección de flores y frutos, se ha utilizado YOLOv7 con Swin Transformer y GS-ELAN (Bai et al., 2024). Además, la identificación de fresas maduras, inmaduras y flores en imágenes aéreas capturadas por drones se ha realizado mediante RTSD-Net, basado en YOLOv4-Tiny (Zhang et al., 2022), y una versión modificada de YOLOv8 (C. Wang, Han, et al., 2024). Para detectar fresas verdes, semi maduras y maduras en camellones dentro de túneles de polietileno, se

empleó un algoritmo basado en MobileNetv2 (Mejia et al., 2023). En cuanto a la detección bajo cuatro condiciones de oclusión, se utilizó DSW-YOLO, una variante de YOLOv7 (Du et al., 2023). Finalmente, YOLOv4s fue implementado para clasificar cuatro niveles de madurez y flores en camellones a campo abierto (Tituaña et al., 2024).

A pesar de los avances en detección de fresas con modelos de aprendizaje profundo, persisten desafíos que requieren optimización. Este estudio compara el desempeño de los ajustes realizados a los modelos YOLOv8n, YOLOv8s, YOLOv9t, YOLOv9s, YOLOv10n, YOLOv10s, YOLOv11n, YOLOv11s y RTDETR-l en macro túneles con oclusión e iluminación variables, clasificando las fresas en siete grados de madurez según la Norma Mexicana NMX-FF-062-SCFI-2002.

3.3 MATERIALES Y MÉTODOS

La capacidad de generalización de un modelo de visión por computadora está influenciada por las imágenes utilizadas durante su entrenamiento, la diversidad de escenarios y las variaciones de iluminación capturadas. Por tal motivo, como se muestra en la Figura 12, el primer paso es la disponibilidad de un conjunto de imágenes que representen las diferentes condiciones del objeto a detectar y clasificar. Con el propósito de tener un mayor número de imágenes e incrementar los escenarios de las condiciones del cultivo de fresas, para el proceso de entrenamiento, además de la base de datos generada, se utilizó el conjunto de imágenes StrawDI_Db1 de Pérez-Borrero et al. (2020).

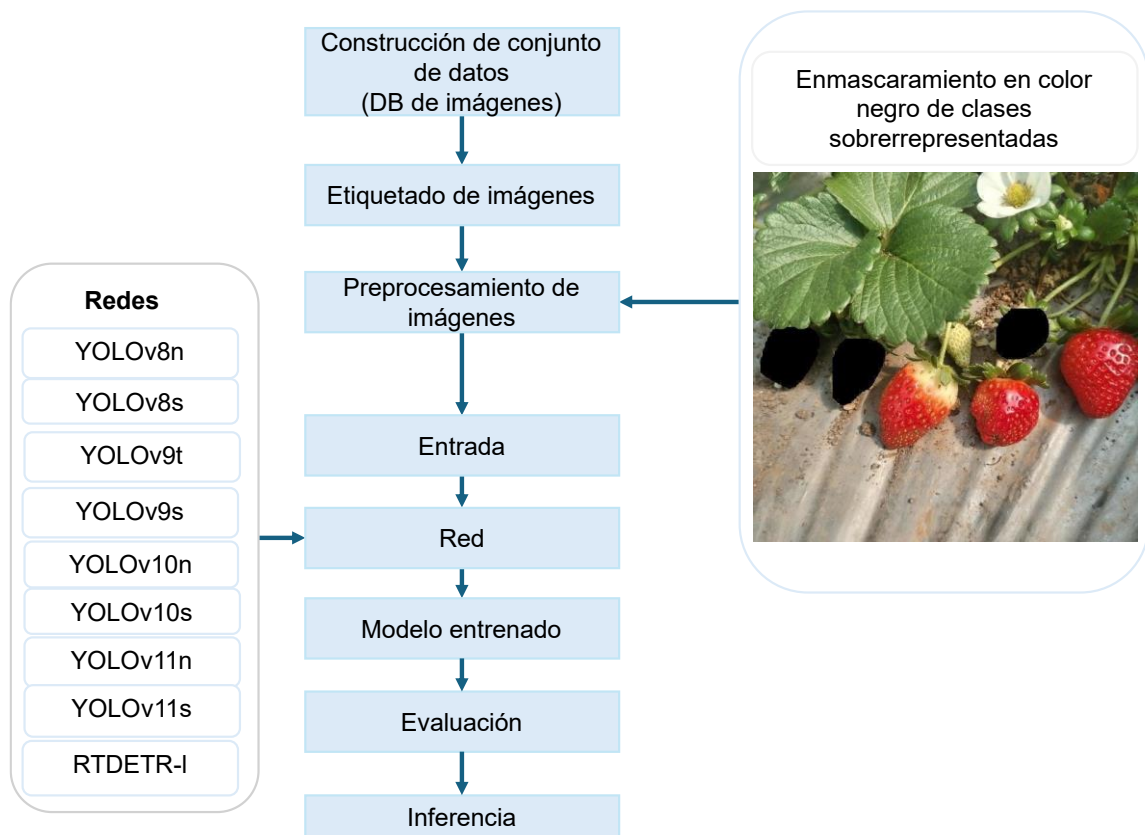


Figura 12. Diagrama de flujo del entrenamiento de la red para la detección de frutos de fresas

3.3.1 Construcción del conjunto de datos (imágenes anotadas)

3.3.1.1 Adquisición de imágenes

Para construir una base de imágenes para los procesos de entrenamiento y evaluación de las CNN empleadas, se tomaron imágenes de plantas de fresas cultivadas en camellones con acolchado gris y en macetas con plástico negro bajo condiciones en macro túneles, situadas en el municipio de Santa María Tlahuitoltepec, Oaxaca, México. La captura de imágenes se realizó bajo diversas condiciones climáticas exteriores que incluyeron días soleados, semi nublados y nublados. De esta forma se obtuvieron imágenes con diversas condiciones de iluminación y una amplia gama de grados de madurez, tamaños, apariencias y oclusiones (Figura 13 y Figura 14).

Se identificaron siete grados de madurez conforme a la Norma Mexicana NMX-FF-062-SCFI-2002 (Figura 15), la cual define los criterios adecuados para la comercialización y el consumo en fresco dentro del territorio nacional, excluyendo las fresas destinadas al procesamiento industrial.

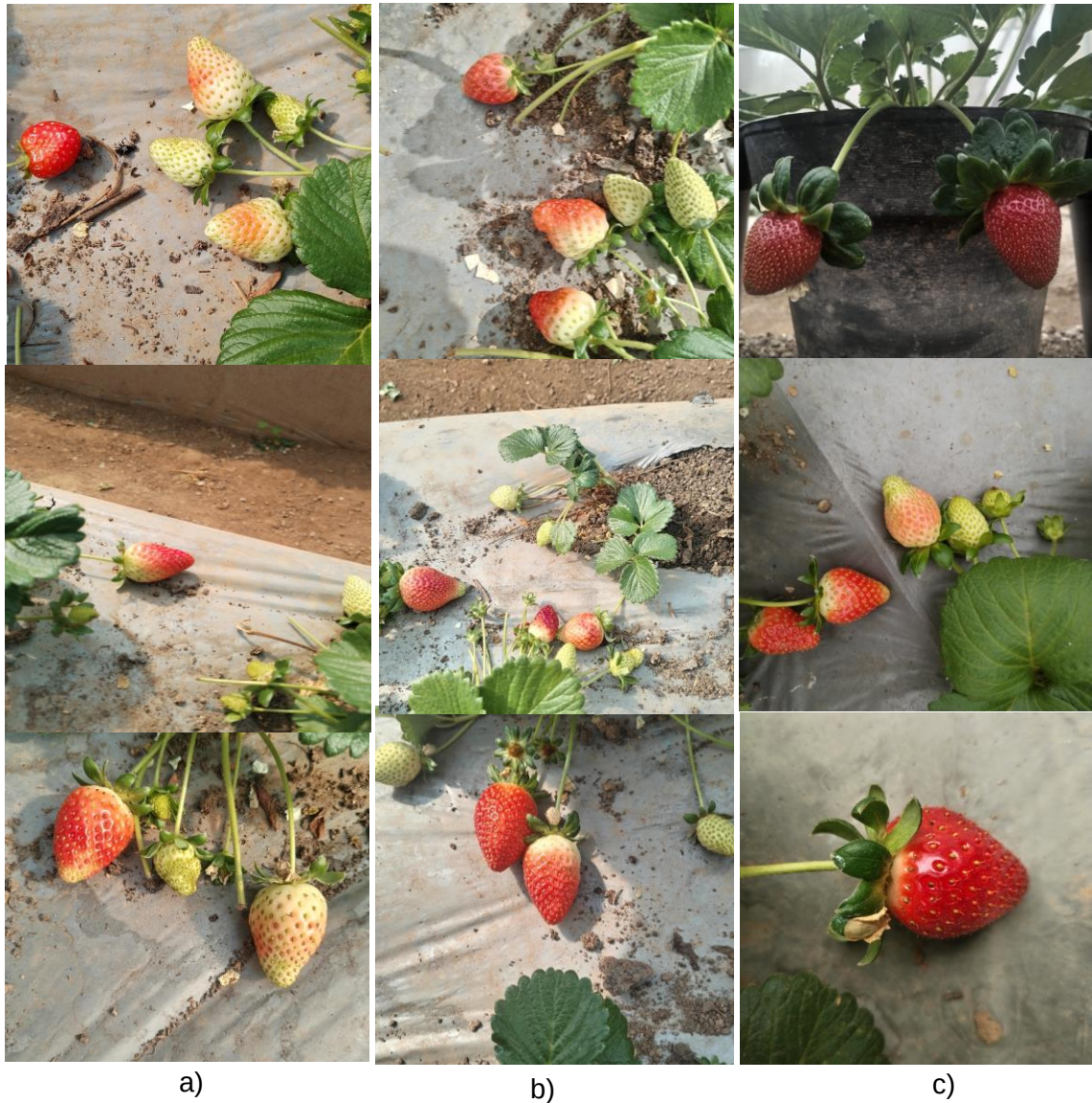


Figura 13. Imágenes muestra con diferentes grados de iluminación, a) luz a través del túnel, b) luz directa y c) exterior nublado

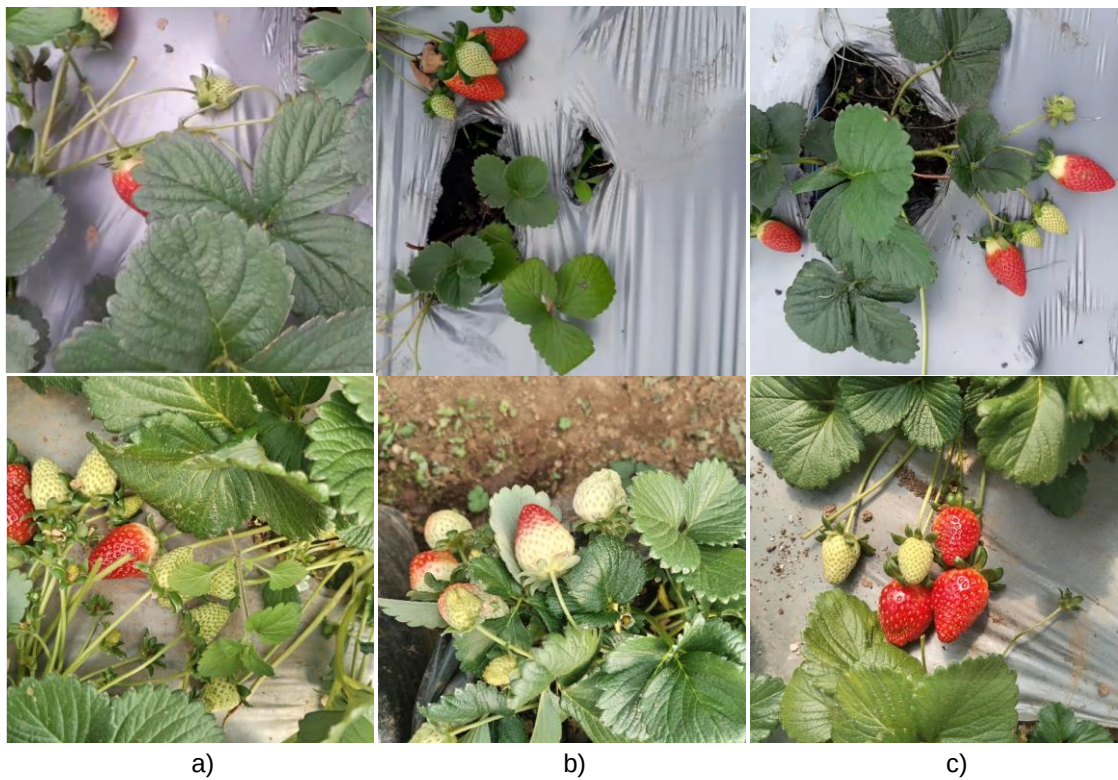


Figura 14. Imágenes de muestra con diferentes grados de oclusión, a) alta b) moderada y c) sin oclusión

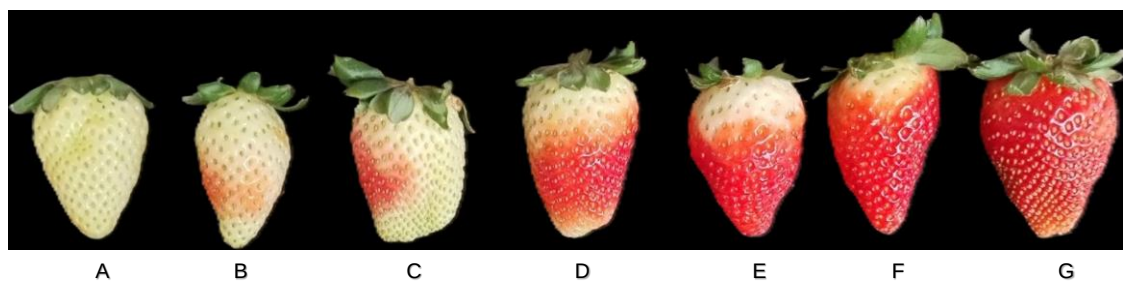


Figura 15. Grados de madurez en fresas, de acuerdo con la Norma Mexicana NMX-FF-062-SCFI-2002

Se obtuvieron 3,340 imágenes, para ello tal como se muestra en la Figura 16, se utilizó la cámara de un celular Honor x8 modelo tfy-lx3 con una resolución de 3456 x 3456 píxeles en formato JPG. Se consideraron dos formas principales de captura, variando la distancia verticalmente de 0.15 a 0.30 m sobre el camellón y en otra variando también los ángulos de captura entre 0 y 90 grados.

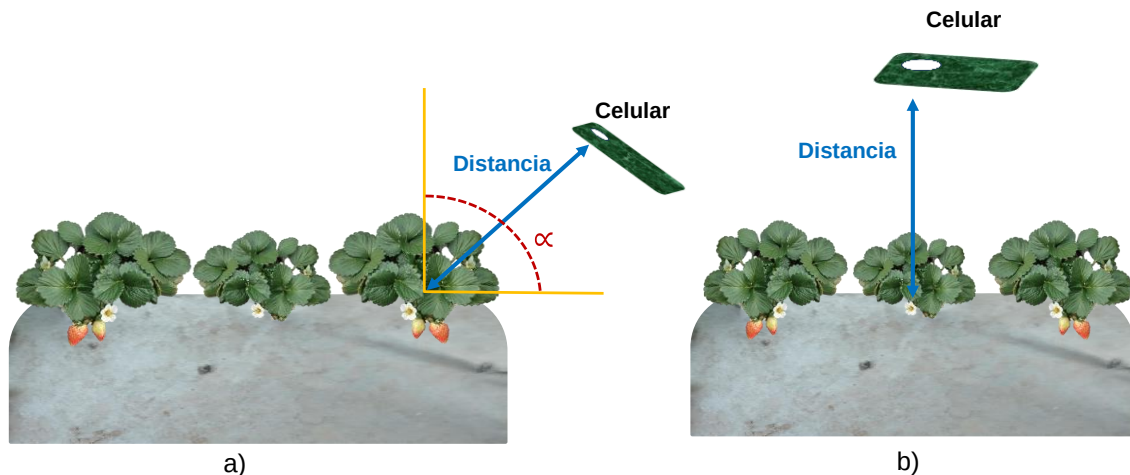


Figura 16. Captura de imágenes, (a) al borde del camellón y (b) centrado

El segundo conjunto de imágenes utilizadas (3,100 de la base de datos StrawDI_Db1) tienen una resolución de 1008 x 756 píxeles en formato PNG. Como lo mencionan Pérez-Borrero et al. (2020) están cultivadas sobre un acolchado negro y obtenidas de 20 plantaciones, dentro de un área aproximada de 150 hectáreas, en la provincia de Huelva, España.

3.3.1.2 Preparación y etiquetado de imágenes.

Dado que el estudio utilizó modelos YOLO preentrenados con el conjunto de datos COCO (T.-Y. Lin et al. (2014), todas las imágenes (DB propia y StrawDI_Db1) se redimensionaron con una resolución de 640x640 píxeles siguiendo la recomendación de Ultralytics (2024a). Con el propósito de tener anotaciones de todas las clases lo más balanceada posible, se combinaron ambas bases de datos.

Para evitar la sobre representación de algunas clases específicas, algunas imágenes fueron excluidas y otras se les aplicó un enmascaramiento para ocultar en color negro aquellos frutos con exceso de anotaciones totales (Figura 17). De esta forma, se obtuvo un total de 2,151 imágenes, se destinó el 70% para el entrenamiento, 20% para la validación y un 10% para la prueba. El etiquetado de imágenes se realizó en la plataforma online Roboflow (Dwyer et al., 2024a),

aplicando el formato de etiqueta YOLOv5 pytorch y las anotaciones se guardaron en formato de archivo txt.

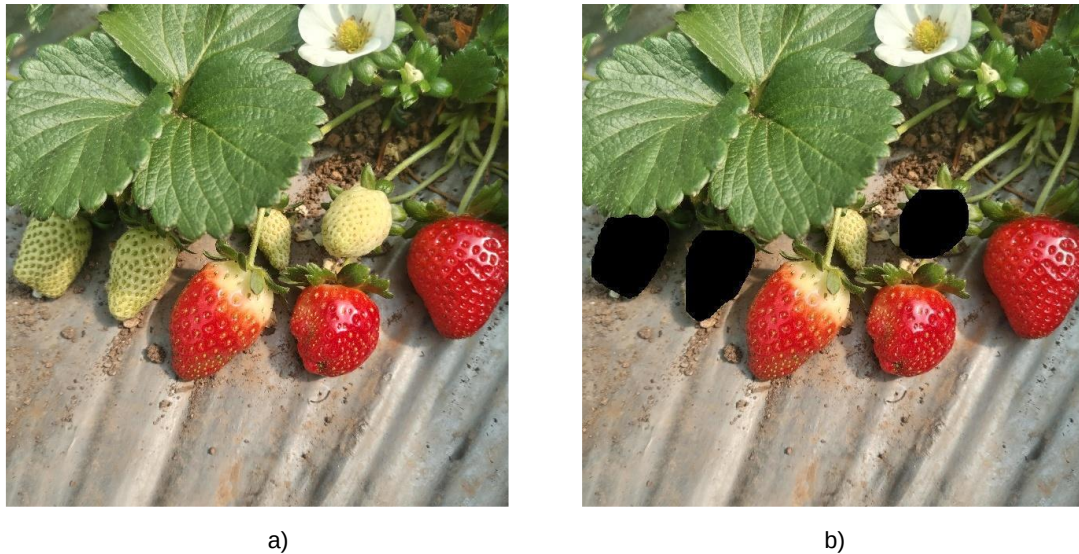


Figura 17. Enmascaramiento en color negro de frutos sobrerrepresentados, a) original y b) enmascarado

3.3.2 Selección de algoritmos de aprendizaje profundo

Los modelos de detección de objetos se dividen principalmente en algoritmos de una etapa o de dos etapas. Según Bochkovskiy et al. (2020) los algoritmos de una etapa tienen una arquitectura compuesta por la entrada, la red troncal, cuello y la predicción densa. Mientras que los algoritmos de dos etapas tienen además una etapa de predicción dispersa. En este estudio se optó por utilizar algoritmos de una etapa, específicamente modelos de la familia YOLO, debido a su rapidez, simplicidad y eficacia en dispositivos con recursos limitados.

Como menciona Terven et al. (2023), YOLO ha experimentado una evolución notable desde YOLOv1 hasta YOLOv8, introduciendo mejoras clave como las cajas de anclaje, arquitecturas multi-escala, técnicas de aumentación de datos y mecanismos de atención, lo que ha permitido equilibrar con éxito la velocidad y precisión en diversas aplicaciones, incluidas aquellas en vehículos autónomos y vigilancia por video. El modelo YOLO-NAS, presentado en 2023, destacó por

superar en precisión a sus predecesores, gracias al uso de búsqueda de arquitectura neuronal (NAS), bloques básicos fáciles de cuantificar y avanzados esquemas de entrenamiento, adaptándose así a diferentes resoluciones y dispositivos, lo que refuerza su idoneidad para la detección en tiempo real. C.-Y. Wang et al. (2024) presentaron YOLOv9, un modelo que ofrece mejoras sustanciales en la eficiencia de parámetros y carga computacional, logrando alta precisión en el conjunto de datos MS COCO mediante el uso de información de gradiente programable (PGI) y la red de agregación de capas eficientes generalizadas (GELAN). El modelo YOLOv10, desarrollado por A. Wang et al. (2024), optimizó la detección de objetos en tiempo real mediante asignaciones duales consistentes sin necesidad de *NMS*, lo que permite generar múltiples predicciones por objeto en la fase de entrenamiento y una única predicción en la inferencia, superando en precisión, velocidad y menor latencia a otros modelos. Finalmente, el modelo YOLOv11 lanzado en septiembre del 2024 por Ultralytics, introduce mejoras en la extracción de características, optimiza velocidad y eficiencia, logra mayor precisión con menos parámetros, presenta adaptabilidad a diversos entornos y además soporta una amplia variedad de tareas. La amplia adopción y continua mejora de YOLO en la literatura justifica plenamente su elección en el presente trabajo.

3.3.3 Proceso de entrenamiento

Se seleccionaron los modelos YOLOv8n, YOLOv8s, YOLOv9t, YOLOv9s, YOLOv10n, YOLOv10s, YOLOv11n y YOLOv11s. Todos ellos preentrenados en el conjunto de datos COCO T.-Y. Lin et al. (2014). Ultralytics proporciona hiperparámetros optimizados para estos modelos, que les permite lograr un equilibrio óptimo entre velocidad y precisión. La configuración de algunos de estos hiperparámetros es: imagen de entrada con 640 x 640 píxeles, un tamaño del lote de 16, optimizador SGD, una tasa de aprendizaje inicial de 0.01, un coeficiente de decaimiento de pesos de 0.0005, un factor de momento de 0.937 y se estableció un máximo de 500 épocas de entrenamiento. Sin embargo, para prevenir el sobreajuste, se implementó un mecanismo de paciencia de 100

épocas; es decir, si no se observaba una mejora en las métricas de validación durante este período, el proceso de entrenamiento finalizaba automáticamente.

Las tareas de entrenamiento, validación e inferencia se realizaron utilizando la librería Ultralytics 8.3.3 (Ultralytics, 2024b), implementado en un entorno Python 3.9.19 con la biblioteca de aprendizaje profundo PyTorch 2.4.1 y soporte para CUDA 12.4. Se utilizó una GPU NVIDIA GeForce RTX 4090, con 24 GB de memoria VRAM. El entorno de ejecución contaba con 24 núcleos de CPU, 127.8 GB de RAM.

3.3.4 Métricas de evaluación

En la detección de objetos la métrica estándar para reportar el desempeño de un modelo es el *mAP* (mean Average Precision, por sus siglas en inglés), esta métrica mide tanto la precisión como la exhaustividad (*recall*), reflejando la capacidad del modelo para identificar correctamente objetos y detectar el mayor número de objetos reales posible. Utiliza una curva de precisión-exhaustividad, calculando el área bajo la curva para equilibrar el aumento en objetos detectados sin sacrificar precisión. Además, para múltiples categorías, la *AP* calcula la precisión promedio en cada clase y promedia estos valores, ofreciendo así una evaluación completa del rendimiento del modelo en cada categoría y en general (Terven et al., 2023)

El *mAP* utiliza la Intersección sobre Unión (*IoU*) que mide la superposición entre las cajas delimitadoras predichas y las cajas de verdad, siendo un indicador clave para evaluar la precisión en la localización de objetos. El Umbral de *IoU* define el nivel mínimo de superposición necesario para considerar una predicción como exitosa, lo cual es útil para clasificar las detecciones en verdadero positivo (*TP*), falso positivo (*FP*) o falso negativo (*FN*). Un *TP* es una detección correcta donde el *IoU* es igual o superior al umbral. Un *FP* es una detección incorrecta donde el objeto no está presente o la localización es imprecisa. Finalmente, un *FN* es un objeto existente que no ha sido detectado por el modelo. La precisión (*P*) refleja el porcentaje de predicciones positivas correctas del modelo, mientras que el

recall (R) indica el porcentaje de verdaderos positivos detectados. El puntaje $F1$ combina ambas métricas para evaluar el rendimiento, especialmente en situaciones de clases desbalanceadas.

Típicamente el mAP usa un umbral de IoU de 0.5 de modo que otra forma de expresarlo es $mAP@0.50$, por otra parte, para una evaluación mas rigurosa se emplea el $mAP@0.50-0.95$ el cual calcula el mAP en umbrales de IoU de 0.5 a 0.95 con incrementos de 0.05 y luego los promedia.

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP^i \quad (3)$$

$$F1\ score = \frac{2 \cdot (P \cdot R)}{(P + R)} \quad (4)$$

$$IoU = \frac{A \cap B}{A \cup B} \quad (5)$$

Donde n denota el número de imágenes en el conjunto de prueba.

3.4 RESULTADOS Y DISCUSIÓN

3.4.1 Entrenamiento

Finalizado el entrenamiento, YOLOv9t tuvo el tamaño en MB más ligero, mientras que YOLOv10s y YOLOv9s lograron la convergencia más rápida(Cuadro 3).

YOLOv9t a pesar de haber tardado más en alcanzar la convergencia logró el más alto valor en la métrica $mAP@0.50$ mientras que, YOLOv10n mostró el nivel más bajo, con una diferencia entre ambos de 2.7 % (Figura 18).

Cuadro 3. Comparativa de arquitecturas de los modelos resultantes

Modelo	Épocas	Época de convergencia	Capas	Parámetros	Tamaño (MB)
YOLOv8s	250	150	186	9,830,373	20.0
YOLOv11s	209	109	238	9,415,509	19.2
YOLOv10s	173	73	293	8,040,378	16.6
YOLOv9s	173	73	504	6,196,357	13.3
YOLOv10n	296	196	285	2,697,146	5.8
YOLOv8n	335	235	186	2,685,733	5.7
YOLOv11n	229	129	238	2,583,517	5.5
YOLOv9t	344	244	504	1,731,189	4.2

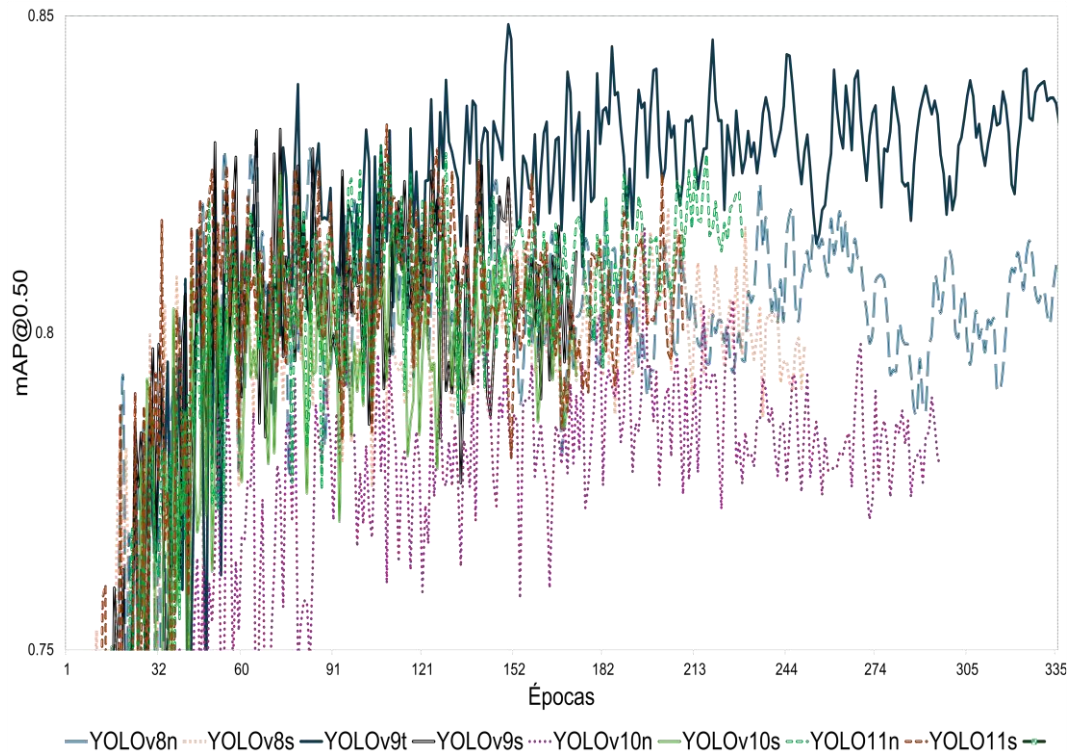


Figura 18. Cambios en las curvas $mAP@0.50$ durante el entrenamiento de varios modelos

Los valores alcanzados en la métrica $mAP@0.50-0.95$ a pesar de usar altos umbrales de IoU, en promedio tuvieron solo un valor 5.7 % menor que la métrica $mAP@0.50$, esto indica que los modelos tienen una alta capacidad para localizar correctamente a las fresas en la imagen. Los valores alcanzados por los modelos en el $mAP@0.50-0.95$ tiene una desviación estándar de 1%, 0.8 % para el $mAP@0.50$ y 0.7 % para el puntaje $F1$ (Cuadro 4).

Cuadro 4. Resultado de las métricas de detección obtenidas durante el entrenamiento

Modelo	P	R	$F1$	$mAP@0.50$	$mAP@0.50-0.95$
YOLOv9t	0.786	0.773	0.779	0.844	0.792
YOLOv11s	0.724	0.809	0.764	0.833	0.774
YOLOv9s	0.724	0.812	0.765	0.832	0.774
YOLOv11n	0.709	0.812	0.757	0.828	0.773
YOLOv8n	0.708	0.830	0.764	0.824	0.773
YOLOv8s	0.755	0.766	0.760	0.823	0.768
YOLOv10s	0.730	0.787	0.757	0.825	0.760
YOLOv10n	0.737	0.777	0.756	0.817	0.759

3.4.2 Evaluación

Para determinar la capacidad de generalización de cada modelo es importante emplear imágenes no usadas durante el entrenamiento y evaluación, por ello se utilizó todo el conjunto de imágenes de prueba. Además, ya que normalmente se reporta un solo valor general de $mAP@0.50$ y $mAP@0.50-0.95$ por cada modelo, es necesario utilizar un umbral de confianza lo más bajo posible para que la curva precisión-recall pueda calcularse sin filtros previos en la predicción, por esta razón se fijó un umbral de IoU de 0.6 y un umbral de confianza de 0.001 (Cuadro 5). El modelo YOLOv9t logró el mejor rendimiento en $mAP@0.50-0.95$, por otra parte, la diferencia promedio entre el $mAP@0.50$ y $mAP@0.50-0.95$ obtenido por todos los modelos fue de 7.5 % que indica alta capacidad de localización de fresas.

Cuadro 5. Métricas generales de cada modelo obtenidas en la evaluación

Modelo	<i>P</i>	<i>R</i>	<i>F1</i>	<i>mAP@0.50</i>	<i>mAP@0.50-0.95</i>
YOLOv9t	0.729	0.798	0.762	0.823	0.763
YOLOv10n	0.736	0.809	0.771	0.816	0.758
YOLOv11n	0.736	0.791	0.763	0.830	0.757
YOLOv8s	0.732	0.799	0.764	0.822	0.757
YOLOv8n	0.752	0.787	0.769	0.817	0.757
YOLOv11s	0.751	0.753	0.752	0.813	0.744
YOLOv10s	0.734	0.755	0.744	0.815	0.734
YOLOv9s	0.752	0.755	0.753	0.799	0.734

Se realizó una exhaustiva evaluación de todos los modelos, para determinar el mayor umbral de confianza a la cual los modelos ofrecen alto equilibrio entre precisión y recall, se usaron umbrales de confianza de 0.05 a 0.95 con incrementos de 0.05. La caída de los valores *mAP@0.50-0.95* (Figura 20) con respecto a los valores del *mAP@0.50* (Figura 19) en promedio fueron un 4.8%. Con un umbral de confianza mayor que 0.8 la caída del rendimiento en todos los modelos es cada vez más pronunciada en ambas gráficas. El rendimiento de los modelos en los umbrales de confianza de 0.05 a 0.8 tiene una desviación estándar de 2.2 % para los valores *mAP@0.50* y 1.8 % para los valores *mAP@0.50-0.95*. Al evaluar todos los modelos en el puntaje *F1* en los umbrales de confianza de 0.05 a 0.95 con incrementos de 0.05, los resultados en promedio disminuyeron apenas un 0.2% usando un umbral de *IoU* de 0.8 frente a un umbral *IoU* de 0.4, por lo que es representativo realizar la evaluación del puntaje *F1* usando un umbral *IoU* de 0.8 (Figura 21).

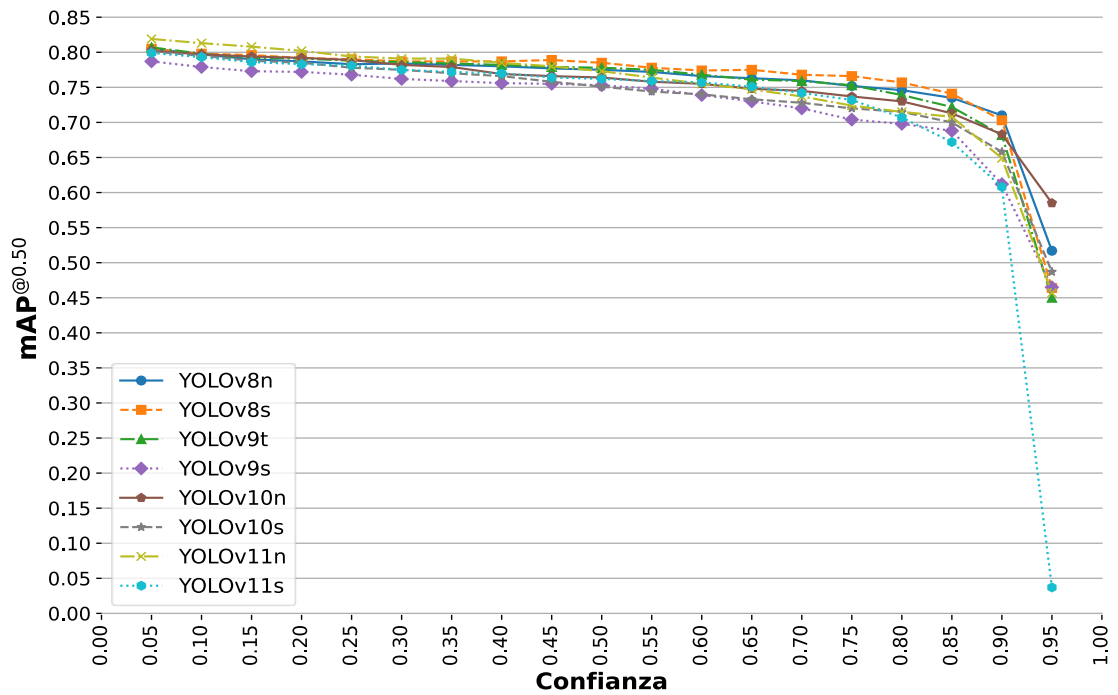


Figura 19. Impacto del umbral de confianza en el $mAP@0.50$

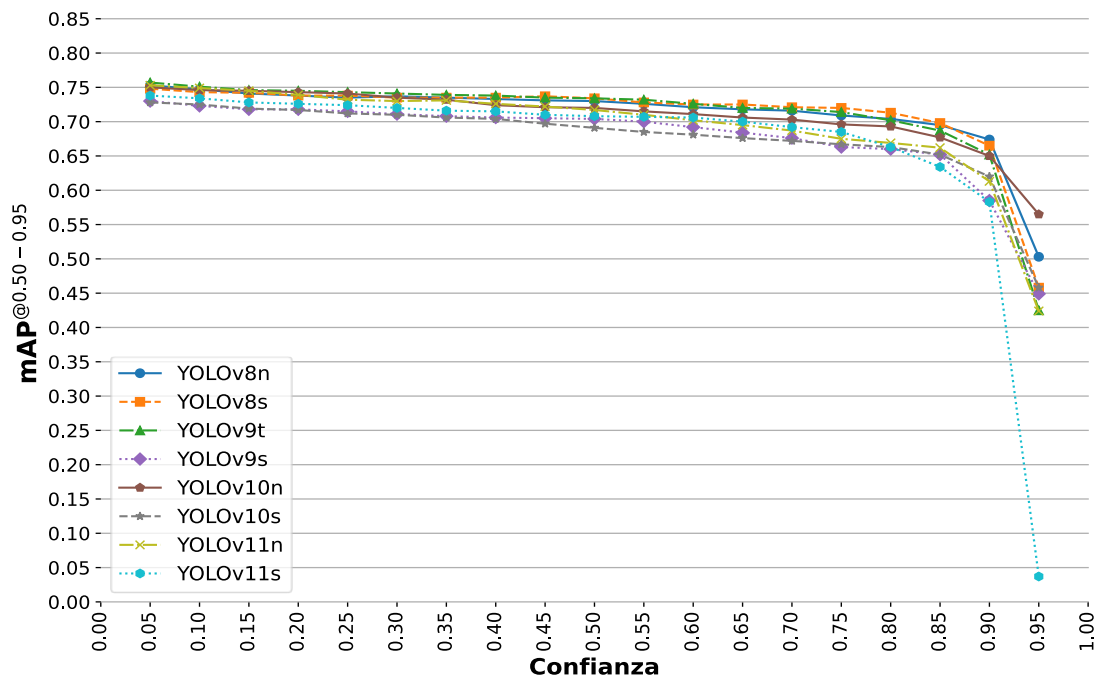


Figura 20. Impacto del umbral de confianza en el $mAP@0.50-0.95$

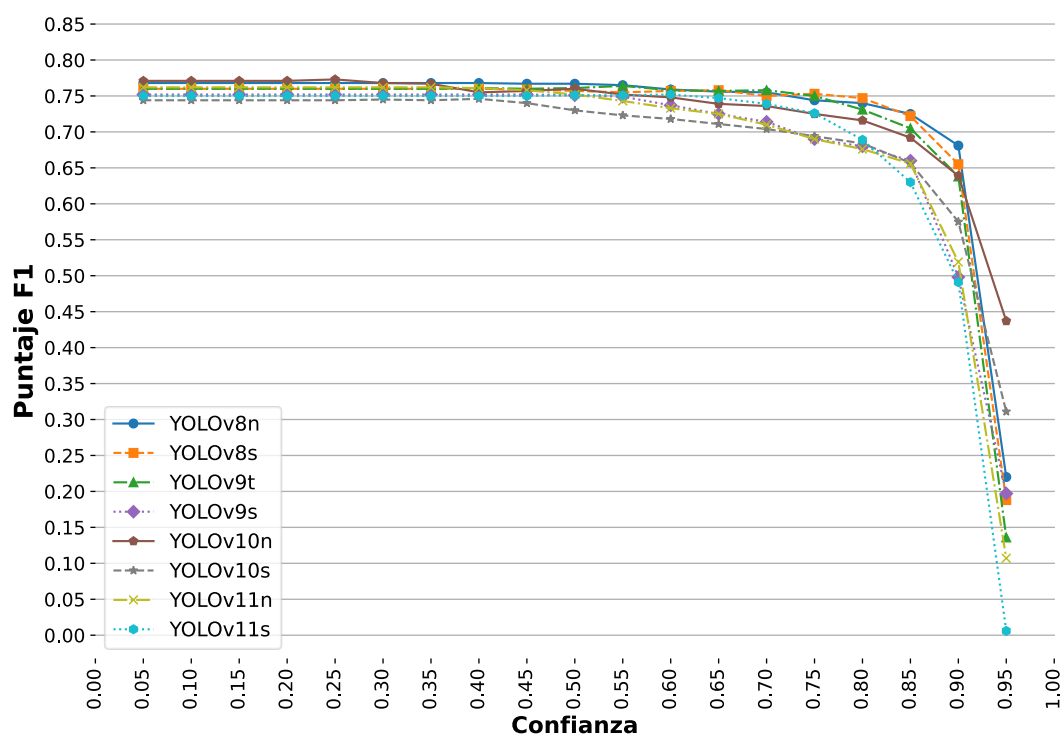


Figura 21. Evaluación del puntaje $F1$ en el conjunto de datos de prueba con confianza de 0.05 a 0.95

Para visualizar el rendimiento en un caso concreto, todos los modelos se evaluaron con un umbral de IoU y confianza en 0.7 (Cuadro 6).

Cuadro 6. Resultado de las métricas de detecciones evaluadas en el conjunto de datos de prueba

Modelo	P	R	$F1$	$mAP@0.50$	$mAP@0.50-0.95$
YOLOv8s	0.795	0.713	0.752	0.769	0.722
YOLOv9T	0.810	0.714	0.759	0.763	0.720
YOLOv8n	0.786	0.726	0.755	0.760	0.716
YOLOv10n	0.816	0.671	0.736	0.745	0.703
YOLOv11s	0.794	0.693	0.74	0.743	0.693
YOLOv11n	0.791	0.646	0.711	0.738	0.688
YOLOv9s	0.774	0.664	0.715	0.722	0.678
YOLOv10s	0.794	0.633	0.704	0.728	0.672

En los resultados de la evaluación, YOLOv8s, YOLOv9t y YOLOv8n alcanzaron los mejores rendimientos, con valores de $F1$ y $mAP@0.50$ muy similares. En el $mAP@0.50-0.95$ aunque YOLOv8s alcanzó un rendimiento 0.2 % mayor que YOLOv9t y 0.6 % mayor que YOLOv8n, lo cierto es que su tamaño en MB es 3.5 veces más grande que YOLOv8n y 4.8 veces más que YOLOv9t.

YOLOv9t experimentó una reducción de solo un 4.3 % en su valor $mAP@0.50-0.95$ con respecto a su $mAP@0.50$, mientras que YOLOv8n tuvo una disminución similar del 4.4 %. Además, considerando que la velocidad de procesamiento en cuadros por segundo (FPS, por sus siglas en inglés) es un factor clave para seleccionar un modelo, se llevaron a cabo predicciones utilizando todo el conjunto de imágenes de prueba (Cuadro 7).

Cuadro 7. Velocidad de predicción en cada uno de los modelos

Modelo	Preprocesamiento (ms)	Inferencia (ms)	Post procesamiento (ms)	FPS
YOLOv8n	1.4	4.3	1.2	144.93
YOLOv8s	1.6	4.5	1.2	136.99
YOLOv10s	1.4	6.1	0.4	126.58
YOLOv11n	1.5	5.3	1.1	126.58
YOLOv11s	1.6	5.5	1.1	121.95
YOLOv10n	2.0	6.1	0.4	117.65
YOLOv9t	1.5	9.9	1.0	80.65
YOLOv9s	2.4	14.8	1.8	52.63

El cálculo de FPS incluye el tiempo de preprocesamiento, inferencia y posprocesamiento, evaluado en imágenes con dimensiones de (1, 3, 640, 640). Las predicciones se realizaron utilizando el mismo hardware empleado durante el entrenamiento, con un umbral de IoU y una confianza de 0.7.

Debido a que YOLOv8n presentó la mayor velocidad en FPS, con un desempeño en las métricas de detección de fresas bastante competitivo, un número de parámetros bastante reducido y un tamaño de 5.7 MB, es el modelo mejor equilibrado. Para ver su capacidad de clasificación de los datos se presenta su matriz de confusión normalizada (Figura 22), el cual es resultado de una evaluación con un umbral de IoU y una confianza de 0.7. La matriz de confusión normalizada es aquella donde los valores de cada fila o columna se han escalado para que sumen 1, permitiendo comparar el desempeño del modelo entre clases con tamaños de muestra diferentes. Los valores altos en la diagonal indican mejor desempeño del modelo por clase, en este caso YOLOv8n puede clasificar con una precisión del 84 % fresas totalmente maduras (G), y con una precisión igual o superior al 70 % los grados de madurez A, B y C. El valor más bajo se presentó al clasificar el grado de madurez D con una precisión del 62 %.

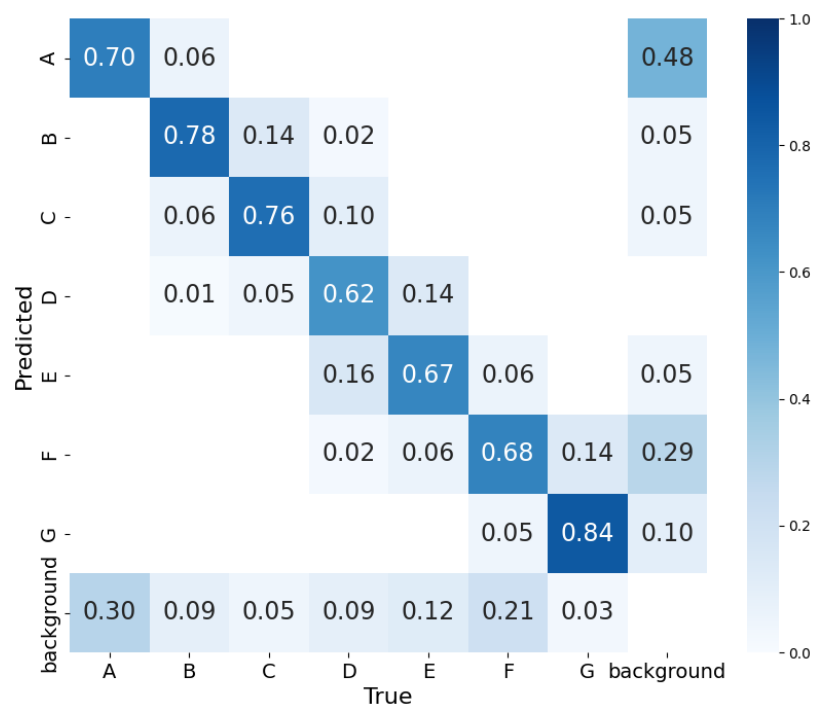


Figura 22. Matriz de confusión normalizada del modelo YOLOv8n evaluada en el conjunto de datos de prueba

La Figura 23 muestra las salidas de algunas predicciones del modelo en el conjunto de imágenes de prueba, con un umbral *IoU* de 0.7 y confianza del mismo

valor. Las letras arriba de las cajas indican la clase predicha y los números son el porcentaje de confianza en la etiqueta asignada. En general las confianzas en todas las detecciones superan el 80%, las cajas predichas son bastante precisas.

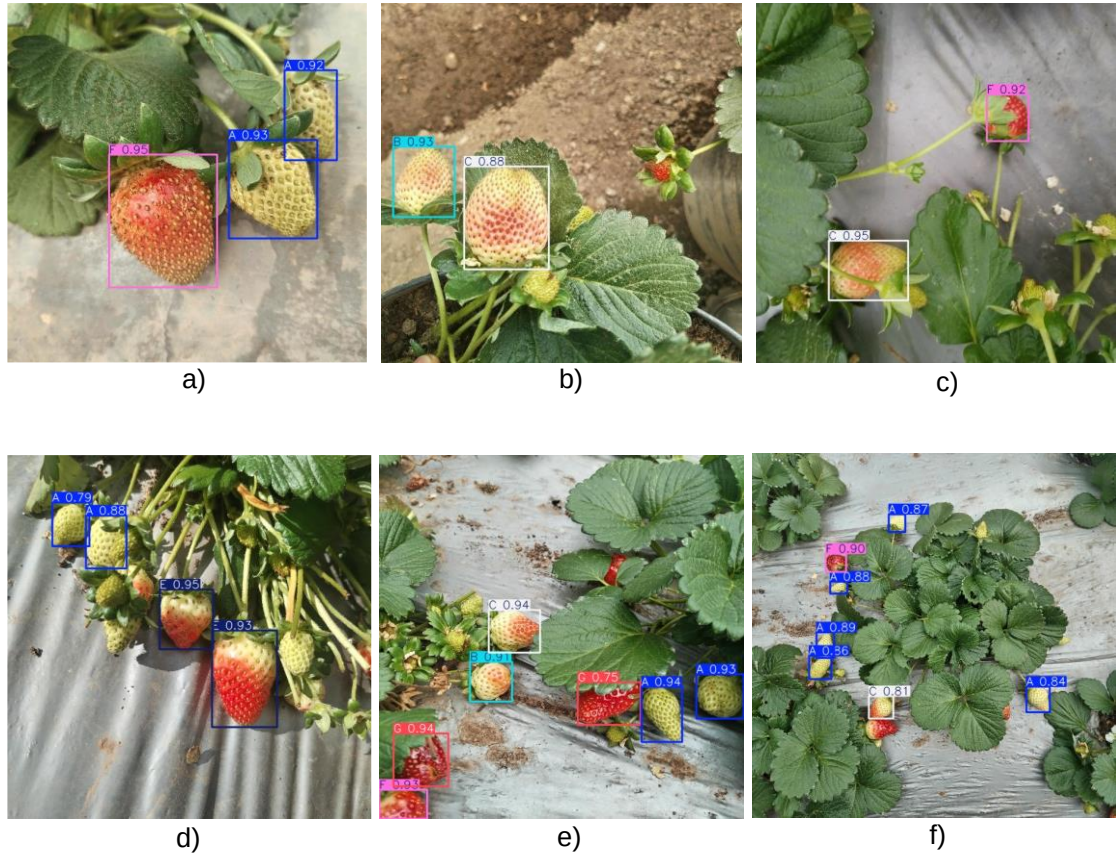


Figura 23. Salidas de las predicciones del modelo YOLOv8n

3.5 CONCLUSIÓN

En este trabajo se desarrolló un modelo de detección de fresas, realizando ajustes a los modelos preentrenados: YOLOv8n, YOLOv8s, YOLOv9t, YOLOv9s, YOLOv10n, YOLOv10s, YOLOv11n y YOLOv11s. Se recolectó una base de datos de imágenes de fresas cultivadas en camellones bajo macro túneles, con diversos escenarios de iluminación y oclusión.

Al evaluar la capacidad de generalización de los modelos en el conjunto de imágenes de prueba con un umbral de *IoU* 0.7 y confianza del mismo valor,

YOLOv8s, YOLOv9t y YOLOv8n alcanzaron los mejores desempeños en las métricas de detección. La red YOLOv8n alcanzó la mayor velocidad con hasta 144.93 cuadros por segundo (FPS), seguida por YOLOv8s con 136.99 FPS y YOLOv9t con 80.65 FPS. En la métrica $mAP@0.50-0.95$, YOLOv8n alcanzó un rendimiento de 71.6 % que es un 0.6% menor que YOLOv8s y 0.4% menor que YOLOv9t. En las métricas $mAP@0.50$ y F1 YOLOv8n logró un valor de 0.76 y 0.755 respectivamente, en comparación YOLOv8s y YOLOv9t resultaron ser superiores en menos de un 1 %.

YOLOv9t tuvo una caída en su valor $mAP@0.50-0.95$ con respecto al $mAP@0.50$ de solo 4.3 %, y YOLOv8n una caída de 4.4 % lo que indica un comportamiento robusto para localizar y delimitar con alta precisión las fresas presentes en las imágenes.

YOLOv8n con un tamaño de 5.7 MB, presenta alta velocidad de procesando de cuadros por segundo con alta capacidad de generalización reflejado en un rendimiento bastante competitivo para detectar y clasificar fresas en sistemas de producción bajo macro túneles.

3.6 LITERATURA CITADA

- Bai, Y., Yu, J., Yang, S., & Ning, J. (2024). An improved YOLO algorithm for detecting flowers and fruits on strawberry seedlings. *Biosystems Engineering*, 237, 1–12. <https://doi.org/10.1016/j.biosystemseng.2023.11.008>
- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection*.
- Du, X., Cheng, H., Ma, Z., Lu, W., Wang, M., Meng, Z., Jiang, C., & Hong, F. (2023). DSW-YOLO: A detection method for ground-planted strawberry fruits under different occlusion levels. *Computers and Electronics in Agriculture*, 214, 108304. <https://doi.org/10.1016/j.compag.2023.108304>
- Dwyer, B., Nelson, J., Hansen, T., & et. al. (2024). *Computer vision*. Roboflow 3.0 Object Detection (Fast).
- Gao, Z., Shao, Y., Xuan, G., Wang, Y., Liu, Y., & Han, X. (2020). Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with

- deep learning. *Artificial Intelligence in Agriculture*, 4, 31–38. <https://doi.org/10.1016/j.aiia.2020.04.003>
- Huang, Y., Qian, Y., Wei, H., Lu, Y., Ling, B., & Qin, Y. (2023). A survey of deep learning-based object detection methods in crop counting. *Computers and Electronics in Agriculture*, 215, 108425. <https://doi.org/10.1016/j.compag.2023.108425>
- Li, J., Zhu, Z., Liu, H., Su, Y., & Deng, L. (2023). Strawberry R-CNN: Recognition and counting model of strawberry based on improved faster R-CNN. *Ecological Informatics*, 77, 102210. <https://doi.org/10.1016/j.ecoinf.2023.102210>
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2014). *Microsoft COCO: Common Objects in Context*.
- Mejia, G., Montes de Oca, A., & Flores, G. (2023). Strawberry localization in a ridge planting with an autonomous rover. *Engineering Applications of Artificial Intelligence*, 119, 105810. <https://doi.org/10.1016/j.engappai.2022.105810>
- Pérez-Borrero, I., Marín-Santos, D., Gegúndez-Arias, M. E., & Cortés-Ancos, E. (2020). A fast and accurate deep learning method for strawberry instance segmentation. *Computers and Electronics in Agriculture*, 178, 105736. <https://doi.org/10.1016/j.compag.2020.105736>
- Shen, S., Duan, F., Tian, Z., & Han, C. (2024). A Novel Deep Learning Method for Detecting Strawberry Fruit. *Applied Sciences*, 14(10), 4213. <https://doi.org/10.3390/app14104213>
- Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, 5(4), 1680–1716. <https://doi.org/10.3390/make5040083>
- Tituaña, L., Gholami, A., He, Z., Xu, Y., Karkee, M., & Ehsani, R. (2024). A small autonomous field robot for strawberry harvesting. *Smart Agricultural Technology*, 8, 100454. <https://doi.org/10.1016/j.atech.2024.100454>
- Ultralytics. (2024a). *Model training with ultralytics YOLO*. Ultralytics. <https://docs.ultralytics.com/modes/train/>
- Ultralytics. (2024b). *Ultralytics YOLOv8*. Ultralytics YOLO Docs. <https://docs.ultralytics.com/models/>
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). *YOLOv10: Real-Time End-to-End Object Detection*.

- Wang, C., Ding, F., Wang, Y., Wu, R., Yao, X., Jiang, C., & Ling, L. (2024). Real-Time Detection and Instance Segmentation of Strawberry in Unstructured Environment. *Computers, Materials & Continua*, 78(1), 1481–1501. <https://doi.org/10.32604/cmc.2023.046876>
- Wang, C., Han, Q., Li, C., Li, J., Kong, D., Wang, F., & Zou, X. (2024). Assisting the Planning of Harvesting Plans for Large Strawberry Fields through Image-Processing Method Based on Deep Learning. *Agriculture*, 14(4), 560. <https://doi.org/10.3390/agriculture14040560>
- Wang, C., Wang, H., Han, Q., Zhang, Z., Kong, D., & Zou, X. (2024). Strawberry Detection and Ripeness Classification Using YOLOv8+ Model and Image Processing Method. *Agriculture*, 14(5), 751. <https://doi.org/10.3390/agriculture14050751>
- Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information.
- Xiaogang, L., Cheng, F., Jianian, L., Yanli, G., Yuyang, Z., & Qiliang, Y. (2020). Identification method of strawberry based on convolutional neural network. *Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery*, 8. <https://doi.org/10.6041/j.issn.1000-1298.2020.02.026>
- Yang, S., Wang, W., Gao, S., & Deng, Z. (2023). Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-Swin Transformer. *Computers and Electronics in Agriculture*, 215, 108360. <https://doi.org/10.1016/j.compag.2023.108360>
- Zhang, Y., Yu, J., Chen, Y., Yang, W., Zhang, W., & He, Y. (2022). Real-time strawberry detection using deep neural networks on embedded system (rtsd-net): An edge AI application. *Computers and Electronics in Agriculture*, 192, 106586. <https://doi.org/10.1016/j.compag.2021.106586>
- Zhou, X., Zhang, Y., Jiang, X., Riaz, K., Rosenbaum, P., Lefsrud, M., & Sun, S. (2024). Advancing tracking-by-detection with MultiMap: Towards occlusion-resilient online multiclass strawberry counting. *Expert Systems with Applications*, 255, 124587. <https://doi.org/10.1016/j.eswa.2024.124587>

4. IDENTIFICACIÓN DE PUNTOS CLAVE EN FRESAS CON ALGORITMOS DE APRENDIZAJE PROFUNDO

Eleuterio Gonzalez-Jesus¹; Efrén Fitz-Rodríguez^{1*}, Raquel Salazar Moreno¹; José Armando Ramírez Arias¹; Luis Arturo Soriano-Avenidaño¹;

¹Posgrado de Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo. Texcoco, Estado de México.

*Correo-e: efitzr@chapingo.mx

4.1 RESUMEN

En la detección y caracterización de frutos mediante visión artificial en diversos cultivos, es importante estimar la pose de los frutos, ya que esto permite conocer su orientación, tamaño y la ubicación del pedúnculo. Este estudio compara diferentes algoritmos de aprendizaje profundo, centrándose en los modelos YOLOv8 y YOLOv11, en sus variantes nano (n), small (s), medium (m), large (l) y extra large (x). Se evaluó su capacidad para detectar dos puntos clave en los frutos de fresa, los cuales, al conectarse, representan el eje principal del fruto (longitud). Además, se analizó su rendimiento en la clasificación de siete grados de madurez de fresas cultivadas en camellones bajo macrotúneles. En la evaluación, la diferencia entre el promedio de todos los modelos en $mAP@0.50$ y $mAP@0.50-0.95$, evaluados en umbrales de confianza desde 0.05 hasta 0.95 con intervalos de 0.05, fue de solo un 0.8 %, lo que indica una alta capacidad de detección de puntos clave. En la evaluación con un umbral IoU de 0.8 y una confianza del mismo valor, la red YOLOv11l logró los mejores resultados en las métricas de localización de puntos clave. En comparación, la red YOLOv8n obtuvo un valor 1.4 % inferior en $F1$ y 1.8 % inferior en las métricas $mAP@0.50$ y $mAP@0.50-0.95$. Aunque YOLOv11l alcanzó mejores valores de detección de puntos clave y clasificación de fresas, YOLOv8n lo superó en velocidad de ejecución, siendo 227.3 % más rápido.

Palabras clave: Sistema de monitoreo de cultivos, Visión por computadora, Redes neuronales convolucionales, Medición por visión.

4.1 ABSTRACT

IDENTIFICATION OF KEY POINTS IN STRAWBERRIES WITH DEEP LEARNING ALGORITHMS

In the detection and characterization of fruits through computer vision in various crops, it is important to estimate the pose of the fruits, as this allows for determining their orientation, size, and the location of the peduncle. This study compares different deep learning algorithms, focusing on the YOLOv8 and YOLOv11 models, in their nano (n), small (s), medium (m), large (l), and extra large (x) variants. Their ability to detect two key points in strawberry fruits was evaluated, which, when connected, represent the main axis of the fruit (length). Additionally, their performance in classifying seven ripeness stages of strawberries grown in raised beds under macro tunnels was analyzed. In the evaluation, the difference between the average of all models in $mAP@0.50$ and $mAP@0.50-0.95$, evaluated at confidence thresholds ranging from 0.05 to 0.95 with intervals of 0.05, was only 0.8%, indicating a high capability of key point detection. In the evaluation with an IoU threshold of 0.8 and the same confidence threshold, the YOLOv11l model achieved the best results in key point localization metrics. In comparison, the YOLOv8n model had a 1.4% lower $F1$ score and 1.8% lower $mAP@0.50$ and $mAP@0.50-0.95$ metrics. Although YOLOv11l achieved better key point detection and strawberry classification values, YOLOv8n outperformed it in execution speed, being 227.3% faster.

Keywords: Crop monitoring system, Computer vision, Convolutional neural networks, Vision-based measurement.

4.2. INTRODUCCION

La recolección de frutas con sistemas robóticos se caracteriza por la complejidad del entorno de cultivo, que implica obstrucciones en línea directa con el fruto objetivo, y que incluye: ramas, hojas, frutos vecinos, variabilidad en tamaños, formas y ubicación aleatoria. Esto en general plantea dos desafíos principales para la visión artificial: 1) el reconocimiento y detección precisa de las frutas y la determinación exacta del punto de cosecha (Xiong et al., 2018; Zhang et al., 2024). La estimación de la pose de frutos, se ha aplicado en diferentes cultivos de frutos, por ejemplo: en la detección de la posición del tallo en pimientos verdes japoneses (Eizentals & Oka, 2016) y en la detección de racimos de jitomate (Rong, Dai, et al., 2022, Rong, Wang, et al., 2022).

En los cultivos de fresa, los algoritmos de estimación de pose se han integrado en robots cosechadores para optimizar el desprendimiento del fruto (Lin et al., 2019; Yu et al., 2020). Se han desarrollado métodos para determinar el eje de fresas maduras usando redes Mask-RCNN y R-YOLO, lo que mejora la precisión en la identificación del punto de corte (Yu et al., 2019;2020). También se han aplicado enfoques basados en la similitud de la forma de la fresa con una cometa simple para estimar su diámetro y longitud (Oo & Aung, 2018). La estimación de pose (conjunto de puntos clave) es ideal cuando se requiere identificar y localizar partes específicas de un objeto en una escena y comprender su posición relativa (Ultralytics, 2024b). Recientes investigaciones reportan la detección de puntos clave mediante algoritmos de aprendizaje profundo como HRNet combinado con sensores LiDAR para medir el ancho y la altura de fresas (fruto) (Jeong et al., 2024), Deep-ToMaToS para la estimación de pose 6D del tomate y el tallo lateral (Kim et al., 2022) y OpenPose para la orientación 2D de tomates (Kim et al., 2023).

Esta investigación utilizó fresas cultivadas en camellones bajo macrotúneles para determinar la orientación del eje y su longitud. La detección de puntos clave se hizo mediante los modelos YOLOv8 y YOLOv11 en versiones nano (n), small (s), medium (m), large (l) y extra large (x), se consideraron siete clases de madurez conforme a la Norma Mexicana NMX-FF-062-SCFI-2002.

4.3 MATERIALES Y MÉTODOS

4.3.1 Construcción del conjunto de datos

Se utilizaron 2,870 imágenes, combinando imágenes de la base de datos de imágenes recolectadas y el conjunto de imágenes StrawDI_Db1 de Pérez-Borrero et al. (2020). El método de colección y preprocesamiento de imágenes están descritos en el capítulo 3.3.1. En ambas investigaciones se consideraron siete clases de madurez de acuerdo con la Norma Mexicana NMX-FF-062-SCFI-2002.

Etiquetado de imágenes

De las 2,870 imágenes, el 70 % se usó para el proceso de entrenamiento, el 20 % para el proceso de validación y el restante 10% para la fase de prueba. El etiquetado de las imágenes se realizó en la plataforma Roboflow (Dwyer et al., 2024b), y consistió en unir dos puntos claves en el fruto mediante una línea a lo largo del eje de la fresa apuntando siempre hacia el pedúnculo (Figura 24). La cantidad de etiquetas para las siete clases de madurez se distribuyen como se muestra en el Cuadro 8. El formato de etiqueta utilizado fue YOLOv5 Pytorch y formato de archivo txt.



Figura 24. Etiquetado de imágenes con esqueleto de orientación en las fresas

Cuadro 8. Balance de etiquetas para las siete clases de grado madurez

Clases	A	B	C	D	E	F	G
No. de Etiquetas	697	572	589	686	564	745	828

4.3.2 Algoritmo de estimación de pose

La arquitectura YOLO, es un algoritmo de aprendizaje profundo caracterizado por su alta velocidad de inferencia y constante mejora en la precisión mediante nuevas actualizaciones. Actualmente YOLOv11 es el estado del arte en los modelos de detección de objetos, está diseñado para 5 tareas diferentes incluido la detección de pose. En el contexto de la detección de pose, YOLOv8n es el predecesor directo de YOLOv11, ambos desarrollados por Ultralytics (2024b).

La estimación de pose identifica la ubicación de puntos clave en una imagen, que representan partes de un objeto, como articulaciones o puntos de referencia, en coordenadas 2D o 3D. La salida del modelo es un conjunto coordenadas (puntos clave) junto con sus puntuaciones de confianza.(Ultralytics, 2024b).

4.3.3 Proceso de entrenamiento

Los modelos implementados fueron YOLOv8 y YOLO11 en sus variantes (n, s, m, l y x) desarrollados por Ultralytics (2024b) y pre entrenados con el conjunto de datos COCO (T.-Y. Lin et al., 2014).

Las tareas de entrenamiento, validación e inferencia se realizaron con la biblioteca Ultralytics 8.3.3 (Ultralytics, 2024b) en un entorno Python 3.9.19, utilizando PyTorch 2.4.1 con soporte para CUDA 12.4. Se empleó una GPU NVIDIA GeForce RTX 4090, que cuenta con 24 GB de memoria VRAM, y el sistema de ejecución disponía de 24 núcleos de CPU y 127.8 GB de RAM.

Los ajustes de los hiperparámetros incluyen: imagen de entrada con 640 x 640 píxeles, tasa de aprendizaje inicial de 0.01, optimizador SGD, tamaño del lote de 16, factor de momento de 0.937, coeficiente de decaimiento de pesos de 0.0005 y una paciencia de 100 épocas; es decir, ante la ausencia de alguna mejora en las métricas de validación durante este período, el proceso de entrenamiento finaliza en automático.

4.3.4 Métricas de evaluación

Las métricas para evaluar el rendimiento de los modelos en la detección de puntos clave, son el $mAP@0.50$ y $mAP@0.50-0.95$, las mismas que implementa Ultralytics. Las métricas como: la intersección sobre unión (IoU), el umbral de

IoU, la precisión (*P*), Recall (*R*), detecciones verdadero positivo (*TP*), falso positivo (*FP*) y falso negativo (*FN*) se describen en el capítulo 3.3.4. La Media de la Precisión Promedio ($mAP@0.50$) toma en consideración la precisión y el recall evaluada en umbral *IoU* de 0.5, mientras que el $mAP@0.50-0.95$ considera el promedio de la evaluación de un rango de umbrales *IoU* que va desde 0.5 a 0.95 con intervalos de 0.05, lo cual brinda una visión global del rendimiento del modelo ante diversas exigencias de localización de objetos o en este caso de puntos clave, ambas métricas se calculan con el umbral de confianza lo mas cercano posible a cero. La métrica del puntaje *F1*, resulta útil para evaluar la capacidad del modelo en la clasificación multiclase, combina precisión y recall en una única puntuación, donde un valor cercano a 1 es mejor y un valor cercano a 0 es peor.

4.4 RESULTADOS Y DISCUSIONES

4.4.1 Entrenamiento

En el entrenamiento los modelos YOLOv11 en sus distintas variantes resultaron más pequeños en comparación con sus homólogos de YOLOv8 (Cuadro 9). El rendimiento alcanzado por cada modelo se presenta en el Cuadro 10, donde $BX = mAP@0.50$ y $BY = mAP@0.50-0.95$.

Cuadro 9. Comparativa de arquitecturas de los modelos resultantes

Modelo	Épocas	Época de convergencia	Capas	Parámetros	Tamaño (MB)
YOLOv11n	246	146	257	2,655,343	5.7
YOLOv8n	397	297	205	2,757,559	5.9
YOLOv11s	315	215	257	9,701,991	19.8
YOLOv8s	211	111	205	10,116,855	20.6
YOLOv11m	328	228	322	20,884,855	42.3
YOLOv8m	272	172	255	23,767,927	48.0
YOLOv11l	370	270	483	26,134,135	53.0
YOLOv8l	293	193	305	40,287,607	81.1
YOLOv11x	312	212	483	58,745,143	118.3
YOLOv8x	364	264	305	62,929,047	126.4

Cuadro 10. Rendimiento alcanzado durante el entrenamiento de los modelos

Modelos	Cajas delimitadoras					Puntos clave				
	<i>P</i>	<i>R</i>	<i>F1</i>	<i>BX</i>	<i>BY</i>	<i>P</i>	<i>R</i>	<i>F1</i>	<i>BX</i>	<i>BY</i>
YOLOv11n	0.814	0.819	0.816	0.878	0.787	0.820	0.825	0.822	0.886	0.876
YOLOv11l	0.794	0.837	0.815	0.871	0.804	0.796	0.839	0.817	0.877	0.869
YOLOv8n	0.804	0.823	0.813	0.872	0.796	0.810	0.829	0.819	0.879	0.868
YOLOv11s	0.813	0.822	0.817	0.872	0.797	0.817	0.826	0.821	0.878	0.868
YOLOv8s	0.794	0.836	0.814	0.869	0.779	0.800	0.845	0.822	0.877	0.867
YOLOv11x	0.783	0.821	0.802	0.868	0.790	0.773	0.845	0.807	0.874	0.866
YOLOv8l	0.776	0.840	0.807	0.867	0.787	0.780	0.845	0.811	0.874	0.864
YOLOv11m	0.831	0.787	0.808	0.866	0.791	0.832	0.797	0.814	0.871	0.863
YOLOv8m	0.782	0.821	0.801	0.860	0.782	0.786	0.824	0.805	0.868	0.857
YOLOv8x	0.768	0.837	0.801	0.854	0.774	0.774	0.843	0.807	0.860	0.851

Las cajas delimitadoras se refieren a los cuadros que se utilizan propiamente en la detección de objetos. En cambio, los puntos clave en este trabajo se refiere a los dos puntos que conforman el eje de la fresa, es decir, la orientación. Los modelos están ordenados en forma descendente con base en el $mAP@0.50-0.95$ pertenecientes a los puntos clave. Se escogió la métrica $mAP@0.50-0.95$ porque refleja el rendimiento en un rango más amplio y estricto que el $mAP@0.50$, debido a que considera diferentes umbrales de intersección sobre unión (*IoU*) y se presenta el de los puntos clave porque es el interés principal en este estudio. En cuanto a los puntos clave, todos los modelos alcanzaron una diferencia promedio entre el $mAP@0.50$ y el $mAP@0.50-0.95$ de solo 0.95 % es decir, una caída del rendimiento bastante pequeña. Por otra parte, YOLOv11n alcanzó el mejor rendimiento en las métricas *F1*, $mAP@0.50$ y $mAP@0.50-0.95$, pero todos los modelos en la métrica $mAP@0.50-0.95$ alcanzaron la convergencia con un valor superior al 0.80 (Figura 25).

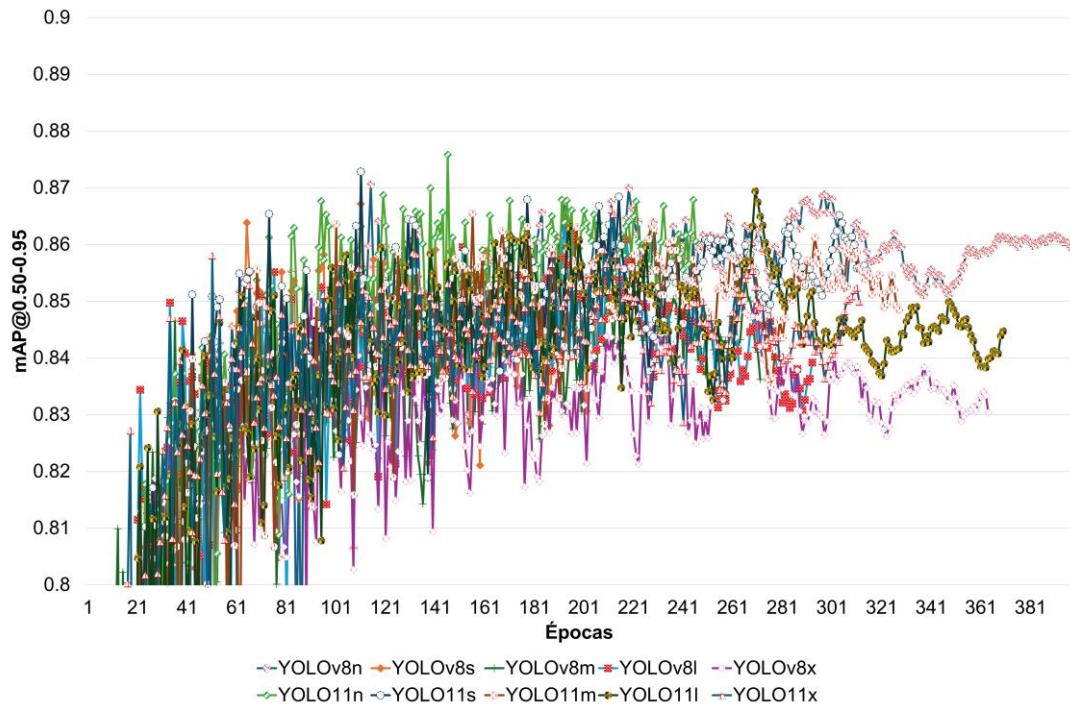


Figura 25. Rendimiento del $mAP@0.50-0.95$ para los puntos clave durante el entrenamiento

4.4.2 Evaluación

Si bien el $mAP@0.50$ y $mAP@0.50-0.95$ general de cada modelo se calcula usando un umbral de confianza tan cercano al cero como sea posible, lo cierto es que en este estudio no es de principal interés, en cambio se busca el más alto umbral de confianza con un elevado desempeño en las métricas de detección de puntos clave. En todo el proceso de evaluación se usó todo el conjunto de imágenes de prueba, ya que, al ser desconocidos para los modelos permite determinar su capacidad de generalización. Se realizó una extensa evaluación del $mAP@0.50$ en cada modelo, con umbrales de confianza desde 0.05 hasta 0.95, a intervalos de 0.05. Los resultados se muestran en la Figura 26; observando que a partir de un valor superior a 0.80 en el umbral de confianza, los valores en el $mAP@0.50$ presentaron caídas cada vez más abruptas, mientras que antes de eso casi todos los modelos mantuvieron un valor superior o muy cercano a 0.80.

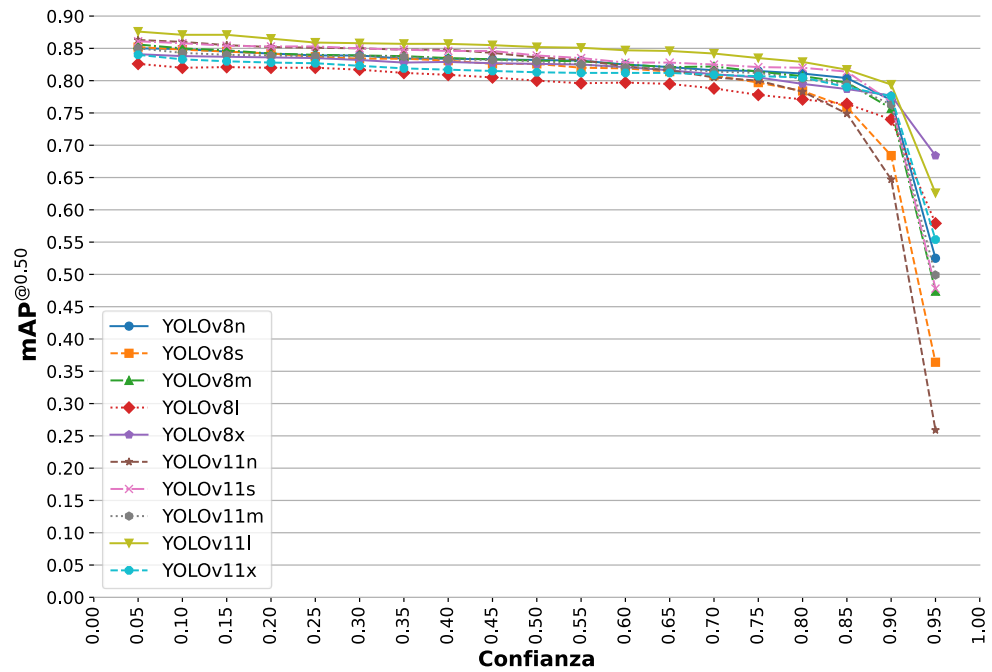


Figura 26. Desempeño del $mAP@0.50$ según el Umbral de Confianza

Para una evaluación exhaustiva de la capacidad de los modelos para localizar con precisión los puntos clave en las fresas a altos umbrales de IoU , se consideraron umbrales de confianza que varían de 0.05 a 0.95 en incrementos de 0.05 para el $mAP@0.50-0.95$ (

Figura 27).

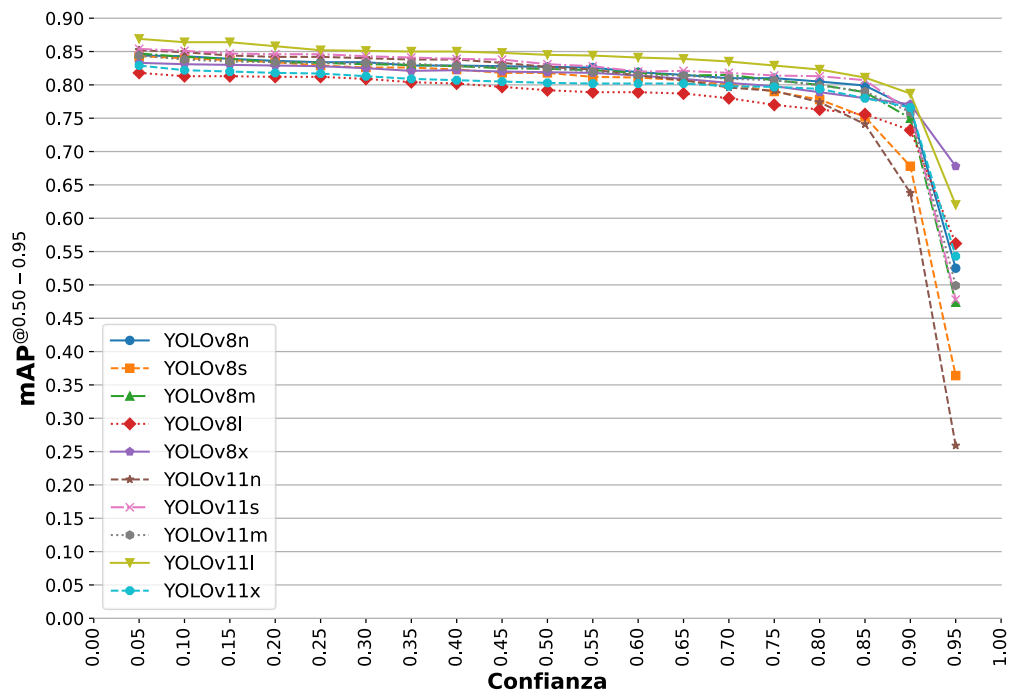


Figura 27. Desempeño del $mAP@0.50-0.95$ según el Umbral de Confianza

Los valores obtenidos entre el $mAP@0.50$ de la Figura 26 y el $mAP@0.50-0.95$ de la Figura 27 tienen una diferencia promedio de solo 0.8 %, por lo tanto, los modelos son robustos ante diferentes exigencias de umbrales IoU , es decir, tienen alta capacidad para localizar puntos claves en las fresas

El promedio del $mAP@0.50-0.95$ obtenido por todos los modelos con un umbral de confianza de 0.05 cae un 4.9 %, 6.2 %, 10.2 % y 34.3 % al evaluarse con un umbral de confianza de 0.8, 0.85, 0.9 y 0.95 respectivamente.

También, se evaluaron los puntajes $F1$ en umbrales de confianza de 0.05 a 0.95 con incrementos de 0.05 y con dos umbrales IoU : 0.4 y 0.8. Normalmente se esperaría que entre más alto sea el umbral IoU el valor $F1$ disminuiría considerablemente. Sin embargo, al pasar del umbral de IoU de 0.4 a 0.8 el rendimiento en $F1$ cayó solo un 0.2%. Por tal motivo, es conveniente y suficiente graficar los puntajes $F1$ obtenidos con el umbral de $IoU = 0.8$ (Figura 28), que confirma la robustez de los modelos para localizar los puntos clave y evitar etiquetar erróneamente objetos que no existen. En otras palabras, todos los modelos presentan alto equilibrio entre precisión y recuperación utilizando un umbral de confianza de hasta 0.8, en umbrales de confianza superiores algunos modelos muestran mayores caídas en el rendimiento.

Es de destacar que con un umbral de confianza de 0.9 y un umbral IoU de 0.8, a excepción de YOLOv11n y YOLOv8s, el resto de los modelos alcanzan valores en $F1$, $mAP@0.50$ y $mAP@0.50-0.95$ superiores al 0.7 e incluso varios tuvieron un valor muy cerca de 0.8.

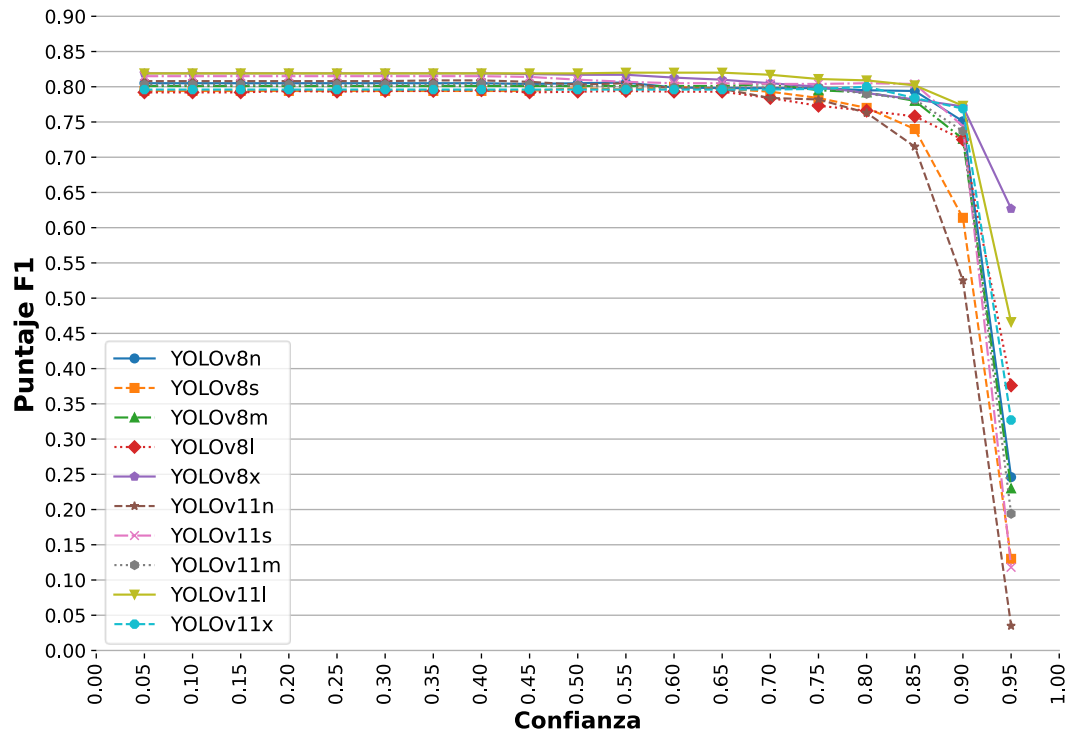


Figura 28. Puntuaciones $F1$ para puntos claves en umbrales de confianza de 0.05 a 0.95

Para evaluar el rendimiento de todos los modelos en un caso concreto, se fijó un umbral de confianza de 0.8, valor elegido con base en las gráficas de la Figuras 4, 5 y 6, ya que es el máximo valor donde los modelos difieren mínimamente, el umbral IoU asignado fue de 0.8. En los resultados (Cuadro 11), ordenados en forma descendente según su $mAP@0.50-0.95$, la desviación estándar fue de 1.8 %, la diferencia entre el mejor y peor modelo en las métricas $F1$, $mAP@0.50$ y $mAP@0.50-0.95$ fueron de 4.3, 5.8 y 6 %, respectivamente. YOLOv11l obtuvo los mejores resultados en la detección de puntos clave. Por otra parte, se realizaron predicciones en todas las imágenes de prueba con un IoU de 0.8 y confianza del mismo valor (Cuadro 12), donde YOLOv8n logró la mayor velocidad de procesamiento de cuadros por segundo (FPS).

Cuadro 11. Desempeño de las métricas de detección de puntos clave en fresas con umbral de confianza y IoU de 0.8

Modelo	P	R	F1	mAP@0.50	mAP@0.50-0.95
YOLOv11l	0.842	0.779	0.809	0.829	0.823
YOLOv11s	0.85	0.765	0.805	0.82	0.813
YOLOv8n	0.835	0.759	0.795	0.811	0.805
YOLOv8m	0.847	0.743	0.792	0.807	0.8
YOLOv11m	0.844	0.742	0.79	0.803	0.799
YOLOv11x	0.834	0.768	0.8	0.805	0.794
YOLOv8x	0.845	0.743	0.791	0.795	0.789
YOLOv8s	0.847	0.706	0.77	0.784	0.778
YOLOv11n	0.88	0.674	0.763	0.783	0.774
YOLOv8l	0.827	0.713	0.766	0.771	0.763

Cuadro 12. Predicción en el conjunto de datos de prueba con umbral de *IoU* y confianza

Modelo	milisegundos			FPS
	Preprocesamiento	Inferencia	Post procesamiento	
YOLOv8n	1.5	5.0	1.2	129.87
YOLOv8s	1.6	4.9	1.2	129.87
YOLOv11n	1.6	6.1	1.2	112.36
YOLOv11s	1.6	6.4	1.3	107.53
YOLOv8m	1.5	6.7	1.2	106.38
YOLOv11m	1.7	8.6	1.3	86.21
YOLOv8l	1.7	8.8	1.4	84.03
YOLOv8x	1.8	13.2	1.7	59.88
YOLOv11l	1.8	14.2	1.5	57.14
YOLOv11x	1.8	15.9	1.6	51.81

Con base en el Cuadro 11, YOLOv11l lideró los resultados en las métricas de detección, seguida por YOLOv11s y YOLOv8n. La diferencia del rendimiento entre los tres modelos es mínima, donde los valores de YOLOv8n en comparación con YOLOv11l fueron 1.4 % inferior en la métrica $F1$ y un 1.8 % en $mAP@0.50$ y $mAP@0.50-0.95$. En cuanto a la velocidad de ejecución, en comparación con YOLOv11l, YOLOv11s fue 188.19 % más rápido y YOLOv8n un 227.28 %.

Ya que YOLOv8n presenta un desempeño muy similar a YOLOv11s en la detección de puntos clave con una mayor velocidad en FPS, YOLOv11s puede relegarse a una tercera posición, quedando de esta forma YOLOv11l y YOLOv8n como los dos mejores modelos, uno con mayor precisión y otro con mayor velocidad.

La evaluación de los modelos reportados en el Cuadro 11 también proporciona los datos correspondientes a la clasificación de las siete clases de madurez (Figura 29), donde la precisión promedio de clasificación es de 0.754 para YOLOv8n y 0.771 para YOLOv11l, es decir, una diferencia de 1.7 %.

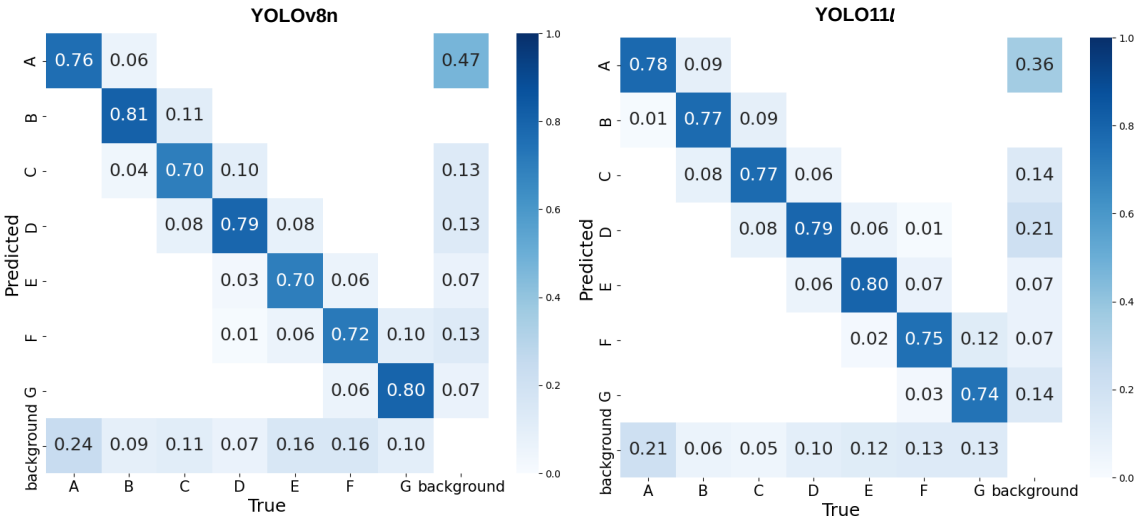


Figura 29. Matriz de confusión normalizada de YOLOv8n y YOLOv11l con un umbral de IoU y confianza de 0.8

Para una mejor visualización del desempeño de los modelos YOLOv8n y YOLOv11l, la Figura 30 y Figura 31 muestran dos ejemplos de salida de las

predicciones en el conjunto de imágenes de prueba, con un umbral de confianza de 0.6 y un umbral IoU de 0.9. La línea blanca une los dos puntos clave establecidos durante el etiquetado de imágenes y las letras en blanco representan la clase de madurez asignada. Por otra parte, los puntos en rojo y amarillo son las predicciones hechas por los modelos y los cuadros de diversos colores indican la clase predicha y su porcentaje de confianza. Se observa que ambos modelos, YOLOv8n y YOLOv11l realizaron la misma cantidad de detecciones y asignaron las mismas etiquetas, las diferencias en el porcentaje de confianza a veces favorecen a un modelo y otras veces al otro. También se observa que los puntos clave predichos por ambos modelos son bastante similares al asignado durante el etiquetado, con diferencias apenas perceptibles, esto refuerza los datos encontrados durante la evaluación de los valores $mAP@0.50$ y $mAP@0.50-0.95$, indicando una alta localización de los puntos clave.

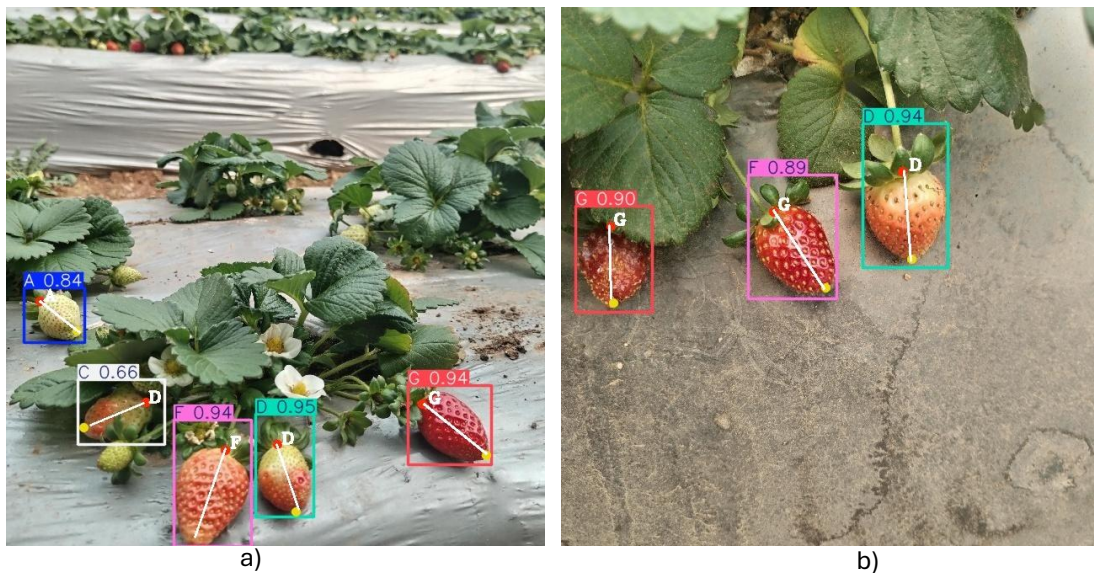


Figura 30. Comparativo de la predicción del modelo YOLOv8n y las etiquetas reales



Figura 31. Comparativo de la predicción del modelo YOLOv11l y las etiquetas reales

Debido al desempeño competitivo en la detección de puntos clave y capacidad de clasificación junto con una alta velocidad, YOLOv8n puede considerarse como el modelo mejor equilibrado. Sin embargo, para una mejor precisión en las métricas detección de puntos clave, YOLOv11l es mejor opción, aunque con una velocidad de ejecución relativamente baja.

4.5 CONCLUSION

En la evaluación de la capacidad de los algoritmos de aprendizaje profundo, YOLOv8 y YOLOv11, en sus variantes nano, small, medium, large y extra large, para detectar puntos clave y clasificarlas en una de siete clases de madurez, se encontró que todos los modelos presentaron un comportamiento bastante robusto ante diferentes umbrales de IoU, pues la diferencia entre el promedio de todos los modelos en $mAP@0.50$ y $mAP@0.50-0.95$ evaluados en umbrales de confianza desde 0.05 hasta 0.95 con intervalos de 0.05, fue de solo un 0.8 %, lo que significa una alta capacidad de detección de puntos clave y alta capacidad para evitar detectar erróneamente objetos que no existen.

El promedio del $mAP@0.50-0.95$ obtenido por todos los modelos con un umbral de confianza de 0.05 cayó un 4.9 %, 6.2 %, 10.2 % y 34.3 % al evaluarse con un umbral de confianza de 0.8, 0.85, 0.9 y 0.95 respectivamente.

En la evaluación, utilizando un umbral de IoU de 0.8 y confianza del mismo valor, YOLOv11l lideró los resultados en las métricas de detección de puntos clave, seguida por YOLOv11s y YOLOv8n. Las diferencias en las métricas fueron mínimas ya que los valores de YOLOv8n en comparación con YOLOv11l fueron 1.4 % inferior en la métrica $F1$ y un 1.8 % en $mAP@0.50$ y $mAP@0.50-0.95$.

En la velocidad de procesamiento de cuadros por segundo, en comparación con YOLOv11l, YOLOv11s fue 188.19 % más rápido y YOLOv8n un 227.28%.

Con el umbral IoU y confianza de 0.8, en la clasificación de las siete clases de madurez de las fresas, el promedio de precisión para YOLOv8n fue de 0.754 y 0.771 para YOLOv11l.

Dada la alta velocidad de procesamiento de cuadros por segundo de YOLOv8n con alto desempeño en la detección de puntos clave y clasificación, puede considerarse mejor opción que YOLOv11s, incluso mejor que YOLOv11l si la prioridad es una alta velocidad de procesamiento.

4.6 LITERATURA CITADA

- Dwyer, B., Nelson, J., Hansen, T., & et. al. (2024). *Computer vision*. Roboflow 3.0 Keypoint Detection (Fast).
- Eizentals, P., & Oka, K. (2016). 3D pose estimation of green pepper fruit for automated harvesting. *Computers and Electronics in Agriculture*, 128, 127–140. <https://doi.org/10.1016/j.compag.2016.08.024>
- Jeong, H., Moon, H., Jeong, Y., Kwon, H., Kim, C., Lee, Y., Yang, S., & Kim, S. (2024). Automated Technology for Strawberry Size Measurement and Weight Prediction Using AI. *IEEE Access*, 12, 14157–14167. <https://doi.org/10.1109/ACCESS.2024.3356118>
- Kim, J., Pyo, H., Jang, I., Kang, J., Ju, B., & Ko, K. (2022). Tomato harvesting robotic system based on Deep-ToMaToS: Deep learning network using transformation loss for 6D pose estimation of maturity classified tomatoes with side-stem. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2022.107300>

- Kim, T., Lee, D.-H., Kim, K.-C., & Kim, Y.-J. (2023). 2D pose estimation of multiple tomato fruit-bearing systems for robotic harvesting. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2023.108004>
- Lin, G., Tang, Y., Zou, X., Xiong, J., & Li, J. (2019). Guava Detection and Pose Estimation Using a Low-Cost RGB-D Sensor in the Field. *Sensors*, 19(2), 428. <https://doi.org/10.3390/s19020428>
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2014). *Microsoft COCO: Common Objects in Context*.
- Oo, L. M., & Aung, N. Z. (2018). A simple and efficient method for automatic strawberry shape and size estimation and classification. *Biosystems Engineering*, 170, 96–107. <https://doi.org/10.1016/j.biosystemseng.2018.04.004>
- Pérez-Borrero, I., Marín-Santos, D., Gegúndez-Arias, M. E., & Cortés-Ancos, E. (2020). A fast and accurate deep learning method for strawberry instance segmentation. *Computers and Electronics in Agriculture*, 178, 105736. <https://doi.org/10.1016/j.compag.2020.105736>
- Rong, J., Dai, G., & Wang, P. (2022). A peduncle detection method of tomato for autonomous harvesting. *Complex and Intelligent Systems*, 8, 2955–2969. <https://doi.org/10.1007/s40747-021-00522-7>
- Rong, J., Wang, P., Wang, T., Hu, L., & Yuan, T. (2022). Fruit pose recognition and directional orderly grasping strategies for tomato harvesting robots. *Computers and Electronics in Agriculture*, 202, 14. <https://doi.org/10.1016/j.compag.2022.107430>
- Ultralytics. (2024). *Ultralytics YOLOv8*. Ultralytics YOLO Docs. <https://docs.ultralytics.com/models/>
- Xiong, J., Lin, R., Liu, Z., He, Z., Tang, L., Yang, Z., & Zou, X. (2018). The recognition of litchi clusters and the calculation of picking point in a nocturnal natural environment. *Biosystems Engineering*, 166, 44–57. <https://doi.org/10.1016/j.biosystemseng.2017.11.005>
- Yu, Y., Zhang, K., Liu, H., Yang, L., & Zhang, D. (2020). Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot. *IEEE Access*, 8, 13. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9119372>
- Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture*, 163, 104846. <https://doi.org/10.1016/j.compag.2019.06.001>

Zhang, J., Kang, N., Qu, Q., Zhou, L., & Zhang, H. (2024). Automatic fruit picking technology: a comprehensive review of research advances. *Artificial Intelligence Review*, 57(3), 54. <https://doi.org/10.1007/s10462-023-10674-2>