



UNIVERSIDAD AUTÓNOMA CHAPINGO

POSGRADO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

"DETECCIÓN AUTOMATIZADA DE FLORESCENCIA EN PIÑA (*Ananas comosus*) MEDIANTE LA COMBINACIÓN DE APRENDIZAJE PROFUNDO Y TECNOLOGÍA DE DRONES"



TESIS



Que como requisito parcial para obtener el grado de:

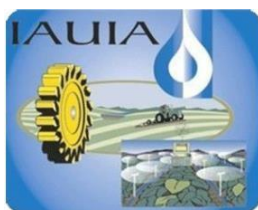
MAESTRO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

Presenta:

Martínez Nicolás Carlos

Bajo la supervisión de:

Dr. López Canteñs Gilberto de Jesús



Chapingo, Estado de México, noviembre 2024

DETECCIÓN AUTOMATIZADA DE FLORESCENCIA EN PIÑA (*Ananas comosus*) MEDIANTE LA COMBINACIÓN DE APRENDIZAJE PROFUNDO Y TECNOLOGÍA DE DRONES

Tesis realizada por el C. Martínez Nicolás Carlos 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:


DR. GILBERTO DE JESÚS LÓPEZ CANTEÑS

ASESOR:


DR. IRINEO LORENZO LÓPEZ CRUZ

ASESOR:


DR. NOÉ VELÁZQUEZ LOPEZ

DATOS BIOGRÁFICOS

DATOS PERSONALES

Nombre Martínez Nicolás Carlos

Fecha de nacimiento 09 de abril de 1999

Lugar de nacimiento Santo Tomas de Arriba, Tlacolula, Oaxaca

CURP MANC9904HOCRCR03

Profesión Ingeniero Mecánico Agrícola

Cédula profesional



DESARROLLO ACADÉMICO

Bachillerato: Preparatoria Agrícola Chapingo,
Texcoco, Estado de México.

Licenciatura: Universidad Autónoma Chapingo,
Texcoco, Estado de México

CONTENIDO

RESUMEN GENERAL	VIII
GENERAL ABSTRACT	IX
1. Capítulo 1: INTRODUCCIÓN GENERAL	10
1.1 Introducción	10
1.2 Objetivo general.....	12
1.3 Objetivos específicos	12
1.4 Distribución del contenido de la Tesis.....	13
2. Capítulo 2: REVISIÓN DE LITERATURA.....	14
2.1 Cultivo de la Piña.....	14
La agricultura de precisión.....	16
2.1.1 Percepción remota.....	17
Uso de drones en la agricultura	18
2.1.2 Ventajas del uso de drones en la agricultura	18
2.1.3 Sensores ópticos y su implementación en drones.....	19
2.2 Inteligencia Artificial en la agricultura “Aprendizaje profundo”	21
2.2.1 Redes neuronales convolucionales	24
2.2.2 Ventajas y desafíos de las CNN	29
2.2.3 Detección de objetos	29
2.2.4 Red YOLO (You Only Look Once)	31
2.2.5 La calibración y evaluación de CNN	38
Aplicaciones del aprendizaje profundo en la agricultura	41
2.3 Literatura citada	46

3.	Capítulo 3 ARTÍCULO CIENTÍFICO.....	55
3.1	Resumen	55
3.2	Abstract.....	56
3.3	Introducción	57
3.4	Materiales y Métodos.....	60
3.4.1	Adquisición de datos.....	60
3.4.2	Procesamiento de conjuntos de datos y su etiquetado	61
3.4.3	Selección de la red de detección y su entrenamiento	64
3.4.4	Métodos de evaluación	65
3.4.5	Recursos Implementados	66
3.5	Resultados.....	67
3.5.1	Entrenamiento	67
3.5.2	Evaluación	71
3.6	Discusión	73
3.7	Conclusiones	74
3.8	Literatura citada	75
4.	ANEXOS.....	85

LISTA DE FIGURAS

Figura 2.1: <i>Espectro electromagnético</i>	17
Figura 2.2: <i>Imágenes según tipos de cámara y espectros (RGB y NIR)</i>	19
Figura 2.3: <i>Comparativa de imagen multiespectral e hiperspectral.</i>	20
Figura 2.4: <i>Relación entre inteligencia artificial, aprendizaje automático (ML), red neuronal y aprendizaje profundo (DL).</i>	21
Figura 2.5: <i>Comparativa pipeline Aprendizaje automático y Aprendizaje profundo. Recuperado de</i>	22
Figura 2.6: <i>Los componentes de la CNN</i>	25
Figura 2.7: <i>Arquitectura de las CNN aplicada</i>	26
Figura 2.8: <i>Ilustración de la operación de agrupamiento máximo</i>	27
Figura 2.9: <i>Método de agrupamiento Promedio</i>	27
Figura 2.10: <i>Capa completamente conectada</i>	28
Figura 2.11: <i>Estructura de CNN</i>	29
Figura 2.12: <i>Funcionamiento R-CNN.</i>	30
Figura 2.13: <i>Comparativa R-CNN, R-CNN rápida, R-CNN más rápida y R-CNN enmascarada</i>	31
Figura 2.14: <i>Arquitectura YOLOv3 que muestra los bloques residuales y las capas de sobremuestreo para mejorar la eficiencia de la detección de objetos a través de diferentes escalas</i>	33
Figura 2.15: <i>Arquitectura detallada de YOLOv5, incluida la columna vertebral de CSPDarknet, el cuello de PANet y el cabezal de capa YOLO</i>	34
Figura 2.16: <i>Arquitectura detallada de YOLOv8</i>	35
Figura 2.17: <i>Arquitectura YOLOv9</i>	36
Figura 2.18: <i>Arquitectura YOLOv10</i>	36
Figura 2.19: <i>Arquitectura YOLO11 que muestra los nuevos bloques C3k2 y el módulo C2PSA</i>	37
Figura 2.20: <i>Evolución de los algoritmos YOLO a lo largo de los años</i>	38

Figura 3.1: a) Mosaico generado mediante Agisoft; b) imagen seccionada a partir del ortomosaico para su etiquetado 640x640.....	61
Figura 3.2: Etiquetado de la imagen (640x640) mediante la herramienta Roboflow	62
Figura 3.3: Distribución de etiquetas para su entrenamiento.....	62
Figura 3.4: mAP@0.50 calculado para el conjunto de prueba durante el entrenamiento con las confianzas previamente mencionadas.	69
Figura 3.5: <i>Evaluación de métricas de validación para el entrenamiento del modelo YOLO V8n</i>	70
Figura 3.6: Evaluación de métricas de validación para el entrenamiento del modelo YOLO V10n.....	71
Figura 3.7: Evaluación de métricas de validación para el entrenamiento del modelo YOLO V11n	71
Figura 3.8: Detecciones de floración de la piña utilizando con umbral 0.45	72

LISTA DE CUADROS

Cuadro 2.1: Datos de producción, superficie y rendimiento de los 10 países más productores de piña del mundo	14
Cuadro 2.2: Principales entidades productoras de piña en México	15
Cuadro 3.1: Características Dron Mavic 3 RGB	61
Cuadro 3.2: Ajuste de profundidad y el ancho de capas para YOLO	65
Cuadro 3.3: Hiperparámetros de entrenamiento para los algoritmos elegidos	65
Cuadro 3.4: Recursos de hardware y software implementados	66
Cuadro 3.5: Tiempo de entrenamiento de cada Modelo.....	67
Cuadro 3.6: Parámetros de entrenamiento para los modelos YOLOv8n, YOLOv10n y YOLOv11n	68
Cuadro 3.7: Métricas del conjunto de prueba para la confianza y IoU previamente mencionados.....	70
Cuadro 3.8: Comparación de Desempeño de Modelos YOLO en Detección de Floración	72

RESUMEN GENERAL

DETECCIÓN DE FLORESCENCIA EN PIÑA (ANANAS COMOSUS) MEDIANTE LA COMBINACIÓN DE APRENDIZAJE PROFUNDO Y TECNOLOGÍA DE DRONES

La detección precisa de la floración en cultivos de piña es esencial para la planificación de la cosecha y la gestión agrícola. Esta investigación tiene como objetivo implementar un sistema automatizado de detección de floración de piña utilizando drones y redes neuronales profundas, optimizando el monitoreo de cultivos y aumentando la eficiencia en su gestión. Se realizó el vuelo en la localidad de Juan Rodríguez Clara, Veracruz sobre un campo de piña en etapa de floración, utilizando un dron Mavic 3 RGB con una resolución espacial (GSD) de 0.25 cm/píxel. Se obtuvieron ortomosaicos a partir de los cuales se desarrolló una base de datos compuesta por más de 6,000 imágenes y 500,000 etiquetas de floración clasificadas en diversas condiciones lumínicas y climáticas, con el fin de evaluar la robustez de los algoritmos. Para el análisis de las imágenes, se implementó el modelo YOLOv11, que alcanzó una mAP@0.5 de 97.0% y una mAP@0.5:0.95 de 80.47%, con un nivel de confianza de 0.25, lo que indica un rendimiento alto en la detección de flores de piña. Se ajustaron parámetros clave como la tasa de aprendizaje inicial (0.001), el tamaño de lote (16) y el peso de la penalización por regularización (0.0005) para maximizar la precisión y la robustez del modelo. Se evaluaron los rendimientos del modelo YOLOv11 utilizando métricas clave como la precisión y el recall, que alcanzaron valores de 95.1% y 93.4%, respectivamente. El modelo se probó con diferentes umbrales de IoU (Intersección sobre Unión), siendo el óptimo 0.45, y en intervalos de confianza, lo que permitió refinar la capacidad de identificación del modelo en condiciones reales de campo. Los resultados muestran que la combinación de tecnología de drones y algoritmos de aprendizaje profundo, con el modelo YOLOv11 implementado, permite detectar y contar de manera automatizada la floración en piña con una alta precisión, proporcionando información valiosa para la planificación de la cosecha y la toma de decisiones agrícolas.

Palabras claves: Imágenes aéreas, CNN, Floración de piña, Drones, Detección, YOLO.

GENERAL ABSTRACT

PINEAPPLE (ANANAS COMOSUS) FLOWERING DETECTION USING A COMBINATION OF DEEP LEARNING AND DRONE TECHNOLOGY

Accurate detection of flowering in pineapple crops is essential for crop planning and farm management. This research aims to implement an automated pineapple flowering detection system using drones and deep neural networks, optimizing crop monitoring and increasing crop management efficiency. The flight was performed in the town of Juan Rodríguez Clara, Veracruz over a pineapple field in flowering stage, using a Mavic 3 RGB drone with a spatial resolution (GSD) of 0.25 cm/pixel. Orthomosaics were obtained from which a database of more than 6,000 images and 500,000 flowering labels classified under different light and weather conditions was developed to evaluate the robustness of the algorithms. For image analysis, the YOLOv11 model was implemented, which achieved an mAP@0.5 of 97.0% and an mAP@0.5:0.95 of 79.9%, with a confidence level of 0.394, indicating high performance in pineapple blossom detection. Key parameters such as initial learning rate (0.001), batch size (16) and regularization penalty weight (0.0005) were adjusted to maximize the accuracy and robustness of the model. YOLOv11 model performances were evaluated using key metrics such as precision and recall, which reached values of 95.1% and 93.4%, respectively. The model was tested at different IoU (Intersection over Union) thresholds, the optimum being 0.45, and at confidence intervals, which allowed refining the model's identification capability in real field conditions. The results show that the combination of drone technology and deep learning algorithms, with the YOLOv11 model implemented, allows automated detection and counting of pineapple flowering with high accuracy, providing valuable information for crop planning and agricultural decision making.

Key words: Aerial images, CNN, Pineapple flowering, Drones, Detection, YOLO.

Capítulo 1: INTRODUCCIÓN GENERAL

1.1 Introducción

Los cultivos especiales, como la piña (*Ananas comosus* L.), presentan desafíos únicos y requieren enfoques sofisticados para maximizar la productividad. Los productores de cultivos de gran superficie, como el maíz o la soja, tienen acceso a mapas de rendimiento basados en GPS y aplican con precisión insumos como fertilizantes y agua teniendo en cuenta la variabilidad del campo. Los productores de cultivos de especialidad no tienen acceso a estos datos ya que sus cultivos tienden a cosecharse a mano. Debido a lo anterior, los productores de especialidad han estado en desventaja, teniendo que tomar decisiones sin un nivel de conocimiento especializado.

Los productores de los cultivos de alto valor toman una serie de decisiones clave en cada ciclo de cultivo. Para la piña, los datos que respaldan estas decisiones generalmente se limitan a observaciones visuales del terreno. Pero estas observaciones son de la periferia donde la variabilidad espacial y temporal, la etapa de crecimiento y desarrollo no se pueden determinar o cuantificar en todo el campo. Caminar por el campo es difícil ya que las plantas crecen muy juntas hasta 30 000 plantas por ha. Esta falta de información completa y en tiempo real sobre las condiciones del campo puede conducir a malas decisiones que resulten en muy poca o demasiada agua, fertilización, pesticidas y reguladores de crecimiento, o una mala planificación y programación de los recursos de siembra y cosecha, incluidos los equipos y la mano de obra.

La floración natural de la piña afecta el desarrollo y la calidad de la fruta e impacta en la cosecha (Bartholomew et al., 2002; Sanewski et al., 2018). Los productores de piña utilizan productos químicos que producen etileno para inducir la floración, de modo que la mayoría de las plantas dentro de un campo produzcan frutos de alta calidad listos para cosechar casi al mismo tiempo (Bartholomew et al., 2002).

La situación ideal sería que un productor cosechara todo el campo en una sola pasada cuando hay poca variación en la floración, aumentando significativamente la productividad y eliminando el costo de cosechas adicionales.

Las tecnologías agrícolas continúan avanzando diariamente, creando nuevas oportunidades para mejorar la productividad y seguridad de los agricultores. Entre estas herramientas se encuentra el dron, que ha estado disponible durante algún tiempo, pero que recientemente ha ganado popularidad en el ámbito agrícola. La mayor accesibilidad de los drones y el incremento de la digitalización en la agricultura han permitido que estas tecnologías se conviertan en aliadas fundamentales para la producción agrícola a nivel global (Buenas Prácticas Agrícolas para proteger el agua, 2019).

Los ortomosaicos obtenidos mediante drones ofrecen imágenes aéreas con una resolución espacial a escala de centímetros. Estas ortofotos, que destacan por su alta definición y bajo costo, permiten la interpretación, análisis y cuantificación de objetos dispuestos horizontalmente, utilizando bibliotecas de aprendizaje automático para el reconocimiento de imágenes y el análisis de conglomerados (Montoya, 2020).

Para extraer datos cuantificables a partir de imágenes capturadas por sensores remotos, es necesario realizar un procesamiento que permita convertir estas imágenes en unidades e índices que tengan un significado biológico. Este proceso se logra mediante el análisis y tratamiento de las distintas variables representadas por los píxeles en las diferentes bandas espectrales, dependiendo del tipo de sensor empleado.

Las imágenes más utilizadas son las de tipo RGB, que capturan solo una parte del espectro visible dentro del espectro electromagnético completo. Estas imágenes son similares a las que obtienen las cámaras de los teléfonos móviles actuales (Padmavathi, 2016).

El aprendizaje profundo, un subconjunto del aprendizaje automático, se ha convertido en una herramienta poderosa para el manejo de grandes volúmenes de datos heterogéneos. La comunidad científica de plantas ha adoptado progresivamente esta tecnología para facilitar la interpretación de grandes conjuntos de datos, recopilados a través de trabajos de fenotipado y genotipado de alto rendimiento (Cruz, 2019). Actualmente existen proyectos y trabajos en la literatura relacionados con el cálculo basal y objetos en los cuales se utilizan diversas metodologías y técnicas para realizar esta tarea, muchas veces recurriendo al aprendizaje profundo, en el que se resalta la propiedad principal de aprender características no lineales de manera automática con base a muestras ingresadas.

1.2 Objetivo general

Implementar un sistema de detección de la floración en cultivos de piña mediante el uso de drones y algoritmos de aprendizaje profundo, con el fin de optimizar el monitoreo agrícola, mejorar la precisión en la identificación de floración y contribuir a la toma de decisiones estratégicas para maximizar la productividad y rentabilidad del cultivo

1.3 Objetivos específicos

- Desarrollar una base de datos con imágenes georreferenciadas de cultivos de piña en etapa de floración, para su uso en el entrenamiento y prueba de algoritmos de detección.
- Evaluar la precisión y efectividad de los modelos YOLOv8, YOLOv10 y YOLOv11 en la detección de florecencia en piña, analizando métricas de rendimiento como la precisión, el recall y el mAP.

1.4 Distribución del contenido de la Tesis

Este documento se estructura en tres capítulos: I) Introducción general, II) Revisión de literatura, y III) Artículo científico.

En el capítulo II, Revisión de literatura, se aborda la importancia del monitoreo del cultivo de la piña y las metodologías empleadas para la detección y conteo de la floración. Además, se describe el uso de redes neuronales convolucionales, centrándose en el funcionamiento de modelos como YOLOv8, YOLOv10 y YOLOv11 para la detección de objetos en aplicaciones agrícolas.

El capítulo III incluye el manuscrito correspondiente al trabajo de investigación, en el cual se presentan y cumplen los objetivos planteados, demostrando la efectividad de la combinación de drones y algoritmos de aprendizaje profundo en la detección automatizada de la floración en cultivos de piña.

Capítulo 2: REVISIÓN DE LITERATURA

2.1 Cultivo de la Piña

A nivel global, la piña (*Ananas comosus*) es una de las frutas tropicales más consumidas, en el mundo, con una producción anual que supera los 26 millones de toneladas. Países como Costa Rica, Filipinas y Tailandia lideran la producción y exportación de esta fruta (Cuadro 2.1), abasteciendo mercados en Europa, América del Norte y Asia (FAO, 2023). Este panorama refleja la importancia de la piña en el comercio internacional, posicionándose como un producto estratégico en las economías tropicales.

A la par de este panorama global, México se posiciona como un actor relevante en la producción de piña, ya que este cultivo aporta a su economía, situándolo entre los principales productores a nivel mundial. De acuerdo con los datos de la Secretaría de Agricultura y Desarrollo Rural (SADER), en 2020, México alcanzó 1.2 millones de toneladas de piña. El estado de Veracruz lidera la producción nacional con el 64.9%, seguido por Oaxaca con un 12.8% y Tabasco con un 6.2% (Cuadro 2.2) (Hidroponía, 2017). Estos datos muestran la relevancia del cultivo de piña como impulsor clave en estos estados del país, ampliando las oportunidades de exportación.

Cuadro 2.1: Datos de producción, superficie y rendimiento de los 10 países más productores de piña del mundo.

PAÍS	PRODUCCIÓN (t)	SUPERFICIE (ha)	RENDIMIENTO (t ha ⁻¹)
Costa Rica	3,418,155	45,000	76
Filipinas	2,730,985	66,190	41
Brasil	2,650,479	71,553	37
Tailandia	2,113,380	86,038	25
China	2,005,555	81,946	24
Indonesia	1,805,506	14,704	123
India	1,706,000	103,000	17

Nigeria	1,664,510	199,891	8
México	999,593	20,771	48
Colombia	900,395	21,641	42

Fuente: FAOSTAT

Cuadro 2.2: Principales entidades productoras de piña en México

Entidad federativa	Producción (toneladas)
Veracruz	634,227
Oaxaca	139,867
Tabasco	55,583
Quintana Roo	50,932
Jalisco	40,455
Nayarit	39,344
Colima	29,214
Chiapas	7,713
Campeche	1,277
Guerrero	526

Fuente: (SIAL – SAGARPA, 2019)

Este cultivo no solo contribuye significativamente al Producto Interno Bruto (PIB) agrícola, sino que también genera empleo y desarrollo en las regiones productoras, fortaleciendo las economías locales y mejorando la calidad de vida de las comunidades rurales (Infoagro, 2022).

Sin embargo, la producción de piña enfrenta desafíos que afectan su eficiencia y rentabilidad. Uno de los principales problemas es la floración desincronizada, que resulta en cosechas irregulares y dificulta la planificación de la producción. La floración natural puede ser desigual (Nkhoma, 2020) , para mitigar este problema, los productores recurren a la inducción artificial de la floración mediante la aplicación de agentes químicos como el etileno. Esta práctica, aunque efectiva,

requiere un monitoreo constante para evitar efectos adversos y garantizar una producción uniforme (Uriza, 2011).

La agricultura de precisión

La agricultura de precisión representa un enfoque de gestión agrícola que integra la recopilación, procesamiento y análisis de datos temporales, espaciales e individuales sobre cultivos y ganado. Esta estrategia combina información proveniente de diversas fuentes, incluyendo sensores remotos, estaciones meteorológicas, dispositivos IoT y datos históricos, con el objetivo de generar decisiones informadas y específicas para cada área o unidad productiva. Su propósito principal es optimizar el uso de recursos naturales y agrícolas, maximizando la productividad y calidad de los cultivos mientras se minimizan los costos y el impacto ambiental.

Este enfoque, fundamentado en el análisis de la variabilidad dentro de las parcelas agrícolas, permite personalizar prácticas como la fertilización, riego, control de plagas y cosecha. Así, se logra un equilibrio entre la eficiencia operativa, la rentabilidad económica y la sostenibilidad ambiental, contribuyendo a sistemas agrícolas resilientes y adaptados a los desafíos actuales, como el cambio climático y la seguridad alimentaria (ISPA, 2024).

La agricultura de precisión integra diversas tecnologías avanzadas que, en conjunto, permiten una gestión agrícola más eficiente y sostenible. Entre estas tecnologías destacan la percepción remota proximal, aérea y satelital, las cuales proporcionan datos en tiempo real sobre el estado de los cultivos. Estas herramientas permiten monitorear variables como la salud de las plantas, la humedad del suelo y la presencia de plagas o enfermedades, facilitando la toma de decisiones informadas y oportunas. Además, la inteligencia artificial desempeña un papel crucial en el análisis y procesamiento de grandes volúmenes de datos, incluyendo imágenes obtenidas por sensores remotos. Mediante algoritmos avanzados, la IA puede identificar patrones, predecir rendimientos y optimizar prácticas agrícolas, contribuyendo así a la eficiencia

operativa ya la sostenibilidad ambiental de los sistemas agrícolas (Sabir et al., 2024).

2.1.1 Percepción remota

La percepción remota implica la obtención de información sobre objetos o áreas a través de dispositivos que no están en contacto directo con ellos, utilizando sensores que capturan datos en diferentes bandas del espectro electromagnético (Figura 2.1) (Jensen, 2007). En la agricultura de precisión, esta tecnología permite monitorear extensas áreas de cultivo, proporcionando datos valiosos sobre el estado de las plantas, la humedad del suelo y la presencia de plagas o enfermedades.

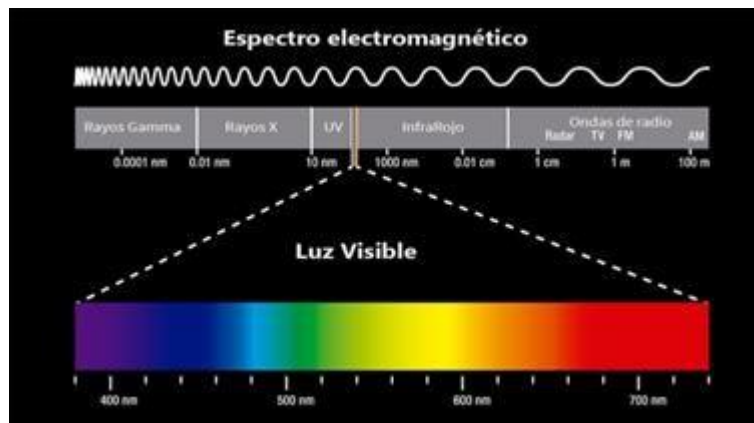


Figura 2.1: Espectro electromagnético. Recuperado Jensen (2007).

Una de las aplicaciones más destacadas de la percepción remota en la agricultura es el uso de drones equipados con sensores ópticos. Los drones, o vehículos aéreos no tripulados, ofrecen una plataforma flexible y económica para la captura de imágenes aéreas de alta resolución. Estos dispositivos pueden sobrevolar campos agrícolas y, mediante cámaras RGB (Red, Green, Blue, por sus siglas en inglés), multiespectrales, hiperespectrales o térmicas, recopilar información detallada sobre la salud y el desarrollo de los cultivos (Zhang & Kovacs, 2012).

Los sensores ópticos instalados en los drones detectan la radiación reflejada por las plantas en diferentes longitudes de onda. Esta información se utiliza para

estimar índices de vegetación, como por ejemplo NDVI (Índice de Vegetación de Diferencia Normalizada), que indican vigor y salud de las plantas (Rouse et al., 1976). Además, los sensores térmicos pueden identificar áreas con estrés hídrico al detectar variaciones en la temperatura del dosel vegetal (González-Dugo et al., 2013).

Uso de drones en la agricultura

El uso de drones en la agricultura ha transformado las prácticas tradicionales, ofreciendo soluciones que abarcan desde el monitoreo de cultivos hasta la aplicación precisa de insumos como fertilizantes y pesticidas. Este caso de estudio se enfoca en el uso de los drones dedicados al monitoreo, los cuales permiten a los agricultores optimizar recursos, mejorar la eficiencia y aumentar la productividad de sus campos.

Gracias a su capacidad para capturar imágenes aéreas, esta tecnología permite una observación continua del estado de los cultivos, facilitando la detección temprana de problemas y la implementación de medidas correctivas oportunas (Hunt et al., 2010). Esta capacidad es esencial para detectar problemas como estrés hídrico, deficiencias nutricionales o infestaciones de plagas en etapas tempranas, (Zhang & Kovacs, 2012). Por ejemplo, el DJI Mavic 3M está diseñado para la topografía aérea agrícola, equipado con cámaras multispectrales y RGB que ofrecen datos precisos para el análisis del crecimiento de los cultivos (DJI, 2024).

2.1.2 Ventajas del uso de drones en la agricultura

Las ventajas del uso de drones y sensores ópticos en la agricultura de precisión son múltiples:

Cobertura amplia y detallada: Los drones pueden cubrir grandes extensiones de terreno en poco tiempo, proporcionando imágenes de alta resolución que superan las limitaciones de las observaciones satelitales en términos de resolución espacial y temporal (Lillesand et al., 2015).

Reducción de costos: Al identificar con precisión las áreas que requieren intervención, se optimiza el uso de insumos como agua, fertilizantes y pesticidas, reduciendo costos y minimizando el impacto ambiental (Mulla, 2013).

Toma de decisiones informada: La integración de los datos obtenidos por drones y sensores ópticos en sistemas de información geográfica (SIG) permite a los agricultores analizar patrones y tendencias, mejorando la planificación y gestión de los cultivos (Bongiovanni & Lowenberg, 2004).

Minimización del Impacto Ambiental: La aplicación precisa de insumos reduce la escorrentía y la contaminación ambiental, promoviendo prácticas agrícolas más sostenibles (Huang et al., 2009).

2.1.3 Sensores ópticos y su implementación en drones.

Los sensores ópticos han transformado la agricultura de precisión, especialmente cuando se integran en drones para el monitoreo y gestión de cultivos. Estos sensores capturan información detallada sobre el estado de las plantas y el suelo, permitiendo a los agricultores tomar decisiones informadas para optimizar la producción.

Tipos de sensores ópticos utilizados en drones agrícolas:

Cámaras multiespectrales: Capturan imágenes en diferentes bandas del espectro electromagnético, incluyendo el visible y el infrarrojo cercano (Figura 2.2).

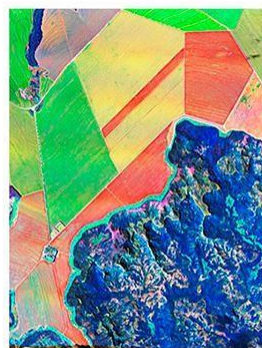


Figura 2.2: Imágenes según tipos de cámara y espectros (RGB y NIR). Recuperado de Pino (2019).

Cámaras hiperespectrales: Ofrecen una resolución espectral más detallada que las multiespectrales, capturando información en cientos de bandas estrechas (Figura 2.3). Esto facilita la identificación precisa de enfermedades, estrés hídrico y deficiencias nutricionales en los cultivos (Redagrícola, 2023).



Multispectral Images
From 2 to 10 bands



Hyperspectral Images
Hundreds of bands

Figura 2.3: Comparativa de imagen multiespectral e hiperespectral. Recuperado de Redagrícola, (2023).

Sensores térmicos: Miden la radiación infrarroja emitida por las plantas y el suelo, proporcionando datos sobre la temperatura de la superficie. Son útiles para detectar estrés hídrico, identificar áreas con riego insuficiente o excesivo y monitorear la salud general de los cultivos (Tecnoagro, 2024).

Lidar (Light Detection and Ranging): Utiliza pulsos láser para generar mapas tridimensionales del terreno y la vegetación. Es valioso para evaluar la estructura del dosel, estimar la biomasa y analizar la topografía del terreno, lo que ayuda en la planificación de riego y manejo del suelo (Tecnoagro, 2024).

La combinación de imágenes multiespectrales y térmicas permite detectar enfermedades, plagas y estrés hídrico en etapas tempranas, facilitando intervenciones oportunas y reduciendo pérdidas. Los sensores ópticos ayudan a identificar diferencias en la composición y humedad del suelo, permitiendo una aplicación precisa de fertilizantes y riego, lo que optimiza el uso de recursos. Mediante el análisis de imágenes multitemporales, es posible estimar el

crecimiento de los cultivos y predecir rendimientos, lo que facilita la planificación de la cosecha y la logística (Buriticá, 2024)

La integración de sensores ópticos en drones ha proporcionado a la agricultura datos precisos que mejoran la toma de decisiones y promueven prácticas agrícolas más sostenibles y eficientes.

2.2 Inteligencia Artificial en la agricultura “Aprendizaje profundo”

El aprendizaje profundo (deep learning), una rama avanzada del aprendizaje automático (Figura 2.4), utiliza redes neuronales artificiales de múltiples capas para modelar y analizar patrones complejos presentes en los datos. En el sector agrícola, esta tecnología ha transformado significativamente diversas prácticas, contribuyendo a la optimización de procesos y al incremento en la eficiencia de la producción agrícola (CulturaAI, 2024 Mahajan et al., 2023).

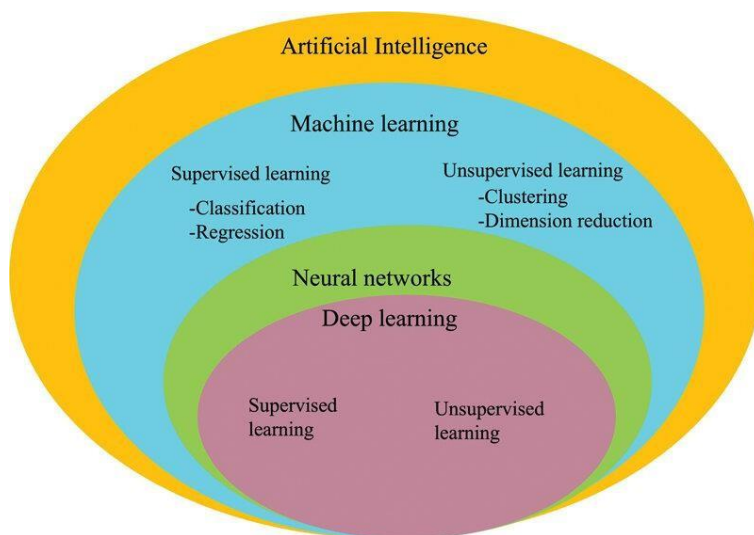


Figura 2.4: Relación entre inteligencia artificial, aprendizaje automático (ML), red neuronal y aprendizaje profundo (DL). Recuperado de Mahajan et al. (2023).

En el aprendizaje profundo, el pipeline recibe como entrada los datos de las imágenes sin procesar (Figura 2.5), para luego extraer de forma automática las características que proveen la mejor clasificación o regresión, de acuerdo con el problema que se esté trabajando. Dicha extracción se lleva a cabo durante el proceso de entrenamiento y corresponde a un proceso jerárquico, donde se

emplea un banco extenso de filtros no lineales antes de tomar decisiones (Wolfe, 2021).

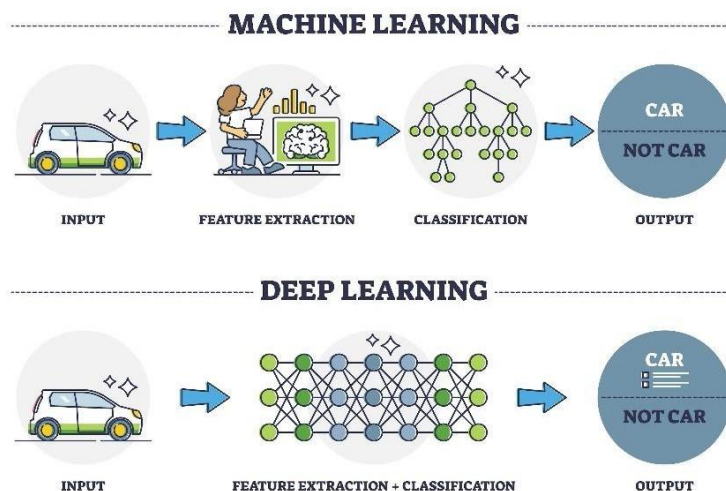


Figura 2.5: Comparativa pipeline Aprendizaje automático y Aprendizaje profundo. Recuperado de (Fernandez, 2023).

Los enfoques de aprendizaje profundo requieren una gran cantidad de datos etiquetados para maximizar su rendimiento y, por lo tanto, se puede exigir mucho a los anotadores humanos para recopilar dichos datos. Para compensar estas demandas, se ha realizado un trabajo significativo en enfoques de aprendizaje débil, semisupervisado y auto supervisado (Xie et al., 2018). Los trabajos poco supervisados que entrenan de forma incremental un modelo en una selección de datos, corrigen cualquier predicción errónea utilizando un anotador humano u "oráculo", y luego vuelven a entrenar el modelo en el conjunto más grande de datos correctamente anotados (CINVESTAV, 2020).

En la agricultura, el aprendizaje profundo se ha aplicado en diversas áreas, como la detección de enfermedades en cultivos, la estimación de rendimientos y la clasificación de especies vegetales. Se han desarrollado modelos de redes neuronales convolucionales (CNN) para identificar enfermedades en hojas de plantas a partir de imágenes, logrando precisiones superiores al 90% (Torres, 2023). Asimismo, se han implementado sistemas basados en aprendizaje

profundo para monitorear el crecimiento de los cultivos mediante imágenes satelitales, permitiendo a los agricultores tomar decisiones informadas sobre el manejo de sus tierras (Wolfe, 2021). A pesar de sus ventajas, el aprendizaje profundo en la agricultura enfrenta desafíos, como la necesidad de grandes volúmenes de datos etiquetados y la complejidad en la interpretación de los modelos. Sin embargo, con el avance de técnicas como el aprendizaje semisupervisado y la mejora en la calidad de los datos, se espera que estas limitaciones se reduzcan, ampliando las aplicaciones del aprendizaje profundo en el sector agrícola (Malambo et al., 2019).

En el ámbito del aprendizaje profundo aplicado a la agricultura, es esencial comprender las diferencias entre los enfoques supervisado, no supervisado y semisupervisado, ya que cada uno presenta características y aplicaciones específicas que pueden influir en la eficacia de las soluciones implementadas.

- Aprendizaje Supervisado

El aprendizaje supervisado se basa en la utilización de conjuntos de datos etiquetados, donde cada entrada está asociada a una salida deseada. Este enfoque es ampliamente empleado en tareas de clasificación y regresión dentro del sector agrícola. Por ejemplo, en la detección de enfermedades en cultivos, se entrenan modelos con imágenes de plantas etiquetadas como sanas o enfermas, permitiendo al sistema identificar patrones y realizar predicciones precisas en nuevas muestras (IA, 2024). No obstante, la principal limitación de este método radica en la necesidad de grandes volúmenes de datos etiquetados, cuya obtención puede ser costosa y laboriosa.

- Aprendizaje No Supervisado

A diferencia del supervisado, el aprendizaje no supervisado trabaja con datos no etiquetados, buscando descubrir estructuras o patrones ocultos sin una salida específica predefinida. En la agricultura, este enfoque es útil para tareas como el clustering de tipos de suelo o la detección de anomalías en el crecimiento de cultivos. Por ejemplo, mediante algoritmos de agrupamiento, es posible

segmentar imágenes satelitales de campos agrícolas para identificar áreas con diferentes necesidades de riego (IA, 2024). Aunque este método no requiere datos etiquetados, la interpretación de los resultados puede ser más compleja y menos precisa en comparación con el aprendizaje supervisado.

- **Aprendizaje Semisupervisado**

El aprendizaje semisupervisado combina elementos de los enfoques supervisado y no supervisado, utilizando una pequeña cantidad de datos etiquetados junto con una gran cantidad de datos no etiquetados. Este método es especialmente valioso en contextos agrícolas donde la obtención de datos etiquetados es limitada o costosa. Por ejemplo, en la clasificación de especies de plantas, se puede entrenar un modelo inicialmente con un conjunto reducido de imágenes etiquetadas y posteriormente mejorar su rendimiento incorporando un volumen mayor de imágenes no etiquetadas (Aitutoriales, 2024). Este enfoque permite mejorar la precisión del modelo y reducir la dependencia de datos etiquetados, aunque su implementación puede ser más compleja debido a la necesidad de equilibrar adecuadamente ambos tipos de datos.

En este proyecto de investigación, se optó por un enfoque supervisado debido a la disponibilidad de datos etiquetados del cultivo de piña. Este enfoque permite entrenar modelos con precisión en la detección y clasificación de características específicas del cultivo. Para ello, se implementaron redes neuronales convolucionales (CNN), conocidas por su eficacia en el procesamiento de imágenes y reconocimiento de patrones (Rio-Alvarez et al., 2024).

2.2.1 Redes neuronales convolucionales

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) han revolucionado el procesamiento de imágenes y el reconocimiento de patrones, encontrando aplicaciones significativas en la agricultura de precisión. Estas redes están compuestas por capas convolucionales, de agrupamiento (pooling) y totalmente conectadas, que permiten la extracción jerárquica de características de las imágenes de entrada (Taye, 2023). Entender los diferentes elementos que

conforman una CNN y sus usos resulta esencial para apreciar los avances en su arquitectura. En la Figura 2.6 ilustran las diversas secciones que componen la CNN.

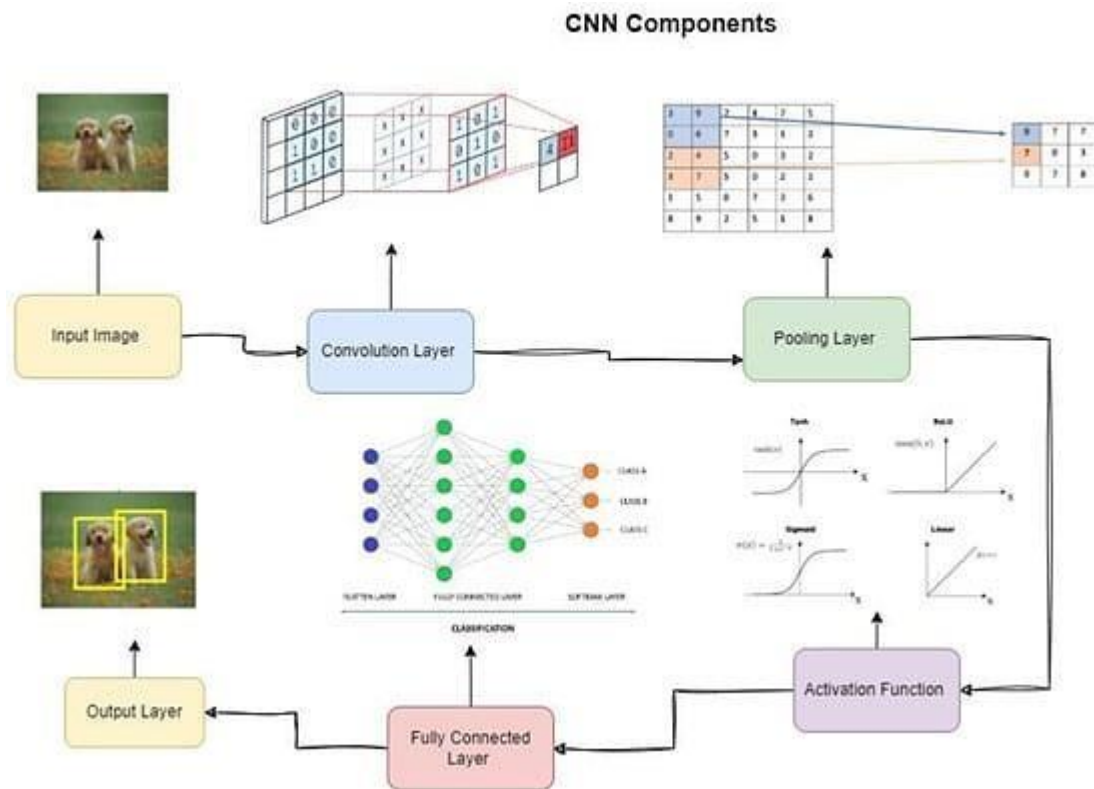


Figura 2.6: Los componentes de la CNN. Recuperado de (Taye, 2023).

- Imagen de entrada

Los píxeles son los elementos básicos que constituyen una imagen digital. Representan el sistema binario para visualizar datos, con valores que varían entre 0 y 255, organizados en un arreglo similar a una matriz para formar la estructura de la imagen. El brillo y el color de cada píxel están determinados por su valor específico (Bouchard, 2020).

Al observar una imagen por primera vez, el cerebro humano procesa una gran cantidad de información. De manera similar, las capas de una CNN se entrenan para identificar primero patrones simples, como líneas y curvas, antes de avanzar hacia formas más complejas, como rostros y objetos. Esto permite afirmar que

las CNN pueden proporcionar capacidades visuales a las computadoras (Hu et al., 2021).

- Capas Convolucionales

Las capas convolucionales son componentes esenciales en las CNN, ya que permiten la extracción automática de características locales de las imágenes de entrada, como bordes, texturas y formas. Estas capas operan aplicando filtros, también conocidos como kernels, que se desplazan a lo largo de la imagen, realizando operaciones de convolución para generar mapas de características que resaltan patrones específicos presentes en la imagen (Figura 2.7).

Estructura y Funcionamiento de las Capas Convolucionales

Una capa convolucional consta de múltiples filtros, cada uno de los cuales es una matriz de pesos que se inicializa aleatoriamente y se ajusta durante el proceso de entrenamiento. Estos filtros tienen dimensiones más pequeñas que la imagen de entrada y se aplican de manera deslizante sobre la totalidad de la imagen. En cada posición, se calcula el producto punto entre los valores del filtro y los valores correspondientes de la imagen, sumando los resultados para obtener un valor escalar que conforma el mapa de características de salida (Taye, 2023).

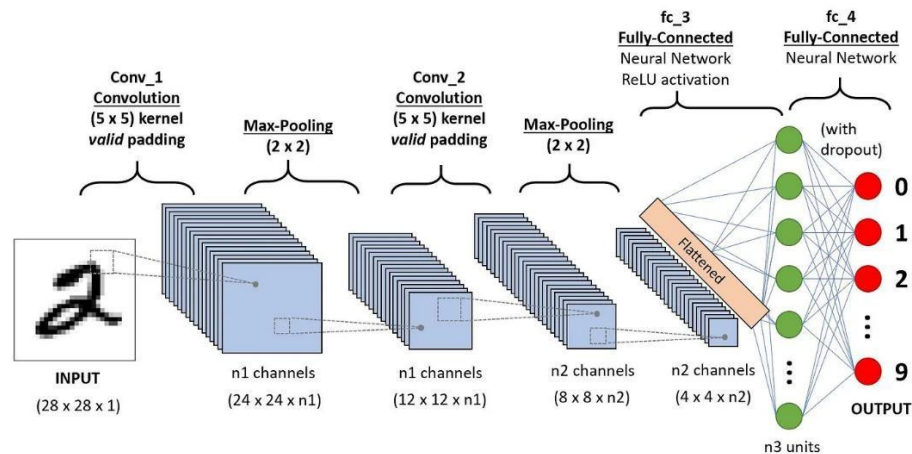


Figura 2.7: Arquitectura de las CNN aplicada. Recuperada de Bouchard (2020).

- Capas de Agrupamiento (Pooling)

Las capas de agrupamiento reducen la dimensionalidad de los mapas de características, disminuyendo la cantidad de parámetros y la carga computacional. Las técnicas más comunes son el max-pooling (Figura 2.8) y el average-pooling (Figura 2.9), que seleccionan el valor máximo o promedio dentro de una ventana de la imagen, respectivamente. Este proceso conserva las características más relevantes y proporciona invariancia a pequeñas traslaciones o distorsiones en la imagen (Zafar et al., 2022).

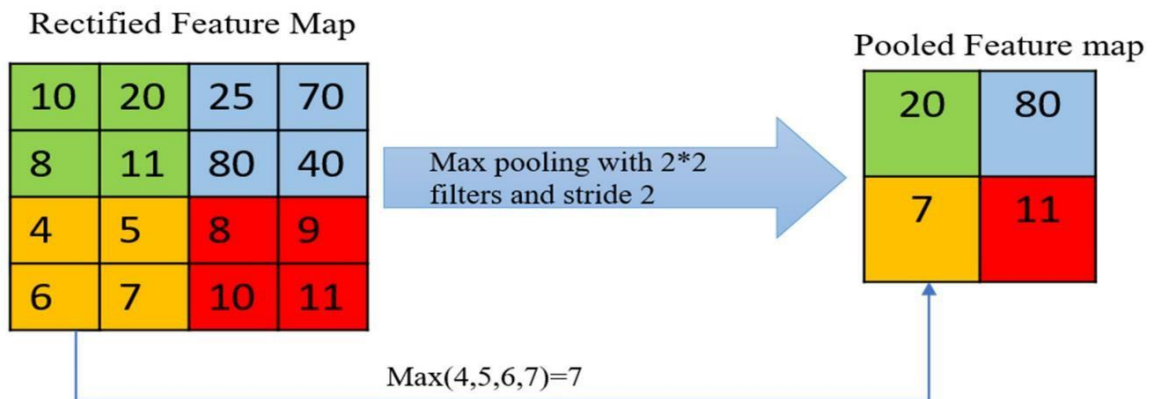


Figura 2.8: Ilustración de la operación de agrupamiento máximo. Recuperado de Zafar et al. (2022).

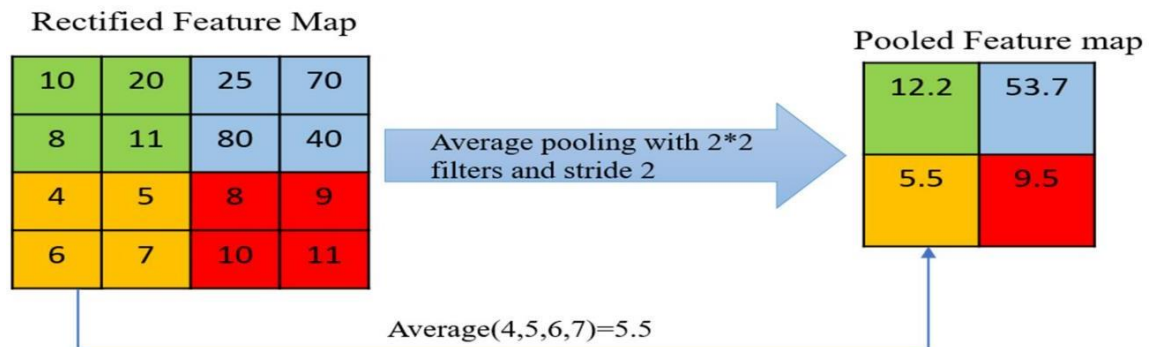


Figura 2.9: Método de agrupamiento Promedio. Recuperado de Zafar et al. (2022).

Dado que existen múltiples métodos de agrupación, Zafar et al., (2022) ha elaborado una tabla comparativa (Anexo 1) que examina las ventajas y desventajas de las operaciones de pooling en diversas arquitecturas basadas en CNN.

- Capas Totalmente Conectadas

Las capas totalmente conectadas, ubicadas generalmente al final de la red, conectan cada neurona con todas las neuronas de la capa anterior (Figura 2.10). Estas capas integran las características extraídas por las capas anteriores y realizan la clasificación o regresión final, dependiendo de la tarea específica (Taye, 2023).

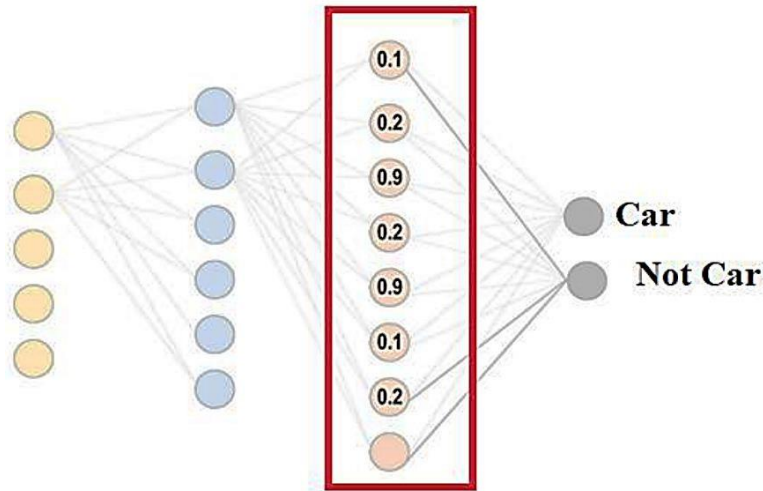


Figura 2.10: Capa completamente conectada. Recuperado de Taye (2023).

Estructura de una CNN para clasificación en n categorías: Los datos originales pasan por dos capas de convolución (Convolución 1 y Convolución 2), seguidas de dos capas de agrupamiento máximo (Max Pooling 1 y Max Pooling 2). El resultado se envía a una capa completamente conectada, donde la función de activación Softmax comprime los vectores de salida a valores entre 0 y 1 antes de dirigirlos a la capa de salida (Figura 2.11). Cada valor resultante representa la probabilidad de pertenecer a una de las n categorías; un valor más alto indica una mayor probabilidad de pertenencia a esa categoría (Zhao et al., 2024).

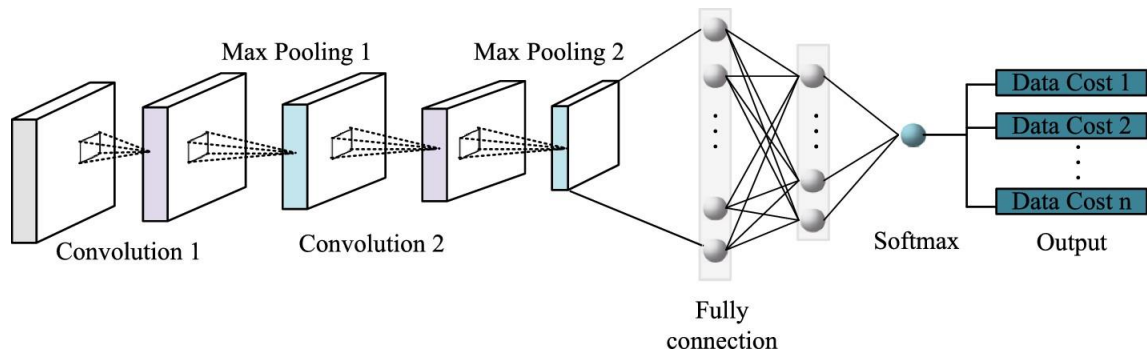


Figura 2.11: Estructura de CNN. Recuperado de Zhao et al. (2024).

2.2.2 Ventajas y desafíos de las CNN

Las CNN ofrecen ventajas significativas, como la capacidad de aprender características directamente de los datos sin necesidad de ingeniería manual, y su eficacia en el procesamiento de grandes volúmenes de datos. Sin embargo, presentan desafíos, como la necesidad de grandes conjuntos de datos etiquetados para un entrenamiento efectivo y la complejidad en la interpretación de los modelos resultantes (Zhao et al., 2024).

2.2.3 Detección de objetos

La detección de objetos es una disciplina fundamental en la visión por computadora que se centra en identificar y localizar instancias de objetos pertenecientes a diversas categorías dentro de imágenes o secuencias de video. Esta tarea es esencial en aplicaciones como la conducción autónoma, la vigilancia, la robótica y, particularmente, en la agricultura de precisión, donde la identificación precisa de cultivos, plagas y enfermedades es crucial para optimizar la producción agrícola (Gené-Mola et al., 2019).

- Evolución de la Detección de Objetos

Históricamente, los métodos de detección de objetos se basaban en técnicas de procesamiento de imágenes y aprendizaje automático tradicionales, que requerían la extracción manual de características y el uso de clasificadores como máquinas de vectores de soporte (SVM). Sin embargo, con el advenimiento de las redes neuronales convolucionales (CNN), la detección de objetos ha

experimentado avances significativos, permitiendo la extracción automática de características y mejorando la precisión y eficiencia de los modelos (Khan et al., 2020).

- Modelos Basados en CNN para la Detección de Objetos

Los modelos de detección de objetos basados en CNN se dividen principalmente en dos categorías: de dos etapas y de una sola etapa (Figura 2.12).

Modelos de Dos Etapas: Estos modelos, como la familia de R-CNN (Region-based Convolutional Neural Networks), primero generan propuestas de regiones que podrían contener objetos y luego clasifican estas regiones. Aunque ofrecen alta precisión, suelen ser más lentos debido a su complejidad computacional (Vina, 2024).

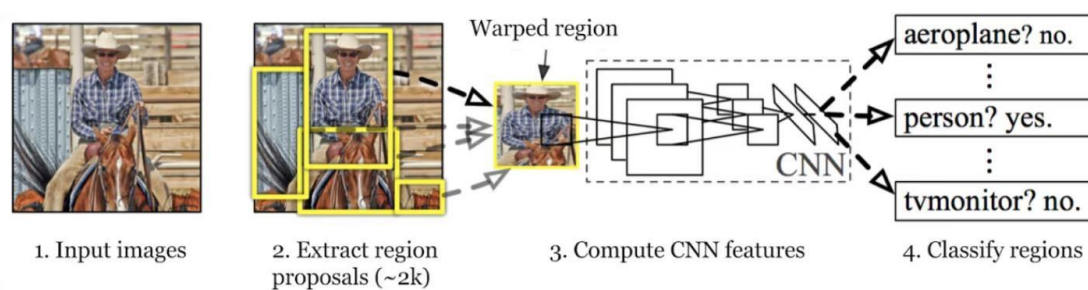


Figura 2.12: *Funcionamiento R-CNN. Recuperado de Vina (2024).*

Aunque R-CNN es una técnica destacada, tiene ciertas desventajas como una alta complejidad computacional y tiempos de inferencia lentos, lo que lo hace inadecuado para aplicaciones en tiempo real. Además, separar las propuestas de regiones y la clasificación en pasos distintos puede resultar en un rendimiento menos eficiente. Para abordar estos desafíos, a lo largo de los años han surgido varios modelos de detección de objetos (Figura 2.13). Fast R-CNN combina en un solo paso las propuestas de regiones y la extracción de características de la CNN, acelerando así el proceso. Faster R-CNN introduce una Red de Propuestas de Región (RPN) para agilizar la generación de propuestas, mientras que Mask

R-CNN añade segmentación a nivel de píxel para obtener detecciones más detalladas (Vina, 2024).

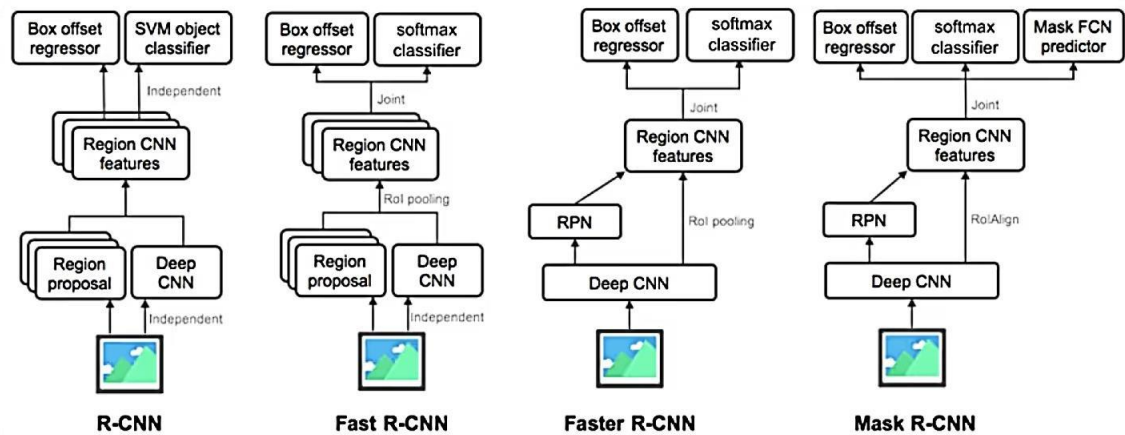


Figura 2.13: Comparativa R-CNN, R-CNN rápida, R-CNN más rápida y R-CNN enmascarada. Recuperado de Vina (2024).

De manera paralela al desarrollo de Faster R-CNN, surgió YOLO (You Only Look Once) como una innovación en la detección de objetos en tiempo real. Los modelos YOLO realizan la predicción de cuadros delimitadores y probabilidades de clase en un único paso dentro de la red (Arrendajo, 2024).

Modelos de Una Sola Etapa: Ejemplos destacados incluyen YOLO (You Only Look Once) y SSD (Single Shot MultiBox Detector). Estos modelos realizan la detección y clasificación en una sola pasada a través de la red, lo que los hace más rápidos y adecuados para aplicaciones en tiempo real, aunque con una ligera disminución en precisión en comparación con los modelos de dos etapas (Mounika et al., 2021).

2.2.4 Red YOLO (You Only Look Once)

El algoritmo YOLO (You Only Look Once) se considera uno de los algoritmos de detección de objetos más destacados. Logra una velocidad y precisión de última generación, y sus diversas aplicaciones lo han hecho indispensable en numerosos campos e industrias. Numerosos investigadores han mostrado interés en este algoritmo de detección de objetos publicando artículos que revisan su

evolución, ajustan sus modelos y comparan su rendimiento con otros algoritmos de visión por computadora. Este interés generalizado subraya el importante papel de YOLO en el avance del campo de la visión por computadora.

La serie de modelos YOLO ha transformado la detección de objetos en tiempo real, estableciéndose como un estándar por su equilibrio entre velocidad y precisión en diversas aplicaciones. A continuación, se presenta un análisis técnico detallado de las versiones más relevantes de YOLO, destacando sus innovaciones arquitectónicas, métricas de rendimiento y aplicaciones prácticas.

- YOLOv1

Introducido por Redmon et al. (2015), YOLOv1 propuso un enfoque unificado para la detección de objetos, procesando imágenes completas en una sola pasada a través de una red neuronal convolucional (CNN). Dividía la imagen en una cuadrícula de celdas, donde cada celda predecía cuadros delimitadores y probabilidades de clase. Este diseño permitió una detección rápida, aunque presentaba limitaciones en la precisión, especialmente para objetos pequeños o agrupados.

- YOLOv2

En 2016, YOLOv2 Redmon et al. (2016) introdujo mejoras significativas, como la normalización por lotes y el uso de cuadros delimitadores ancla, lo que aumentó la precisión y permitió la detección de más de 9000 clases de objetos. Además, incorporó técnicas como el ajuste de anclas y la agrupación de dimensiones, optimizando la detección de objetos de diferentes tamaños.

- YOLOv3

Lanzado en 2018, YOLOv3 Redmon et al. (2018) implementó una arquitectura más profunda y la detección en múltiples escalas, mejorando la precisión en la detección de objetos de diversos tamaños. Utilizó una red troncal Darknet-53 y

centraron en optimizar el rendimiento para dispositivos móviles y aplicaciones de baja latencia, respectivamente.

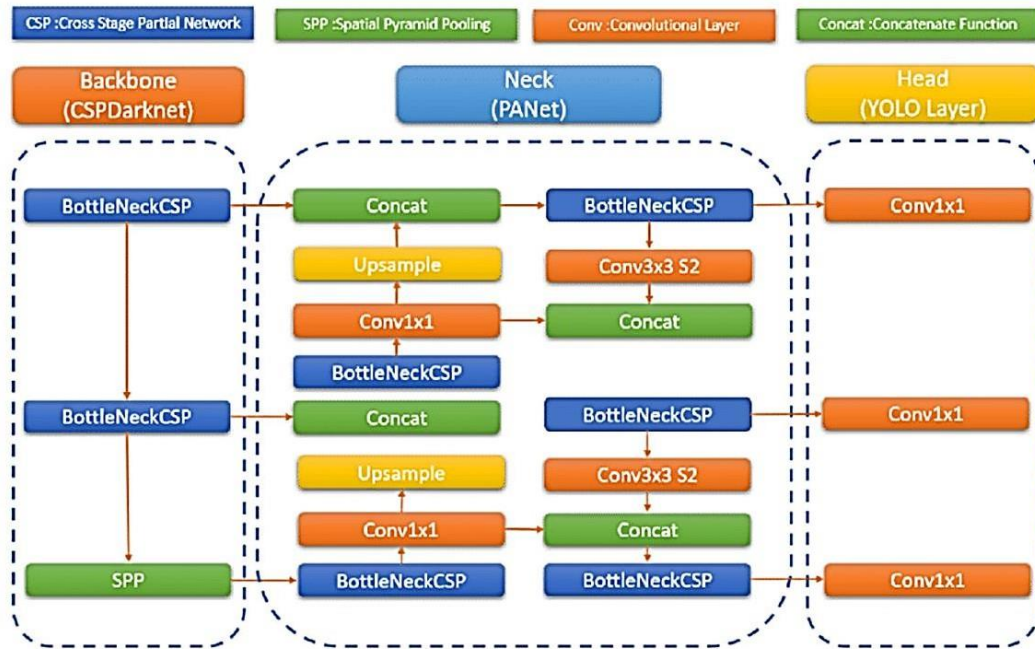


Figura 2.15: Arquitectura detallada de YOLOv5, incluida la columna vertebral de CSPDarknet, el cuello de PANet y el cabezal de capa YOLO. Recuperado de Egi et al. (2022).

- YOLOv8 a YOLOv11

YOLOv8 y sus sucesores incorporaron módulos de segmentación semántica y pose estimation, expandiendo el alcance de YOLO más allá de la detección de objetos tradicionales. YOLOv9 Wang et al., (2024) destacó por la Información de Gradiente Programable (PGI), optimizando el flujo de gradientes durante el entrenamiento, mientras que YOLOv10 Wang et al., (2024) eliminó la necesidad de Non-Maximum Suppression (NMS), agilizando el procesamiento. Finalmente, YOLOv11 Glenn et al., (2024) implementó bloques C2PSA para mejorar la atención espacial, aumentando la precisión en objetos pequeños y superpuestos.

YOLO V8 muestra las múltiples capas convolucionales de la red troncal para extraer características jerárquicas (Figura 2.16), la red de pirámide de

características (FPN) para mejorar la detección a diferentes escalas y el cabezal de la red para realizar predicciones finales, incorporando bloques convolucionales y bloques de muestreo ascendente para refinar las características (Jegham et al., 2024).

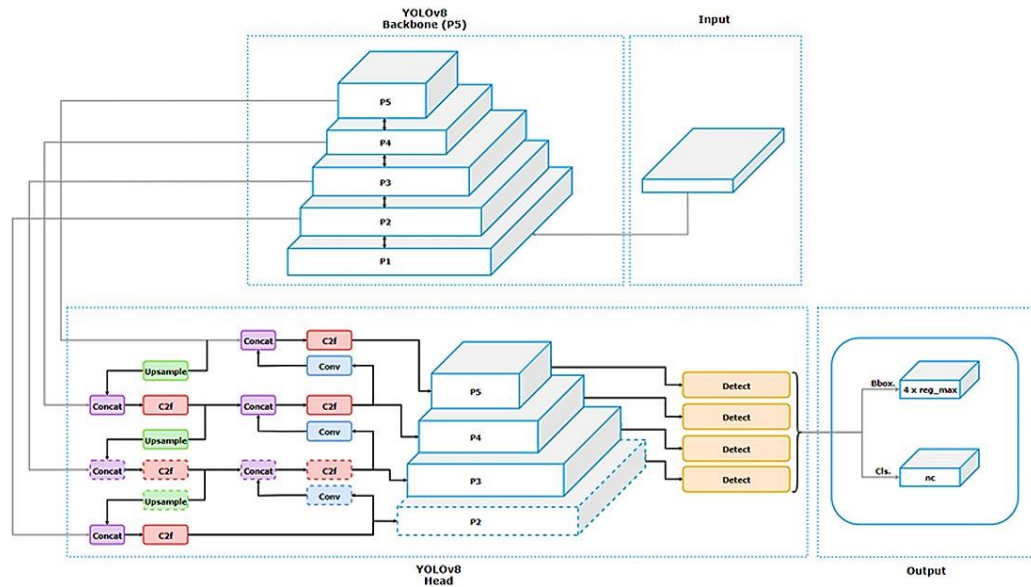


Figura 2.16: Arquitectura detallada de YOLOv8. Recuperado de Jegham et al. (2024).

YOLOv9 con módulos CSPNet, ELAN y GELAN (Figura 2.17). CSPNet optimiza el flujo de gradiente y reduce la complejidad computacional mediante la partición de mapas de características. ELAN mejora la eficiencia del aprendizaje mediante la agregación lineal de características, y GELAN amplía este concepto mediante la integración de características de varias profundidades y vías, lo que ofrece una mayor flexibilidad y precisión en la extracción de características (Chien-Yao, 2024).

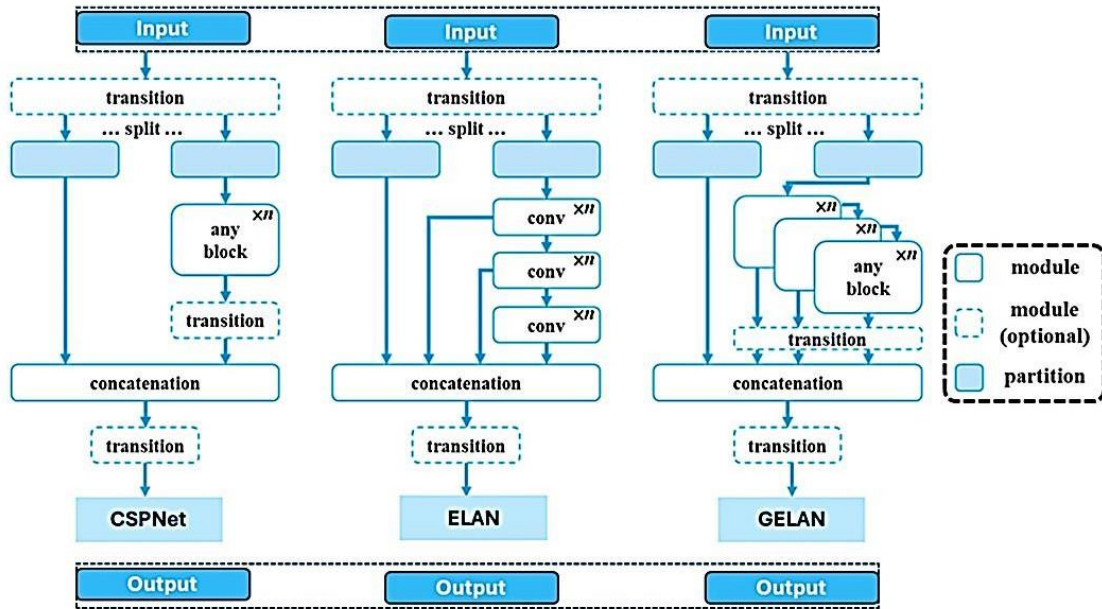


Figura 2.17: Arquitectura YOLOv9. Recuperado de Jegham et al. (2024).

YOLOv10 muestra la estrategia de asignación de etiquetas duales para mejorar la precisión y la capa PAN para mejorar la representación de características, junto con el cabezal de uno a muchos para tareas de regresión y clasificación, y el cabezal uno a uno para una localización precisa.

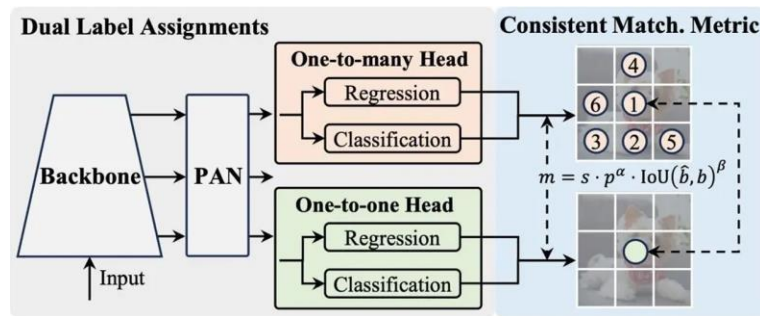


Figura 2.18: Arquitectura YOLOv10. Recuperado de Jegham et al. (2024).

YOLO v11 (Figura 2.19) (Jegham et al. ,2024) es la última innovación de la serie YOLO desarrollada por Ultralytics, que se basa en los desarrollos de sus predecesores, especialmente YOLOv8. Esta iteración ofrece cinco modelos a escala, desde nano hasta extragrandes, que se adaptan a diversas aplicaciones. Al igual que YOLOv8, YOLO11 incluye numerosas aplicaciones, como la

detección de objetos, la segmentación de posturas, la clasificación de imágenes, la estimación de poses y la detección de objetos orientados (OBB).

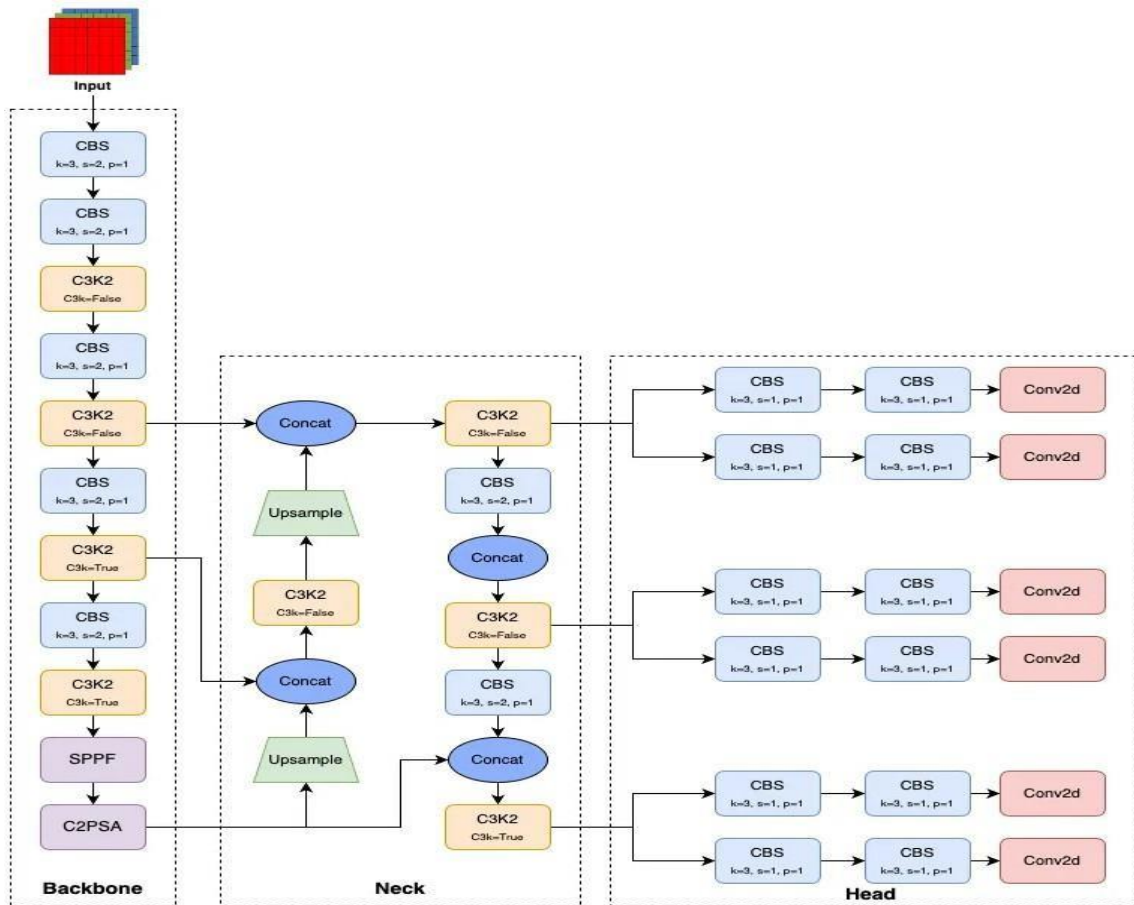


Figura 2.19: Arquitectura YOLO11 que muestra los nuevos bloques C3k2 y el módulo C2PSA. Recuperado de Jegham et al. (2024).

La evolución de los algoritmos YOLO, resumida en la Figura 2.20, ha marcado hitos significativos en la detección de objetos en tiempo real, comenzando con YOLOv1 en 2015, que introdujo la detección unificada de objetos. Desde entonces, cada versión ha incorporado mejoras clave, como mayor precisión, velocidad y capacidades avanzadas. Por ejemplo, YOLOv4 escaló el rendimiento mediante la red CSPDarknet53 y YOLOv5 optimizó aún más el uso práctico en múltiples tareas. Las versiones más recientes, como YOLOv9 y YOLOv11, han añadido innovaciones como segmentación semántica y el uso de atención espacial. Este progreso ha hecho de YOLO una herramienta esencial en

aplicaciones como conducción asistida, agricultura y monitoreo de seguridad, posicionándose como un estándar en detección de objetos eficiente y accesible (Jegham et al., 2024).

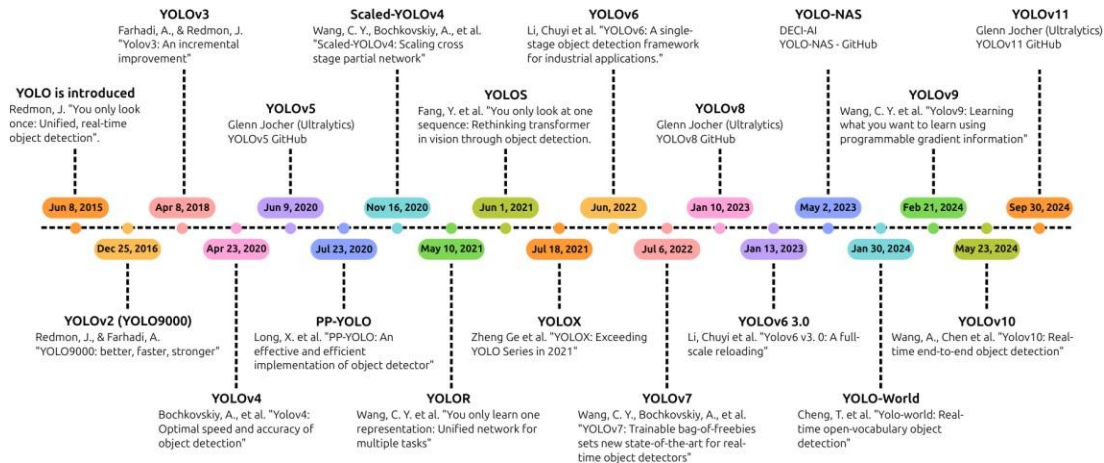


Figura 2.20: Evolución de los algoritmos YOLO a lo largo de los años. Recuperado de Jegham et al. (2024).

2.2.5 La calibración y evaluación de CNN

La calibración y evaluación de modelos de aprendizaje profundo, especialmente en CNN son procesos fundamentales para garantizar su rendimiento y generalización. Estos procesos implican el ajuste de hiperparámetros y la implementación de técnicas de evaluación robustas.

• Calibración de Modelos

La calibración se refiere a la capacidad de un modelo para proporcionar estimaciones de probabilidad que reflejen con precisión la incertidumbre en sus predicciones. Un modelo bien calibrado asigna probabilidades que corresponden a la frecuencia real de ocurrencia de los eventos. Se han desarrollado diversas técnicas de calibración, entre las que destacan:

- **Escalado de Temperatura (Temperature Scaling):** Es una técnica posentrenamiento que ajusta las salidas del modelo mediante un parámetro de temperatura, sin alterar la precisión de las predicciones. Esta

técnica ha demostrado ser efectiva para mejorar la calibración en modelos de clasificación de imágenes (Bohdal et al., 2023).

- Calibración Platt: Originalmente diseñada para máquinas de vectores de soporte, esta técnica se ha adaptado para redes neuronales, ajustando las salidas del modelo a través de una regresión logística en un conjunto de validación.
- Calibración Isotónica: Utiliza una función no paramétrica para mapear las salidas del modelo a probabilidades calibradas, siendo útil cuando las relaciones entre las salidas y las probabilidades verdaderas son complejas.

- Evaluación de Modelos

La evaluación rigurosa de modelos es esencial para determinar su desempeño y capacidad de generalización. Las métricas comunes incluyen:

Precisión (Accuracy): Proporción de predicciones correctas sobre el total de predicciones realizadas.

Precisión y Exhaustividad (Precision and Recall): La precisión mide la proporción de verdaderos positivos entre las predicciones positivas, mientras que la exhaustividad indica la proporción de verdaderos positivos entre todos los casos reales positivos.

Puntuación F1 (F1-Score): Media armónica de la precisión y la exhaustividad, proporcionando un equilibrio entre ambas métricas.

Curva ROC y AUC (Receiver Operating Characteristic and Area Under the Curve): Evalúan la capacidad del modelo para distinguir entre clases, siendo el AUC una medida del rendimiento general del modelo.

Además de estas métricas, es crucial evaluar la calibración del modelo. La métrica más utilizada para este propósito es el Error de Calibración Esperado (ECE), que cuantifica la discrepancia entre las probabilidades predichas y las

frecuencias reales observadas. Un ECE bajo indica una buena calibración del modelo (Bohdal et al., 2023).

- Ajuste de Hiperparámetros

Los hiperparámetros son variables que configuran la estructura y el proceso de aprendizaje de un modelo, pero que no se aprenden directamente de los datos. Incluyen parámetros como la tasa de aprendizaje, el tamaño del lote (batch size), el número de capas y neuronas en una red, entre otros. El ajuste adecuado de estos hiperparámetros es crucial para el rendimiento óptimo del modelo.

Las técnicas comunes para la optimización de hiperparámetros incluyen:

- Búsqueda en Cuadrícula (Grid Search): Explora exhaustivamente todas las combinaciones posibles de hiperparámetros dentro de un conjunto definido. Aunque es completa, puede ser computacionalmente costosa.
- Búsqueda Aleatoria (Random Search): Selecciona combinaciones de hiperparámetros de manera aleatoria dentro de un espacio definido, siendo más eficiente en términos de tiempo que la búsqueda en cuadrícula.
- Optimización Bayesiana: Utiliza modelos probabilísticos para predecir el rendimiento de diferentes combinaciones de hiperparámetros, enfocándose en áreas prometedoras del espacio de búsqueda. Esta técnica ha demostrado ser efectiva en la optimización de modelos complejos (Roch, 2024).
- Algoritmos Evolutivos: Simulan procesos de evolución natural para iterativamente seleccionar y recombinar conjuntos de hiperparámetros, buscando mejorar el rendimiento del modelo.

El ajuste de hiperparámetros es un proceso iterativo que requiere una evaluación constante del modelo en un conjunto de validación para evitar el sobreajuste y garantizar una buena generalización.

Aplicaciones del aprendizaje profundo en la agricultura

Se pretende integrar a esta investigación el marco de estimación de densidad mediante aprendizaje profundo. Una vez que se completa la inferencia de densidad para una nueva imagen, extraer ubicaciones discretas de cada inflorescencia utilizando un algoritmo de búsqueda de picos. Estos puntos se envían a anotadores humanos para que los corrijan y el modelo se vuelve a entrenar con los nuevos datos; esto permite que el rendimiento del modelo mejore a medida que ve más y más datos al tiempo que reduce la carga de los anotadores humanos (Ghosal et al., 2019).

Dentro del dominio más amplio del conteo de entidades, los enfoques se clasifican en una de tres categorías: conteo por detección, conteo por regresión y conteo por estimación de densidad (Sindagi & Patel, 2018).

Con el éxito del aprendizaje profundo, muchos de estos enfoques tradicionales han sido reemplazados por algoritmos de detección y segmentación basados en redes neuronales (Redmon & Farhadi, 2018), pero estos nuevos métodos aún buscan resolver el problema del conteo a través de la localización precisa de todas las entidades deseadas en la imagen. Los inconvenientes clave de estos métodos son que tienden a ser computacionalmente pesados, la complejidad del tiempo a menudo aumenta con la cantidad de entidades presentes, a menudo tienen un límite superior de entidades detectables antes de encontrar problemas de memoria y tienden a tener problemas a medida que la oclusión se vuelve más pronunciada. o las entidades se hacen pequeñas. Además, los métodos de detección, como Faster R-CNN y YOLO (Redmon et al., 2018) anotaciones de cuadro delimitador y Mask R-CNN (He et al., 2017) requiere además una máscara de instancia densa.

Los enfoques de estimación de densidad han demostrado ser muy exitosos (Xu & Qiu, 2016), especialmente cuando se combinan con arquitecturas profundas (Boominathan & Babu, 2016; Sam et al., 2019) cuando deseamos la localización además de un recuento final. Muchos de estos aprovechan las redes neuronales totalmente convolucionales (CNN) para predecir una densidad (Ma et al., 2019) a

través de la imagen; esta densidad se puede integrar para proporcionar el recuento de una región. Además, estos métodos tienden a superar a los métodos basados en la detección en escenarios muy ocluidos. También requieren solo anotaciones de puntos simples que se pueden adquirir mucho más rápidamente que las anotaciones de cuadro delimitador o máscara de instancia que necesitan los métodos de detección. Además, debido a que el mapa de densidad de salida es en sí mismo una imagen de un solo canal, no una colección de cuadros delimitadores, la complejidad computacional es independiente del número de entidades presentes. Nuestro enfoque sigue estos métodos, ya que la inflorescencia puede estar ocluida por otras partes de la planta, y el número de inflorescencias en una imagen dada puede ser extremadamente grande.

Tanto la visión por computadora tradicional como los enfoques basados en el aprendizaje profundo se han utilizado para una variedad de aplicaciones agrícolas basadas en el conteo. El trabajo de Guo et al., 2018 y Malambo et al., 2019 utilizaron técnicas basadas en detección para cabezas de sorgo en un campo. Del mismo modo, (Gené et al., 2020) utilizó Mask-RCNN para identificar y segmentar por completo las manzanas en los árboles de un huerto. Para contar palmeras a partir de imágenes de dron, (Li et al., 2017) utilizó un enfoque de detección basado en CNN, (Osco et al., 2020) utilizaron un enfoque muy similar para contar el número de árboles de cítricos en una arboleda. Donde buscaron contar cada árbol presente, en este trabajo se tratará de contar solo aquellas plantas que han comenzado a florecer.

La detección y manejo de la floración en la piña también enfrentan desafíos relacionados con factores ambientales. Variables como la temperatura, la humedad y la luz influyen en el proceso de floración. Por ejemplo, temperaturas bajas pueden inducir una floración prematura no deseada, mientras que condiciones de sequía pueden retrasarla. Por ello, es esencial que los productores implementen sistemas de monitoreo climático y adopten prácticas de manejo que mitiguen los efectos de las condiciones ambientales adversas (Nkhoma, 2024).

La identificación precisa y oportuna de la floración es esencial para optimizar la programación de la cosecha y maximizar la calidad del fruto. Tradicionalmente, este proceso se realiza de manera manual, lo que implica un consumo significativo de tiempo y recursos humanos, además de estar sujeto a errores humanos (INIFAD, 2020). La implementación de tecnologías avanzadas, como el uso de drones equipados con sensores multiespectrales y la aplicación de redes neuronales convolucionales para la detección automática de la floración, representa una solución innovadora para este desafío. Estas herramientas permiten un monitoreo más preciso y eficiente, reduciendo costos y mejorando la toma de decisiones en el manejo del cultivo.

La utilización de los drones en el cultivo de piña ha sido limitado, sin embargo Mapsens (2023) realizó un estudio en Veracruz implementando drones para generar mapas de salud de las plantas de piña, lo que permitió a los agricultores tomar decisiones informadas sobre riego y fertilización. Además, la cartografía aérea mediante drones ha proporcionado mapas detallados que destacan variaciones en la topografía y la humedad del suelo, optimizando estrategias de riego y reduciendo costos operativos (Calderón, 2024).

En el ámbito de la agricultura de precisión aplicada al cultivo de la piña, diversos estudios han explorado el uso de redes neuronales convolucionales (CNN) para mejorar la detección y el conteo de flores de piña, lo cual es crucial para optimizar la producción y la calidad del fruto.

Un estudio titulado "Large-Scale Counting and Localization of Pineapple Inflorescence Through Deep Density-Estimation" implementó una técnica de estimación de densidad profunda para contar y localizar inflorescencias de piña en imágenes aéreas. Este enfoque permitió una detección más precisa y eficiente en comparación con métodos tradicionales, facilitando la planificación de la cosecha y la gestión del cultivo. El modelo implementado U-net alcanzó resultados destacados. Los parámetros de evaluación obtenidos incluyen un error absoluto (MAE) de 11.5 y una desviación porcentual media absoluta (MAPD) de 6.37% en el conjunto de prueba (Hobbs et al., 2021).

Otro trabajo, "A Deep Object Detection Method for Pineapple Fruit and Flower Recognition in Cluttered Background", desarrolló un método de detección de objetos basado en CNN para reconocer flores y frutos de piña en entornos con fondos complejos. El estudio destacó la eficacia del modelo en condiciones de iluminación variables y con oclusiones, mejorando la precisión en la identificación de las etapas de desarrollo del cultivo, ya que alcanzó una precisión de 65.6% y una capacidad de recuperación de 74.1%, lo que indica una adecuada capacidad para identificar correctamente los elementos presentes. Además, obtuvo un mAP de 64% y un F1 score de 0.69, reflejando un equilibrio favorable entre precisión y recall, incluso bajo condiciones desafiantes como iluminación variable, tamaños de objeto diferente y oclusión significativas. (Wang et al., 2020).

Además, el estudio "A-pruning: a lightweight pineapple flower counting network based on filter pruning" presentó una red ligera para el conteo de flores de piña, utilizando técnicas de poda de filtros para optimizar la red YOLOv5. Este modelo logró una precisión del 71.7% con 1.7 millones de parámetros y una velocidad de procesamiento de 178 cuadros por segundo, demostrando su eficiencia en aplicaciones en tiempo real (Yu et al., 2024).

La combinación de técnicas de poda de filtros y extracción de características multiescala podría ofrecer soluciones más eficientes y precisas para el monitoreo del cultivo de piña. Estos avances en el uso de CNN y técnicas de aprendizaje profundo en la agricultura de precisión resaltan el potencial de estas tecnologías para mejorar la gestión y productividad del cultivo de la piña.

La combinación de drones y redes neuronales ha demostrado ser una herramienta poderosa en la agricultura de precisión, permitiendo a los agricultores monitorear sus cultivos de manera más efectiva y tomar decisiones basadas en datos precisos. Estas tecnologías continúan evolucionando, ofreciendo nuevas oportunidades para mejorar la eficiencia y sostenibilidad en la producción agrícola.

En este trabajo de investigación, se decidió emplear las arquitecturas YOLOv8, YOLOv10 y YOLOv11 en sus versiones nano para abordar la detección de flores

en cultivos de piña, dada su capacidad demostrada para combinar precisión y velocidad en tareas de detección de objetos en tiempo real. Estas versiones representan los avances más recientes en la serie de modelos YOLO (You Only Look Once), adaptándose a las exigencias específicas de este estudio, que requiere alta sensibilidad para la identificación precisa de flores en condiciones agrícolas variadas.

2.3 Literatura citada

- Aitutoriales. (2024). *Aprendizaje automático no supervisado: algoritmos, tipos con ejemplo*. Aitutoriales. <https://aitutoriales.com/aprendizaje-automatico-no-supervisado-algoritmos-tipos-con-ejemplo/>
- Arrendajo. (2024). *Introducción a YOLO: detección de objetos en tiempo real*. HashDork. <https://hashdork.com/es/yolo/>
- Bartholomew, D. P., Rohrbach, K. G., & Evans, D. O. (2002). Pineapple Cultivation in Hawaii. *Fruits and Nuts*, 1–8.
- Bohdal, O., Yang, Y., & Hospedales, T. (2023). Meta-calibration: learning of model calibration using differentiable expected calibration error. *Transactions on Machine Learning Reseach*.
- Bongiovanni, R., & Lowenberg, J. D. (2004). Precision agriculture and sustainability. *Precision Agriculture*, 5, 359–387. <https://doi.org/10.1023/B:PRAG.0000040806.39604.aa>
- Boominathan, L., & Babu, R. V. (2016). CrowdNet: a deep convolutional network for dense crowd counting. *ACM Multimedia*. <https://doi.org/10.1145/2964284.2967300>
- Bouchard, L. (2020). *Introduction to convolutional neural networks (CNNs) / the most popular deep learning architecture*. Medium. <https://medium.com/what-is-artificial-intelligence/introduction-to-convolutional-neural-networks-cnns-the-most-popular-deep-learning-architecture-b938f62f133f>
- Buriticá, A. (2024). *Uso de drones y sensores: tecnologías para monitorear cultivos y detectar problemas de plagas*. Agricultura.Croper. <https://blog.croper.com/uso-de-drones-y-sensores-tecnologias-para-monitorear-cultivos-y-detectar-problemas-de-plagas/>

- Calderón, J. G. (2024). *Optimizando la agricultura de piña con drones: un vuelo hacia la eficiencia*. Agrotec. <https://agroteccr.com/optimizando-la-agricultura-de-pina-con-drones-un-vuelo-hacia-la-eficiencia/>
- CINVESTAV. (2020). *Aplican aprendizaje profundo a la agricultura para reducir costos y aumentar su calidad*. Centro de Investigación y de Estudios Avanzados Del Instituto Politécnico Nacional. <https://conexion.cinvestav.mx/Publicaciones/aplican-aprendizaje-profundo-a-la-agricultura-para-reducir-costos-y-aumentar-su-la-calidad>
- Cruz, J. C. (2019). *Uso del aprendizaje profundo para el estudio de fenotipo en plantas*. Medium. <https://medium.com/@jckruz777/usos-del-aprendizaje-profundo-para-el-estudio-de-fenotipo-en-plantas-2fa3f02f46e5>
- CulturaAI. (2024). *El aprendizaje profundo y sus aplicaciones*. CulturaAI. <https://culturaai.com/el-aprendizaje-profundo-y-sus-aplicaciones/>
- DJI. (2024). *Mavic 3 multispectral*. DJI. <https://ag.dji.com/es/mavic-3-m>
- Egi, Y., Hajyzadeh, M., & Eyceyurt, E. (2022). Drone-computer communication based tomato generative organ counting model using YOLO V5 and deep-sort. *Agriculture*, 12(9). <https://doi.org/10.3390/agriculture12091290>
- FAO. (2023). *Pineapple production statistics*. Food and Agriculture Organization. <https://vida-sustentable.com/como-se-cultiva-la-pina/>
- Fernandez, O. (2023). *Python DL: Deep Learning en python*. Codigospython. <https://codigospython.com/python-dl-deep-learning-en-python/>
- Gené, J. M., Sanz, R. C., Rosell, J. R. P., Ramon, J. M., Ruiz, J. H., Vilaplana, V., & Gregorio, E. (2020). Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. *Computers and Electronics in Agriculture*, 169. <https://doi.org/10.1016/j.compag.2019.105165>
- Gené-Mola, J., Vilaplana, V., Rosell-Polo, J. R., Morros, J.-R., Ruiz-Hidalgo, J., & Gregorio, E. (2019). *Uso de redes neuronales convolucionales para la*

detección remota de frutos con cámaras RGB-D. 1081–1087.
https://doi.org/10.26754/c_agroing.2019.com.3325

Geoinnova. (2019). *Composiciones RGB de imágenes satélite.* 19 De Octubre.
<https://geoinnova.org/blog-territorio/composiciones-rgb-de-imagenes-satelite/>

Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., Singh, A. K., Singh, A., Hirafuji, M., Ninomiya, S., Ganapathysubramanian, B., Sarkar, S., & Guo, W. (2019). A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics*.
<https://doi.org/10.34133/2019/1525874>

Gonzalez-Dugo, V., Zarco-Tejada, P., Nicolás, E., Nortes, P. A., Alarcón, J. J., Intrigliolo, D. S., & Fereres, E. (2013). Using high resolution UAV thermal imagery to assess the variability in the water status of five fruit tree species within a commercial orchard. *Precision Agriculture*, 14(6), 660–678.
<https://doi.org/10.1007/s11119-013-9322-9>

Guo, J., Li, X., Li, Z., Hu, L., Yang, G., Zhao, C., Fairbairn, D., Watson, D., & Ge, M. (2018). Multi-GNSS precise point positioning for precision agriculture. *Precision Agriculture*, 19(5), 895–911. <https://doi.org/10.1007/s11119-018-9563-8>

Guo, W., Zheng, B., Potgieter, A. B., Diot, J., Watanabe, K., Noshita, K., Jordan, D. R., Wang, X., Watson, J., Ninomiya, S., & Chapman, S. C. (2018). Aerial imagery analysis – quantifying appearance and number of sorghum heads for applications in breeding and agronomy. *Frontiers in Plant Science*, 9(1544), 1–9. <https://doi.org/10.3389/fpls.2018.01544>

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *IEEE*, 2961–2969. <https://doi.org/10.1109/ICCV.2017.322>

Hidroponía, C. (2017). *Producción de piña en México.* Hydroenvironment.
<https://hidroponia.mx/produccion-de-pina-en-mexico/>

- Hobbs, J., Prakash, P., Paull, R., Hovhannisyan, H., Markowicz, B., & Rose, G. (2021). Large-scale counting and localization of pineapple inflorescence through deep density-estimation. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2020.599705>
- Hu, C., Wang, F., Ai, C., Wang, X., & Wang, Y. (2021). An improved new convolutional neural network method for inverting the pore pressure in oil reservoir by surface vertical deformation. *Lithosphere*. <https://doi.org/10.2113/2021/5597238>
- Huang, Y., Hoffmann, W. C., Lan, Y., Wu, W., & Fritz, B. K. (2009). Development of a spray system for an unmanned aerial vehicle platform. *Applied Engineering in Agriculture*, 25(6), 803–809. <https://doi.org/10.13031/2013.29229>
- Hunt, E. R., Dean Hively, W., Fujikawa, S. J., Linden, D. S., Daughtry, C. S. T., & McCarty, G. W. (2010). Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring. *Remote Sensing*, 2, 290–305. <https://doi.org/10.3390/rs2010290>
- IA, T. sobre. (2024). *Aprendizaje supervisado vs. no supervisado: diferencias y aplicaciones*. Todo Sobre IA. <https://iadatamind.com/aprendizaje-supervisado-vs-no-supervisado-diferencias-y-aplicaciones/>
- Infoagro. (2022). *Producción de piña en México y el mundo*. Reevista Infoagro México. <https://mexico.infoagro.com/produccion-de-pina-en-mexico-y-el-mundo/>
- INIFAD. (2020). *La tecnología potencializa al “rey de los frutos”: la piña*. INIFAD. <https://www.gob.mx/inifap/articulos/la-tecnologia-potencializa-al-rey-de-los-frutos-la-pina>
- ISPA. (2024). *Precision Ag definition*. International Society of Precision Agriculture. <https://www.ispag.org/about/definition>

- Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. (2024). Evaluating the evolution of YOLO (You Only Look Once) models: a comprehensive benchmark study of YOLO11 and its predecessors. *Computer Science*. <https://doi.org/10.48550/arXiv.2411.00201>
- Jensen, J. R. (2007). *Remote Sensing of Vegetation*. 1–33.
- Jocher, G., Chaurasia, A., & Qiu, J. (2023). *Ultralytics YOLOv8*. GitHub Repository. <https://github.com/ultralytics/ultralytics>
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(27). <https://doi.org/10.1186/s40537-019-0192-5>
- Jung, F. R., Traut, K., Gauggel, C., Brooke, A. S., & Umali, R. (2016). Pineapple plantation information system-application of geo-informatics and unmanned aerial sensing technology for efficient land use management. *Acta Horticulturae*, 1111, 255–266. <https://doi.org/10.17660/ActaHortic.2016.1111.37>
- Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. In *Artificial intelligence review* (Vol. 53). Springer Netherlands. <https://doi.org/10.1007/s10462-020-09825-6>
- Li, W., Fu, H., Yu, L., & Cracknell, A. (2017). Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sensing*, 9(1). <https://doi.org/10.3390/rs9010022>
- Ma, Z., Wei, X., Hong, X., & Gong, Y. (2019). Bayesian loss for crowd count estimation with point supervision. *IEEE*, 6142–6151. <https://doi.org/10.1109/ICCV.2019.00624>
- Mahajan, A., Chakrabarty, N., Majithia, J., Ahuja, A., Agarwal, U., Suryavanshi, S., Biradar, M., Sharma, P., Raghavan, B., Arafath, R., & Shukla, S. (2023). Multisystem imaging recommendations/guidelines: in the pursuit of precision

- oncology. *Indian Journal of Medical and Paediatric Oncology*, 44(1), 2–25.
<https://doi.org/10.1055/s-0043-1761266>
- Malambo, L., Popescu, S., Ku, N., Rooney, W., Zhou, T., & Moore, S. (2019). A deep learning semantic segmentation-based approach for field-level sorghum panicle counting. *Remote Sensing*, 11(24), 1–19.
<https://doi.org/10.3390/rs11242939>
- Mapsens. (2023). *Inventario de piñas y control de salud con drones en Veracruz (México)*. Mapsens. <https://www.mapsens.com/proyecto-mapsens-agro-uso-de-mapsens-para-cultivo-de-pinas-en-mexico/>
- Montoya, S. (2020). *Conteo geoespacial de cultivos a partir de drones con Phyton, Scikit Learn y Scikit Image-tutorial-gidahatari*. Gidahatari. <https://gidahatari.com/ih-es/conteo-geoespacial-de-cultivos-a-partir-de-ortofotos-de-drones-con-python-scikit-learn-y-scikit-image-tutorial>
- Mounika, P., Narayana, C. L., Kondapalli, D., & Ramana, V. (2021). A comparative study of YOLO and SSD object detection algorithms. *International Journal of Creative Research Thoughts*, 9(11), 821–827.
- Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), 358–371. <https://doi.org/10.1016/j.biosystemseng.2012.08.009>
- Nkhoma, T. (2020). *Floración, polinización y poda de la piña*. Wikifarmer. <https://wikifarmer.com/es/floracion-polinizacion-y-poda-de-la-pina/>
- Nkhoma, T. (2024). *Pineapple cultivation guidelines*. Wikifarmer. <https://wikifarmer.com/es/floracion-polinizacion-y-poda-de-la-pina/>
- Oscó, L. P., de Arruda, M. dos S., Marcato Junior, J., da Silva, N. B., Ramos, A. P. M., Moryia, É. A. S., Imai, N. N., Pereira, D. R., Creste, J. E., Matsubara, E. T., Li, J., & Gonçalves, W. N. (2020). A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*,

- 160(December 2019), 97–106.
<https://doi.org/10.1016/j.isprsjprs.2019.12.010>
- Pino, E. V. (2019). Los drones una herramienta para una agricultura eficiente: un futuro de alta tecnología. *Idesia*, 37, 75–84. <https://doi.org/10.4067/s0718-34292019005000402>
- Redagícola. (2023). *Nuevos sensores ópticos en agricultura de precisión*. Redagícola. <https://redagricola.com/nuevos-sensores-opticos-en-agricultura-de-precision/>
- Redmon, J., & Farhadi, A. (2018). *YOLOv3: an incremental improvement*. Arxiv. <https://arxiv.org/abs/1804.02767>
- Reményi, I., Domián, B., & Kárász, Z. (2023). Improved Object Detection by Utilizing the Image Stream. *Communications in Computer and Information Science*, 575–584. https://doi.org/10.1007/978-3-031-41774-0_45
- Rio-Alvarez, A., Marcos, P. G., González, P. P., Serrano-Pertierra, E., Novelli, A., Fernández-Sánchez, M. T., & González, V. M. (2024). Evaluating deep learning techniques for optimal neurons counting and characterization in complex neuronal cultures. *Medical and Biological Engineering and Computing*. <https://doi.org/10.1007/s11517-024-03202-z>
- Roch, E. M. (2024). *Optimización bayesiana: afinando hiperparámetros como un profesional*. Lovtechnology. <https://lovtechnology.com/optimizacion-bayesiana-afinando-hiperparametros-como-un-profesional/>
- Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1976). Monitoring vegetation systems in the great plains with erts. *Remote Sensing Center*, A20(1), 309–317. <https://doi.org/10.1021/jf60203a024>
- Sanewski, G. M., Bartholomew, D. P., & Paull, R. E. (2018). *The pineapple: botany, production and uses* (2nd editio). CABI. <https://doi.org/10.1079/9781786393302.0000>

- Sindagi, V. A., & Patel, V. M. (2018). A survey of recent advances in CNN-based single image crowd counting and density estimation. *Pattern Recognition Letters*, 107, 3–16. <https://doi.org/10.1016/j.patrec.2017.07.007>
- Taye, M. M. (2023). Theoretical understanding of convolutional neural network: concepts, architectures, applications, future directions. *Computation*, 11(3). <https://doi.org/10.3390/computation11030052>
- Tecnoagro. (2024). *Drones agrícolas: aplicaciones y beneficios en la gestión de cultivos*. Tecnoagro. <https://tecnoagro.com.mx/no.-177/drones-agricolas-aplicaciones-y-beneficios-en-la-gestion-de-cultivos>
- Torres, J. T. (2023). Deep learning for multispectral imaging based predictions in agriculture. *Concienciadigital*, 6, 75–87. <https://doi.org/10.33262/concienciadigital.v6i4.1.2734>
- Uriza, D. E. Á. (2011). Paquete tecnológico piña MD2 (Ananas comosus var. comosus): establecimiento y mantenimiento. *Centro de Investigacion Regional Golfo Centro*.
- Vina, A. (2024). *¿Qué es R-CNN? un resumen rápido*. Ultralytics. <https://www.ultralytics.com/es/blog/what-is-r-cnn-a-quick-overview>
- Wang, C., Zhou, J., Xu, C., & Bai, X. (2020). A deep object detection method for pineapple fruit and flower recognition in cluttered background. *Springer Nature*, 218–227.
- Wolfe, J. (2021). Aprendizaje profundo en aplicaciones de teledetección agrícola. *Telematica*.
- Xie, W., Noble, J. A., & Zisserman, A. (2018). Microscopy cell counting and detection with fully convolutional regression networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*. <https://doi.org/10.1080/21681163.2016.1149104>

- Xu, B., & Qiu, G. (2016). Crowd density estimation based on rich features and random projection forest. *IEEE*, 1–8. <https://doi.org/10.1109/WACV.2016.7477682>
- Yu, G., Cai, R., Luo, Y., Hou, M., & Deng, R. (2024). A-pruning: a lightweight pineapple flower counting network based on filter pruning. *Complex and Intelligent Systems*, 10, 2047–2066. <https://doi.org/10.1007/s40747-023-01261-7>
- Zafar, A., Aamir, M., Mohd Nawi, N., Arshad, A., Riaz, S., Alruban, A., Dutta, A. K., & Almotairi, S. (2022). A comparison of pooling methods for convolutional neural networks. *Applied Sciences*, 12(8643). <https://doi.org/10.3390/app12178643>
- Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: a review. *Precision Agriculture*, 13, 693–712. <https://doi.org/10.1007/s11119-012-9274-5>
- Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. In *Artificial Intelligence Review*. Springer Netherlands. <https://doi.org/10.1007/s10462-024-10721-6>

Capítulo 3 ARTÍCULO CIENTÍFICO

DETECCIÓN DE FLORESCENCIA EN PIÑA (*Ananas comosus*) MEDIANTE APRENDIZAJE PROFUNDO Y TECNOLOGÍA DE DRONES

DETECTION OF FLOWERSCENCE IN PINEAPPLE (*Ananas comosus*) USING DEEP LEARNING AND DRONE TECHNOLOGY

3.1 Resumen

La detección de la floración en cultivos de piña (*Ananas comosus*) es un desafío clave para optimizar la planificación de la cosecha y la gestión agrícola. Este estudio tuvo como objetivo desarrollar un método eficiente y preciso para la identificación de la floración, empleando drones y redes neuronales profundas. La investigación se llevó a cabo en terrenos agrícolas proporcionados por DIMASUR, en la localidad de Juan Rodríguez Clara, Veracruz. Se utilizó el dron DJI Mavic 3 RGB, equipado con una cámara de resolución espacial de 0.25 cm/píxel (GSD), logrando capturar imágenes detalladas de los cultivos en floración. La resolución espacial del vuelo permitió una cobertura uniforme y precisa, generando un conjunto de datos compuesto por más de 6,000 imágenes y 500,000 etiquetas bajo diversas condiciones de luz y clima. Se evaluaron los modelos YOLOv8n, YOLOv10n y YOLOv11n, comparando su rendimiento en términos de precisión y consistencia. El modelo YOLOv11n se destacó como la mejor opción, alcanzando una precisión de 95.25% y un recall de 93.25%, con valores de mAP@0.5 de 97.21% y mAP@0.5-0.95 de 80.47%. Estos resultados demostraron que YOLOv11n proporciona un equilibrio ideal entre precisión y sensibilidad sin necesidad de recurrir a modelos más robustos, lo cual reduce significativamente la demanda computacional y los tiempos de procesamiento.

Palabras clave: Detección, Dron, Monitoreo, Aprendizaje Profundo, YOLO, Piña.

3.2 Abstract

The detection of flowering in pineapple (*Ananas comosus*) crops is a key challenge for optimizing crop planning and farm management. This study aimed to develop an efficient and accurate method for flowering identification using drones and deep neural networks. The research was carried out on agricultural land provided by DIMASUR, in the town of Juan Rodríguez Clara, Veracruz. The DJI Mavic 3 RGB drone was used, equipped with a camera with a spatial resolution of 0.25 cm/pixel (GSD), capturing detailed images of flowering crops. The spatial resolution of the flight allowed uniform and accurate coverage, generating a dataset composed of more than 6,000 images and 500,000 labels under various light and weather conditions. The YOLOv8n, YOLOv10n and YOLOv11n models were evaluated, comparing their performance in terms of accuracy and consistency. The YOLOv11n model stood out as the best option, achieving an accuracy of 95.25% and recall of 93.25%, with mAP@0.5 values of 97.21% and mAP@0.5-0.95 of 80.47%. These results demonstrated that YOLOv11n provides an ideal balance between accuracy and sensitivity without the need for more robust models, significantly reducing computational demand and processing times.

Keywords: Detection, Drone, Monitoring, Deep Learning, YOLO, Pineapple.

3.3 Introducción

Los productores de estos cultivos de alto valor toman una serie de decisiones clave en cada ciclo de cultivo. Para la piña, los datos que respaldan estas decisiones generalmente se limitan a observaciones visuales del terreno. Pero estas observaciones son de la periferia donde la variabilidad espacial y temporal, la etapa de crecimiento y desarrollo no se pueden determinar o cuantificar en todo el campo. Caminar por el campo es difícil ya que las plantas crecen muy juntas hasta 30 000 plantas por ha. Esta falta de información completa y en tiempo real sobre las condiciones del campo puede conducir a malas decisiones que resulten en muy poca o demasiada agua, fertilización, pesticidas y reguladores de crecimiento, o una mala planificación y programación de los recursos de siembra y cosecha, incluidos los equipos y la mano de obra.

La floración natural de la piña afecta el desarrollo y la calidad de la fruta e impacta en la cosecha (Bartholomew, 2014 Sanewski et al., 2018). Los productores de piña utilizan productos químicos que producen etileno para inducir la floración, de modo que la mayoría de las plantas dentro de un campo produzcan frutos de alta calidad listos para cosechar casi al mismo tiempo (Bartholomew, 2014). La situación ideal sería que un productor cosechara todo el campo en una sola pasada cuando hay poca variación en la floración, aumentando significativamente la productividad y eliminando el costo de cosechas adicionales.

Sin embargo, debería considerarse que las características de las hojas de la planta de piña, como sus bordes serrados y afilados, pueden provocar daños físicos a los cosechadores durante la recolección, lo que representa un desafío adicional en términos de ergonomía y seguridad laboral. Este aspecto subraya la importancia de implementar estrategias tecnológicas, como la detección automatizada de floración mediante drones y algoritmos avanzados, para mejorar no solo la eficiencia en la cosecha, sino también las condiciones de trabajo en el campo.

Las tecnologías agrícolas avanzan constantemente, ofreciendo nuevas oportunidades para mejorar tanto la productividad como la seguridad de los agricultores. Una de las herramientas tecnológicas que, aunque ha estado disponible durante algún tiempo, ha cobrado mayor relevancia en la agricultura en años recientes es el dron. Su creciente accesibilidad, junto con el avance de la digitalización en el sector agrícola, ha posicionado a los drones como aliados clave para optimizar la producción agrícola a nivel global. (Jung et al., 2016).

Las ortofotos generadas por drones ofrecen imágenes aéreas de alta resolución espacial, con una precisión escala de centímetros. Estas imágenes, que destacan por su alta calidad y costo accesible, permiten la interpretación, análisis y cuantificación precisa de objetos en una disposición horizontal. Además, su integración con bibliotecas de aprendizaje automático facilita el reconocimiento de imágenes y el análisis de patrones o conglomerados, ampliando las posibilidades en aplicaciones avanzadas de análisis de datos. (Montoya, 2020).

Tanto la visión por computadora tradicional como los enfoques basados en el aprendizaje profundo se han utilizado para una variedad de aplicaciones agrícolas basadas en el conteo. Guo et al. (2018) , Malambo et al. (2019) utilizaron técnicas basadas en detección para cabezas de sorgo en un campo. Del mismo modo, Gené et al. (2020) utilizó Mask-RCNN para identificar y segmentar por completo las manzanas en los árboles de un huerto. Para contar palmeras a partir de imágenes de dron, Li et al. (2017) utilizó un enfoque de detección basado en CNN. Osco et al. (2020) utilizaron un enfoque muy similar para contar el número de árboles de cítricos en una arboleda.

Además, se han llevado a cabo investigaciones específicas sobre la detección y el conteo en cultivos de piña utilizando enfoques avanzados de aprendizaje profundo. Un estudio titulado "Large-Scale Counting and Localization of Pineapple Inflorescence Through Deep Density-Estimation" implementó una técnica de estimación de densidad profunda para contar y localizar inflorescencias de piña a partir de imágenes aéreas. Este enfoque permitió una detección más precisa y eficiente en comparación con métodos tradicionales,

facilitando la planificación de la cosecha y la gestión del cultivo. El modelo alcanzó resultados destacados. Los parámetros de evaluación obtenidos incluyen un error absoluto (MAE) de 11.5 y una desviación porcentual media absoluta (MAPD) de 6.37% en el conjunto de prueba (Hobbs et al., 2021).

Otro trabajo, "A Deep Object Detection Method for Pineapple Fruit and Flower Recognition in Cluttered Background", desarrolló un modelo basado en CNN para la detección de flores y frutos de piña en entornos con fondos complejos. El estudio destacó la eficacia del modelo en condiciones de iluminación variables y con oclusiones, mejorando la precisión en la identificación de las etapas de desarrollo del cultivo, ya que alcanzó una precisión de 65.6% y una capacidad de recuperación de 74.1%, lo que indica una adecuada capacidad para identificar correctamente los elementos presentes. Además, obtuvo un mAP de 64% y un F1 score de 0.69, reflejando un equilibrio favorable entre precisión y recall, incluso bajo condiciones desafiantes como iluminación variable, tamaños de objeto diferente y oclusión significativas. (Wang et al., 2020).

Por otro lado, el estudio "A-pruning: a lightweight pineapple flower counting network based on filter pruning" presentó una red ligera basada en técnicas de poda de filtros para optimizar el modelo YOLOv5 en el conteo de flores de piña. Este modelo alcanzó una precisión del 71.7% con solo 1.7 millones de parámetros y una velocidad de procesamiento de 178 cuadros por segundo, mostrando su eficacia en aplicaciones en tiempo real (Yu et al., 2024).

Estos estudios demuestran el potencial de combinar técnicas como la poda de filtros y la extracción multiescala de características para crear soluciones eficientes y precisas en el monitoreo del cultivo de piña. El avance de estas tecnologías en la agricultura de precisión subraya su capacidad para mejorar la gestión de cultivos y la productividad, consolidando su importancia en sistemas agrícolas modernos.

Las imágenes más ampliamente utilizadas en diversos campos son las imágenes RGB, las cuales registran únicamente una parte del espectro visible dentro del espectro electromagnético completo. Estas imágenes son comparables a las que

generan las cámaras integradas en los teléfonos móviles actuales, lo que las hace familiares y accesibles para múltiples aplicaciones. (Geoinnova, 2019).

El aprendizaje profundo, una rama especializada dentro del aprendizaje automático, se ha establecido como una herramienta poderosa y adaptable para gestionar grandes volúmenes de datos heterogéneos. En el ámbito de la ciencia de las plantas, esta tecnología ha sido adoptada de manera creciente para simplificar y optimizar el análisis de extensos conjuntos de datos, generados principalmente a través de investigaciones de fenotipado y genotipado de alto rendimiento. Su capacidad para identificar patrones complejos ha revolucionado la interpretación y el uso de datos en este campo. (Cruz, 2019)

En este contexto, el estudio tuvo como objetivo desarrollar un método preciso y eficiente para la detección de la floración en piña mediante aprendizaje profundo y tecnología de drones.

3.4 Materiales y Métodos

3.4.1 Adquisición de datos

El conjunto de datos utilizado en esta investigación se obtuvo de una plantación de piña de la variedad Golden Sweet (MD-2), la cual presentaba un estado de inducción floral con etileno con tres meses de desarrollo. La superficie total del cultivo abarcó 1.8 ha y fue proporcionada por la empresa DIMASUR, ubicada en la comunidad de Juan Rodríguez Clara, Veracruz, con coordenadas 17° 19' 42.19" N, 95° 19' 59.19" W (115 msnm).

La adquisición de imágenes se realizó a las 10:00 horas bajo condiciones climáticas estables, empleando un dron DJI Mavic 3 equipado con una cámara RGB (Cuadro 3.1). El vuelo se llevó a cabo a una altitud de 14 m, lo que permitió obtener un tamaño de píxel en el suelo (GSD) de 0.28 cm/píxel, capturando imágenes detalladas de las flores de piña con una resolución de 5280 × 3956 píxeles. Las misiones de vuelo se realizaron con una superposición del 80% en las imágenes, tanto en direcciones frontales como laterales, lo que garantizó una cobertura completa y uniforme del área de estudio, evitando zonas sin datos.

Cuadro 3.1: Características Dron Mavic 3 RGB

Cámara RGB	
Sensor de imagen	CMOS 4/3. Píxeles efectivos: 20 MP
Objetivo	Campo de visión: 84°
	Distancia focal equivalente: 24 mm
	Apertura: de f/2.8 a f/11
	Enfoque: de 1 m a ∞
Fuente: DJI Mavic 3M - Specifications – DJI (2024)	

3.4.2 Procesamiento de conjuntos de datos y su etiquetado

Se llevó a cabo la generación de ortomosaicos (Figura 3.1) utilizando el software de fotogrametría Agisoft Metashape, a partir de las imágenes capturadas por el dron DJI Mavic 3 RGB durante las misiones de vuelo. Este proceso permitió integrar las imágenes aéreas en un único mosaico con una resolución espacial de 0.25 cm/píxel (GSD), asegurando un nivel de detalle adecuado para los análisis posteriores con un total de 1355 imágenes con resolución de 5280 × 3956 píxeles fueron seleccionadas y escaladas a 640 × 640 píxeles para estandarizar su tamaño una vez generado el ortomosaico.



Figura 3.1: a) Mosaico generado mediante Agisoft; b) imagen seccionada a partir del ortomosaico para su etiquetado 640x640

Las imágenes se anotaron manualmente utilizando la herramienta *Roboflow*, donde cada flor de piña fue identificada y etiquetada con la categoría "flor_de_piña" (Figura 3.2). Este proceso generó archivos ".txt" que contenían las coordenadas de las cajas delimitadoras generando una máscara en la cual se especifican las coordenadas en formato YOLO

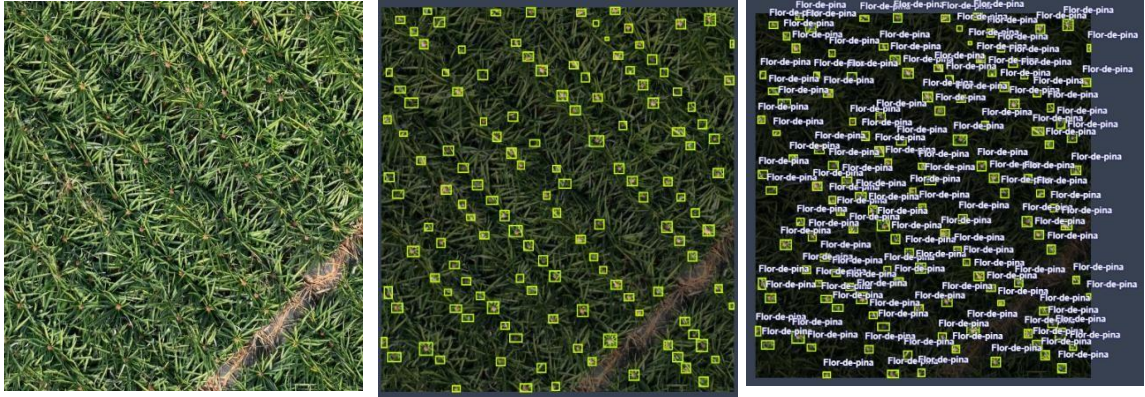


Figura 3.2: Etiquetado de la imagen (640x640) mediante la herramienta Roboflow

Finalmente, el conjunto de datos anotado se dividió en subconjuntos de entrenamiento, validación y prueba (Figura 3.3) en una proporción de 8.8: 0.8: 0.4 asignando 5808 imágenes al entrenamiento, 564 a la validación y 288 a la prueba, garantizando un balance adecuado para el entrenamiento y evaluación del modelo.

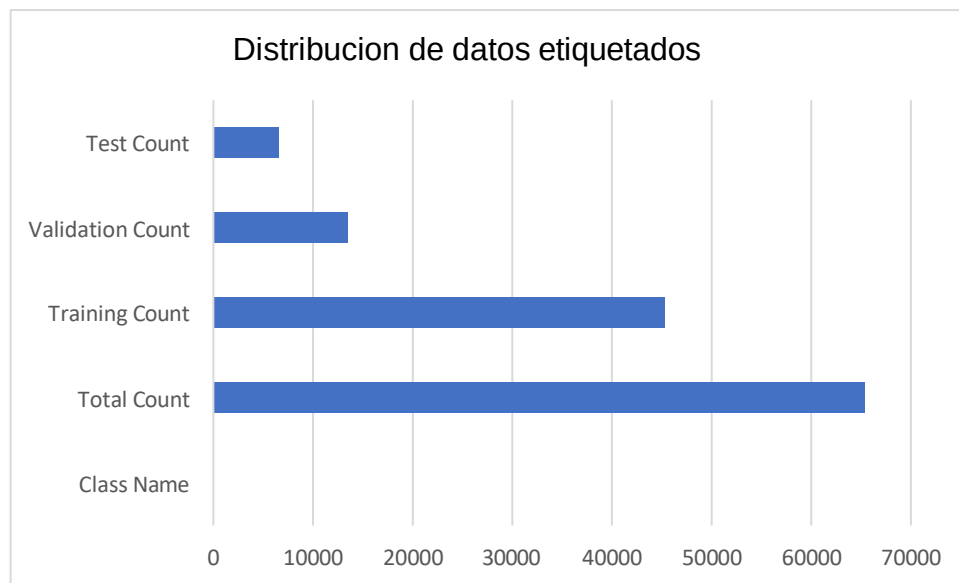


Figura 3.3: Distribución de etiquetas para su entrenamiento

Para obtener el total de etiquetas mencionado, en la fase de preprocesamiento, se aplicó una orientación automática a las imágenes para asegurar la consistencia en la disposición del contenido, y se ajustó automáticamente el

contraste mediante Contrast Stretching, lo cual optimizó la visibilidad de los detalles críticos en condiciones de iluminación variables. La disposición de las imágenes se estructuró en mosaicos de 2 filas por 2 columnas, mejorando la eficiencia del procesamiento.

Adicionalmente, se implementaron técnicas de aumento de datos (augmentations) para incrementar la variabilidad de las imágenes de entrenamiento, generando tres variantes por cada ejemplo de entrenamiento. Estas técnicas incluyeron reflejos horizontales y verticales, ajustes en la saturación con variaciones entre -33% y +33%, y cambios en el brillo que oscilaron entre -21% y +21%. Estas modificaciones aumentaron la robustez de la base de datos.

La elección de una proporción de 88% para el conjunto de entrenamiento, 8% para validación y 4% para prueba se justificó en función de la necesidad de maximizar la cantidad de datos disponibles para el entrenamiento del modelo, sin comprometer la calidad del proceso de validación y evaluación. Este enfoque se basa en estudios que destacan que, en conjuntos de datos con un volumen considerable de imágenes, es beneficioso asignar la mayor parte de los datos al entrenamiento, siempre que los subconjuntos de validación y prueba sean suficientemente representativos para evitar sesgos (Torres, 2023).

Adicionalmente, este tipo de partición es utilizado en tareas donde el volumen de datos anotados es relativamente grande y permite garantizar una cobertura robusta durante el aprendizaje del modelo (Johnson & Khoshgoftaar, 2019). La proporción de 88% para el entrenamiento permite al modelo aprender un amplio rango de características presentes en los datos, mientras que el 8% destinado a validación asegura un ajuste fino de los hiperparámetros durante el entrenamiento. Por su parte, el 4% asignado a prueba proporciona una evaluación independiente y confiable del desempeño final del modelo.

Aunque configuraciones más tradicionales como 70-15-15 o 80-10-10 son comunes, en este caso particular, la cantidad de datos disponibles permitió utilizar subconjuntos más pequeños para validación y prueba, manteniendo un

balance entre la representatividad de los datos y la eficiencia en el entrenamiento. Esto asegura que el modelo tenga acceso a una mayor cantidad de ejemplos durante el aprendizaje, optimizando su capacidad de generalización en nuevas imágenes.

3.4.3 Selección de la red de detección y su entrenamiento

Las redes neuronales convolucionales (CNN) han revolucionado el procesamiento de imágenes en diversas aplicaciones, incluyendo la agricultura de precisión. Estas arquitecturas están compuestas por capas convolucionales que extraen características locales de las imágenes, capas de pooling que reducen la dimensionalidad y capas completamente conectadas que realizan la clasificación final. Esta estructura permite a las CNN aprender representaciones jerárquicas de las imágenes, desde patrones simples hasta complejos, mejorando significativamente la precisión en tareas como detección de objetos y segmentación semántica (Zhao et al., 2024).

En este trabajo de investigación, se implementaron las arquitecturas YOLOv8, YOLOv10 y YOLOv11 para la detección de la floración en plantaciones de piña. Estas versiones de YOLO han demostrado mejoras significativas en precisión y eficiencia en tareas de detección de objetos. YOLOv8 introdujo una arquitectura anclada que simplifica el proceso de entrenamiento y mejora la precisión en la detección de objetos pequeños (Jocher et al., 2023). Por su parte, YOLOv10 eliminó la necesidad de la supresión no máxima (NMS) mediante asignaciones duales consistentes, lo que redujo la latencia de inferencia y mejoró la eficiencia computacional (Wang et al., 2024). Finalmente, YOLOv11 incorporó bloques C3k2 y SPPF, optimizando la extracción de características y aumentando la precisión en la detección de objetos en diversas condiciones (Khanam & Hussain, 2024). Estas mejoras arquitectónicas y metodológicas hacen que estas versiones de YOLO sean adecuadas para aplicaciones en agricultura y en este caso para la detección de florecencia el cultivo de piña.

La adaptación de las arquitecturas YOLOv8, YOLOv10 y YOLOv11 para este trabajo se realizó ajustando la profundidad y el ancho de las capas de red con

base en los requerimientos específicos de detección de floración en cultivos de piña. Estos ajustes implicaron la modificación de los bloques principales de la red (Backbone y Neck) multiplicando el número de bloques y filtros por factores predefinidos Cuadro 3.2.

Cuadro 3.2: Ajuste de profundidad y el ancho de capas para YOLO

Versión	Profundidad de la arquitectura	Ancho de capas
YOLOv8-n	0.33	0.25
YOLOv10-n	0.33	0.50
YOLOv11-n	0.33	0.25

La implementación de YOLOv8, YOLOv10 y YOLOv11 se realizó utilizando la biblioteca PyTorch, aprovechando sus optimizaciones para tareas de detección de objetos. Los modelos se entrenaron con los hiperparámetros ajustados para la base de datos COCO, adaptándolos al conjunto de datos de floración en cultivos de piña. Los detalles de los hiperparámetros utilizados se resumen en Cuadro 3.3.

Cuadro 3.3: Hiperparámetros de entrenamiento para los algoritmos elegidos

Algoritmo	Tamaño de imagen	Lote	Optimizador	Tasa de aprendizaje	Decaimiento (% iteraciones)	Max (iteraciones)	Pesos pre-entrenados
YOLOv8-n	640 × 640 × 3	16	Adam	0.01	Automático	700	COCO
YOLOv10-n	640 × 640 × 3	16	Adam	0.01	Automático	600	COCO
YOLOv11-n	640 × 640 × 3	16	Adam	0.01	Automático	500	COCO

3.4.4 Métodos de evaluación

Se emplearon métricas cuantitativas para evaluar el desempeño de los modelos, incluyendo el Recall, la precisión y la precisión promedio media (mean Average Precision, mAP). El valor de mAP@0.5 representa la precisión promedio cuando el umbral de Intersección sobre Unión (IoU) se establece en 0.5. Por su parte, los

valores de recuperación y precisión se determinaron utilizando las siguientes fórmulas:

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

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

F1: indicador que mezcla los conceptos de precisión y recall.

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (3)$$

Donde, True Positive (TP) representa el número de veces que el modelo identificó correctamente las flores de piña. False Positive (FP) se refiere a las ocasiones en que el modelo identificó erróneamente elementos como flores de piña cuando no lo eran. Por otro lado, True Negative (TN) indica el número de veces que el modelo no clasificó un elemento como flor de piña y esta predicción fue correcta, mientras que False Negative (FN) corresponde a las ocasiones en que el modelo no identificó una flor de piña cuando sí lo era. La precisión promedio (Average Precision, AP) puede calcularse como el área bajo la curva de precisión y recuperación (Precision-Recall, P-R).

3.4.5 Recursos Implementados

Para el desarrollo de este estudio, se utilizó una infraestructura de hardware y software adecuada para entrenar y evaluar los modelos ligeros de detección como se muestra en el Cuadro 3.4. Este entorno permitió integrar de manera efectiva los algoritmos de detección YOLOv8-n, YOLOv10-n y YOLOv11-n, asegurando un flujo de trabajo eficiente desde la preparación de datos hasta la evaluación de los resultados.

Cuadro 3.4: Recursos de hardware y software implementados

Configuración	Parámetro
Sistema operativo	Windows 11
Tipo de sistema operativo	64 bits

CPU	Intel(R) Core (TM) i7-12700H
GPU	GeForce RTX 3050 Ti 4GB
Versión de Python	Pitón 3.12
Entorno acelerado	CUDA 12.4
Marco de referencia	PyTorch
Versión de PyTorch	2.5.1 + CU124
IDE	PyCharm

3.5 Resultados

3.5.1 Entrenamiento

Se registraron detalles sobre la duración del entrenamiento, el número de iteraciones realizadas y la GPU asignada previamente mencionada (Cuadro 3.5).

Cuadro 3.5: Tiempo de entrenamiento de cada Modelo

Algoritmo	GPU	Tiempo de entrenamiento (hrs)	Iteraciones	Factor de error
YOLO V8_n	Nvidia RTX	6.2	700	$\leq 0,35$
YOLO V10-n	3050 Ti 4GB	5.9	600	$\leq 0,40$
YOLO V11-n		5.6	500	$\leq 0,30$

El número de iteraciones distintas entre los modelos se explica por la configuración del parámetro de paciencia (patience) establecida en cada caso, lo cual permitió detener el entrenamiento automáticamente cuando no se obtiene una mejora significativa en las métricas de validación durante un número consecutivo de épocas. En este sentido, el factor de error está asociado a la tolerancia definida por este parámetro, optimizando así el tiempo de entrenamiento y evitando sobreajustes en los modelos. Este enfoque resulta particularmente útil para garantizar que cada modelo alcance su punto óptimo de rendimiento. consumir recursos computacionales innecesarios.

Los parámetros de entrenamiento utilizados para los modelos YOLOv8n, YOLOv10n y YOLOv11n (Cuadro 3.6) reflejan ajustes específicos para optimizar

el desempeño en la tarea de detección de floración en cultivos de piña. Los valores de epochs y patience varían entre los modelos, ajustándose a las necesidades de convergencia y estabilidad del aprendizaje. El tamaño de lote (batch size) y la resolución de las imágenes (image size) se mantuvieron constantes en 16 y 640 píxeles, respectivamente, para garantizar uniformidad en el procesamiento.

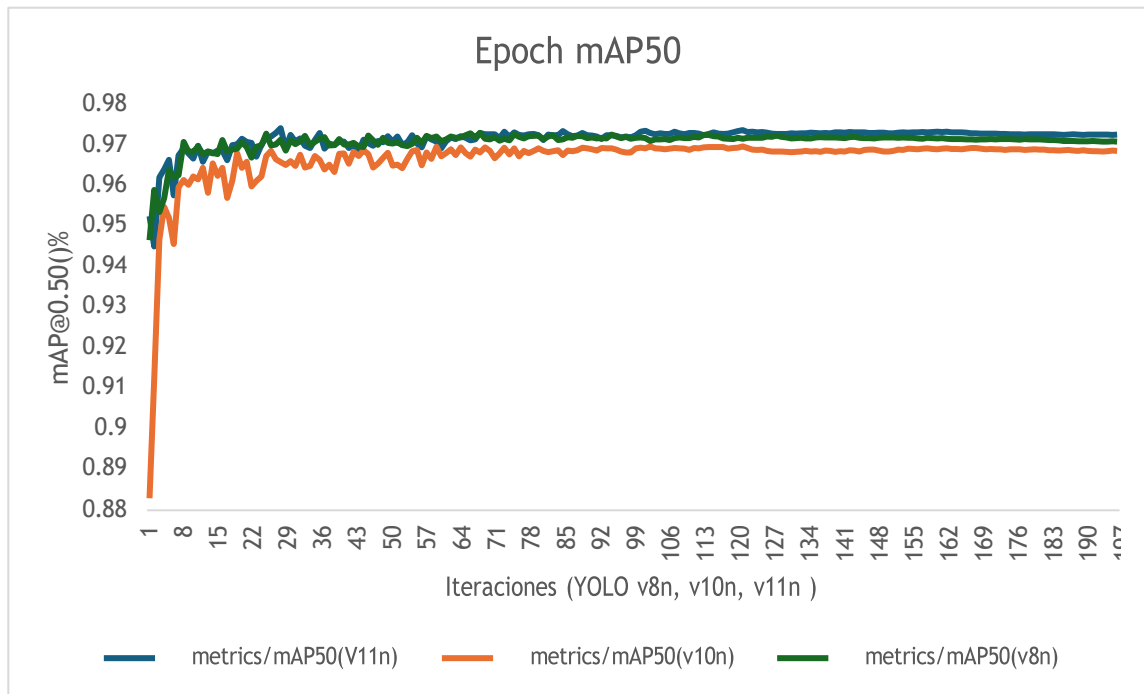
El optimizador configurado como "auto" permitió que cada modelo seleccionara automáticamente la estrategia más adecuada para minimizar la función de pérdida. Las variaciones en el confidence threshold y el IoU threshold entre los modelos reflejan ajustes en los criterios de detección, donde YOLOv10n priorizó una confianza mayor (0.25) y un IoU más estricto (0.5) para minimizar falsos positivos, mientras que YOLOv8n y YOLOv11n se enfocaron en un balance entre precisión y recuperación con valores de 0.2 y 0.45, respectivamente.

Cuadro 3.6: Parámetros de entrenamiento para los modelos YOLOv8n, YOLOv10n y YOLOv11n

Parámetros	Valor		
	YOLOv8n	YOLOv10n	YOLOv11n
Modelo	YOLOv8n	YOLOv10n	YOLOv11n
Epochs	700	500	500
Patience	250	250	250
Batch Size	16	16	16
Image Size	640	640	640
Optimizer	auto	auto	auto
Confidence Threshold (conf)	0.2	0.25	0.2
IoU Threshold (iou)	0.45	0.5	0.45
Learning Rate (lr)	0.001	0.01	0.001
Weight Decay	0.0005	0.0005	0.0005
Momentum	0.937	0.937	0.937
Label Smoothing	0	0	0.1

En cuanto a la métrica mAP@0.5 que se puede apreciar en las Figura 3.4, YOLOv8n alcanzó un promedio de 97.11% a partir de las 130 iteraciones. Por su parte, YOLOv10n obtuvo un mAP@0.5 promedio de 96.71% a partir de las 150 iteraciones, mostrando menor consistencia en comparación con YOLOv8n. Finalmente, YOLOv11n presentó el mejor desempeño con un mAP@0.5

promedio de 97.21% desde las primeras 115 iteraciones más preciso y estable entre los tres evaluados.



Respecto al $mAP@0.50$, que mide la precisión promedio al identificar flores con un umbral de coincidencia del 50%, YOLOv11n destacó con 97.47%, seguido por YOLOv8n (97.35%) y YOLOv10n (97.01%). Esto confirma que YOLOv11n tiene la mayor capacidad de detección a un nivel de coincidencia estándar. Para $mAP@0.50-0.95$, que evalúa la precisión promedio bajo condiciones más estrictas, YOLOv11n también obtuvo el mejor resultado (80.47%), seguido de YOLOv10n (80.15%) y YOLOv8n (80.04%), destacándose por su consistencia en la detección de flores.

Cuadro 3.7: Métricas del conjunto de prueba para la confianza y IoU previamente mencionados

Modelo	Precisión	Recall	F1-score	$mAP@0.50$	$mAP50-95$
YOLO v8n*	95.38%	92.98%	94.16%	97.35%	80.04%
YOLO v10n**	95.11%	92.55%	93.80%	97.01%	80.15%
YOLO V11n***	95.95%	93.69%	94.80%	97.47	80.47

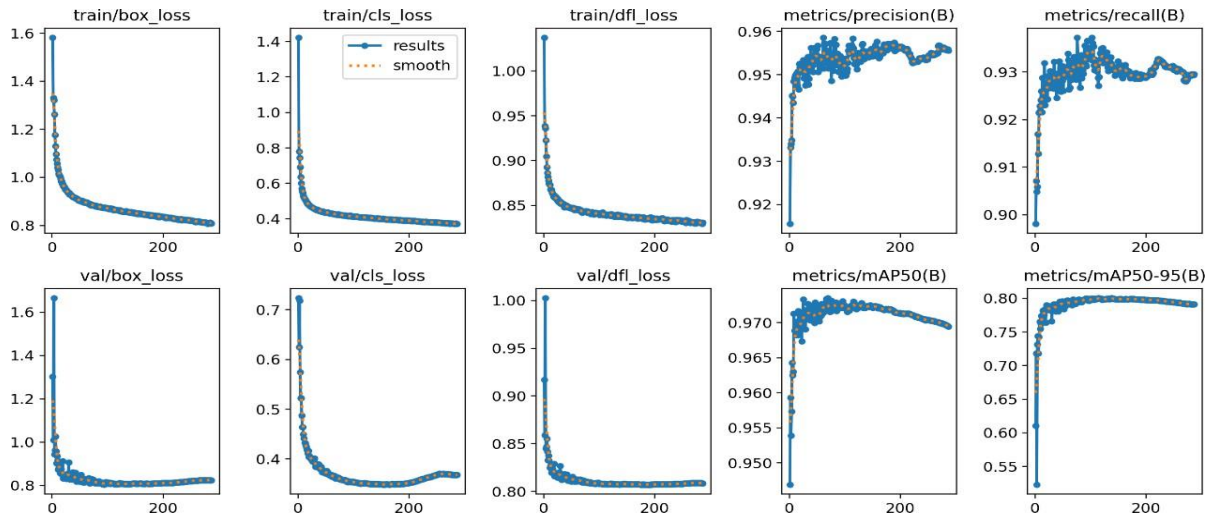


Figura 3.5: Evaluación de métricas de validación para el entrenamiento del modelo YOLO V8n

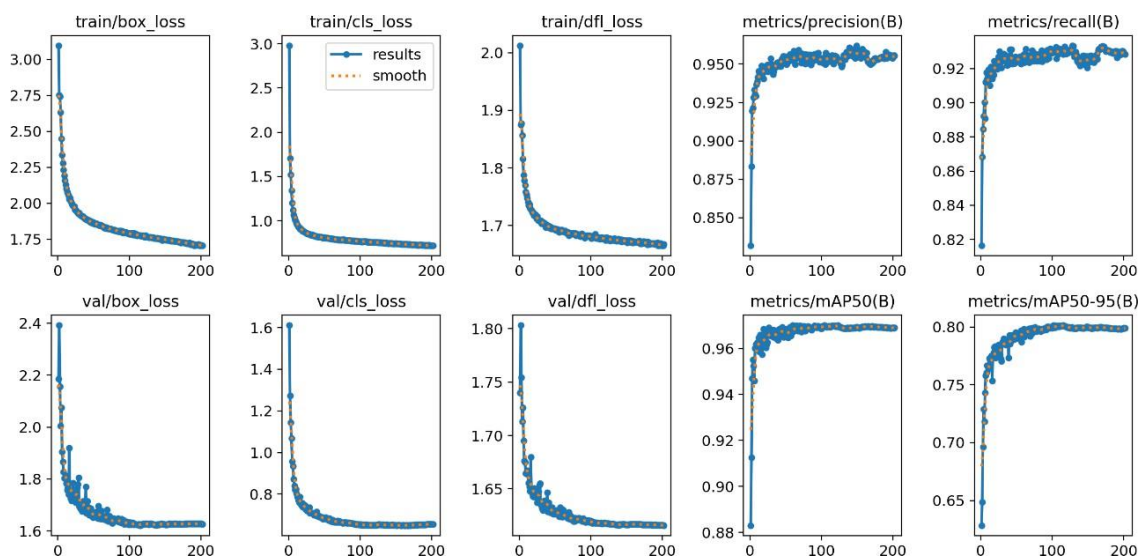


Figura 3.6: Evaluación de métricas de validación para el entrenamiento del modelo YOLO V10n

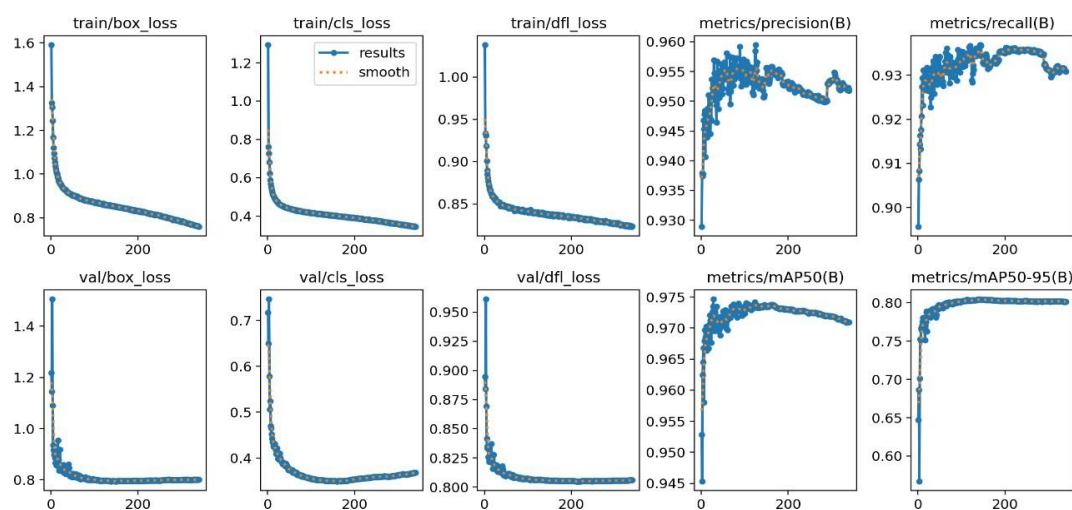


Figura 3.7: Evaluación de métricas de validación para el entrenamiento del modelo YOLO V11n

3.5.2 Evaluación

En esta investigación, se evaluó el impacto de los umbrales de Intersección sobre Unión (IoU) en la precisión de detección de floración (Figura 3.8) utilizando el modelo YOLOv11n en cultivos de piña. Para ello, se analizaron resultados a

umbrales de IoU de 0.45 y 0.25, considerando imágenes capturadas con un GSD de 0.28 cm/píxel.

Al establecer un umbral de IoU de 0.45, se observó un incremento en los True Positives (TP), reflejando detecciones precisas de las flores, aunque acompañado de un ligero aumento en los False Positives (FP) debido a la mayor sensibilidad del modelo. En contraste, con un umbral más bajo de IoU (0.25), los TP se mantuvieron consistentes, pero los FP incrementaron significativamente debido a la relajación en los criterios de coincidencia. Esto evidencia que la selección del umbral de IoU afecta directamente la precisión y recuperación del modelo siendo así el 0.45 el óptimo para este modelo.



Figura 3.8: Detecciones de floración de la piña utilizando con umbral 0.45

El Cuadro 3.8 representa los resultados de las métricas precisión, Recall, F1, mAP50, correspondientes a cada modelo CNN evaluado. Los valores de TP, FP y FN están normalizados en un rango de 0 a 1, mientras que FP se expresa en relación con la suma de TP + FN.

Cuadro 3.8: Comparación de Desempeño de Modelos YOLO en Detección de Floración

Modelo	Precisión	Recall	F1	mAP50	TP	FP	FN	TN
YOLO v8n	95.38%	92.98%	94.16%	97.35%	0.96	0.00	0.04	1.00
YOLO v10n	95.11%	92.55%	93.80%	97.01%	0.95	0.00	0.05	1.00
YOLO V11n	95.95%	93.69%	94.80%	97.47	0.96	0.00	0.04	1.00

3.6 Discusión

La detección de flores de piña en cultivos mediante el uso de modelos de aprendizaje profundo como YOLOv8n, YOLOv10n y YOLOv11n ha demostrado ser una solución eficiente para abordar los desafíos inherentes a la falta de sincronización en la floración. A diferencia de los métodos tradicionales de inspección visual descritos por Hobbs et al. (2020), que solo permitían una visión limitada del cultivo y se caracterizaban por una alta carga de trabajo manual, nuestros modelos han alcanzado precisiones promedio (mAP@0.5) superiores al 97%, con YOLOv11n sobresalió con un mAP@0.5 de 97.47%. Este desempeño supera significativamente los valores reportados por métodos de estimación de densidad basados en aprendizaje profundo, como el propuesto por Hobbs et al. (2020), que obtuvo un error absoluto medio (MAE) de prueba de 11.5 y un porcentaje absoluto medio de desviación (MAPD) de 6.37%.

Comparando con el trabajo de Chen et al. (2020), donde un modelo convencional logró una mAP de 0.64 y una puntuación F1 de 0.69, nuestros modelos basados en YOLO muestran una mejora significativa, con valores de F1 superiores al 94%. Esto resalta la capacidad de los modelos YOLO para manejar mejor las complejidades del entorno agrícola, como la variabilidad en las condiciones de iluminación y la oclusión causada por hojas y frutos. Por ejemplo, Fan Meng et al. (2023) destacan que los modelos basados en transformadores alcanzan precisiones del 92.54%, pero con tiempos de inferencia más altos, lo que limita su aplicación en tiempo real. En contraste, nuestros modelos, especialmente YOLOv11n, no solo ofrecen alta precisión, sino también estabilidad en métricas

como el Recall (93.69%) y la precisión (95.95%), siendo de utilidad para aplicaciones de fenotipado en el cultivo de la piña.

El enfoque de optimización de modelos ligeros, como el YOLOv5_E propuesto por Guoyan Yu et al. (2023), se enfoca en la eficiencia en dispositivos móviles, alcanzando una precisión del 71.7% y una velocidad de 178 FPS. Sin embargo, la reducción en la precisión del modelo es evidente cuando se compara con los resultados obtenidos en esta investigación. Aunque el enfoque ligero tiene ventajas en cuanto a velocidad y capacidad de integración en plataformas móviles, los modelos YOLOv8N, YOLOv10N Y YOLOv11N ofrecen un balance óptimo entre precisión y robustez en la detección.

Finalmente, nuestros hallazgos resaltan que, al igual que en el estudio de (Hobbs et al., 2021) SE CITA HOBBS ET AL (2021), la detección automatizada basada en aprendizaje profundo permite a los productores optimizar sus prácticas de manejo y reducir costos operativos. Sin embargo, a diferencia de los enfoques anteriores, el uso de modelos YOLO en este trabajo muestra una capacidad superior para manejar grandes volúmenes de datos y capturar variaciones en las flores de piña, lo que permite una detección más precisa. Estos resultados confirman que los modelos YOLOv8n, YOLOv10n y YOLOv11n no solo son herramientas valiosas para la agricultura de precisión, sino que también representan un avance significativo en la gestión de cultivos de piña, superando ampliamente las limitaciones de métodos tradicionales y otros modelos reportados en la literatura reciente.

3.7 Conclusiones

En este trabajo se desarrolló con éxito una base de datos robusta y georreferenciada compuesta por imágenes aéreas de cultivos de piña en etapa de floración, capturadas con un dron DJI Mavic 3 RGB. Esta base de datos sirvió como insumo principal para entrenar y evaluar los modelos YOLOv8, YOLOv10 y YOLOv11, permitiendo analizar su desempeño en la detección de florescencia en condiciones agrícolas reales. Los resultados obtenidos destacan que los tres

modelos alcanzaron métricas de precisión superiores al 95%, con YOLOv11 mostrando la mejor combinación entre precisión, recall y mAP@0.5, confirmando su idoneidad para aplicaciones en la gestión y planificación agrícola de cultivos de piña.

Los principales desafíos abordados en esta investigación incluyeron la variabilidad en las condiciones de iluminación, la oclusión por hojas y la superposición de flores, problemas típicos de los entornos agrícolas. Adicionalmente, se enfrentaron retos relacionados con la selección de hiperparámetros adecuados para optimizar los modelos y garantizar su robustez en escenarios adversos. Comparado con trabajos previos, donde los modelos obtenían métricas significativamente más bajas, los resultados aquí presentados validan la superioridad de YOLOv8, YOLOv10 y YOLOv11 en el manejo de estas complejidades, evidenciando su potencial para superar las limitaciones de métodos tradicionales y otros algoritmos de aprendizaje profundo.

Para investigaciones futuras, se sugiere explorar el uso de modelos más ligeros y optimizados para dispositivos móviles, como redes optimizadas por poda estructural o variantes híbridas que integren mecanismos de atención para mejorar la precisión en tiempo real. También sería beneficioso ampliar la base de datos incluyendo imágenes capturadas en distintas estaciones del año y bajo diferentes condiciones climáticas, lo que permitiría una generalización más amplia de los modelos. Finalmente, se recomienda integrar estos sistemas de detección con plataformas de gestión agrícola para facilitar la toma de decisiones y maximizar la productividad en cultivos de piña.

3.8 Literatura citada

Aitutoriales. (2024). *Aprendizaje automático no supervisado: algoritmos, tipos con ejemplo*. Aitutoriales. <https://aitutoriales.com/aprendizaje-automatico-no-supervisado-algoritmos-tipos-con-ejemplo/>

Arrendajo. (2024). *Introducción a YOLO: detección de objetos en tiempo real*. HashDork. <https://hashdork.com/es/yolo/>

- Bartholomew, D. P., Rohrbach, K. G., & Evans, D. O. (2002). Pineapple Cultivation in Hawaii. *Fruits and Nuts*, 1–8.
- Bohdal, O., Yang, Y., & Hospedales, T. (2023). Meta-calibration: learning of model calibration using differentiable expected calibration error. *Transactions on Machine Learning Reseach*.
- Bongiovanni, R., & Lowenberg, J. D. (2004). Precision agriculture and sustainability. *Precision Agriculture*, 5, 359–387. <https://doi.org/10.1023/B:PRAG.0000040806.39604.aa>
- Boominathan, L., & Babu, R. V. (2016). CrowdNet: a deep convolutional network for dense crowd counting. *ACM Multimedia*. <https://doi.org/10.1145/2964284.2967300>
- Bouchard, L. (2020). *Introduction to convolutional neural networks (CNNs) / the most popular deep learning architecture*. Medium. <https://medium.com/what-is-artificial-intelligence/introduction-to-convolutional-neural-networks-cnns-the-most-popular-deep-learning-architecture-b938f62f133f>
- Buriticá, A. (2024). *Uso de drones y sensores: tecnologías para monitorear cultivos y detectar problemas de plagas*. Agricultura.Croper. <https://blog.croper.com/uso-de-drones-y-sensores-tecnologias-para-monitorear-cultivos-y-detectar-problemas-de-plagas/>
- Calderón, J. G. (2024). *Optimizando la agricultura de piña con drones: un vuelo hacia la eficiencia*. Agrotec. <https://agroteccr.com/optimizando-la-agricultura-de-pina-con-drones-un-vuelo-hacia-la-eficiencia/>
- CINVESTAV. (2020). *Aplican aprendizaje profundo a la agricultura para reducir costos y aumentar su calidad*. Centro de Investigación y de Estudios Avanzados Del Instituto Politécnico Nacional. <https://conexion.cinvestav.mx/Publicaciones/aplican-aprendizaje-profundo-a-la-agricultura-para-reducir-costos-y-aumentar-su-la-calidad>

- Cruz, J. C. (2019). *Uso del aprendizaje profundo para el estudio de fenotipo en plantas*. Medium. <https://medium.com/@jckruz777/usos-del-aprendizaje-profundo-para-el-estudio-de-fenotipo-en-plantas-2fa3f02f46e5>
- CulturaAI. (2024). *El aprendizaje profundo y sus aplicaciones*. CulturaAI. <https://culturaai.com/el-aprendizaje-profundo-y-sus-aplicaciones/>
- DJI. (2024). *Mavic 3 multispectral*. DJI. <https://ag.dji.com/es/mavic-3-m>
- Egi, Y., Hajyzadeh, M., & Eyceyurt, E. (2022). Drone-computer communication based tomato generative organ counting model using YOLO V5 and deep-sort. *Agriculture*, 12(9). <https://doi.org/10.3390/agriculture12091290>
- FAO. (2023). *Pineapple production statistics*. Food and Agriculture Organization. <https://vida-sustentable.com/como-se-cultiva-la-pina/>
- Fernandez, O. (2023). *Python DL: Deep Learning en python*. Codigospython. <https://codigospython.com/python-dl-deep-learning-en-python/>
- Gené, J. M., Sanz, R. C., Rosell, J. R. P., Ramon, J. M., Ruiz, J. H., Vilaplana, V., & Gregorio, E. (2020). Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. *Computers and Electronics in Agriculture*, 169. <https://doi.org/10.1016/j.compag.2019.105165>
- Gené-Mola, J., Vilaplana, V., Rosell-Polo, J. R., Morros, J.-R., Ruiz-Hidalgo, J., & Gregorio, E. (2019). *Uso de redes neuronales convolucionales para la detección remota de frutos con cámaras RGB-D*. 1081–1087. https://doi.org/10.26754/c_agroing.2019.com.3325
- Geoinnova. (2019). *Composiciones RGB de imágenes satélite*. 19 De Octubre. <https://geoinnova.org/blog-territorio/composiciones-rgb-de-imagenes-satelite/>
- Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., Singh, A. K., Singh, A., Hirafuji, M., Ninomiya, S., Ganapathysubramanian, B., Sarkar, S., & Guo, W. (2019). A weakly supervised deep learning

- framework for sorghum head detection and counting. *Plant Phenomics*.
<https://doi.org/10.34133/2019/1525874>
- Gonzalez-Dugo, V., Zarco-Tejada, P., Nicolás, E., Nortes, P. A., Alarcón, J. J., Intrigliolo, D. S., & Fereres, E. (2013). Using high resolution UAV thermal imagery to assess the variability in the water status of five fruit tree species within a commercial orchard. *Precision Agriculture*, 14(6), 660–678.
<https://doi.org/10.1007/s11119-013-9322-9>
- Guo, J., Li, X., Li, Z., Hu, L., Yang, G., Zhao, C., Fairbairn, D., Watson, D., & Ge, M. (2018). Multi-GNSS precise point positioning for precision agriculture. *Precision Agriculture*, 19(5), 895–911. <https://doi.org/10.1007/s11119-018-9563-8>
- Guo, W., Zheng, B., Potgieter, A. B., Diot, J., Watanabe, K., Noshita, K., Jordan, D. R., Wang, X., Watson, J., Ninomiya, S., & Chapman, S. C. (2018). Aerial imagery analysis – quantifying appearance and number of sorghum heads for applications in breeding and agronomy. *Frontiers in Plant Science*, 9(1544), 1–9. <https://doi.org/10.3389/fpls.2018.01544>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *IEEE*, 2961–2969. <https://doi.org/10.1109/ICCV.2017.322>
- Hidroponía, C. (2017). *Producción de piña en México*. Hydroenvironment.
<https://hidroponia.mx/produccion-de-pina-en-mexico/>
- Hobbs, J., Prakash, P., Paull, R., Hovhannisyan, H., Markowicz, B., & Rose, G. (2021). Large-scale counting and localization of pineapple inflorescence through deep density-estimation. *Frontiers in Plant Science*.
<https://doi.org/10.3389/fpls.2020.599705>
- Hu, C., Wang, F., Ai, C., Wang, X., & Wang, Y. (2021). An improved new convolutional neural network method for inverting the pore pressure in oil reservoir by surface vertical deformation. *Lithosphere*.
<https://doi.org/10.2113/2021/5597238>

- Huang, Y., Hoffmann, W. C., Lan, Y., Wu, W., & Fritz, B. K. (2009). Development of a spray system for an unmanned aerial vehicle platform. *Applied Engineering in Agriculture*, 25(6), 803–809. <https://doi.org/10.13031/2013.29229>
- Hunt, E. R., Dean Hively, W., Fujikawa, S. J., Linden, D. S., Daughtry, C. S. T., & McCarty, G. W. (2010). Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring. *Remote Sensing*, 2, 290–305. <https://doi.org/10.3390/rs2010290>
- IA, T. sobre. (2024). *Aprendizaje supervisado vs. no supervisado: diferencias y aplicaciones*. Todo Sobre IA. <https://iadatamind.com/aprendizaje-supervisado-vs-no-supervisado-diferencias-y-aplicaciones/>
- Infoagro. (2022). *Producción de piña en México y el mundo*. Reevista Infoagro México. <https://mexico.infoagro.com/produccion-de-pina-en-mexico-y-el-mundo/>
- INIFAD. (2020). *La tecnología potencializa al “rey de los frutos”: la piña*. INIFAD. <https://www.gob.mx/inifap/articulos/la-tecnologia-potencializa-al-rey-de-los-frutos-la-pina>
- ISPA. (2024). *Precision Ag definition*. International Society of Precision Agriculture. <https://www.ispag.org/about/definition>
- Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. (2024). Evaluating the evolution of YOLO (You Only Look Once) models: a comprehensive benchmark study of YOLO11 and its predecessors. *Computer Science*. <https://doi.org/10.48550/arXiv.2411.00201>
- Jensen, J. R. (2007). *Remote Sensing of Vegetation*. 1–33.
- Jocher, G., Chaurasia, A., & Qiu, J. (2023). *Ultralytics YOLOv8*. GitHub Repository. <https://github.com/ultralytics/ultralytics>

- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(27). <https://doi.org/10.1186/s40537-019-0192-5>
- Jung, F. R., Traut, K., Gauggel, C., Brooke, A. S., & Umali, R. (2016). Pineapple plantation information system-application of geo-informatics and unmanned aerial sensing technology for efficient land use management. *Acta Horticulturae*, 1111, 255–266. <https://doi.org/10.17660/ActaHortic.2016.1111.37>
- Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. In *Artificial intelligence review* (Vol. 53). Springer Netherlands. <https://doi.org/10.1007/s10462-020-09825-6>
- Li, W., Fu, H., Yu, L., & Cracknell, A. (2017). Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sensing*, 9(1). <https://doi.org/10.3390/rs9010022>
- Ma, Z., Wei, X., Hong, X., & Gong, Y. (2019). Bayesian loss for crowd count estimation with point supervision. *IEEE*, 6142–6151. <https://doi.org/10.1109/ICCV.2019.00624>
- Mahajan, A., Chakrabarty, N., Majithia, J., Ahuja, A., Agarwal, U., Suryavanshi, S., Biradar, M., Sharma, P., Raghavan, B., Arafath, R., & Shukla, S. (2023). Multisystem imaging recommendations/guidelines: in the pursuit of precision oncology. *Indian Journal of Medical and Paediatric Oncology*, 44(1), 2–25. <https://doi.org/10.1055/s-0043-1761266>
- Malambo, L., Popescu, S., Ku, N., Rooney, W., Zhou, T., & Moore, S. (2019). A deep learning semantic segmentation-based approach for field-level sorghum panicle counting. *Remote Sensing*, 11(24), 1–19. <https://doi.org/10.3390/rs11242939>

- Mapsens. (2023). *Inventario de piñas y control de salud con drones en Veracruz (México)*. Mapsens. <https://www.mapsens.com/proyecto-mapsens-agro-uso-de-mapsens-para-cultivo-de-pinas-en-mexico/>
- Montoya, S. (2020). *Conteo geoespacial de cultivos a partir de drones con Phyton, Scikit Learn y Scikit Image-tutorial-gidahatari*. Gidahatari. <https://gidahatari.com/ih-es/conteo-geoespacial-de-cultivos-a-partir-de-ortofotos-de-drones-con-python-scikit-learn-y-scikit-image-tutorial>
- Mounika, P., Narayana, C. L., Kondapalli, D., & Ramana, V. (2021). A comparative study of YOLO and SSD object detection algorithms. *International Journal of Creative Research Thoughts*, 9(11), 821–827.
- Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), 358–371. <https://doi.org/10.1016/j.biosystemseng.2012.08.009>
- Nkhoma, T. (2020). *Floración, polinización y poda de la piña*. Wikifarmer. <https://wikifarmer.com/es/floracion-polinizacion-y-poda-de-la-pina/>
- Nkhoma, T. (2024). *Pineapple cultivation guidelines*. Wikifarmer. <https://wikifarmer.com/es/floracion-polinizacion-y-poda-de-la-pina/>
- Osco, L. P., de Arruda, M. dos S., Marcato Junior, J., da Silva, N. B., Ramos, A. P. M., Moryia, É. A. S., Imai, N. N., Pereira, D. R., Creste, J. E., Matsubara, E. T., Li, J., & Gonçalves, W. N. (2020). A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160(December 2019), 97–106. <https://doi.org/10.1016/j.isprsjprs.2019.12.010>
- Pino, E. V. (2019). Los drones una herramienta para una agricultura eficiente: un futuro de alta tecnología. *Idesia*, 37, 75–84. <https://doi.org/10.4067/s0718-34292019005000402>

- Redagrícola. (2023). *Nuevos sensores ópticos en agricultura de precisión*. Redagrícola. <https://redagricola.com/nuevos-sensores-opticos-en-agricultura-de-precision/>
- Redmon, J., & Farhadi, A. (2018). *YOLOv3: an incremental improvement*. Arxiv. <https://arxiv.org/abs/1804.02767>
- Reményi, I., Domián, B., & Kárász, Z. (2023). Improved Object Detection by Utilizing the Image Stream. *Communications in Computer and Information Science*, 575–584. https://doi.org/10.1007/978-3-031-41774-0_45
- Rio-Alvarez, A., Marcos, P. G., González, P. P., Serrano-Pertierra, E., Novelli, A., Fernández-Sánchez, M. T., & González, V. M. (2024). Evaluating deep learning techniques for optimal neurons counting and characterization in complex neuronal cultures. *Medical and Biological Engineering and Computing*. <https://doi.org/10.1007/s11517-024-03202-z>
- Roch, E. M. (2024). *Optimización bayesiana: afinando hiperparámetros como un profesional*. Lovtechnology. <https://lovtechnology.com/optimizacion-bayesiana-afinando-hiperparametros-como-un-profesional/>
- Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1976). Monitoring vegetation systems in the great plains with erts. *Remote Sensing Center*, A20(1), 309–317. <https://doi.org/10.1021/jf60203a024>
- Sanewski, G. M., Bartholomew, D. P., & Paull, R. E. (2018). *The pineapple: botany, production and uses* (2nd editio). CABI. <https://doi.org/10.1079/9781786393302.0000>
- Sindagi, V. A., & Patel, V. M. (2018). A survey of recent advances in CNN-based single image crowd counting and density estimation. *Pattern Recognition Letters*, 107, 3–16. <https://doi.org/10.1016/j.patrec.2017.07.007>
- Taye, M. M. (2023). Theoretical understanding of convolutional neural network: concepts, architectures, applications, future directions. *Computation*, 11(3). <https://doi.org/10.3390/computation11030052>

- Tecnoagro. (2024). *Drones agrícolas: aplicaciones y beneficios en la gestión de cultivos*. Tecnoagro. <https://tecnoagro.com.mx/no.-177/drones-agricolas-aplicaciones-y-beneficios-en-la-gestion-de-cultivos>
- Torres, J. T. (2023). Deep learning for multispectral imaging based predictions in agriculture. *Concienciadigital*, 6, 75–87. <https://doi.org/10.33262/concienciadigital.v6i4.1.2734>
- Uriza, D. E. Á. (2011). Paquete tecnológico piña MD2 (Ananas comosus var. comosus): establecimiento y mantenimiento. *Centro de Investigacion Regional Golfo Centro*.
- Vina, A. (2024). *¿Qué es R-CNN? un resumen rápido*. Ultralytics. <https://www.ultralytics.com/es/blog/what-is-r-cnn-a-quick-overview>
- Wang, C., Zhou, J., Xu, C., & Bai, X. (2020). A deep object detection method for pineapple fruit and flower recognition in cluttered background. *Springer Nature*, 218–227.
- Wolfe, J. (2021). Aprendizaje profundo en aplicaciones de teledetección agrícola. *Telematica*.
- Xie, W., Noble, J. A., & Zisserman, A. (2018). Microscopy cell counting and detection with fully convolutional regression networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*. <https://doi.org/10.1080/21681163.2016.1149104>
- Xu, B., & Qiu, G. (2016). Crowd density estimation based on rich features and random projection forest. *IEEE*, 1–8. <https://doi.org/10.1109/WACV.2016.7477682>
- Yu, G., Cai, R., Luo, Y., Hou, M., & Deng, R. (2024). A-pruning: a lightweight pineapple flower counting network based on filter pruning. *Complex and Intelligent Systems*, 10, 2047–2066. <https://doi.org/10.1007/s40747-023-01261-7>

- Zafar, A., Aamir, M., Mohd Nawi, N., Arshad, A., Riaz, S., Alruban, A., Dutta, A. K., & Almotairi, S. (2022). A comparison of pooling methods for convolutional neural networks. *Applied Sciences*, *12*(8643). <https://doi.org/10.3390/app12178643>
- Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: a review. *Precision Agriculture*, *13*, 693–712. <https://doi.org/10.1007/s11119-012-9274-5>
- Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. In *Artificial Intelligence Review*. Springer Netherlands. <https://doi.org/10.1007/s10462-024-10721-6>

ANEXOS

Anexo 1: Ventajas y desventajas de los diferentes enfoques de agrupamiento en CNN.

Tipo de Pooling	Ventajas	Desventajas
Promedio máximo cerrado	Responsivo en estilo. Es adaptativo en el sentido de que la fracción de volumen puede fluctuar en función de las propiedades de la región de agrupación.	Produce parámetros de entrenamiento adicionales.
Promedio máximo mixto	Agrupamiento estocástico. Facilita el problema de evitar el sobreajuste.	Una vez aprendida, la proporción de la mezcla realmente no responde ni se adapta a los atributos de la región que se está integrando.
Agrupamiento piramidal	Flexibilidad para gestionar entradas de cualquier tamaño. Contenedores espaciales con múltiples niveles. Capacidad de respuesta a las escalas de imagen de una entrada.	El paso de entrenamiento de redes profundas implica una implementación compleja.
Agrupamiento estocástico	Procedimiento estocástico. Es concebible utilizar activaciones no máximas. Viabilidad de integrar cualquier método de regularización, incluyendo abandonos, aumento de datos, tangente de pérdida, etc. Parece que no hay ningún hiperparámetro para especificar. Menor complejidad computacional.	Complicado de interpretar. Extrañas a las palabras activaciones negativas. Debido a la falta de datos de entrenamiento, se produce un sobreajuste porque las activaciones fuertes funcionan principalmente en la actualización del proceso. Desafío de escalamiento.
Agrupamiento de árboles	Flexible para adaptarse a la naturaleza. Diferenciable en perspectiva tanto entre parámetros como entre insumos. Eficaz en los niveles inferiores de la red.	Ineficiente debido a las gruesas capas de la red.
Agrupamiento máximo fraccional	Método estocástico. Elección de la región de agrupamiento mediante pseudoaleatoriedad o aleatoriedad. Uso apropiado de ampliación de datos y selección pseudoaleatoria. La agrupación máxima fraccional superpuesta en lugar de disjunta resultó ser más eficiente.	La selección arbitraria de la zona de agrupación afecta significativamente el rendimiento del modelo además del aumento de datos. La agrupación máxima fraccional desarticulada conduce a una degradación significativa.
Piscina S3	Fácil de aprender y utilizar. Cálculos rápidos durante el entrenamiento. Se extruye en la medida de las distorsiones. Implemente el aumento de datos en los niveles de la capa de agrupación para brindarle un sólido rendimiento de generalización.	Dependiendo del diseño para el cual se esté empleando, el tamaño de la cuadrícula debe especificarse adecuadamente en cada capa de agrupamiento. Un tamaño de cuadrícula mayor potencialmente resulta en un mayor error de prueba.

Agrupamiento de promedios basado en rangos	En comparación con el agrupamiento máximo, aumenta considerablemente la carga computacional. Para tareas de reconocimiento de objetos, se implementa. Nos permite reutilizar el mapa de características de la red de convolución. Proporciona la oportunidad de entrenar sistemas de detección de objetos de principio a fin, acortando significativamente los períodos de prueba y entrenamiento.	Pueden surgir problemas de rendimiento al generar muchas regiones de interés. La frecuencia de cálculo no alcanza las expectativas. La capacitación de extremo a extremo, o capacitar cada aspecto del sistema de una sola vez, no es practicable, pero podría producir resultados mucho mejores.
--	---	--

Fuente: (Zafar et al., 2022)