UNIVERSIDAD AUTÓNOMA CHAPINGO

POSGRADO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

DENSIDAD DE PLANTAS DE MAÍZ EN PRESENCIA DE MALEZA CON REDES NEURONALES PROFUNDAS

TESIS

Que como requisito parcial para obtener el grado de:

MAESTRO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

Presenta:

CANEK MOTA DELFIN

Bajo la supervisión de:

GILBERTO DE JESÚS LÓPEZ CANTEÑS, DR.

Chapingo, Estado de México, noviembre de 2022

DENSIDAD DE PLANTAS DE MAÍZ EN PRESENCIA DE MALEZA CON REDES NEURONALES PROFUNDAS

Tesis realizada por el C. Canek Mota Delfin 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. EUGENIO ROMANTCHIK KRIUCHKOVA

AGRADECIMIENTOS

A la Universidad Autónoma Chapingo, que me ha brindado la oportunidad y el apoyo en el transcurso de mis estudios profesionales y de posgrado.

Al posgrado en Ingeniería Agrícola y Uso Integral del Agua (IAUIA), por conceder los medios para culminar mis estudios de maestría.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo financiero durante mis estudios de posgrado.

Al Dr. Gilberto de Jesús López Canteñs, por compartirme su entusiasmo, conocimientos y motivación para el desarrollo de este proyecto de investigación.

Al Dr. Irineo Lorenzo López Cruz, Dr. Eugenio Romantchick Kriuchkova y al M.I. Juan Carlos Olguín Rojas por respaldar y brindarme las herramientas necesarias para culminación de esta investigación.

A mis amigos y compañeros que me han prestado un gran apoyo moral y humano motivándome a siempre seguir adelante.

Al Ing. Yulinali Valente Morales, Ing. Mixtli Quetzali Cruz Mota, Ing. Ivette Itzel Avedaño Torres y Ing. Prisca Esperanza Jimenez Santiago por su apoyo en la validación del etiquetado de la base de datos empleada en esta investigación.

DATOS BIBLIOGRÁFICOS

DATOS PERSONALES

Nombre	Canek Mota Delfin
Fecha de nacimiento	17 de agosto de 1996
Lugar de nacimiento	Juan Rodríguez Clara, Veracruz
CURP	MODC960817HVZTLN07
Profesión	Ingeniero Mecánico Agrícola
Cédula profesional	12294326

DESARROLLO ACADÉMICO

Bachillerato:	Centro de bachillerato Tecnológico Agropecuario No. 85, Juan Rodríguez Clara, Veracruz
Licenciatura:	Universidad Autónoma Chapingo, Texcoco, Estado de México

CONTENIDO

LISTA DE CUADROS

LISTA DE FIGURAS

RESUMEN GENERAL

DENSIDAD DE PLANTAS DE MAÍZ EN PRESENCIA DE MALEZA CON REDES NEURONALES PROFUNDAS

La densidad de población de un cultivo se puede cuantificar mediante detección y el conteo remoto de plantas y está directamente correlacionado al rendimiento del cultivo. La obtención precisa de esta información ayuda a los agricultores a gestionar y controlar su producción. Sin embargo, las metodologías basadas en imágenes aéreas aún son un reto, debido a la complejidad de las condiciones del campo. En este contexto, se propuso el establecimiento de una base de datos que contiene imágenes aéreas del cultivo de maíz con malezas con el objetivo de implementar y evaluar la robustez de algoritmos de aprendizaje profundo para la detección y conteo de plantas de maíz en tales condiciones. Se realizaron diez misiones de vuelo, seis con una distancia de muestreo en tierra (GSD) de 0.33 cm/píxel en etapas vegetativas de V3 a V7 y cuatro con un GSD de 1.00 cm/píxel para etapas vegetativas V6, V7 y V8. Los detectores comparados fueron YOLOv4, YOLOv4 Tiny, YOLOv4 Tiny 3L, y las versiones de YOLOv5 s, m y l. Se evaluó cada detector en umbrales de intersección sobre la unión (IoU) de 0.25, 0.50 y 0.75 en intervalos de confianza de 0.05. Para niveles de confianza superiores a 0.35, YOLOv4 mostró mayor robustez en la detección ante los demás modelos. Considerando la moda de 0.3 para la confianza que maximiza la métrica F1 y el umbral IoU de 0.25 en todos los modelos, YOLOv5s obtuvo una precisión media promedio (mAP) de 73.1 % con una correlación R^2 de 0.78 y raíz del error cuadrático medio relativo (rRMSE) de 42 % en el conteo de plantas, seguido de YOLOv4 con mAP de 72.0 %, R^2 de 0.81 y rRMSE de 39.5 %. Las detecciones más bajas en todos los detectores se obtuvieron al evaluar las etapas vegetativas V6, V7 y V8 con GSD de 1.00 cm/píxel.

Palabras claves: Imágenes aéreas, CNN, Conteo de plantas, Maíz, Maleza, Detección

Tesis de Maestría en Ingeniería Agrícola y Uso Integral del Agua (IAUIA)
Universidad Autónoma Chapingo
Autor: Canek Mota Delfin
Director de Tesis: Gilberto de Jesús López Canteñs

GENERAL ABSTRACT

DENSITY OF CORN PLANTS IN PRESENCE OF WEEDS WITH DEEP NEURAL NETWORKS

The detection and counting of corn plants are directly correlated to crop yield. Accurate collection of this information helps farmers to manage and control their production. However, aerial imaging methodologies are still a challenge, due to the complexity of field conditions. In this context, it was proposed to establish a database containing aerial images of weed corn crops, with the aim of implementing and evaluating the robustness of deep learning algorithms for the detection and counting of corn plants under these conditions. Ten flight missions were performed, six with a ground sampling distance (GSD) of 0.33 cm/pixel in vegetative stages from V3 to V7 and four with a GSD of 1.00 cm/pixel for vegetative stages V6, V7 and V8. The detectors compared were YOLOv4, YOLOv4 Tiny, YOLOv4 Tiny 3L, and YOLOv5 s, m and l versions. Each detector was evaluated at thresholds of intersection over union (IoU) of 0.25, 0.50 and 0.75 at confidence intervals of 0.05. To confidence levels above 0.35, YOLOv4 showed greater robustness in detection against the other models. Considering the 0.3 mode for the confidence which maximizes the F1 metric and the 0.25 IoU threshold on all models, YOLOv5s obtained a mean average precision (mAP) of 73.1 % with an R^2 correlation of 0.78 and relative mean square error root (rRMSE) of 42 % in plant count, followed by YOLOv4 with mAP of 72.0 %, R^2 of 0.81 and rRMSE of 39.5 %. The lowest detections in all detectors were obtained when evaluating the vegetative stages V6, V7 and V8 with GSD of 1.00 cm/pixel.

Keywords: Aerial images, CNN, Plant count, Corn, Weeds, Detection

Thesis, Universidad Autónoma Chapingo
Author: Canek Mota Delfin
Advisor: Gilberto de Jesús López Canteñs

Capítulo 1

INTRODUCCIÓN GENERAL

1.1. Introducción

La producción de maíz (*Zea mays L.*) en México para el año 2020 superó los 27.4 millones de toneladas (SIAP, 2022). El maíz es uno de los cultivos más importantes del país desde el punto de vista alimentario, político, económico y social (Flores-Cruz, García-Salazar, Mora-Flores & Pérez-Soto, 2014). Los cereales forman parte crucial de la dieta humana y de la alimentación del ganado, por lo que lograr la autosuficiencia en su producción es una forma efectiva de promover la seguridad alimentaria (Panday, Pratihast, Aryal & Kayastha, 2020).

Promover la seguridad alimentaria requiere aumentar la producción agrícola para mantenerse al día con una población en crecimiento, y la agricultura de precisión sostenible es una forma prometedora de asegurar el suministro de productos agrícolas (Delgado, Short, Roberts & Vandenberg, 2019; Mizik, 2022). La agricultura de precisión (AP) implica el uso de tecnologías (Ejemplos: Teledetección, Tecnologías geo-espaciales, Internet de las cosas, Big Data, Inteligencia artificial, entre otras.) con el fin de aumentar la producción de los cultivos con la gestión correcta de los recursos (Bwambale, Abagale & Anornu, 2022; Sishodia, Ray & Singh, 2020).

El monitoreo es una de las cuatro etapas del ciclo (Monitorear-Analizar-Predecir-Actuar) de la AP (Malek, Dhiraj, Upadhyaya & Patel, 2022). Los avances tecnológicos han proporcionado sistemas de monitoreo de los cultivos cada vez más eficientes, como los sistemas aéreos pilotados a distancia (RPAS) equipados

con sensores ópticos que pueden adquirir datos espectrales de la vegetación con alta resolución espacial (~ 1 cm por píxel) (Borgogno-Mondino, 2018; W. Yang et al., 2020). Los sensores ópticos más usados en la teledetección agrícola son de luz visible (RGB), multiespectrales, hiperespectrales y térmicos, con los cuales se ha demostrado relación con parámetros fenológicos y fisiológicos de los cultivos, dando lugar al fenotipado de alto rendimiento basado en imágenes (Awais et al., 2022; Reddy Maddikunta et al., 2021). Bajo este enfoque, se han abordado diferentes problemas en la agricultura; detección de plantas y malezas, conteo de plantas, reconocimiento de áreas sin plantación, fenología y detección de fenotipos, predicción de rendimientos, entre otros (Awais et al., 2022; Osco, Marcato Junior et al., 2021).

Dentro del marco de la AP la densidad de plantación y la ubicación precisa de plantas son factores clave que proporcionan información relacionada con los recursos hídricos, fertilización, infestación de maleza, susceptibilidad a los patógenos, estimación de rendimiento, entre otros (Osco, Marcato Junior et al., 2021; Shi et al., 2022). Esta información es útil para los agricultores con la finalidad de planificar labores agrícolas y estrategias de mejoramiento (Sarabia, Aquino, Ponce, López & Andújar, 2020).

En la actualidad esté problema, se ha abordado a través del análisis de imágenes aéreas, sin embargo, aún se enfrenta a desafíos, como: la oclusión de plantas en cultivos densos, similitudes entre las plantas deseadas y la maleza, plantaciones no uniformes, problemas de sombras e iluminación, ruido en la adquisición de las imágenes, y muchos otros (Osco, Marcato Junior et al., 2021).

1.2. Objetivo general

Evaluar los algoritmos YOLO (You only look once) de última generación en la detección y el conteo de plantas de maíz *(Zea mays)* en imágenes RGB de alta resolución adquiridas por un sistema aéreo pilotado remotamente, considerando la presencia de maleza en el cultivo.

1.3. Objetivos particulares

- Crear una base de datos de imágenes aéreas geo-referenciadas, capturadas en diferentes etapas vegetativas del cultivo de maíz y diferentes condiciones de infestación de maleza.
- Evaluar los algoritmos YOLOv4, YOLOv4-tiny, YOLOv4-tiny-3l, YOLOv5 (s,m,l) en la detección y conteo de plantas, considerando la etapa vegetativa y la distancia de muestreo en tierra.

1.4. Organización de la tesis

Este documento se divide en tres capítulos: I) La introducción general, II) Revisión de literatura y III) el artículo científico.

La revisión de literatura describe la importancia del monitoreo del cultivo del maíz y las metodologías abordadas para el conteo y detección de plantas. Así mismo se describe el uso de las redes neuronales convolucionales, principalmente el funcionamiento de YOLO para la detección de objetos.

El capítulo III incluye el manuscrito correspondiente al trabajo de investigación cumpliendo los objetivos planteados.

Capítulo 2

REVISIÓN DE LITERATURA

2.1. El maíz

El maíz (*Zea mays L.*) es uno de los cultivos más importantes para México y el mundo, proporciona alimento a los humanos, animales, es materia prima para procesos industriales y es fuente de ingresos para poblaciones en desarrollo (Kennett et al., 2020; Ngoune Tandzi & Mutengwa, 2019; Rogers et al., 2021). Solo para México en el año 2020 se superaron los 27.4 millones de toneladas en la producción de maíz grano, ubicando a México como el octavo mayor productor mundial (SIAP, 2022).

El cultivo de maíz en México es uno de los sistemas de producción agrícola más grande y tiene una gran importancia económica, social y cultural (Rivera et al., 2022). Por lo tanto, es crucial garantizar su producción sostenida aprovechando su máximo rendimiento potencial en un entorno dado. De acuerdo con el trabajo de Ngoune Tandzi y Mutengwa (2019) el rendimiento potencial del maíz está dado por la combinación de diversos factores ambientales, el potencial del genotipo y las prácticas agrícolas llevadas a cabo por el agricultor.

Evaluar la idoneidad de las prácticas agrícolas y la relación del genotipo con el medio ambiente usando estadísticas precisas es de suma importancia para determinar y obtener los mayores rendimientos posibles del cultivo. El rendimiento se puede determinar calculando el peso del grano en la cosecha, área cosechada, densidad de plantación y el contenido de humedad del grano (Ngoune Tandzi & Mutengwa, 2019).

La densidad de plantación está directamente relacionada a las prácticas agrícolas, germinación de la semilla, entre otros factores. En la actualidad una de las tecnologías más prometedoras para determinar de forma precisa la densidad de plantación es la fotogrametría (Ngoune Tandzi & Mutengwa, 2019).

2.2. Teledetección en la agricultura de precisión

La agricultura de precisión (AP) es un enfoque basado en la información para la producción agrícola, diseñado para mejorar la calidad, la productividad y la rentabilidad de la producción a largo plazo, al tiempo que se minimiza el impacto en la vida silvestre y el medio ambiente (Raj, Appadurai & Athiappan, 2021).

La AP requiere tres fases, recopilación de datos a través de sensores, análisis de datos para una evaluación de las acciones a realizar y la implementación de esas acciones. La adopción de esta tecnología requiere el desarrollo e implementación de sistemas de posicionamiento e información geográfica, sensores remotos que permitan medir y asociar variables de interés agrícola, maquinaria automatizada y sistemas para la toma de decisiones (Ocampo & Catarina, 2018).

De acuerdo con Weiss, Jacob y Duveiller (2020) la teledetección consiste en la adquisición datos y su análisis sobre un objeto o fenómeno a distancia. Esto implica el uso de ciertos sensores montados en plataformas como, satélites, aviones, sistemas de aeronaves pilotados a distancias o sondas.

La teledetección es una técnica que en los últimos años se ha implementado de manera extensa en el área de la agricultura, ya que ofrece un medio para estimar o relacionar parámetros de interés en los cultivos de manera indirecta con sensores, especialmente sensores ópticos (Ngoune Tandzi & Mutengwa, 2019; Weiss et al., 2020), en varios estudios reportan su uso en diversas aplicaciones.

En una revisión extensa de Khanal, KC, Fulton, Shearer y Ozkan (2020), categorizaron el uso de la teledetección en la agricultura de acuerdo con

las aplicaciones reportadas, desde la planificación de la siembra hasta la poscosecha del cultivo, aplicaciones en el mapeo de topografía (elevación y pendiente), mapeo de drenajes acuíferos, mapeo de la temperatura y la humedad del suelo, evaluación de la compactación del suelo, emergencia y densidad de cultivos, monitoreo de la salud de los cultivos durante la temporada, monitoreo del estrés por nitrógeno, monitoreo de enfermedades del cultivo, identificación y clasificación de malezas, predicción de rendimiento, evaluación de la calidad del grano y por último la evaluación de residuos de cultivos.

La teledetección es una herramienta que tiene un gran potencial de aplicaciones en todos los aspectos que conlleva la agricultura de precisión, sin embargo aún existen limitaciones, necesidades y desafíos a resolver de esta tecnología, principalmente, la resolución de los datos (espacial, espectral y temporal) que está directamente relacionado a la plataforma y sensor óptico empleado, y la capacidad de procesar y relacionar los datos obtenidos con parámetros de interés agrícola (Sishodia et al., 2020).

2.2.1. Sistemas de aeronaves pilotados a distancia

La resolución espacial y temporal de una imagen está relacionada a la plataforma en la cual se monta el sensor de adquisición. La resolución espacial se refiere a la representación del suelo en un píxel, que puede variar según la altura del sensor en relación con el objeto a censar, tamaño del píxel del sensor, ángulo de visión y distancia focal del lente. En la agricultura de precisión y dependiendo la aplicación, se consideran resoluciones espaciales de 0.5 a 10 cm/píxel (Tsouros, Bibi & Sarigiannidis, 2019), en el caso de la detección de plantas de maíz se recomienda una resolución espacial de 0.3 cm/píxel para aplicaciones en el conteo de plantas (Velumani et al., 2021). La resolución temporal está relacionada a la frecuencia con la cual una misma área es muestreada (Sishodia et al., 2020).

Bajo el enfoque agricultura de precisión, cada vez se requiere la recopilación de mayor información y de manera más eficiente, llegando al nivel de la planta, principalmente para cultivos de corto periodo productivos como los cereales.

De acuerdo con las recientes revisiones de literatura de Osco, dos Santos de Arruda et al. (2021), Sishodia et al. (2020) y Weiss et al. (2020) una de las plataformas más prometedoras hasta el momento son las aeronaves pilotadas a distancia para aplicaciones en las cuales se requiere alta resolución de imágenes (espacial, espectral y temporal).

En la agricultura los RPAS, vehículos aéreos no tripulado (UAV) o comúnmente llamados drones, se pueden clasificar de acuerdo al principio de vuelo aerodinámico, carga útil, autonomía de vuelo, entre otros. Utilizando el principio de vuelo aerodinámico los principales RPAS utilizados en la agricultura son de ala fija y ala rotatoria (multirrotor), los RPAS de ala fija logran la sustentación gracias a la velocidad aerodinámica hacia delante y los de ala rotatoria mediante palas posicionadas paralelas al suelo impulsadas por un rotor (Cabreira, Brisolara & Ferreira Jr., 2019; del Cerro, Cruz Ulloa, Barrientos & de León Rivas, 2021; Panday et al., 2020). En la Figura 2.1 se muestran algunos ejemplos de RPAS comerciales utilizados comúnmente en la agricultura.

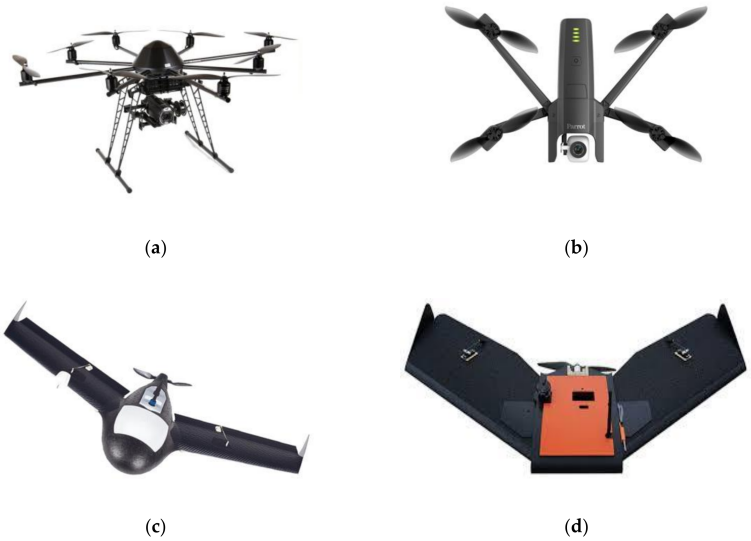


Figura 2.1: *Ejemplos de RPAS comerciales utilizados en la agricultura de precisión: (a) MK Okto XL 2 de ocho rotores (b) Parrot Anafi de cuatro rotores, (c) Gatewing X100 y (d) Tuffwing Mapper. Recuperado de del Cerro, Cruz Ulloa, Barrientos y de León Rivas (2021)*

Aunque ambos tipos de RPAS se pueden utilizar con fines de teledetección en la agricultura, Tsouros et al. (2019) en su revisión de 100 artículos científicos

relacionados al uso de RPAS en la agricultura de precisión encontraron que solo el 22 % utilizó RPAS de ala fija y el 72 % multirrotor, esto debido a que los trabajos presentados en la literatura no realizan mapeos de grandes áreas y por lo tanto los RPAS de ala fija no fueron requeridos. En el Cuadro 2.1 se presentan las principales ventajas y desventajas de los RPAS de ala fija y multirrotor recopiladas de Cabreira et al. (2019), del Cerro et al. (2021) y Panday et al. (2020)

Cuadro 2.1: *Ventajas y desventajas de los RPAS de ala fija y multirrotor*

Clasificación	Ventajas	Desventajas
Multirrotor	▪ Vuelos estacionarios ▪ Fácil de operar ▪ Capacidad de maniobra ▪ Costo relativamente bajo ▪ Despegue y aterrizaje vertical ▪ Gran rendimiento a bajas velocidades ▪ Permiten vuelos a baja altura con mínimo riesgo	▪ Velocidades bajas ▪ Poca eficiencia energética ▪ Difícil mantenimiento
Ala fija	▪ Alta resistencia y altas velocidades ▪ Cubrir grandes áreas en cada vuelo ▪ Mayor carga útil ▪ Mayor eficiencia energética ▪ Fácil mantenimiento	▪ Requieren pista de despegue o aterrizaje ▪ Mayores regulaciones legales ▪ Accesibilidad limitada ▪ Dificíl de maniobrar

La elección correcta de una plataforma RPAS depende de las necesidades del agricultor y de las ventajas que esta ofrezca, en general son preferibles los RPAS multirotor por la facilidad que ofrece respecto a la planeación de vuelo al seguir rutas específicas, capacidad de maniobra y costo relativamente bajos (Tsouros et al., 2019).

Otro aspecto importante mencionado en la literatura respecto al uso de RPAS multirotor es la planificación de vuelos, en el trabajo de Young, Koontz y Weeks (2022) se resolvió el problema de optimización de la toma de imágenes aéreas, se evaluó la calidad de la detección de árboles y se descubrió que, independiente de la altura de vuelo se obtienen mejores resultados con una superposición frontal/lateral del 90/80 % y la vista de la cámara al nadir. La reconstrucción del modelo digital de superficie fue realizado con el software Metashape versión 1.6.5 (Agisoft, LLC) y aunque optimización de los parámetros se realizó para la detección de árboles, estos sirven de referencia para la adquisición de imágenes con el mismo objetivo de detección de plantas.

2.2.2. Sensores ópticos

Existen una gama de sensores ópticos que se han estado empleando en aplicaciones del monitoreo de la vegetación, cámaras de color en espectro visible rojo-verde-azul (RGB), multiespectrales, hiperespectrales y últimamente los sensores térmicos (Panday et al., 2020). Estos sensores permiten capturar información del espectro electromagnético en diferentes longitudes de onda y se ha demostrado que la reflectancia de la vegetación en ciertas longitudes de onda brindan información relacionada a aspectos de interés agrícola (Figura 2.2) (Sishodia et al., 2020).

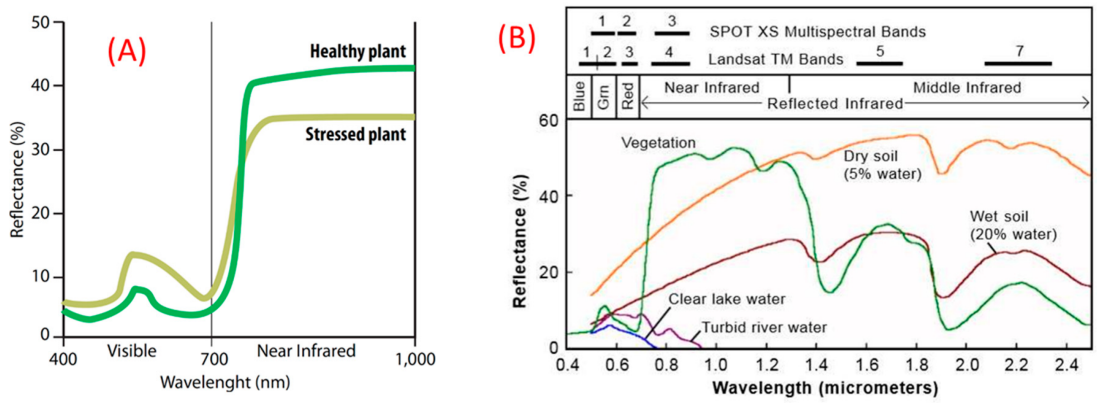


Figura 2.2: *Plantas y su espectro de reflectancia. (A) Planta sana y estresada, (B) reflectancia típica del suelo, agua y vegetación. Recuperado de Sishodia, Ray y Singh (2020)*

Dependiendo el tipo de sensor óptico, estos pueden capturar información en diferentes longitudes de onda. Los sensores RGB son los más comunes empleados en la agricultura de precisión y permiten capturar información en el espectro visible (400-700 nm), sin embargo una de las principales desventaja de este tipo de sensor es que no permiten analizar una amplia variedad de parámetros relacionados con la vegetación (Panday et al., 2020; Tsouros et al., 2019).

Para relacionar aspectos intrínsecos de las plantas con interés agrícola (contenido de agua, carbohidratos, azúcares, entre otros.), es necesario la utilización de sensores que permitan adquirir información en diferentes longitudes de onda aparte de la luz visible, estos sensores son los multiespectrales e hiperespectrales. Los sensores multiespectrales se diferencian de los hiperespectrales principalmente por el ancho y el número de bandas espectrales que estos pueden capturar. Los multiespectrales están formadas por hasta 20 bandas espectrales y los hiperespectrales hasta por cientos contiguos (Panday et al., 2020; Tsouros et al., 2019; Xue & Su, 2017).

Aunque los sensores multiespectrales e hiperespectrales brindan mayor información, aún se limita su uso debido al costo y a la complejidad respecto al análisis de muchas bandas espectrales. En el caso del conteo de plantas o determinación de densidad de plantas, en la literatura se reporta mayormente el uso de sensores RGB, por mencionar unos ejemplos; en el trabajo de Oh et al. (2020) implementan sensores RGB para el recuento de plantas de algodón, del mismo modo Bayraktar, Basarkan y Celebi (2020) emplean una cámara RGB para el recuento y detección de plantas ornamentales y Machefer, Lemarchand, Bonnefond, Hitchins y Sidiropoulos (2020) utilizan sensores RGB para el conteo de plantas de patatas y lechuga.

2.3. Detección y conteo de plantas

Como ya se mencionó desde la introducción la detección y conteo de plantas está estrechamente relacionada con la estimación de rendimientos, infestación de maleza, recursos hídricos, calidad de siembra y otros parámetros que permi-

ten a los agricultores planear labores agrícolas y estrategias de mejoramiento. Diferentes estrategias de procesamiento de imágenes se han empleado para la detección y conteo de plantas de diferentes cultivos, en el Cuadro 2.2 se muestran ejemplos respecto a las metodologías y los resultados obtenidos en el conteo de plantas de los últimos cuatro años.

Cuadro 2.2: *Métodos empleados en el conteo de plantas*

Cultivo	**Sensor**	**Método**	**Resultado**	**Referencia**
Maíz, Remolacha azucarera y Girasol	RGB	DL, Métodos clásicos e Hibrido	DL Precisión/rRMSE: - Maíz $\approx$ 80/<20 % - Remolacha azucarera >80/<10 % - Girasol $\approx$ 80/<40 %	(Daubige et al., 2021)
Espiga de trigo	RGB	DL	- MAE = 3.85 - RMSE = 5.19	(Khaki, Safaei, Pham & Wang, 2022)
Maíz	RGB	Híbridos	Etapa V2-V4 - R^2 = 0.98 - RMSE = 7.7 - rRMSE = 2.6 %	(Che et al., 2022)
Plátano	Cámaras multiespectrales	DL	- Precisión = 89 % - Recuerdo = 97 % - F1 = 0.93	(Aeberli, Johansen, Robson, Lamb & Phinn, 2021)
Arroz	RGB	DL	- Precisión = 93 % - R^2 = 0.94	(Wu et al., 2019)
Algodón	Cámara hiperespectral	Métodos clasicos	- Precisión = 84.1 % - MAPE = 6.8 %	(Feng, Sudduth, Vories & Zhou, 2019)

DL: Aprendizaje profundo, Híbrido: Métodos clásicos + aprendizaje automático

Cultivo	Sensor	Método	Resultado	Referencia
Maíz	RGB	DL	- Precisión = 85.6 % - Recuperación = 90.5 %	(Osco, dos Santos de Arruda et al., 2021)
Palmeras	RGB	DL	- mAP $\approx$ 80 %	(Ammar, Koubaa & Benjdira, 2021)

DL: Aprendizaje profundo, Híbrido: Métodos clásicos + aprendizaje automático

Existen diferentes enfoques que se han abordado en la literatura para la detección y el conteo de plantas, principalmente se resumen en: Métodos tradicionales o clásicos, Híbridos (Métodos tradicionales + aprendizaje automático o aprendizaje profundo) y aprendizaje profundo.

Las metodologías tradicionales son aquellos que se basan en operaciones morfológicas sobre las imágenes, estas operaciones consisten en la manipulación del color determinando índices y en la extracción manual de rasgos descriptivos de la imagen para su posterior segmentación y detección de las plantas. Esta metodología se basa en los supuestos de que las plantas son verdes y se pueden diferenciar del suelo, las plantas se encuentran en una disposición de hileras relativamente espaciadas y hay cierta uniformidad de las plantas a detectar (Daubige et al., 2021).

En la metodología basada en aprendizaje profundo, se utilizan las redes neuronales convolucionales (CNN) con la finalidad de extraer de imágenes rasgos descriptivos de manera automática. De acuerdo con Osco, Marcato Junior et al. (2021) el análisis de imágenes adquiridas por RPAS actualmente depende del aprendizaje profundo, en una revisión de 232 artículos se encontró que el 53.9 % aplican CNN para la detección de objetos en imágenes aéreas, donde el 26.4 % tiene un enfoque agrícola. El uso de CNN en la detección de plantas principalmente depende de la capacidad de estos algoritmos en extraer información de datos de alta dimensionalidad de manera automática y su capacidad de generalización.

Existen una amplia variedad de redes convolucionales con aplicaciones en la detección de plantas que cada vez obtienen mejores resultados, los algoritmos comúnmente mencionados en la literatura son versiones o variantes de RCNN, YOLO, UNET y MobileNet, es de mencionar que en algunos casos se combina técnicas clásicas de procesamiento de imágenes con CNN obteniendo una ligera mejora frente a un algoritmo CNN (Aeberli, Johansen, Robson, Lamb & Phinn, 2021; Daubige et al., 2021; Liu, Sun, Li & Iida, 2020; Wu et al., 2019).

Debido a la complejidad de los ambientes agrícolas, en general las metodologías tradicionales no son robustas a presencia de maleza entre plantas, oclusión y cambios de perspectiva en el entorno (Liu et al., 2020; Varela et al., 2018). El Cuadro 2.3 lista las metodologías empleadas en el conteo y detección de plantas de maíz, sus limitaciones y fortalezas mencionadas en cada investigación de los últimos cinco años.

Cuadro 2.3: *Detección y conteo de plantas de maíz*

Título	**Método**	**Observaciones**	**Referencia**
Digital Counts of Maize Plants by Unmanned Aerial Vehicles (UAVs)	- Decorrelación de color RGB - Umbralización en el espacio de color HSV y Lab	- El conteo de plantas disminuye en presencia de maleza e imágenes borrosas - Alto desempeño en etapas V3-V5	(Gnädinger & Schmidhalter, 2017)
Early-Season Stand Count Determination in Corn via Integration of Imagery from Unmanned Aerial Systems (UAS) and Supervised Learning Techniques	- Indices de color RGB - Filtro Savitzky-Golay - Descriptores geométricos - Árboles de decisiones	- La superposición de las plantas causa subestimación - Problema con presencia de maleza entre plantas - Robusto a maleza entre filas	(Varela et al., 2018)

Título	Método	Observaciones	Referencia
Corn Plant Counting Using Deep Learning and UAV Images	- Segmentación semántica (UNET)	- Mejores resultados en la etapa V4 - Se recomienda probar el método considerando maleza y paja	(Kitano, Mendes, Geus, Oliveira & Souza, 2019)
Application of Color Featuring and Deep Learning in Maize Plant Detection	- Indices de color RGB - Aprendizaje profundo (YOLOv3 y YOLOv3-tiny)	- YOLO es robusto a la presencia de maleza - Se limita a imágenes adquiridas por encima de las plantas de forma manual	(Liu, Sun, Li & Iida, 2020)
Digital Count of Corn Plants Using Images Taken by Unmanned Aerial Vehicles and Cross Correlation of Templates	- Espacio de color CIELAB - Correlación cruzada normalizada entre plantillas representativas de plantas	- Se limitó a un área sin maleza	García-Martínez et al., 2020
Improved crop row detection with deep neural network for early-season maize stand count in UAV imagery	- Segmentación por índices de color RGB - Segmentación semántica (Mask R-CNN)	- No se puede distinguir entre maleza y el cultivo en la etapa con segmentación RGB - Se propone la generación de máscaras de manera automática - Mejores resultados con el algoritmo Mask R-CNN	(Pang et al., 2020)

Título	Método	Observaciones	Referencia
A Convolutional Neural Network-Based Method for Corn Stand Counting in the Field	- Aprendizaje profundo (YOLOv3 y YOLOv3-tiny) - Filtro de kalman	- Cámara posicionada a 0.5m del suelo - Aplicable a etapas V1-V2 - Robusto a maleza	(L. . Wang, Xiang, Tang & Jiang, 2021)
Estimates of Maize Plant Density from UAV RGB Images Using Faster-RCNN Detection Model: Impact of the Spatial Resolution	- Detección por regiones con Faster-RCNN - Aumento de resolución de imagen con Redes Antagónicas Generativas	- Área de estudio desyerbado - Sensibilidad al tamaño de planta y GSD - Se recomienda YOLOv5, YOLOv4 y Retina Net como algoritmos de mejora	(Velumani et al., 2021)

2.4. El aprendizaje de máquina y aprendizaje profundo

El aprendizaje de máquina (ML) forma parte de la inteligencia artificial, el cual consiste en el diseño e implementación de algoritmos que permitan “aprender” relaciones y patrones a partir de ejemplos de manera automática (Janiesch, Zschech & Heinrich, 2021). La Figura 2.3 representa la taxonomía de la inteligencia artificial.

De acuerdo con Reza Keyvanpour y Shirzad (2022) los algoritmos aplicados al procesamiento de imágenes con ámbito agrícola se clasifican en métodos tradicionales y de aprendizaje profundo, donde la principal diferencia es el procedimiento de extracción de información de los datos (Figura 2.4).

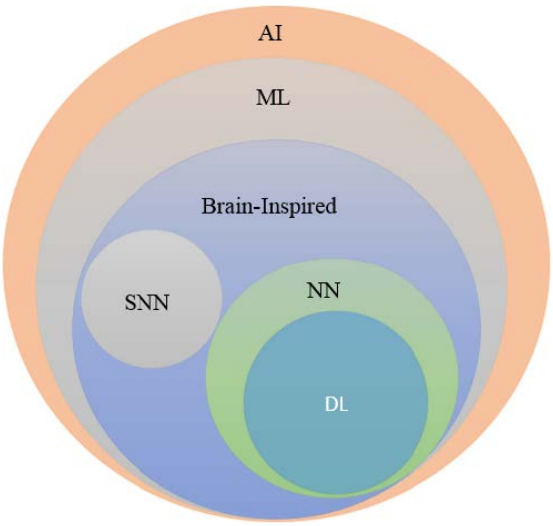


Figura 2.3: *Taxonomía de la inteligencia artificial. AI: Inteligencia artificial, ML: Aprendizaje de máquina, SNN: Redes neuronales de impulso, NN: Redes neuronales artificiales y DL: Aprendizaje profundo. Recuperado de Alom et al. (2018)*

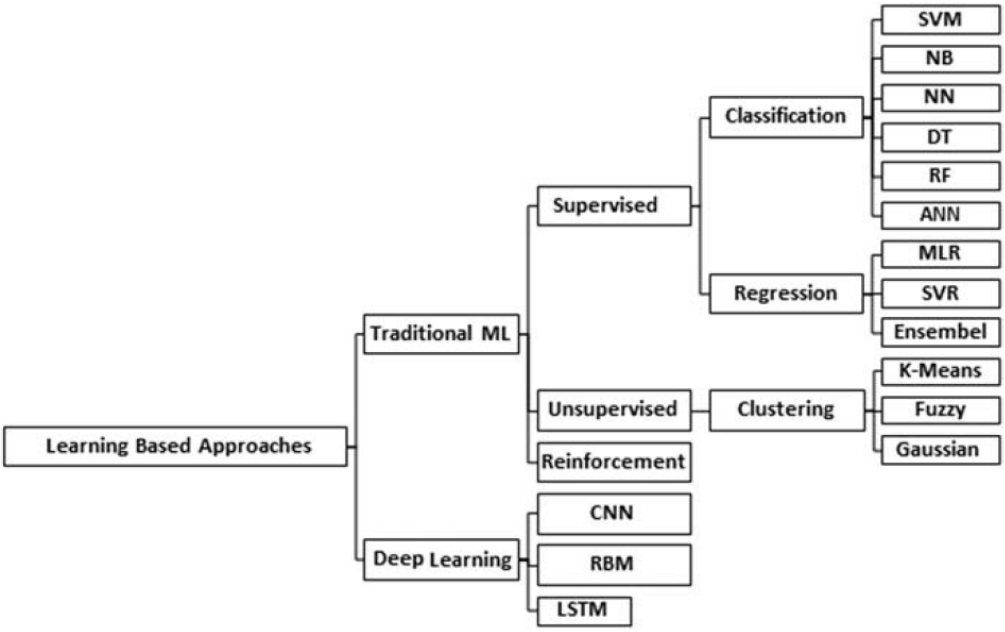


Figura 2.4: *Clasificación de los algoritmos de aprendizaje de máquina aplicados al procesamiento de imágenes con ámbito agrícola. Recuperado de Reza Keyvanpour y Shirzad (2022)*

Aunque en principio, los algoritmos se clasifican en tradicionales y de aprendizaje profundo, en general estos se pueden agrupar de acuerdo a la estrategia utilizada para su proceso de aprendizaje de la siguiente manera; en donde actualmente el DL tiene participación en todas las estrategias que se mencionan:

- **Supervisados:** Los cuales requieren de una base de datos etiquetada con ejemplos de la tarea a resolver (Clasificación o regresión).
- **No supervisados:** Aprenden relaciones entre los datos de entrada para la formación de grupos (clustering) sin la necesidad de una base de datos etiquetada.
- **Aprendizaje por refuerzo:** Los cuales se aplican principalmente al área de control y recomendación, donde el aprendizaje se da por una función de recompensas o penalización de acuerdo a las acciones tomadas por el algoritmo.

A partir del origen de las redes neuronales artificiales y el algoritmo para su entrenamiento "Back-propagation", surgió un subcampo del ML denominado aprendizaje profundo (DL), el cual consiste en la estimación de parámetros (pesos sinápticos) de un modelo neuronal de más de dos capas ocultas, con el fin realizar una tarea especifica Alom et al. (2018) y Janiesch et al. (2021).

La principal característica del DL es que los modelos neuronales reciben de entrada datos sin procesar y estos descubren una representación automática (Feature extraction) de los datos para resolver una tarea especifica (Janiesch et al., 2021). La Figura 2.5 representa los principales características del enfoque de programación explícita, el aprendizaje de máquina y el aprendizaje profundo.

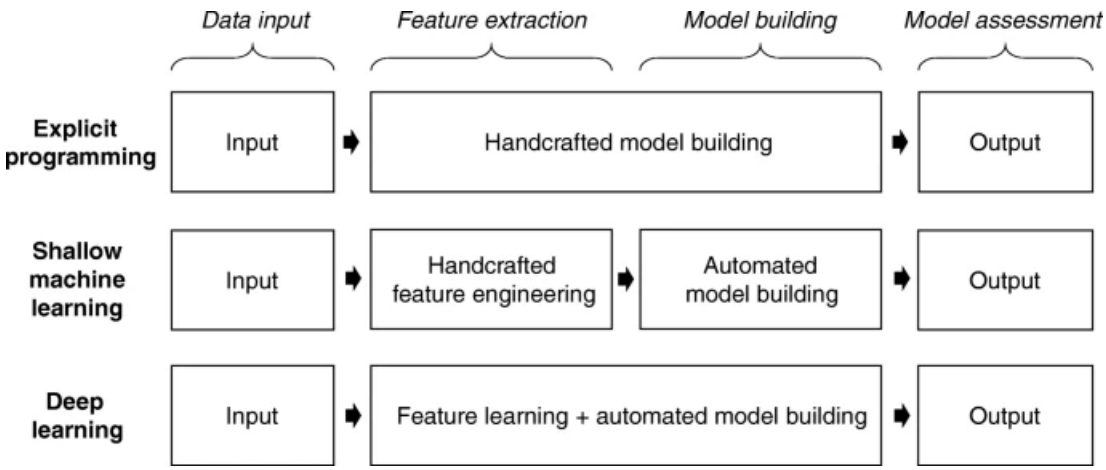


Figura 2.5: *Construcción de modelos analíticos. Recuperado de Janiesch, Zschech y Heinrich (2021)*

Con el uso de los RPAS para el monitoreo de cultivos agrícolas ha surgido la necesidad de técnicas de procesamiento de imágenes cada vez más robustas. La llegada de los algoritmos CNN en el marco de la visión por computadora ha permitido el análisis de estas imágenes de manera relativamente más sencilla, presentando mejores resultados que los algoritmos clásicos de procesamiento de imágenes (Gu et al., 2018; Osco, Marcato Junior et al., 2021).

2.4.1. Redes Neuronales Convolucionales

Las redes neuronales convolucionales están inspiradas en el sistema de la corteza visual animal, generalmente constituidas por la agrupación de la operación matemática lineal entre dos matrices llamada convolución, capas de no linealidad, capas de agrupación y capas completamente conectadas, donde cada una de ellas desempeña una función especifica (Albawi, Mohammed & Al-Zawi, 2017; Alzubaidi, Mostaghimi, Si, Swietojanski & Armstrong, 2022).

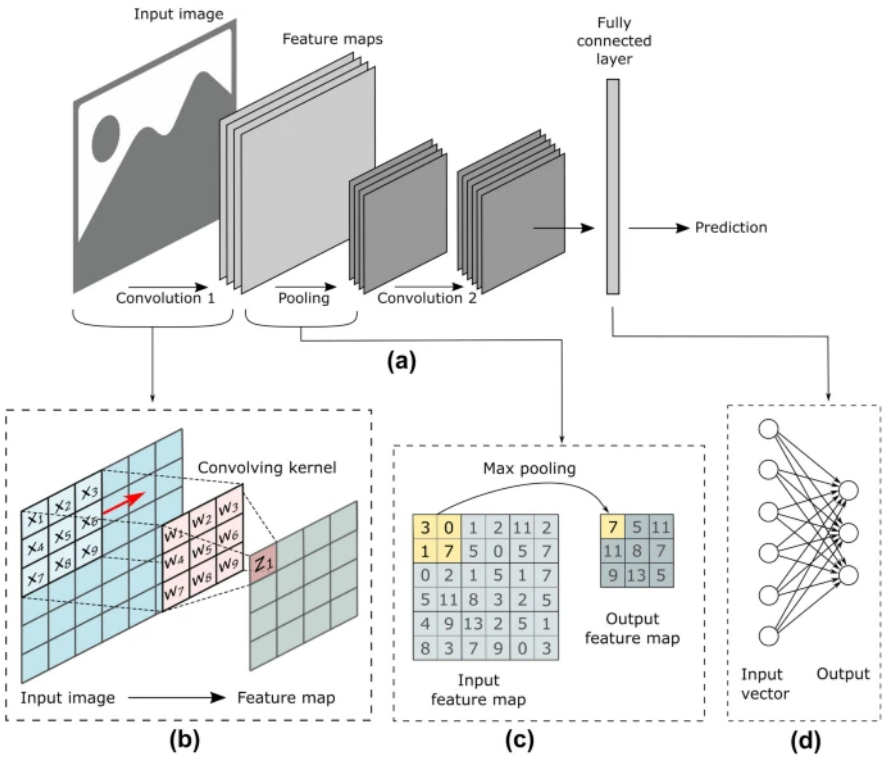


Figura 2.6: *Estructura de una red neuronal convolucional genérica (a). (b) Operación de convolución en una imagen. (c) Operación de agrupación máxima. (d) Capa completamente conectada. Recuperado de Alzubaidi, Mostaghimi, Si, Swietojanski y Armstrong (2022)*

A continuación se describe cada una de las capas que conforman una arquitectura básica de una CNN.

Capa convolucional: Son filtros o kernels agrupados con la finalidad de extraer características de bajo nivel de una entrada, cada filtro con parámetros aprendibles pasa por una capa de activación con funciones de no linealidad o linealidad (sigmoide, tangente hiperbólica, Softmax, Unidad lienal rectificada, entre otras) (Alom et al., 2018). La Ecuación 2.1 representa la operación de convolución de manera general (Alom et al., 2018).

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{2.1}$$

Donde x_j^l es la salidad de la capa convolucional, x_i^{l-1} es la salida de la capa convolucional anterior, k_{ij}^l representa el kernel o filtro con los parámetros aprendibles, M_j representa el mapa de características de entrada y b_j^l representa los sesgos de la capa convolucional.

Capa de agrupación: La principal función de esta capa es de realizar un muestreo descendente en los mapas de características de la capa anterior, rediciendo la dimensión espacial de la entrada. Esto disminuye el número de parámetros entrenables, reduciendo el costo computacional y controlar el sobre ajuste. Las principales operaciones de agrupación son; Average Pooling, Max-Pooling, Mixed Pooling, Lp Pooling, Stochastic Pooling y Spatial Pyramid Pooling (Gholamalinezhad & Khosravi, 2020).

Capas completamente conectadas: Las capas completamente conectadas generalmente hacen referencia a un perceptron multicapa que acondiciona la salida a una tarea de regresión o clasificación. La Ecuación 2.2 representa la operación básica de un perceptron (Alom et al., 2018).

$$y_k = \varphi\left(\sum_{j=1}^{m} w_{kj} x_j + b_k\right) \tag{2.2}$$

Donde w_{kj} representa los pesos sinápticos, x_j la entrada de datos, b_k los sesgos y φ la función de activación lineal o no lineal.

2.4.2. Detección de objetos

La detección es una de las principales tareas de la visión por computadora, que consiste en la localización y clasificación los objetos presentes en imágenes (Xiao et al., 2020), de la cual se derivan dos temas de investigación: detección general de objetos bajo un marco unificado para simular la visión humana y aplicaciones de detección en escenarios específicos (Zou, Shi, Guo & Ye, 2019). Con el surgimiento de las redes neuronales convolucionales y a partir del desarrollo del algoritmo de clasificación AlexNet (Krizhevsky, Sutskever & Hinton, 2012), la detección de objetos se divide en dos métodos: los tradicionales y basados en aprendizaje profundo, el cual comenzó a crecer con la contribución de Girshick, Donahue, Darrell y Malik (2014). En la Figura 2.7 se presentan algunos de los algoritmos más relevantes durante la evolución de la detección de objetos hasta el año 2020.

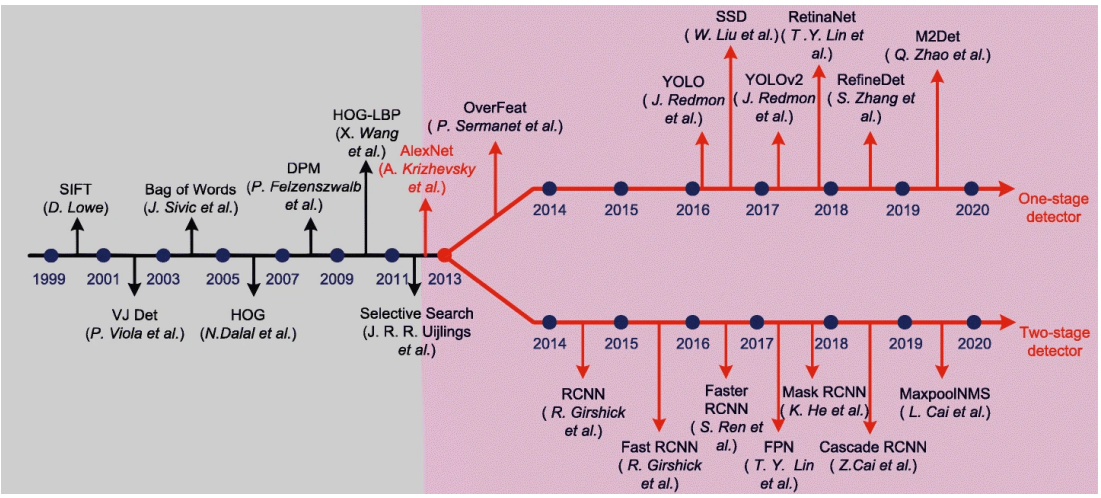


Figura 2.7: *Evolución de la detección de objetos. Recuperado de Xiao et al. (2020)*

De acuerdo con la literatura citada en el Cuadro 2.3, la detección y conteo de plantas de maíz bajo el método de aprendizaje profundo se ha abordado mediante dos enfoques, detección de objetos por regiones dibujando cuadros delimitadores para cada objeto detectado y segmentación semántica que realiza la detección de objetos segmentando las regiones pertenecientes a un mismo objeto a nivel de píxel.

La Figura 2.8 representa las principales redes neuronales convolucionales empleadas en la detección y segmentación de objetos, sin embargo, para fines de esta investigación, solo se abordarán los algoritmos de detección de objetos debido a que los basados en segmentación requieren un proceso adicional para determinar las coordenadas de cada instancia segmentada y además el proceso de etiquetado conlleva a realizar polígonos en las plantas para enmarcar los píxeles pertenecientes a las mismas.

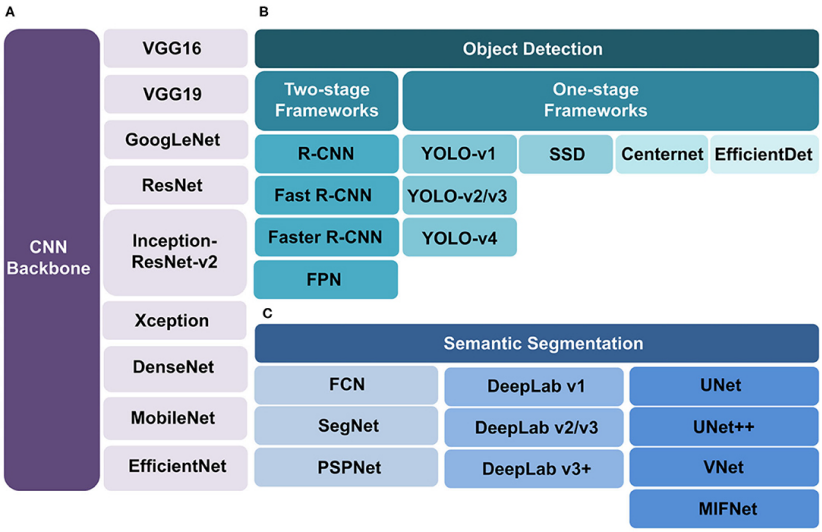


Figura 2.8: *Redes neuronales convolucionales bajo el enfoque de detección y segmentación semántica. Recuperado de R. . Yang y Yu (2021)*

Dentro del desarrollo de los métodos basados en aprendizaje profundo se derivan tres tipos de implementaciones, detectores de una sola etapa, de dos etapas y últimamente basados en transformadores (Xiao et al., 2020; Zaidi et al., 2022; Zou et al., 2019).

Los detectores de dos etapas consisten en la generación de regiones de interés (RoI) usando una red “Region Proposal Network (RPN)” y en la segunda etapa predicen los objetos y los cuadros delimitadores de las regiones propuestas usando otra capa de clasificación (RCNN, Fast RCNN, Faster RCNN, entre otros). En cambio los detectores de una sola etapa realizan la detección y clasificación en un solo paso de la red, retornando los cuadros delimitadores de los objetos y las probabilidades de cada clase (YOLO, RetinaNet, SSD, entre otros) (Diwan, Anirudh & Tembhurne, 2022).

La elección de un algoritmo de detección no es una tarea trivial, ya que cada uno tiene ventajas y desventajas, en cuanto a precisión de la detección y tiempo de inferencia, los detectores de dos etapas superan a los de una etapa en cuanto a Precisión, sin embargo el tiempo de inferencia es mucho mayor y presentan una arquitectura más compleja (Diwan et al., 2022).

2.4.3. YOLO (You Only Look Once)

En el área de teledetección agrícola una de las redes neuronales últimamente utilizada es YOLO, un detector de una sola etapa que en sus últimas versiones ha logrado mejorar su precisión y el tiempo de inferencia superando a otros algoritmos de detección haciendo posibles aplicaciones en tiempo real (Diwan et al., 2022; Singh, Thakur, Goyal & Gupta, 2022; L. . Wang, Xiang, Tang & Jiang, 2021; Q. . Wang et al., 2022).

Como ya se mencionó YOLO es una red neuronal convolucional entrenable de principio a fin. A lo largo del tiempo se han presentado diferentes versiones de YOLO, en este caso se describe la etapa de detección (Head) YOLOv3 y su entrenamiento, la cual se emplea en versiones posteriores (Figura 2.9). La descripción se basa en los trabajos de Li, Huang, Ai, Yi y Xie (2021), Nepal y Eslamiat (2022), Redmon, Divvala, Girshick y Farhadi (2016) y Redmon y Farhadi (2018).

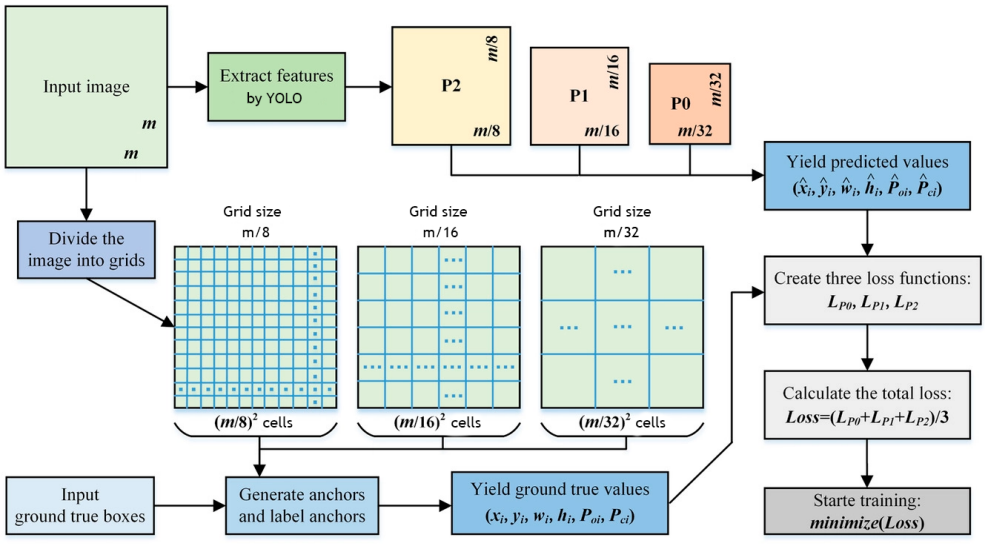


Figura 2.9: *Diagrama del proceso de aprendizaje de YOLOv3. Modificado de Li, Huang, Ai, Yi y Xie (2021)*

YOLOv3 basa su funcionamiento en la división de una imagen de tamaño m x m en rejillas de tamaño S x S, que corresponden a las salidas de predicción (cabezales de predicción) donde cada celda de la rejilla es responsable de la detección de un objeto. Para el caso del cabezal de detección YOLOv3 se cuenta con tres cabezales de detección a diferentes escalas, es decir, la imagen de entrada se divide en tres escalas diferentes, donde S vale m/8, m/16 y m/32.

Cada celda de cada cabezal de detección predice B cuadros delimitadores (anchors boxes) en relación a 3 anchors boxes de tamaño constante para cada cabezal de detección, determinados con el algoritmo de agrupamiento no supervisado K-Means con todas las etiquetas realizadas manualmente (ground true boxes) representando el tamaño del objeto a detectar. Para cada cuadro delimitador se predicen 4 coordenadas (t_x, t_y, t_w y t_h), las primeras 2 relacionadas al centro del cuadro delimitador en relación a la celda y las últimas relacionadas al tamaño de cada cuadro delimitador. La Figura 2.10 representa de manera gráfica la determinación real de cada cuadro delimitador.

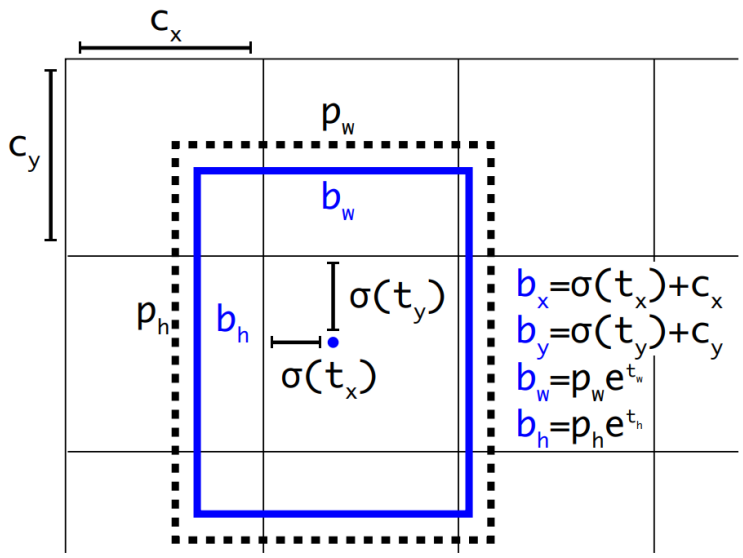


Figura 2.10: *Determinación de la ubicación y tamaño real de cada cuadro delimitador. Donde: b_x y b_y corresponde al centro del cuadro delimitador predicho, b_w y b_h al tamaño, p_w y p_h al tamaño de los cuadros delimitadores determinados por K-Means y la función σ hace referencia a la función sigmoide. Recuperado de Redmon y Farhadi (2018)*

Además de los cuadros delimitadores, cada celda predice la probabilidad de cada cuadro delimitador (P_o) y un vector de probabilidades correspondientes a cada clase de tamaño C (P_c), donde C corresponde al número de clases de objetos a detectar. Las detecciones predichas están condicionadas por el valor de $P_o IoU$ donde IoU corresponde a la intersección sobre la unión de las predicciones entre las etiquetas verdaderas.

El entrenamiento de YOLO se basa en la actualización de los pesos sinápticos, mediante la retropropagación del error entre las predicciones y las etiquetas verdaderas mediante la minimización de la función de coste definida en la Ecuación 2.2 obtenida de Li et al. (2021).

$$
\begin{aligned}
LP_l &= 5\sum_{i=0}^{s^2}\sum_{j=0}^{3} 1_{ij}^{obj}\left[(x_i-\widehat{x_i})^2+(y_i-\widehat{y_i})^2\right] \\
&+ 5\sum_{i=0}^{s^2}\sum_{j=0}^{3} 1_{ij}^{obj}\left[(w_i-\widehat{w_i})^2+\left(h_i-\widehat{hi}^2\right)^2\right] \\
&+ \sum_{i=0}^{s^2}\sum_{j=0}^{3} 1_{ij}^{obj}\left(P_{oi}-\widehat{P_{oi}}^2\right) + 0{,}5\sum_{i=0}^{s^2}\sum_{j=0}^{3}\left(1-1_{ij}^{obj}\right)\left(P_{oi}-\widehat{P_{oi}}^2\right)^2 \\
&+ \sum_{i=0}^{s^2} 1_{ij}^{obj}\left(P_{ci}-\widehat{P_{ci}}\right)^2,
\end{aligned}
\tag{2.3}
$$

Donde $1_{ij}^{obj}=1$ si el objeto aparece en el j-ésimo predictor de cuadros delimitadores en la cuadrícula i, caso contrario $1_{ij}^{obj}=0$.

A continuación se describen dos de las últimas versiones de YOLO. YOLOv4 y YOLOv5 se componen principalmente de tres partes: extractor de funciones (Backbone), agregación de funciones para detección a diferentes escalas (Neck) y predicción/regresión (Head); la diferencia entre estas versiones de YOLO se basa en la modificación de estas tres partes, como cambios en la función de pérdida, función de activación y técnicas de regularización, entre otros.

2.4.4. YOLOv4

La arquitectura la red YOLOv4 se compone de las arquitecturas de otras redes neuronales convolucionales en cada una de sus partes, los detalles de la

red se pueden encontrar en la publicación correspondiente de Bochkovskiy, Wang y Liao (2020). De manera general se describen las siguientes partes correspondientes a la arquitectura.

- Backbone: Utiliza la arquitectura de la red CSPDarknet53 basada en la red Cross Stage Partial de la arquitectura DenseNet, previamente entrenadas con la base de datos ImageNet para la extracción de características. Esta red divide la entrada en dos partes, la primera pasará por capas densas de convolución y la segunda se concatena al final de las capas densas copiando repetidamente la información del gradiente al actualizar los pesos sinápticos, mejorando la capacidad de los campos receptivos o kernels.
- Neck: Se utiliza el bloque de agrupación SSP que produce un tensor de salida constante y la arquitectura PANet para la agregación de características, que permite preservar la información espacial de los mapas de características determinados en capas anteriores.
- Head: Uitliza YOLOv3 para la clasificación/regresión.

La Figura 2.11 muestra de manera gráfica la arquitectura correspondiente a YOLOv4 con una imagen de entrada de tamaño $416 \times 416 \times 3$ y tres cabezales de predicción. Los parámetros a detalle de cada parte de la red se encuentra en el Apéndice A.1.

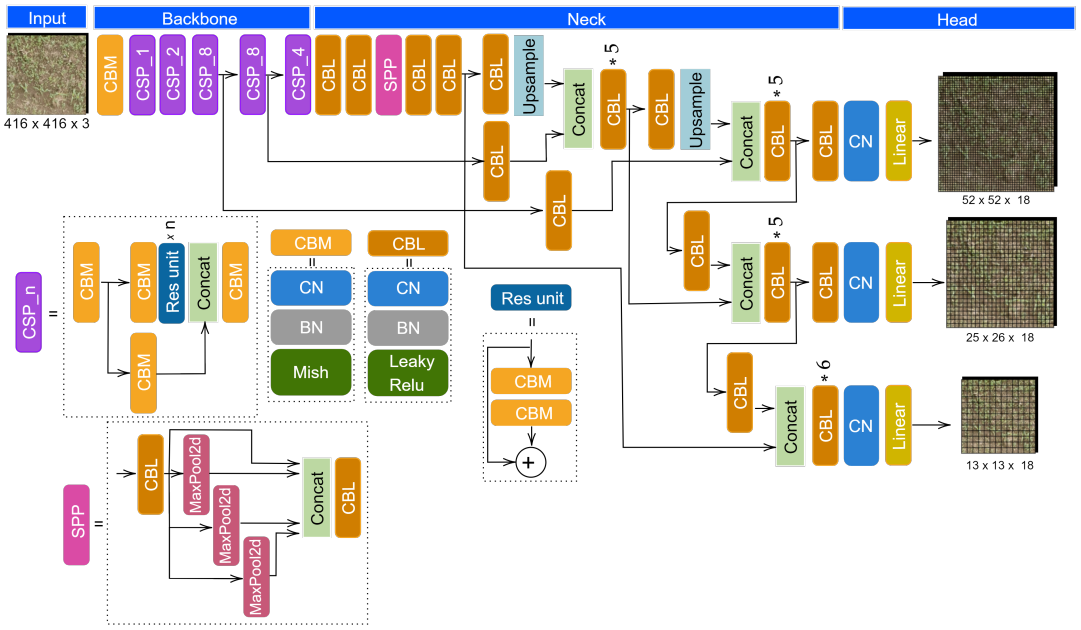


Figura 2.11: *Arquitectura de YOLOv4.*

YOLOv5

Al igual que YOLOv4, YOLOv5 se basa en las arquitecturas de CSPNDarknet53 (columna vertebral), SSP + PANet (cuello) y YOLOv3 Head para la detección de objetos. Los últimos cambios en la arquitectura (V6.0/6.1) están en la primera capa FOCUS por el equivalente de CBS con entradas $[N_{Kernels} = 64, Kernel = 6, Stride = 2, Padding = 2]$ y SPP por un equivalente llamado SPPF, mejorando los tiempos de entrenamiento e inferencia de la red. Se mantiene la estructura para todas las variantes de YOLOv5 , solo se modifica el ancho y la profundidad de la red. La Figura 2.12 muestra el diagrama general de la arquitectura YOLOv5 de las cuales se derivan sus diferentes versiones. Los parámetros a detalle de la red se encuentra en el Apéndice A.2.

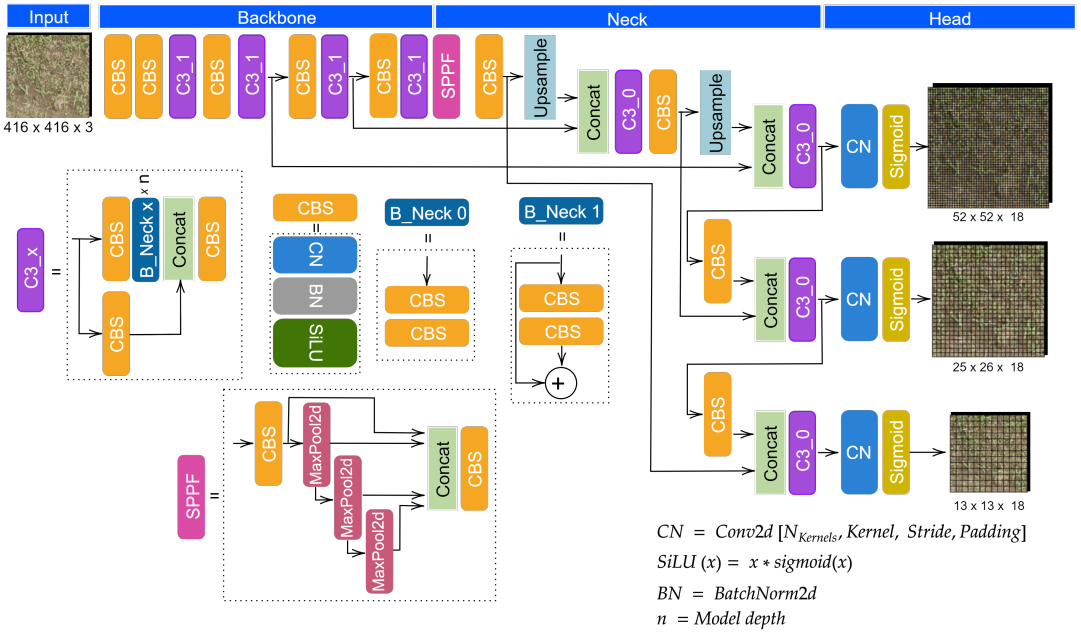


Figura 2.12: *Diagrama de la estructura YOLOv5 para V 6.0/6.1.*

La modificación para la profundidad de la red se realiza tomando el entero positivo de la multiplicación de n (Figura 2.12) por un factor solo en las capas C3 y el ancho de la red al multiplicar el número de kernels por un factor. Los factores están definidos en el Cuadro 2.4.

Cuadro 2.4: *Versiones de YOLOv5*

Versión	Profundidad del modelo	Ancho de capas
YOLOv5-n	0.33	0.25
YOLOv5-s	0.33	0.50
YOLOv5-m	0.67	0.75
YOLOv5-l	1.00	1.00
YOLOv5-x	1.25	1.25

Bibliografía

Aeberli, A. ., Johansen, K. ., Robson, A. ., Lamb, D. W. & Phinn, S. . (2021). Detection of Banana Plants Using Multi-Temporal Multispectral UAV Imagery. *Remote Sensing*, *13*(11), 2123. doi:10.3390/rs13112123. (Vid. págs. 21, 23)

Albawi, S., Mohammed, T. A. & Al-Zawi, S. (2017). Understanding of a convolutional neural network. En *2017 International Conference on Engineering and Technology (ICET)* (pp. 1-6). doi:10.1109/ICEngTechnol.2017.8308186. (Vid. pág. 28)

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... Asari, V. K. (2018). The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. doi:10.48550/ARXIV.1803.01164. (Vid. págs. 26, 27, 29)

Alzubaidi, F. ., Mostaghimi, P. ., Si, G. ., Swietojanski, P. . & Armstrong, R. T. (2022). Automated Rock Quality Designation Using Convolutional Neural Networks. *Rock Mechanics and Rock Engineering*, *55*(6), 3719-3734. doi:10.1007/s00603-022-02805-y. (Vid. pág. 28)

Ammar, A. ., Koubaa, A. . & Benjdira, B. . (2021). Deep-Learning-Based Automated Palm Tree Counting and Geolocation in Large Farms from Aerial Geotagged Images. *Agronomy*, *11*(8), 1458. doi:10.3390/agronomy11081458. (Vid. pág. 22)

Awais, M. ., Li, W. ., Cheema, M. J. M., Zaman, Q. U., Shaheen, A. ., Aslam, B. ., ... Liu, C. . (2022). UAV-based remote sensing in plant stress imagine using high-resolution thermal sensor for digital agriculture practices: a meta-review. *International Journal of Environmental Science and Technology*. doi:10.1007/s13762-021-03801-5. (Vid. pág. 12)

Bayraktar, E. ., Basarkan, M. E. & Celebi, N. . (2020). A low-cost UAV framework towards ornamental plant detection and counting in the wild. *ISPRS Journal of Photogrammetry and Remote Sensing*, *167*, 1-11. doi:10.1016/j.isprsjprs.2020.06.012. (Vid. pág. 20)

Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. doi:10.48550/ARXIV.2004.10934. (Vid. pág. 35)

Borgogno-Mondino, E. (2018). Remote Sensing from RPAS in Agriculture: An Overview of Expectations and Unanswered Questions. En C. Ferraresi & G. Quaglia (Eds.), *Advances in Service and Industrial Robotics* (pp. 483-492). Cham: Springer International Publishing. (Vid. pág. 12).

Bwambale, E. ., Abagale, F. K. & Anornu, G. K. (2022). Smart irrigation monitoring and control strategies for improving water use efficiency in precision agriculture: A review. *Agricultural Water Management*, *260*, 107324. doi:10.1016/j.agwat.2021.107324. (Vid. pág. 11)

Cabreira, T. ., Brisolara, L. . & Ferreira Jr., P. R. (2019). Survey on Coverage Path Planning with Unmanned Aerial Vehicles. *Drones*, *3*(1), 4. doi:10.3390/drones3010004. (Vid. págs. 17, 18)

Che, Y. ., Wang, Q. ., Zhou, L. ., Wang, X. ., Li, B. . & Ma, Y. . (2022). The effect of growth stage and plant counting accuracy of maize inbred lines on LAI and biomass prediction. *Precision Agriculture*. doi:10.1007/s11119-022-09915-1. (Vid. pág. 21)

Daubige, Joudelat, Burger, Comar, de Solan & Baret. (2021). Plant detection and counting from high-resolution RGB images acquired from UAVs: comparison between deep-learning and handcrafted methods with application to maize, sugar beet, and sunflower. *bioRxiv*. doi:10.1101/2021.04.27.441631. (Vid. págs. 21-23)

del Cerro, J. ., Cruz Ulloa, C. ., Barrientos, A. . & de León Rivas, J. . (2021). Unmanned Aerial Vehicles in Agriculture: A Survey. *Agronomy*, *11*(2), 203. doi:10.3390/agronomy11020203. (Vid. págs. 17, 18)

Delgado, J. A., Short, N. M., Roberts, D. P. & Vandenberg, B. . (2019). Big Data Analysis for Sustainable Agriculture on a Geospatial Cloud Framework.

Frontiers in Sustainable Food Systems, *3*. doi:10.3389/fsufs.2019.00054. (Vid. pág. 11)

Diwan, T. ., Anirudh, G. . & Tembhurne, J. V. (2022). Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimedia Tools and Applications*. doi:10.1007/s11042-022-13644-y. (Vid. págs. 31, 32)

Feng, A., Sudduth, K. A., Vories, E. D. & Zhou, J. (2019). Evaluation of cotton stand count using UAV-based hyperspectral imagery. (1900807), 1. doi:10.13031/aim.201900807. (Vid. pág. 21)

Flores-Cruz, L. A., García-Salazar, J. A., Mora-Flores, J. S. & Pérez-Soto, F. (2014). Producción de maíz (Zea mays L.) en el Estado de Puebla: un enfoque de equilibrio espacial para identificar las zonas productoras más competitivas. *Agricultura, sociedad y desarrollo*, *11*, 223-239. (Vid. pág. 11).

García-Martínez, H. ., Flores-Magdaleno, H. ., Khalil-Gardezi, A. ., Ascencio-Hernández, R. ., Tijerina-Chávez, L. ., Vázquez-Peña, M. A. & Mancilla-Villa, O. R. (2020). Digital Count of Corn Plants Using Images Taken by Unmanned Aerial Vehicles and Cross Correlation of Templates. *Agronomy*, *10*(4), 469. doi:10.3390/agronomy10040469. (Vid. pág. 24)

Gholamalinezhad, H. & Khosravi, H. (2020). Pooling Methods in Deep Neural Networks, a Review. doi:10.48550/ARXIV.2009.07485. (Vid. pág. 29)

Girshick, R., Donahue, J., Darrell, T. & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. En *2014 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580-587). doi:10.1109/CVPR.2014.81. (Vid. pág. 30)

Gnädinger, F. . & Schmidhalter, U. . (2017). Digital Counts of Maize Plants by Unmanned Aerial Vehicles (UAVs). *Remote Sensing*, *9*(6), 544. doi:10.3390/rs9060544. (Vid. pág. 23)

Gu, J. ., Wang, Z. ., Kuen, J. ., Ma, L. ., Shahroudy, A. ., Shuai, B. ., ... Chen, T. . (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354-377. doi:10.1016/j.patcog.2017.10.013. (Vid. pág. 28)

Janiesch, C., Zschech, P. & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685-695. doi:10.1007/s12525-021-00475-2. (Vid. págs. 25, 27)

Kennett, D. J., Prufer, K. M., Culleton, B. J., George, R. J., Robinson, M. ., Trask, W. R., . . . Gutierrez, S. M. (2020). Early isotopic evidence for maize as a staple grain in the Americas. *Science Advances*, *6*(23). doi:10.1126/sciadv.aba3245. (Vid. pág. 14)

Khaki, S. ., Safaei, N. ., Pham, H. . & Wang, L. . (2022). WheatNet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. *Neurocomputing*, *489*, 78-89. doi:10.1016/j.neucom.2022.03.017. (Vid. pág. 21)

Khanal, S. ., KC, K. ., Fulton, J. P., Shearer, S. . & Ozkan, E. . (2020). Remote Sensing in Agriculture—Accomplishments, Limitations, and Opportunities. *Remote Sensing*, *12*(22), 3783. doi:10.3390/rs12223783. (Vid. pág. 15)

Kitano, B. T., Mendes, C. C. T., Geus, A. R., Oliveira, H. C. & Souza, J. R. (2019). Corn Plant Counting Using Deep Learning and UAV Images. *IEEE Geoscience and Remote Sensing Letters*, 1-5. doi:10.1109/lgrs.2019.2930549. (Vid. pág. 24)

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. En F. Pereira, C. Burges, L. Bottou & K. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 25), Curran Associates, Inc. (Vid. pág. 30).

Li, G., Huang, X., Ai, J., Yi, Z. & Xie, W. (2021). Lemon-YOLO: An efficient object detection method for lemons in the natural environment. *IET Image Processing*, *15*(9), 1998-2009. doi:10.1049/ipr2.12171. (Vid. págs. 32, 34)

Liu, H. ., Sun, H. ., Li, M. . & Iida, M. . (2020). Application of Color Featuring and Deep Learning in Maize Plant Detection. *Remote Sensing*, *12*(14), 2229. doi:10.3390/rs12142229. (Vid. págs. 23, 24)

Machefer, M. ., Lemarchand, F. ., Bonnefond, V. ., Hitchins, A. . & Sidiropoulos, P. . (2020). Mask R-CNN Refitting Strategy for Plant Counting and Sizing

in UAV Imagery. *Remote Sensing*, *12*(18), 3015. doi:10.3390/rs12183015. (Vid. pág. 20)

Malek, M., Dhiraj, B., Upadhyaya, D. & Patel, D. (2022). A Review of Precision Agriculture Methodologies, Challenges, and Applications. En P. K. Singh, M. H. Kolekar, S. Tanwar, S. T. Wierzchon & R. K. Bhatnagar (Eds.), *Emerging Technologies for Computing, Communication and Smart Cities* (pp. 329-346). Singapore: Springer Nature Singapore. (Vid. pág. 11).

Mizik, T. . (2022). How can precision farming work on a small scale? A systematic literature review. *Precision Agriculture*. doi:10.1007/s11119-022-09934-y. (Vid. pág. 11)

Nepal, U. . & Eslamiat, H. . (2022). Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs. *Sensors*, *22*(2), 464. doi:10.3390/s22020464. (Vid. pág. 32)

Ngoune Tandzi, L. . & Mutengwa, C. S. (2019). Estimation of Maize (Zea mays L.) Yield Per Harvest Area: Appropriate Methods. *Agronomy*, *10*(1), 29. doi:10.3390/agronomy10010029. (Vid. págs. 14, 15)

Ocampo, M. & Catarina, C. (2018). *Agricultura de precisión* (inf. téc. N.º 015). Oficina de Información Científica y Tecnológica para el congreso de la Union (INCyTU). Consultado desde http://foroconsultivo.org.mx/INCyTU/. (Vid. pág. 15)

Oh, S. ., Chang, A. ., Ashapure, A. ., Jung, J. ., Dube, N. ., Maeda, M. ., … Landivar, J. . (2020). Plant Counting of Cotton from UAS Imagery Using Deep Learning-Based Object Detection Framework. *Remote Sensing*, *12*(18), 2981. doi:10.3390/rs12182981. (Vid. pág. 20)

Osco, L. P., dos Santos de Arruda, M. ., Gonçalves, D. N., Dias, A. ., Batistoti, J. ., de Souza, M. ., … Gonçalves, W. N. (2021). A CNN approach to simultaneously count plants and detect plantation-rows from UAV imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *174*, 1-17. doi:10.1016/j.isprsjprs.2021.01.024. (Vid. págs. 17, 22)

Osco, L. P., Marcato Junior, J. ., Marques Ramos, A. P., de Castro Jorge, L. A., Fatholahi, S. N., de Andrade Silva, J. ., … Li, J. . (2021). A review on deep learning in UAV remote sensing. *International Journal of Applied*

Earth Observation and Geoinformation, *102*, 102456. doi:10.1016/j.jag.2021.102456. (Vid. págs. 12, 22, 28)

Panday, U. S., Pratihast, A. K., Aryal, J. . & Kayastha, R. B. (2020). A Review on Drone-Based Data Solutions for Cereal Crops. *Drones*, *4*(3), 41. doi:10.3390/drones4030041. (Vid. págs. 11, 17-20)

Pang, Y. ., Shi, Y. ., Gao, S. ., Jiang, F. ., Veeranampalayam-Sivakumar, A. N., Thompson, L. ., ... Liu, C. . (2020). Improved crop row detection with deep neural network for early-season maize stand count in UAV imagery. *Computers and Electronics in Agriculture*, *178*, 105766. doi:10.1016/j.compag.2020.105766. (Vid. pág. 24)

Raj, E. F. I., Appadurai, M. & Athiappan, K. (2021). Precision Farming in Modern Agriculture. En A. Choudhury, A. Biswas, T. P. Singh & S. K. Ghosh (Eds.), *Smart Agriculture Automation Using Advanced Technologies: Data Analytics and Machine Learning, Cloud Architecture, Automation and IoT* (pp. 61-87). doi:10.1007/978-981-16-6124-2_4. (Vid. pág. 15)

Reddy Maddikunta, P. K., Hakak, S. ., Alazab, M. ., Bhattacharya, S. ., Gadekallu, T. R., Khan, W. Z. & Pham, Q. V. (2021). Unmanned Aerial Vehicles in Smart Agriculture: Applications, Requirements, and Challenges. *IEEE Sensors Journal*, *21*(16), 17608-17619. doi:10.1109/jsen.2021.3049471. (Vid. pág. 12)

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. En *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 779-788). doi:10.1109/CVPR.2016.91. (Vid. pág. 32)

Redmon, J. & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv*. doi:10.48550/ARXIV.1804.02767. (Vid. págs. 32, 33)

Reza Keyvanpour, M. & Shirzad, M. B. (2022). Chapter 14 - Machine learning techniques for agricultural image recognition. En M. A. Khan, R. Khan & M. A. Ansari (Eds.), *Application of Machine Learning in Agriculture* (pp. 283-305). doi:https://doi.org/10.1016/B978-0-323-90550-3.00011-4. (Vid. págs. 25, 26)

Rivera, A. ., Cedillo Ramírez, L. ., Parraguirre Lezama, C. ., Baez Simon, A. ., Laug Garcia, B. . & Romero-Arenas, O. . (2022). Evaluation of Cytotoxic and Genotoxic Risk Derived from Exposure to Pesticides in Corn Producers in Tlaxcala, Mexico. *Applied Sciences*, *12*(18), 9050. doi:10.3390/app12189050. (Vid. pág. 14)

Rogers, A. R., Dunne, J. C., Romay, C. ., Bohn, M. ., Buckler, E. S., Ciampitti, I. A., … Holland, J. B. (2021). The importance of dominance and genotype-by-environment interactions on grain yield variation in a large-scale public cooperative maize experiment. *G3 Genes|Genomes|Genetics*, *11*(2). doi:10.1093/g3journal/jkaa050. (Vid. pág. 14)

Sarabia, R. ., Aquino, A. ., Ponce, J. M., López, G. . & Andújar, J. M. (2020). Automated Identification of Crop Tree Crowns from UAV Multispectral Imagery by Means of Morphological Image Analysis. *Remote Sensing*, *12*(5), 748. doi:10.3390/rs12050748. (Vid. pág. 12)

Shi, R. ., Wang, J. ., Tong, L. ., Du, T. ., Shukla, M. K., Jiang, X. ., … Guo, X. . (2022). Optimizing planting density and irrigation depth of hybrid maize seed production under limited water availability. *Agricultural Water Management*, *271*, 107759. doi:10.1016/j.agwat.2022.107759. (Vid. pág. 12)

SIAP. (2022). Servicio de Información Agroalimentaria y Pesquera: Anuario estadístico de la producción agrícola. Consultado el 26 de octubre de 2022, desde https://nube.siap.gob.mx/cierreagricola/. (Vid. págs. 11, 14)

Singh, S., Thakur, A. K., Goyal, N. & Gupta, K. (2022). Image processing based Wheat spike detection using YOLO. En *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)* (pp. 757-763). doi:10.1109/ICAIS53314.2022.9742888. (Vid. pág. 32)

Sishodia, R. P., Ray, R. L. & Singh, S. K. (2020). Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sensing*, *12*(19), 3136. doi:10.3390/rs12193136. (Vid. págs. 11, 16, 17, 19)

Tsouros, D. C., Bibi, S. . & Sarigiannidis, P. G. (2019). A Review on UAV-Based Applications for Precision Agriculture. *Information*, *10*(11), 349. doi:10.3390/info10110349. (Vid. págs. 16-18, 20)

Varela, S. ., Dhodda, P. ., Hsu, W. ., Prasad, P. V., Assefa, Y. ., Peralta, N. ., … Ciampitti, I. . (2018). Early-Season Stand Count Determination in Corn via Integration of Imagery from Unmanned Aerial Systems (UAS) and Supervised Learning Techniques. *Remote Sensing*, *10*(3), 343. doi:10.3390/rs10020343. (Vid. pág. 23)

Velumani, K. ., Lopez-Lozano, R. ., Madec, S. ., Guo, W. ., Gillet, J. ., Comar, A. . & Baret, F. . (2021). Estimates of Maize Plant Density from UAV RGB Images Using Faster-RCNN Detection Model: Impact of the Spatial Resolution. *Plant Phenomics*, *2021*, 1-16. doi:10.34133/2021/9824843. (Vid. págs. 16, 25)

Wang, L. ., Xiang, L. ., Tang, L. . & Jiang, H. . (2021). A Convolutional Neural Network-Based Method for Corn Stand Counting in the Field. *Sensors*, *21*(2), 507. doi:10.3390/s21020507. (Vid. págs. 25, 32)

Wang, Q. ., Cheng, M. ., Huang, S. ., Cai, Z. ., Zhang, J. . & Yuan, H. . (2022). A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed Solanum rostratum Dunal seedlings. *Computers and Electronics in Agriculture*, *199*, 107194. doi:10.1016/j.compag.2022.107194. (Vid. pág. 32)

Weiss, M. ., Jacob, F. . & Duveiller, G. . (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, *236*, 111402. doi:10.1016/j.rse.2019.111402. (Vid. págs. 15, 17)

Wu, J. ., Yang, G. ., Yang, X. ., Xu, B. ., Han, L. . & Zhu, Y. . (2019). Automatic Counting of in situ Rice Seedlings from UAV Images Based on a Deep Fully Convolutional Neural Network. *Remote Sensing*, *11*(6), 691. doi:10.3390/rs11060691. (Vid. págs. 21, 23)

Xiao, Y. ., Tian, Z. ., Yu, J. ., Zhang, Y. ., Liu, S. ., Du, S. . & Lan, X. . (2020). A review of object detection based on deep learning. *Multimedia Tools and Applications*, *79*(33-34), 23729-23791. doi:10.1007/s11042-020-08976-6. (Vid. págs. 30, 31)

Xue, J. . & Su, B. . (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, *2017*, 1-17. doi:10.1155/2017/1353691. (Vid. pág. 20)

Yang, R. . & Yu, Y. . (2021). Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis. *Frontiers in Oncology*, *11*. doi:10.3389/fonc.2021.638182. (Vid. pág. 31)

Yang, W., Feng, H. ., Zhang, X. ., Zhang, J. ., Doonan, J. H., Batchelor, W. D., . . . Yan, J. . (2020). Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. *Molecular Plant*, *13*(2), 187-214. doi:10.1016/j.molp.2020.01.008. (Vid. pág. 12)

Young, D. J. N., Koontz, M. J. & Weeks, J. . (2022). Optimizing aerial imagery collection and processing parameters for drone-based individual tree mapping in structurally complex conifer forests. *Methods in Ecology and Evolution*, *13*(7), 1447-1463. doi:10.1111/2041-210x.13860. (Vid. pág. 19)

Zaidi, S. S. A., Ansari, M. S., Aslam, A. ., Kanwal, N. ., Asghar, M. . & Lee, B. . (2022). A survey of modern deep learning based object detection models. *Digital Signal Processing*, *126*, 103514. doi:10.1016/j.dsp.2022.103514. (Vid. pág. 31)

Zou, Z., Shi, Z., Guo, Y. & Ye, J. (2019). Object detection in 20 years: A survey. doi:10.48550/ARXIV.1905.05055. (Vid. págs. 30, 31)

Capítulo 3

ARTÍCULO CIENTÍFICO

DETECCIÓN Y CONTEO DE PLANTAS DE MAÍZ EN PRESENCIA DE MALEZA CON REDES NEURONALES CONVOLUCIONALES

DETECTION AND COUNTING OF CORN PLANTS IN THE PRESENCE OF WEEDS WITH CONVOLUTIONAL NEURAL NETWORKS

3.1. Resumen

El maíz es una parte importante de la dieta mexicana. El cultivo requiere un seguimiento constante para asegurar la producción. Para ello, se suele utilizar la densidad de plantas como indicador del rendimiento del cultivo, ya que conocer el número de plantas ayuda a los productores a gestionar y controlar sus parcelas. En este contexto, es necesario detectar y contar las plantas de maíz. Por lo tanto, se creó una base de datos de imágenes RGB aéreas de un cultivo de maíz en condiciones de maleza para implementar y evaluar algoritmos de aprendizaje profundo. Se realizaron diez misiones de vuelo, seis con una distancia de muestreo terrestre (GSD) de 0.33 cm/píxel en los estados vegetativos de V3 a V7 y cuatro con una GSD de 1.00 cm/píxel para los estados vegetativos V6, V7 y V8. Los detectores comparados fueron YOLOv4, YOLOv4-tiny, YOLOv4-tiny-3l y YOLOv5 versiones s, m y l. Cada detector se evaluó en los umbrales de intersección sobre unión (IoU) de 0.25, 0.50 y 0.75 en intervalos de confianza de 0.05. Se observó una fuerte penalización de F1-Score en el umbral de IoU de 0.75 y hubo un aumento del 4.92 % en todos los modelos para un umbral de IoU de 0.25 en comparación con 0.50. Para niveles de confianza superiores a 0.35, YOLOv4 muestra una mayor robustez en la detección en comparación con los otros modelos. Considerando la moda de 0.3 para el nivel de confianza que maximiza la métrica F1-Score y el umbral IoU de 0.25 en todos los modelos, YOLOv5-s obtuvo un mAP de 73.1 % con un coeficiente de determinación (R^2) de 0.78 y un error cuadrático medio relativo (rRMSE) de 42 % en el conteo de plantas, seguido de YOLOv4 con un mAP de 72.0 %, R^2 de 0.81 y rRMSE de 39.5 %.

Palabras clave: Imágenes aéreas, Conteo de plantas, Maleza, Detección, YOLO

3.2. Abstract

Corn is an important part of the Mexican diet. The crop requires constant monitoring to ensure production. For this, plant density is often used as an indicator of crop yield, since knowing the number of plants helps growers to manage and control their plots. In this context, it is necessary to detect and count corn plants. Therefore, a database of aerial RGB images of a corn crop in weedy conditions was created to implement and evaluate deep learning algorithms. Ten flight missions were conducted, six with a ground sampling distance (GSD) of 0.33 cm/pixel at vegetative stages from V3 to V7 and four with a GSD of 1.00 cm/pixel for vegetative stages V6, V7 and V8. The detectors compared were YOLOv4, YOLOv4-tiny, YOLOv4-tiny-3l, and YOLOv5 versions s, m and l. Each detector was evaluated at intersection over union (IoU) thresholds of 0.25, 0.50 and 0.75 at confidence intervals of 0.05. A strong F1-Score penalty was observed at the IoU threshold of 0.75 and there was a 4.92 % increase in all models for an IoU threshold of 0.25 compared to 0.50. For confidence levels above 0.35, YOLOv4 shows greater robustness in detection compared to the other models. Considering the mode of 0.3 for the confidence level that maximizes the F1-Score metric and the IoU threshold of 0.25 in all models, YOLOv5-s obtained a mAP of 73.1 % with a coefficient of determination (R^2) of 0.78 and a relative mean square error (rRMSE) of 42 % in the plant count, followed by YOLOv4 with a mAP of 72.0 %, R^2 of 0.81 and rRMSE of 39.5 %.

Keywords: Aaerial images, Plant count, Weeds, Detection, YOLO

Tesis de Maestría en Ingeniería Agrícola y Uso Integral del Agua (IAUIA)
Universidad Autónoma Chapingo
Autor: Canek Mota Delfin
Director de Tesis: Gilberto de Jesús López Canteñs

3.3. Introducción

La producción de maíz (*Zea mays L.*) en México para el año 2020 superó los 27.4 millones de toneladas (SIAP, 2022). El maíz es uno de los cultivos más importantes del país desde el punto de vista alimentario, político, económico y social (Flores-Cruz, García-Salazar, Mora-Flores & Pérez-Soto, 2014). Los cereales forman una parte crucial de la dieta humana y la alimentación del ganado, por lo que lograr la autosuficiencia en su producción es una forma efectiva de promover la seguridad alimentaria (Panday, Pratihast, Aryal & Kayastha, 2020).

Conocer el número de plantas y monitorear su estado de crecimiento es importante para estimar el rendimiento (Kitano, Mendes, Geus, Oliveira & Souza, 2019; Osco et al., 2021). El conteo manual después de la emergencia de la planta no es práctico en campos de producción a gran escala debido a la cantidad de mano de obra requerida, además de que es intrínsecamente inexacto (Kitano et al., 2019; Panday et al., 2020; Varela et al., 2018). Un enfoque que se ha aplicado en los últimos años es el uso de sistemas aéreos pilotados a distancia (RPAS) equipados con sensores ópticos para la teledetección agrícola (Messina & Modica, 2020). Varios estudios han informado sobre el uso de RPAS para determinar densidades de siembra en diferentes cultivos; por ejemplo, en Oh et al. (2020), informaron sobre su uso para la detección de plantas de algodón, basado en aprendizaje automático con redes neuronales convolucionales (CNNs), en Fan, Lu, Gong, Xie y Goodman (2018), utilizaron CNNs para la detección y conteo de plantas de tabaco, en Valente, Sari, Kooistra, Kramer y Mücher (2020), proponen una CNN (WheatNet) basada en MobileNetV2 con dos salidas para la localización y conteo de espigas a partir de imágenes, de manera similar, en Khaki, Safaei, Pham y Wang (2022), proponen un método integrado de preprocesamiento de imágenes (Excess Green Index y método de Otsu) y CNN para identificar y contar plantas de espinaca.

En la literatura se han reportado tres métodos para el recuento y clasificación de plantas de maíz:

(1) Técnicas clásicas de procesamiento de imágenes. En García-Martínez et al. (2020) se compararon cámaras RGB montadas en RPAS utilizando plantillas y correlación cruzada normalizada, obteniendo R^2 de 0.98, 0.90 y 0.16 para etapas vegetativas V2, V5 y V9 respectivamente. Gnädinger y Schmidhalter (2017), utilizando el procedimiento de realce de contraste decorrstrech con umbralización, obtuvieron coeficientes R^2 de 0.89 para etapas vegetativas V3 y V5. Shuai et al. (2019) emplearon el índice de color exceso de verde (ExG), logrando precisiones del 95 % y un recall del 100 % para el conteo de plantas en la etapa vegetativa V2.

(2) Técnicas clásicas de procesamiento de imágenes más procedimientos de aprendizaje automático. En Gómez-Ramos, Ruíz-Castilla y García-Lamont (2020), utilizaron el análisis de componentes principales (PCA) y el método de umbralización de Otsu para extraer características como entrada a la red neuronal Naive Bayes y clasificadores Random Forest para clasificar plantas de maíz y malezas en imágenes capturadas con dispositivos móviles. Varela et al. (2018) utilizaron índices de color, descriptores geométricos y clasificadores de árbol de decisión para el conteo de maíz, logrando precisiones del 96 % para etapas V2 y V3. Pang et al. (2020) combinaron descriptores geométricos y redes neuronales convolucionales para el conteo de plantas de maíz, logrando precisiones del 95.8 % para etapas vegetativas V5 y V4.

(3) Aprendizaje automático con CNN. En Liu, Sun, Li y Iida (2020) compararon índices de color con arquitecturas CNN, específicamente “You Only Look Once” (YOLO) en sus versiones YOLOv3 y YOLOv3-tiny, para evaluar la detección de plantas de maíz en imágenes capturadas a una altura de 0.3 m del suelo, logrando un 77 % de intersección sobre la unión (IoU). L. . Wang, Xiang, Tang y Jiang (2021) utilizaron una cámara montada en un robot para la detección y conteo de plantas en tiempo real, empleando YOLOv3 y filtro de Kalman logrando precisiones del 98 % en las etapas V2 y V3. Vong, Conway, Zhou, Kitchen y Sudduth (2021) realizaron segmentación semántica con la arquitectura U-NET, obteniendo coeficiente R^2 de 0.95 en la etapa V2. Velumani et al. (2021) evaluaron el desempeño de Faster-RCNN para el conteo de plantas de maíz en diferentes resoluciones espaciales, logrando un valor rRMSE del 8 % con

una distancia de muestreo en tierra (GSD) de 0.3 cm/píxel. Osco et al. (2021) propusieron una arquitectura basada en CNN para segmentar y contar plantas de maíz, logrando puntuaciones F1 de 0.87 para la etapa V3. Daubige et al. (2021) compararon métodos clásicos para el procesamiento de imágenes y Faster-RCNN en el conteo de plantas de maíz, remolacha azucarera y girasol.

En general, los mejores resultados se obtuvieron cuando se utilizaron CNN con métodos de aprendizaje profundo. Aunque existen algunos trabajos que analizan los efectos de las malezas en la detección y conteo de plantas de maíz en imágenes aéreas (Daubige et al., 2021; Gómez-Ramos et al., 2020; Osco et al., 2021), dada la complejidad de los posibles escenarios y las condiciones de los campos de maíz en México, todavía se requieren bases de datos etiquetadas para evaluar la solidez de los algoritmos de detección de objetos de última generación. Por ello, en este trabajo se realizan las siguientes aportaciones: (i) una base de datos con 11,191 imágenes aéreas de dimensión 416 × 416 × 3 etiquetadas, (ii) una comparación de los resultados obtenidos por YOLOv4, YOLOv4-tiny, YOLOv4-tiny-3l , YOLOv5-s, YOLOv5-m y YOLOv5-l en la detección y conteo de plantas de maíz en campos infestados de malezas, considerando el valor de la intersección sobre unión y confianza y (iii) la optimización del nivel de confianza y la intersección over union que maximizan la métrica F1-Score en la evaluación de los modelos.

El documento está organizado de la siguiente manera. La Sección 3.4.1 describe las condiciones y el proceso utilizado para la adquisición de imágenes aéreas, así como el proceso de etiquetado para la formación de la base de datos; La Sección 3.4.2 y la Sección 3.4.3 describen los algoritmos utilizados y las métricas de evaluación. Los resultados y la discusión se proporcionan en la Sección 3.5 y la Sección 3.6 , respectivamente. Finalmente, la Sección 3.7 proporciona conclusiones y ofrece ideas para futuras investigaciones.

3.4. Materiales y Métodos

3.4.1. Conjunto de datos

El área de investigación se dividió en cinco sitios experimentales, ubicados en la Universidad Autónoma Chapingo, Texcoco, Estado de México, ubicados geográficamente en: 19°29′27.3″N Lat., 98°53′06.9″W Long., y 2260 msnm Para una fácil identificación, los campos experimentales se denominaron Irrigación, Xerona, San Juan A1, San Juan A2 y Ranchito X13 (Figura 3.1). Todos los sitios experimentales tenían la variedad de maíz híbrido CP-HS2, excepto el sitio Ranchito X13, que tenía múltiples variedades destinadas a fines de mejoramiento. La Tabla 3.1 describe el arreglo de plantación, variedad y coordenadas geográficas para cada sitio experimental.

Figura 3.1: *Ubicación y distribución de los sitios experimentales.*

Cuadro 3.1: *Áreas de captura de datos.*

Sitio experimental	Coordenadas geográficas Latitud	Longitud	Variedad	Disposición de siembra
Irrigación	19° 28′ 56″N	98° 53′ 28″W	Hibrido CP-HS2	1 Fila
Xerona	19° 29′ 02″N	98° 53′ 57″W	Hibrido CP-HS2	2 Filas
San Juan A1	19° 29′ 32″N	98° 51′ 38″W	Hibrido CP-HS2	1 Fila
San Juan A2	19° 29′ 31″N	98° 51′ 34″W	Hibrido CP-HS2	1 Fila
Ranchito X13	19° 29′ 36″N	98° 52′ 43″W	Variadas [a]	1 Fila

[a] Más de una variedad en la misma área

Adquisición de datos

Además de la adquisición de imágenes, se tomaron muestras del número de hojas y la altura de la planta de maíz. Las imágenes aéreas fueron adquiridas utilizando un RPAS multirotor DJI Mavic Pro (SZ DJI Technology Co., Shenzhen, Guangdong, China), equipado con una cámara RGB modelo FC220, con las siguientes características:

- Sensor CMOS de 1/2.3ćon 12.7 M píxeles totales y 12.3 M píxeles efectivos.
- Objetivo FOV 78.8° 26 mm
- Distancia focal 2.22 mm
- Distorsión <1.5 %
- Rango ISO de 100 a 1600
- Tamaño de imagen: 4000 x 3000 píxeles
- Velocidad del obturador de 8s - 1/8000s

Las misiones de vuelo se planificaron y realizaron en horarios de 10:00 a 14:00 horas, con un traslape entre imágenes del 80 % tanto frontal como lateral y la vista de la cámara hacia el nadir. La altura de vuelo del RPAS fue de 10 m para un GSD de 0.33 cm/píxel y 30 m para un GSD de 1.00 cm/píxel. La Tabla 3.2 resume las misiones de vuelo, el área muestreada, las imágenes capturadas y las condiciones climáticas.

Cuadro 3.2: *Características de las misiones de vuelo.*

Fecha de captura	GSD (cm/px)	Área (m^2)	Imágenes capturadas	Temperatura (°C)	Velocidad de viento (km/h)	Visibilidad (km)
Irrigación						
02/08/2021	0.33	2,883	479	23	7.41	6.70
09/08/2021		3,564	438	18	9.26	11.30
18/08/2021		3,722	436	23	13.00	8.00
26/08/2021		1,950	466	23	7.41	16.10
Xerona						
08/07/2021	0.33	1,663	293	16	7.56	9.66
08/07/2021	1.00	16,106	360	16	7.56	9.66
14/07/2021	0.33	1,667	294	14	5.40	6.66
San Juan A1						
17/06/2021	1.00	15,177	361	16	13.00	12.90
San Juan A2						
01/07/2021	1.00	11,281	222	17	7.56	11.30
Ranchito X13						
24/06/2021	1.00	10,696	306	19	7.41	4.84

Para definir el estado del cultivo se obtuvieron 30 muestras aleatorias del número de hojas y altura de la planta durante cada misión de vuelo, determinando el estado vegetativo expresado con la letra V más el número de hojas verdaderas, siguiendo la metodología descrita en Brewer et al. (2022) y la altura promedio de las plantas de maíz. Como en el trabajo de Daubige et al. (2021), la infestación de malezas se determinó cualitativamente asignando valores de 0, 1 y 2 para áreas libres de malezas, baja presencia de malezas e infestación de malezas, respectivamente. Se incluyeron los subíndices F, P y T para ubicar malezas entre hileras, entre plantas y cobertura total (Tabla 3.3).

Cuadro 3.3: *Caracterización del cultivo.*

Fecha de captura	Días después de siembra	Etapa vegetativa	Altura media de la planta (cm)	Infestación de maleza
		Irrigación		
02/08/2021	16	V3	9.62 ± 1.66	1_T
09/08/2021	23	V4	16.8 ± 4.39	2_T
18/08/2021	32	V5	25.16 ± 6.00	2_T
26/08/2021	40	V6	34.68 ± 8.00	1_T
		Xerona		
08/07/2021	44	V6	52.94 ± 6.64	$0_F, 2_P$
08/08/2021	44	V6	52.94 ± 6.64	$0_F, 2_P$
14/07/2021	50	V7	75.68 ± 10.43	$0_F, 2_P$
		San Juan A1		
17/06/2021	57	V7	49.86 ± 10.39	0_T
		San Juan A2		
01/07/2021	71	V8	92.94 ± 21.67	2_T
		Ranchito X13		
24/06/2021	59	V7	75.78 ± 18.41	0_T

Etiquetado de plantas

En la ortorrectificación de las imágenes se utilizó el software mapper Pix4D (Pix4D SA, Lausanne, Suiza), obteniendo los mejores resultados con los siguientes ajustes de parámetros: en el proceso inicial para la extracción de puntos clave, la imagen completa, calibración alternativa y parámetro interno Se utilizaron optimizaciones con alta prioridad. Para la generación de la nube de puntos y la malla se utilizó la mitad de la imagen. Para cada misión de vuelo, se obtuvo un ortomosaico y se dividió en imágenes de 416 × 416 píxeles para formar la base de datos de plantas de maíz. El etiquetado manual de las plantas (ground Truth Label) se realizó con la herramienta LabelImg (Tzutalin, 2015), respetando el formato exigido por YOLO. Cada etiqueta corresponde al grupo de píxeles en RGB pertenecientes a una planta de maíz, a los que se les asignó el nombre “MAIZ”, la palabra española para maíz. En el proceso de etiquetado se tuvieron en cuenta las siguientes consideraciones: (1) el recuadro

de cada etiqueta cubría toda la planta, (2) en caso de plantas incompletas en los bordes de la imagen, plantas borrosas y plantas con hojas fantasma, etiquetas se consideraban correctas sólo si el centro de la planta era totalmente visible, y (3) se eliminaban las imágenes con errores en su cosido o demasiada complejidad en el rotulado. En la Figura 3.2a–f se muestran ejemplos de etiquetado y condiciones de infestación de malezas en diferentes etapas vegetativas con diferentes distancias de muestreo del suelo.

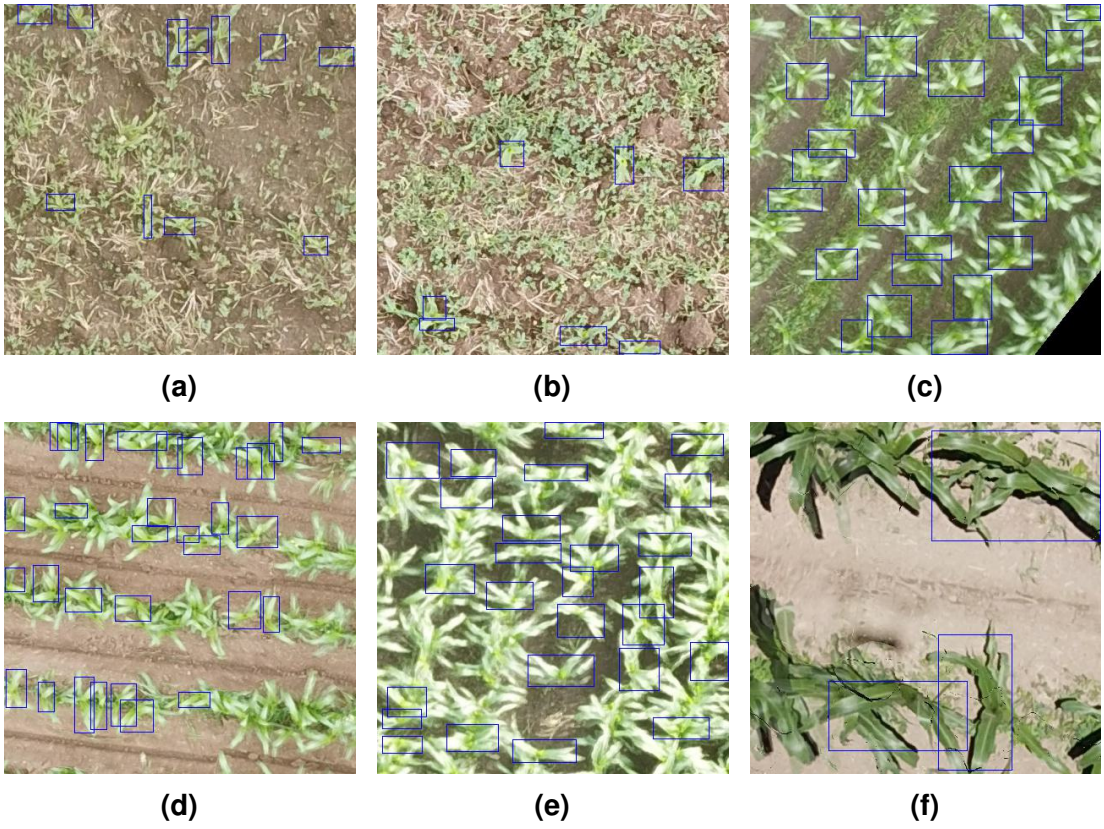


Figura 3.2: *Muestras de imágenes etiquetadas de manera manual. (**a**) y (**b**) corresponde a Irrigación con maíz en etapa $V4_{0,33}$, (**c**) a San Juan A2 con maíz en etapa $V8_{1,00}$, (**d**) a Xerona con maíz en etapa $V6_{1,00}$, (**e**) a Ranchito X13 con maíz en etapa $V7_{1,00}$ y (**f**) a Irrigación con maíz $V6_{0,33}$*

Descripción de la base de datos

La base de datos está compuesta por imágenes con un tamaño de 416 × 416 píxeles, obtenidas completamente al azar de cada ortomosaico con una proporción del 70 % para entrenamiento, 15 % para prueba y 15 % para evaluación, haciendo un total de 11,191 imágenes y 85,419 etiquetas. Considerando cada etapa vegetativa y su resolución espacial GSD en cm/pixel, la Figura 3.3

muestra la distribución de las imágenes correspondientes para entrenamiento, prueba y evaluación.

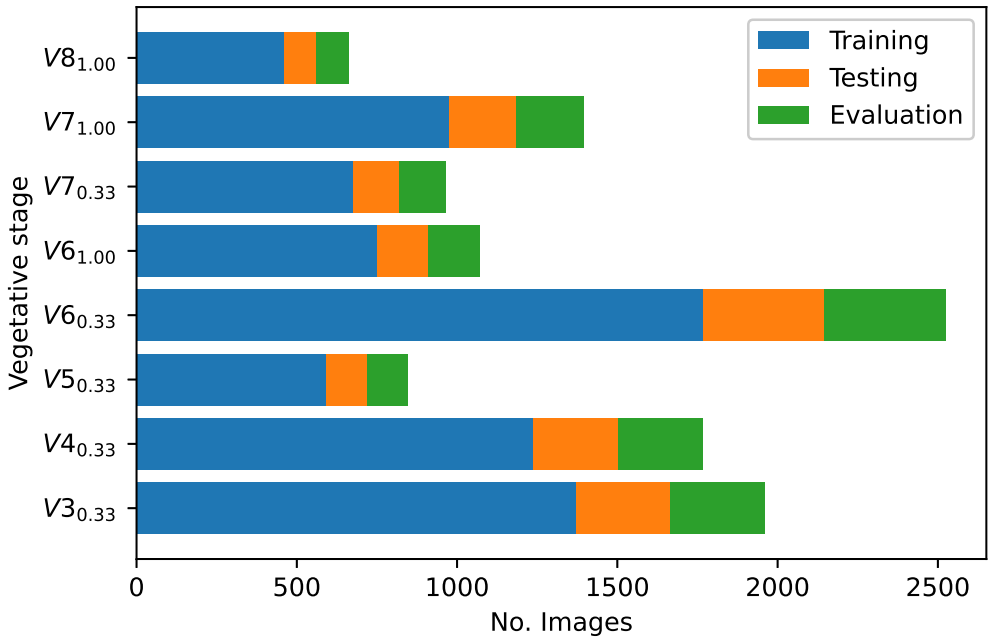


Figura 3.3: *Distribución de imágenes de acuerdo con la etapa vegetativa y GSD*

De acuerdo con las definiciones de Z. . Wang et al. (2021) y la base de datos COCO (Lin et al., 2014) en cuanto al tamaño de las etiquetas, se agruparon en pequeñas (área $<32^2$ píxeles), de tamaño medio (32^2 <área $<96^2$ píxeles) y grandes (área $>96^2$ píxeles). La Figura 3.4 muestra la distribución de tamaños de las etiquetas en la base de datos, donde el 35.58 % de las etiquetas son pequeñas, el 59.58 % medianas y el 4.48 % grandes.

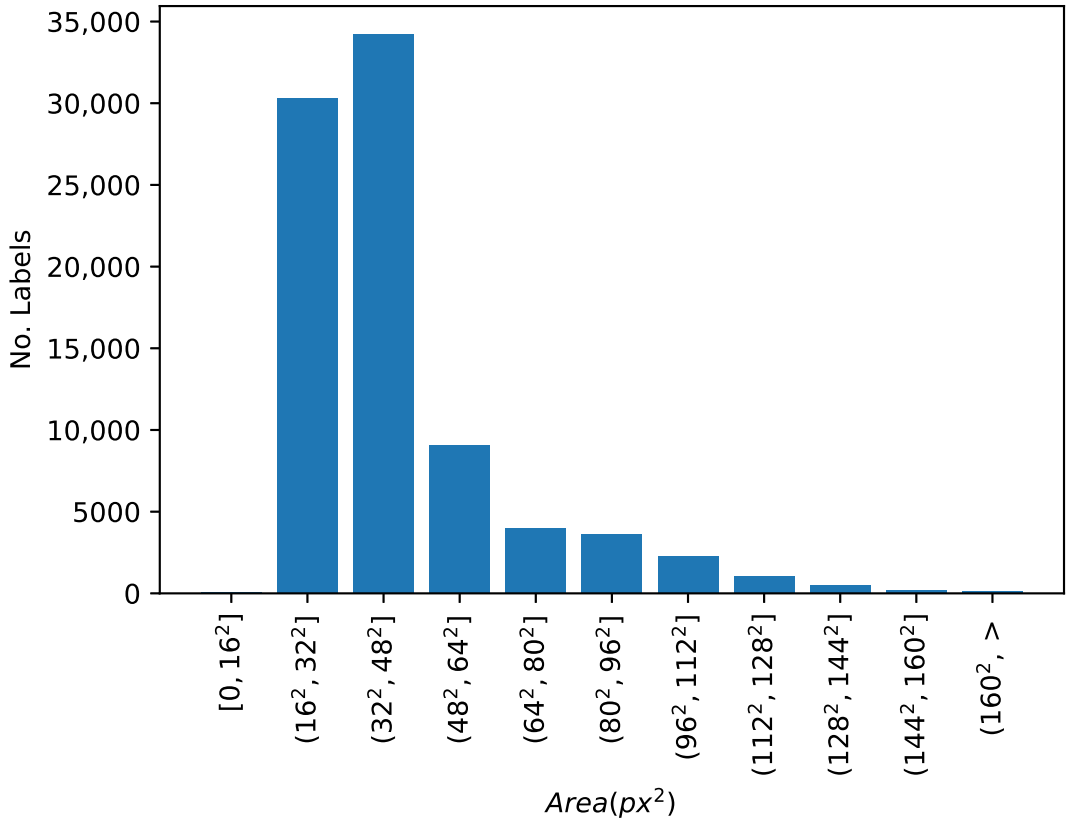


Figura 3.4: *Distribución de las etiquetas de acuerdo con el área en* $pixeles^2$ *para un tamaño de imagen 416 px x 416 px*

3.4.2. Algoritmos de detección y su entrenamiento

Una red neuronal convolucional (CNN) es una variante de la arquitectura del perceptrón multicapa (MLP), inspirada en la corteza visual animal y diseñada para el procesamiento de imágenes, basada en tres tipos principales de capas neuronales: capas convolucionales que aplican operaciones de convolución 2D para encontrar diferentes características de interés en una imagen; capas de reducción de muestreo que reducen la dimensión espacial de las capas convolucionales y capas totalmente conectadas (MLP) que manejan la inferencia de alto nivel en la red (Santos & Papa, 2022; Voulodimos, Doulamis, Doulamis & Protopapadakis, 2018). La extracción automática de características a través de la optimización del filtro convolucional brinda a las CNN una ventaja competitiva sobre los algoritmos tradicionales.

La detección de objetos mediante CNN combina la clasificación de imágenes y la localización de objetos. Generalmente, se basan en propuestas de regiones y la clasificación de cada región en diferentes categorías o como un problema de regresión/clasificación. En general, estos algoritmos se dividen en dos grandes categorías: detectores de dos etapas que realizan la ubicación de objetos en función de la propuesta de regiones y luego se clasifican en diferentes categorías, y detectores que determinan directamente la ubicación y clasificación de objetos en un solo paso en función de la regresión/clasificación (Sozzi, Cantalamessa, Cogato, Kayad & Marinello, 2022).

Analizando los resultados obtenidos por Wenkel, Alhazmi, Liiv, Alrshoud y Simon (2021), al comparar diferentes arquitecturas CNNs en detección de objetos y considerando la propuesta de Velumani et al. (2021), se seleccionaron las arquitecturas YOLOv4 (Bochkovskiy, Wang & Liao, 2020) y YOLOv5 (Jocher et al., 2022) como base para este trabajo de investigación. Estas versiones de YOLO (Redmon, Divvala, Girshick & Farhadi, 2016) se componen principalmente de tres partes: extractor de funciones (Backbone), agregación de funciones para detección a diferentes escalas (Neck) y predicción/regresión (Head); la diferencia entre estas variantes de YOLO se basa en la modificación de estas tres partes, como cambios en la función de pérdida, función de activación y técnicas de regularización, entre otros.

El modelo YOLOv4 utiliza arquitecturas CSPNet + Darknet53 previamente entrenadas con la base de datos ImageNet para la extracción de características, SSP + PANet para la agregación de características y la clasificación/regresión propuesta de YOLOv3 (Redmon & Farhadi, 2018) para la detección de objetos (Nepal & Eslamiat, 2022). La Figura 3.5 muestra el diagrama correspondiente a la arquitectura YOLOv4 con una imagen de entrada de tamaño 416 × 416 × 3 y tres cabezas de predicción a diferentes escalas con 3 cajas cada una, obteniendo un tensor N × N × [3 × (4 offsets de caja delimitadora + 1 predicción de objetividad + n_clases)].

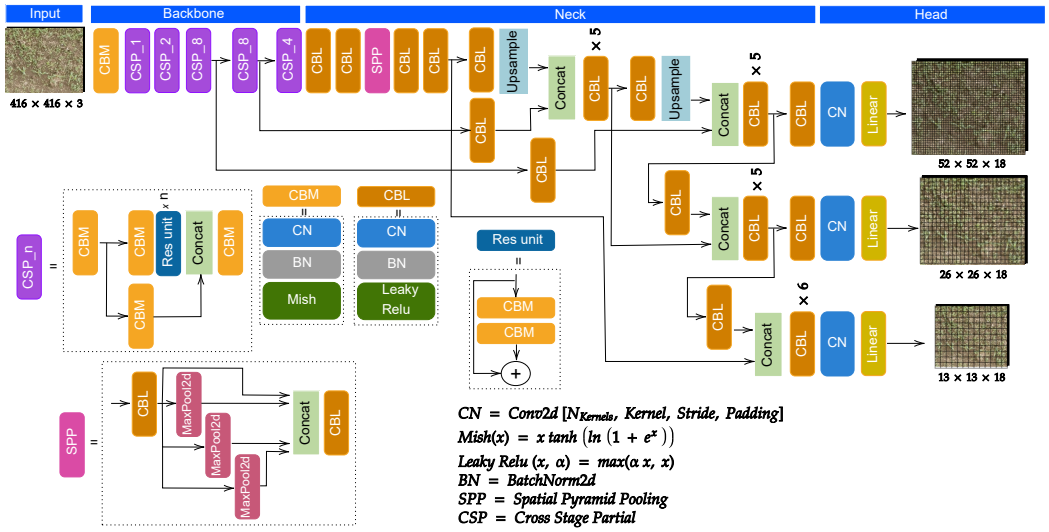


Figura 3.5: *Diagrama de la arquitectura YOLOv4 con una imagen de entrada de 416 × 416 píxeles y 3 canales.*

A partir de la versión completa de YOLOv4, se derivaron las versiones YOLOv4-tiny con dos salidas de predicción a diferentes escalas y YOLOv4-tiny-3l con tres salidas, manteniendo un número reducido de capas respecto a su versión original. La Figura 3.6 muestra el diagrama de la arquitectura YOLOv4-tiny-3l; en nuestro caso, para la versión YOLOv4-tiny, se eliminó la salida de 52 × 52 × 18.

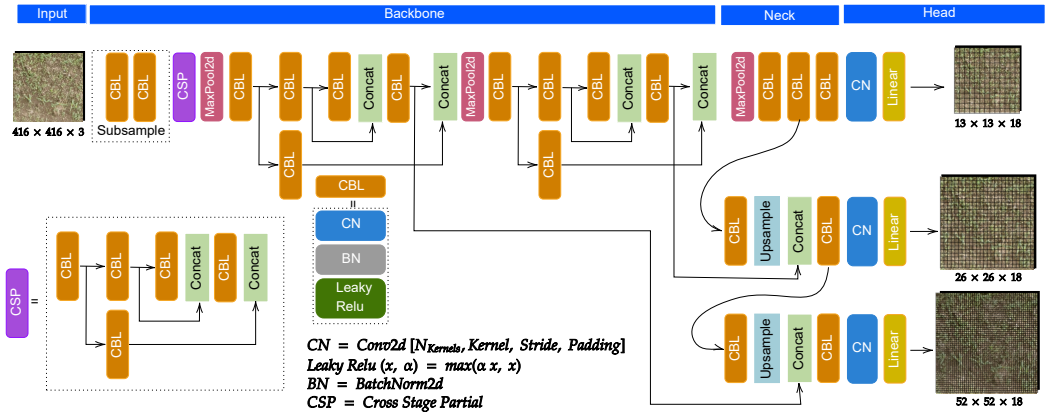


Figura 3.6: *Diagrama de la arquitectura YOLOv4-tiny-3l con una imagen de entrada de 416 × 416 píxeles y 3 canales.*

Se seleccionó la versión YOLOv4-tiny porque tiene tiempos de inferencia más rápidos y la versión YOLOv4-tiny-3l porque se esperaba que proporcionara mejores resultados que YOLOv4-tiny debido a que tiene una salida más en tiempos de inferencia similares. De la misma forma que YOLOv4, la implementación de YOLOv5 presenta versiones n, s, m, l y x con diferente precisión y

velocidad de detección (Sozzi et al., 2022), por lo tanto, se implementaron las versiones s, m y l.

Al igual que YOLOv4, YOLOv5 se basa en las arquitecturas de CSPNet + Darknet53 (columna vertebral), SSP + PANet (cuello) y YOLOv3 Head para la detección de objetos. Los últimos cambios en la arquitectura (V6.0/6.1) están en la primera capa FOCUS por el equivalente de CBS con entradas $[N_{Kernels} = 64, Kernel = 6, Stride = 2, Padding = 2]$ y SPP por un equivalente llamado SPPF, mejorando los tiempos de entrenamiento e inferencia de la red. Se mantiene la estructura para todas las variantes de YOLOv5 (Figura 3.7), solo se modifica el ancho y la profundidad de la red.

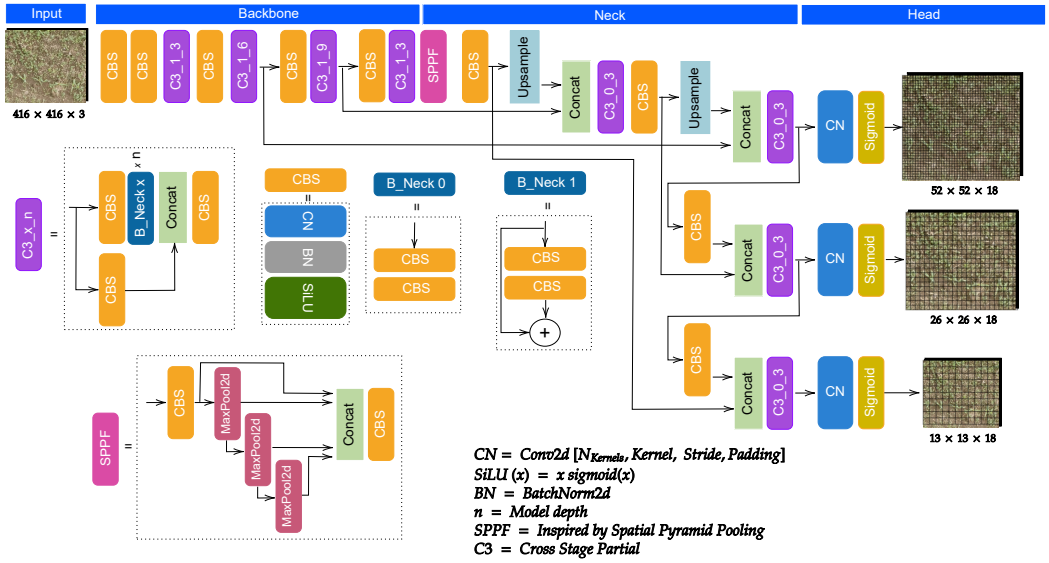


Figura 3.7: *Diagrama de la arquitectura de YOLOv5-l (V6.0/6.1) con una imagen de entrada de 416 × 416 píxeles y 3 canales.*

La modificación de la profundidad de la red se realizó tomando el entero positivo de la multiplicación de los bloques B_Neck por un factor y el ancho de la red multiplicando el número de filtros por un factor, como se muestra en Tabla 3.4.

Cuadro 3.4: *Versiones de YOLOv5*

Versión	Profundidad de la arquitectura	Ancho de capas
YOLOv5-n	0.33	0.25
YOLOv5-s	0.33	0.50
YOLOv5-m	0.67	0.75
YOLOv5-l	1.00	1.00
YOLOv5-x	1.25	1.25

La implementación de YOLOv4 se basó en el marco darknet escrito en el lenguaje de programación C y YOLOv5 en la biblioteca Pytorch implementada en Python, ambas herramientas de código abierto. Para el entrenamiento de los algoritmos se utilizaron los hiperparámetros propuestos en cada implementación de cada CNN optimizada para la base de datos COCO, los cuales se describen en detalle en la Tabla 3.5.

Cuadro 3.5: *Hiperparametros de entrenamiento de los algoritmos.*

Algoritmo	**Tamaño de imagen**	**Lote**	**Optimizador**	**Tasa de aprendizaje**	**Decaemiento (% iteraciones)**	**Max (iteraciones)**	**Pesos preentrenados**
YOLOv4	416 x 416 x 3	64	SGD	0.0013	25, 80 y 90	10000	COCO
YOLOv4-tiny		64	SGD	0.00261	80 y 90	20000	COCO
YOLOv4-tiny-3l		64	SGD	0.00261	80 y 90	20000	COCO
YOLOv5-s		179	Adam	0.01	Automático	200	COCO
YOLOv5-m		99	Adam	0.01	Automático	200	COCO
YOLOv5-l		179	Adam	0.01	Automático	200	COCO

3.4.3. Métricas de evaluación

Se utilizaron las métricas Precision (Pr), Reacall (Rc), Mean average precision (mAP) y F1-Score, comúnmente utilizadas para evaluar los resultados en trabajos de detección de objetos (Padilla, Passos, Dias, Netto & da Silva, 2021). Al considerarse soló una clase, el valor de mAP es igual a Average precision (AP). El cálculo de AP se realizó con interpretación de todos los puntos (APall) (Padilla et al., 2021), adoptada en Pascal Challenge (Everingham et al., 2015). Pr es el porcentaje de predicciones positivas correctas (Padilla et al., 2021).

$$\mathrm{Pr} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{3.1}$$

Rc es el porcentaje de predicciones positivas correctas entre todas las verdades fundamentales dadas (Padilla et al., 2021).

$$\text{Rc} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3.2}$$

Donde:

- TP (Verdadero positivo): una detección correcta de un cuadro delimitador de la verdad fundamental, sí su área de intersección sobre el área de unión (IoU) con el cuadro delimitador etiquetado correspondiente es mayor que un umbral determinado
- FP (Falso positivo): una detección incorrecta de un objeto inexistente o una detección fuera de lugar de un objeto existente.
- FN (Falso negativo): un cuadro delimitador de la verdad fundamental no detectado.

La puntuación F1 se define como la media armónica de la precision y recall de un detector.

$$\text{F1-Score} = 2 * \frac{\text{Pr} * \text{Rc}}{\text{Pr} + \text{Rc}}, \tag{3.3}$$

Para determinar las métricas anteriores, se utilizó un software de código abierto desarrollado por Padilla et al. (2021) y se modificó el código fuente para evaluar todos los modelos CNNs propuestos.

Para medir el rendimiento del conteo de los modelos se utilizó el coeficiente de determinación (R^2) (Yang, Gao, Gao & Zhu, 2021) y el error cuadrático medio relativo (rRMSE) propuesta en Daubige et al. (2021), donde se considera bueno un rRMSE <5 %, satisfactorio entre 5 % <rRMSE <10 %, pobre entre 10 % <rRMSE <20 % y muy pobre rRMSE >20 %.

3.5. Resultados

3.5.1. Entrenamiento

El proceso de entrenamiento de los algoritmos neuronales se realizó fuera de línea, utilizando los servicios de Google Colab Pro, que proporciona un entorno virtual con una Unidad de Procesamiento de Gráficos (GPU). En la Tabla 3.6 se muestra la duración del entrenamiento en horas, número de iteraciones y la GPU asignada en cada modelo entrenado.

Cuadro 3.6: *Tiempo de entrenamiento para cada algoritmo neuronal*

Algoritmo	GPU	Tiempo de entrenamiento (horas)	N Iteraciones
YOLOv4	Tesla T4 - 15GB	27.4	10000
YOLOv4-tiny	Tesla P100-PCIE-16GB	4.7	20000
YOLOv4-tiny-3l	Tesla T4 - 15GB	7.9	20000
YOLOv5-s	Tesla P100-PCIE-16GB	3.6	200
YOLOv5-m	Tesla P100-PCIE-16GB	10	200
YOLOv5-l	Tesla T4 - 15GB	7.16	145

La Figura 3.8 muestra el comportamiento de la métrica mAP para cada modelo en el conjunto de datos de prueba durante el entrenamiento. Se puede ver que, para la red YOLOv4, el mAP se mantuvo en valores del 73 % después de la iteración 2500, sin una mejora significativa en las iteraciones posteriores. En el caso de los modelos YOLOv4-tiny y YOLOv4-tiny-3l, el valor de mAP se mantuvo en el rango de 60 % a 65 % hasta 16,000 iteraciones, alcanzando valores máximos de 68 % y 69 % después de aplicar un decaimiento en la tasa de aprendizaje . Para las versiones YOLOv5-s y YOLOv5-m, el mAP se mantuvo estable después de la época 75 y aumentó ligeramente en épocas posteriores. Para YOLOv5-l, los valores de mAP se estabilizaron y aumentaron hasta la época 75, donde el aprendizaje se mantuvo constante y, al no haber mayor mejora, el proceso se detuvo en la época 145.

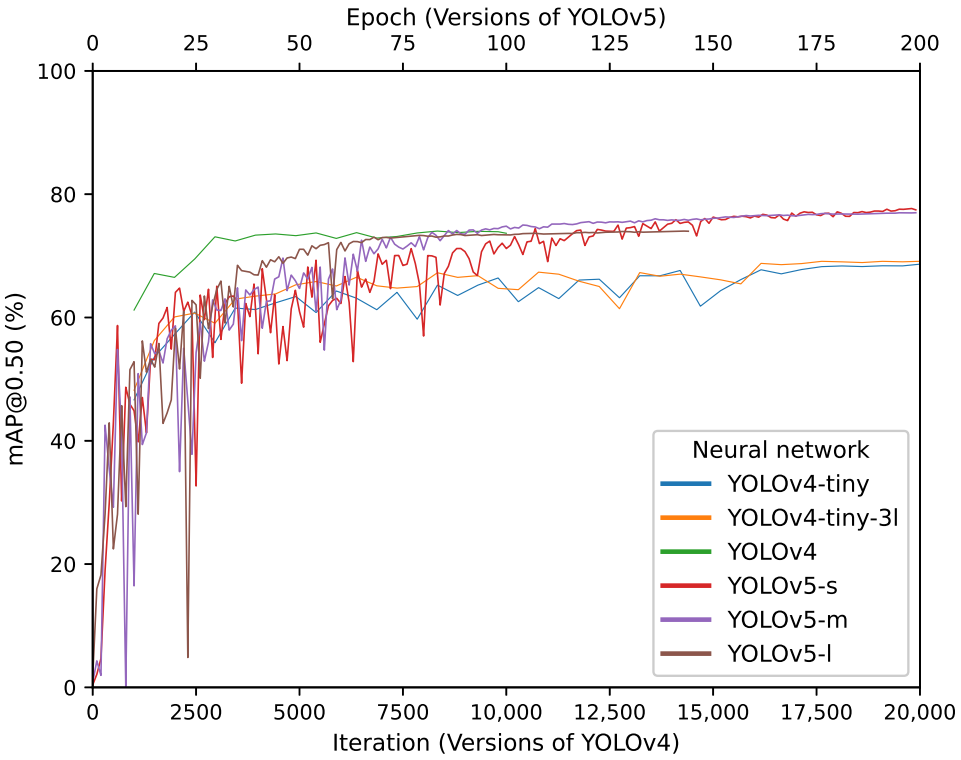


Figura 3.8: *mAP@0.50 calculado para el conjunto de prueba durante el entrenamiento de los algoritmos CNN con una confianza de 0.25.*

La puntuación máxima de mAP@0.50 en el conjunto de datos de prueba fue obtenida por el modelo YOLOv5-s con un valor de 77.6 % y F1-Score de 73.0 %, seguido por el modelo YOLOv5-m. Aunque el modelo YOLOv4 es mejor en cuanto a la métrica Rc, con un valor del 77 %, obtuvo un Pr bajo, lo que penaliza el F1-Score y mAP. La Tabla 3.7 muestra las métricas Pr, Rc, F1-Score y mAP obtenidas por cada modelo en el conjunto de datos de prueba con más detalle.

Cuadro 3.7: *Métricas del conjunto de prueba para una confianza de 0.25 y IoU de 0.50*

Modelo	Pr	Rc	F1	mAP@0.50
YOLOv4	0.650	0.770	0.700	0.736
YOLOv4-tiny	0.620	0.730	0.670	0.686
YOLOv4-tiny-3l	0.680	0.670	0.670	0.691
YOLOv5-s	0.720	0.742	0.730	0.776
YOLOv5-m	0.700	0.748	0.723	0.769
YOLOv5-l	0.683	0.725	0.703	0.740

3.5.2. Evaluación

El F1-Score para cada CNN se determinó en valores umbral IoU de 0.25, 0.50 y 0.75, para valores de confianza en el rango de 0.05 a 1.00 en intervalos de 0.05, obteniendo los resultados que se muestran en la Figura 3.9 para los datos de evaluación. Los F1-Scores máximos se obtuvieron con los mismos valores de confianza para un umbral IoU de 0.50 y 0.25, con un aumento medio del 4.92 % para cada modelo al pasar del umbral IoU de 0.50 a 0.25.

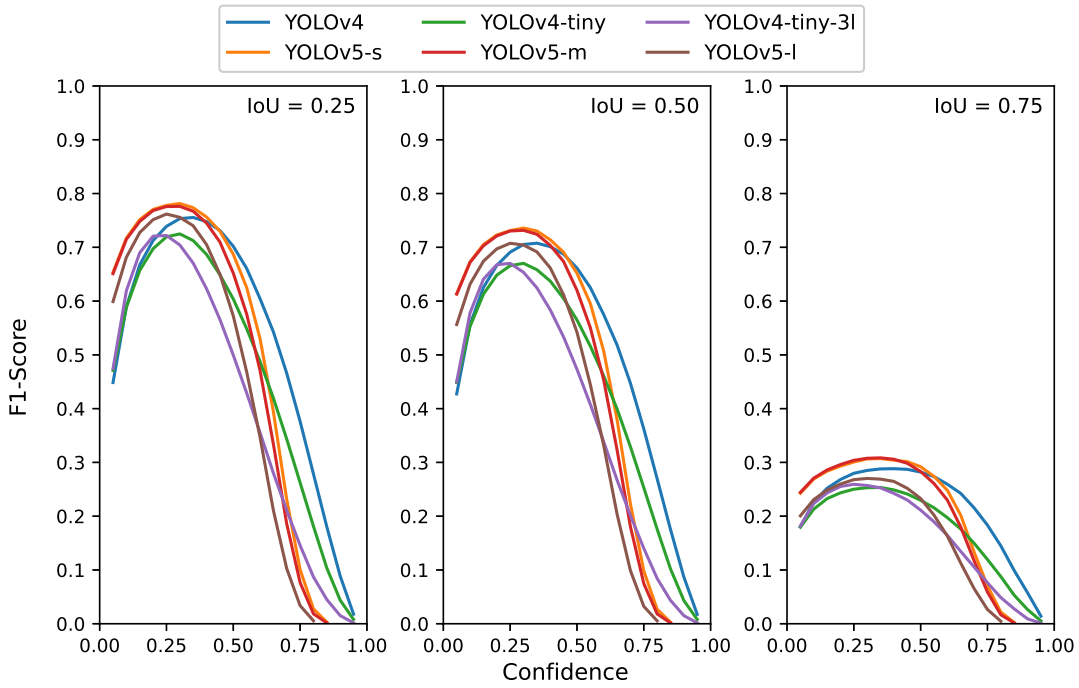
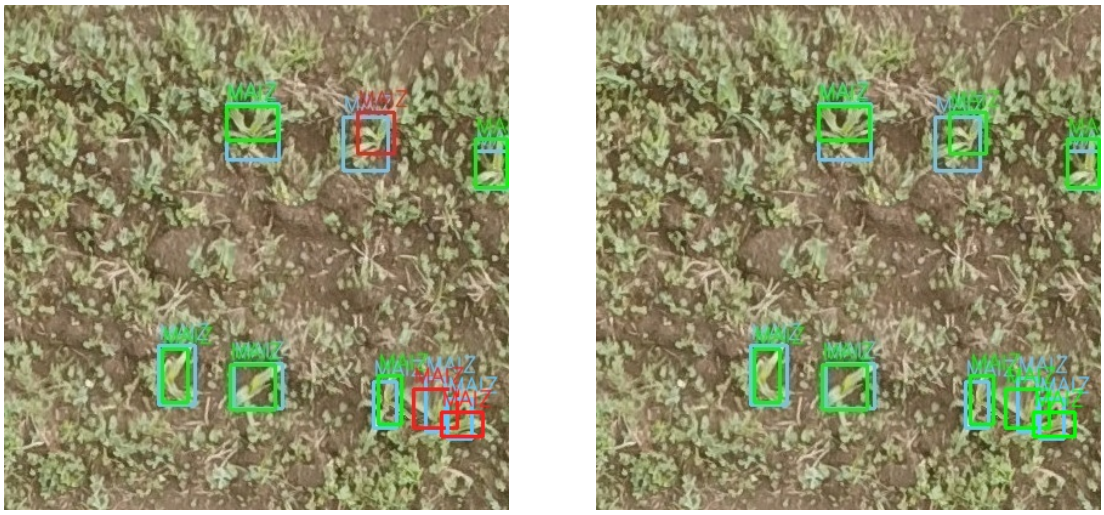


Figura 3.9: *Curvas de puntuación F1 vs confianza en los umbrales IoU 0.25, 0.50 y 0.75 para cada modelo entrenado.*

La Figura 3.10a y 3.10b muestra un ejemplo del impacto en TP y FP al evaluar YOLOv5-l con umbrales IoU de 0.5 y 0.25. Estas imágenes pertenecen a la etapa V4 con GSD de 0.33 cm/píxel.



(a) **(b)**

Figura 3.10: *Detecciones de la arquitectura YOLOv5l para una confianza de 0.3 con umbral IoU de 0.5 (**a**) y 0.25 (**b**). Los recuadros azules representan la verdad fundamental, los verdes TP y rojos FP*

La Tabla 3.8 muestra los resultados para las métricas Pr, Rc, F1, mAP, rRMSE y R^2 para cada modelo CNN evaluado. Los valores de TP, FP y FN se muestran normalizados de 0 a 1 y FP se expresa con relación a la suma de TP + FN. Las puntuaciones más altas de F1 se obtuvieron con los modelos YOLOv5-s y YOLOv5-m, con valores de 0.7814 y 0.776, respectivamente. En cuanto a la métrica Rc, el modelo YOLOv4 supera a otros modelos con un valor de 80.78 %, seguido de YOLOv5-s con un valor de 79.37 %. En términos de conteo de plantas, el modelo YOLOv4 tuvo la correlación más alta con R^2 de 0.81 y rRMSE de 39.55 %, seguido por el modelo YOLOv5-s con R^2 de 0.78 y rRMSE de 42.06 %.

Cuadro 3.8: *Resultados para cada modelo obtenidos en el conjunto de datos de evaluación*

Modelo	**Pr**	**Rc**	**F1**	**mAP**	**TP**	**FP**	**FN**	**rRMSE**	R^2
YOLOv4	0.705	0.807	0.753	0.720	0.807	0.336	0.192	0.395	0.811
YOLOv4-tiny	0.704	0.746	0.724	0.649	0.746	0.312	0.253	0.486	0.711
YOLOv4-tiny-3l	0.770	0.648	0.704	0.580	0.648	0.193	0.351	0.638	0.497
YOLOv5-s	0.769	0.793	0.781	0.731	0.793	0.237	0.206	0.420	0.785
YOLOv5-m	0.777	0.775	0.776	0.716	0.775	0.222	0.224	0.461	0.742
YOLOv5-l	0.763	0.747	0.755	0.685	0.747	0.231	0.252	0.534	0.654

Con el fin de analizar los valores obtenidos en la Tabla 3.8 se trazaron las curvas Recall vs Presicion para cada una de las etapas vegetativas con su resolución espacial, estos resultados se muestran en la Figura 3.11. Para las etapas $V3_{0,33}$ y $V4_{0,33}$ los modelos se comportaron de manera consistente con valores de Pr entre 77 % y 85 %, y Rc por arriba del 90 %, excepto para el modelo YOLOv4-tiny-3l en el cual decae hasta el valor de 85 %. El modelo YOLOv4 y las versiones de YOLOv5 mantuvieron los resultados de 70 % <Pr <80 % y 85 % <Rc <90 % para las etapas vegetativas $V5_{0,33}$, $V6_{0,33}$ y $V7_{0,33}$, donde la puntuación más alta se obtuvo en la etapa $V5_{0,33}$ seguida por $V7_{0,33}$.

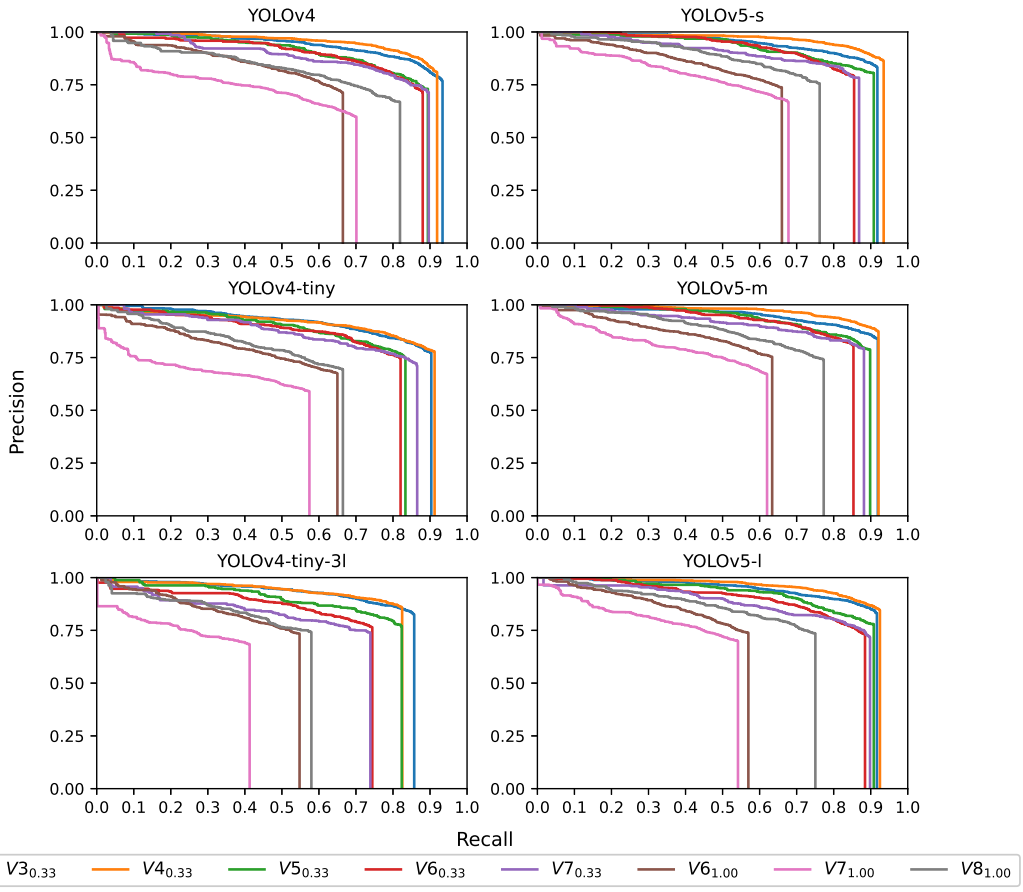


Figura 3.11: *Curvas Recall vs Precision por etapa vegetativa y resolución espacial.*

Comparando la estimación del número de plantas por imagen para cada modelo se obtuvo el rRMSE (Figura 3.12). Los mejores resultados se obtuvieron en las etapas vegetativas V3, V4 y V5 con un GSD de 0.33 cm/píxel con valores de rRMSE entre el 10 y 20 %, error que aumenta en etapas vegetativas superiores a V5.

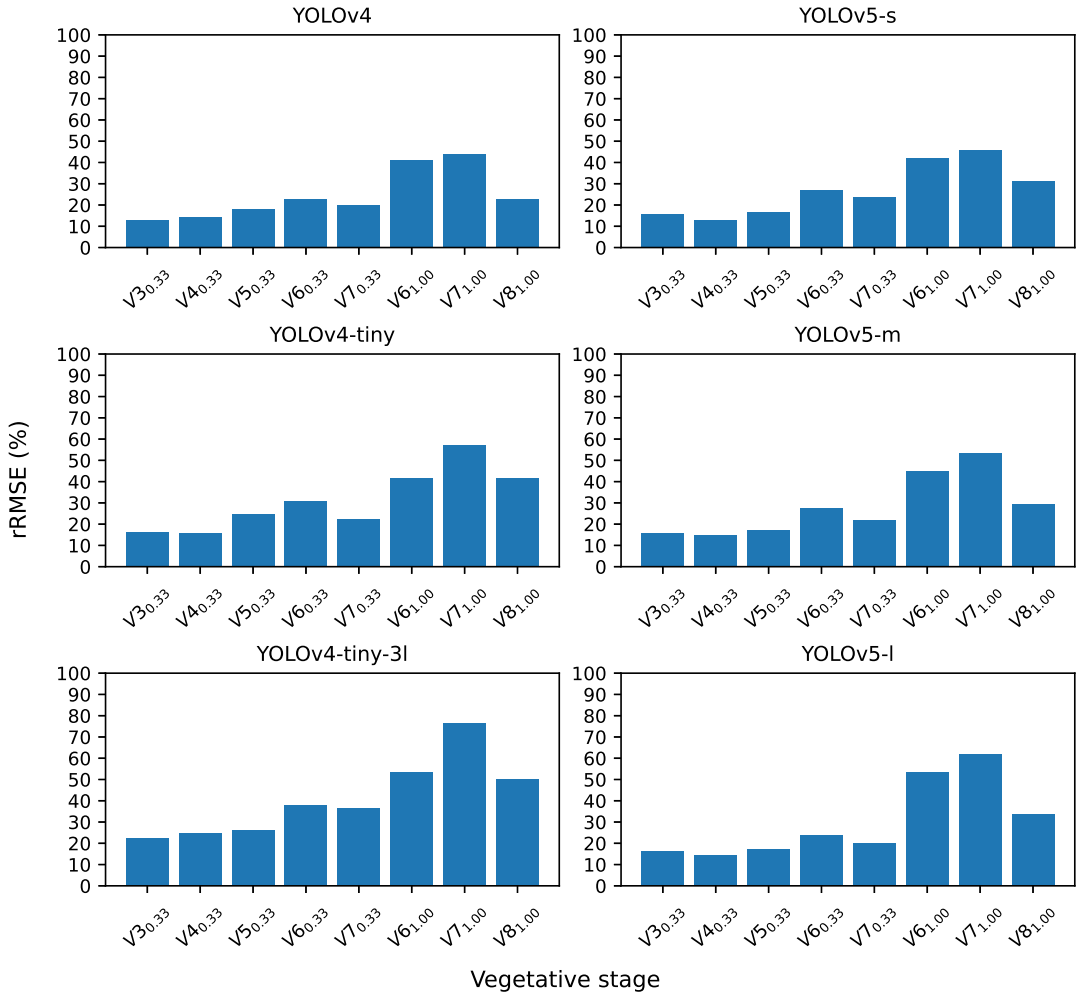


Figura 3.12: *rRMSE obtenido por cada modelo por etapa vegetativa y resolución espacial.*

El coeficiente R^2 determina la relación entre las plantas reales y el número de plantas estimadas por la red, considerando una confianza de 0.3 y un umbral IoU de 0.25 se obtuvieron valores superiores a 0.85 con el modelo YOLOv4, para las etapas vegetativas V3, V4, V5, V6 y V7 con GSD de 0.33 cm/píxel. En la Figura 3.13 se muestran con más detalle los resultados obtenidos para cada una de las arquitecturas CNN.

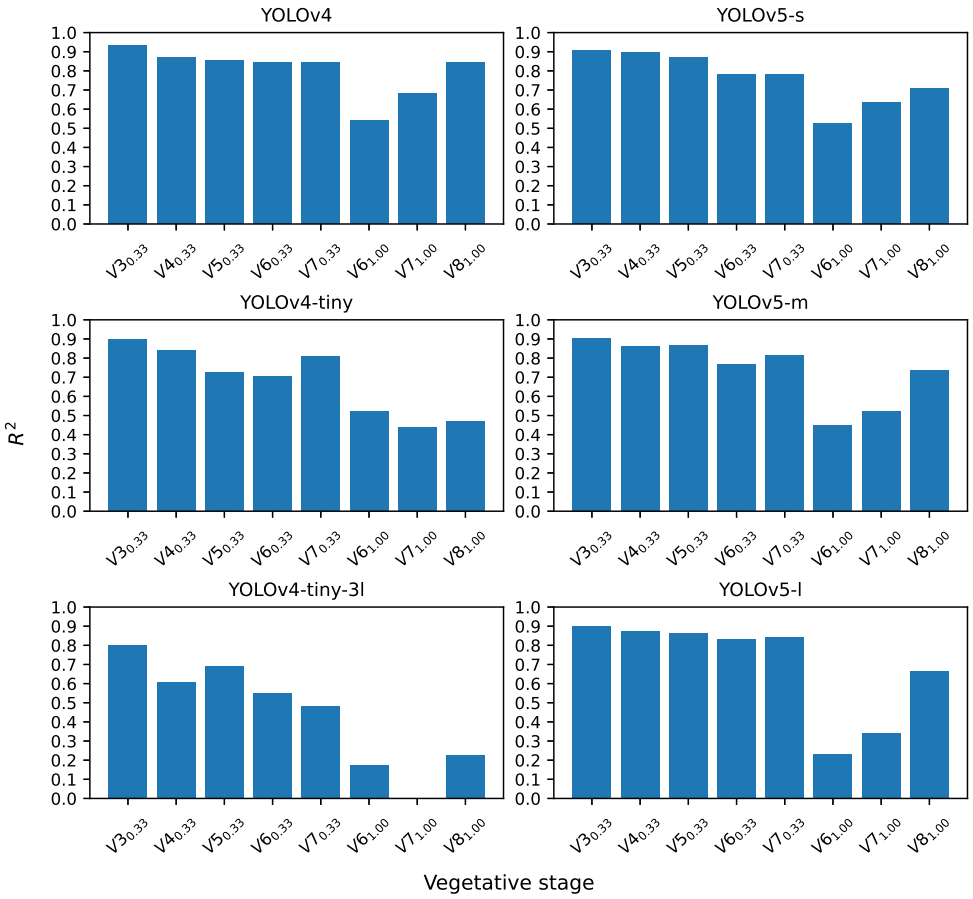


Figura 3.13: *R^2 determinado para cada etapa vegetativa considerando las detecciones con confianza mayor a 0.30 y IoU de 0.25*

Las detecciones se inspeccionaron visualmente en busca de errores. Se observó que los TP en los estados vegetativos V3, V4 y V5 con GSD de 0.33 cm/píxel fueron causados principalmente por hojas de plantas de maíz en los bordes de la imagen, y en algunos casos en V3 se confundieron con malezas, como se muestra en la Figura 3.14a. Para las etapas vegetativas posteriores a V5 y GSD de 1.00 cm/píxel, se observó que los FP se debían principalmente a falta de etiquetas, debido a que estas no se realizaron por la complejidad del etiquetado manual, tal como se muestra en la Figura 3.14b, 3.14c y 3.14d.

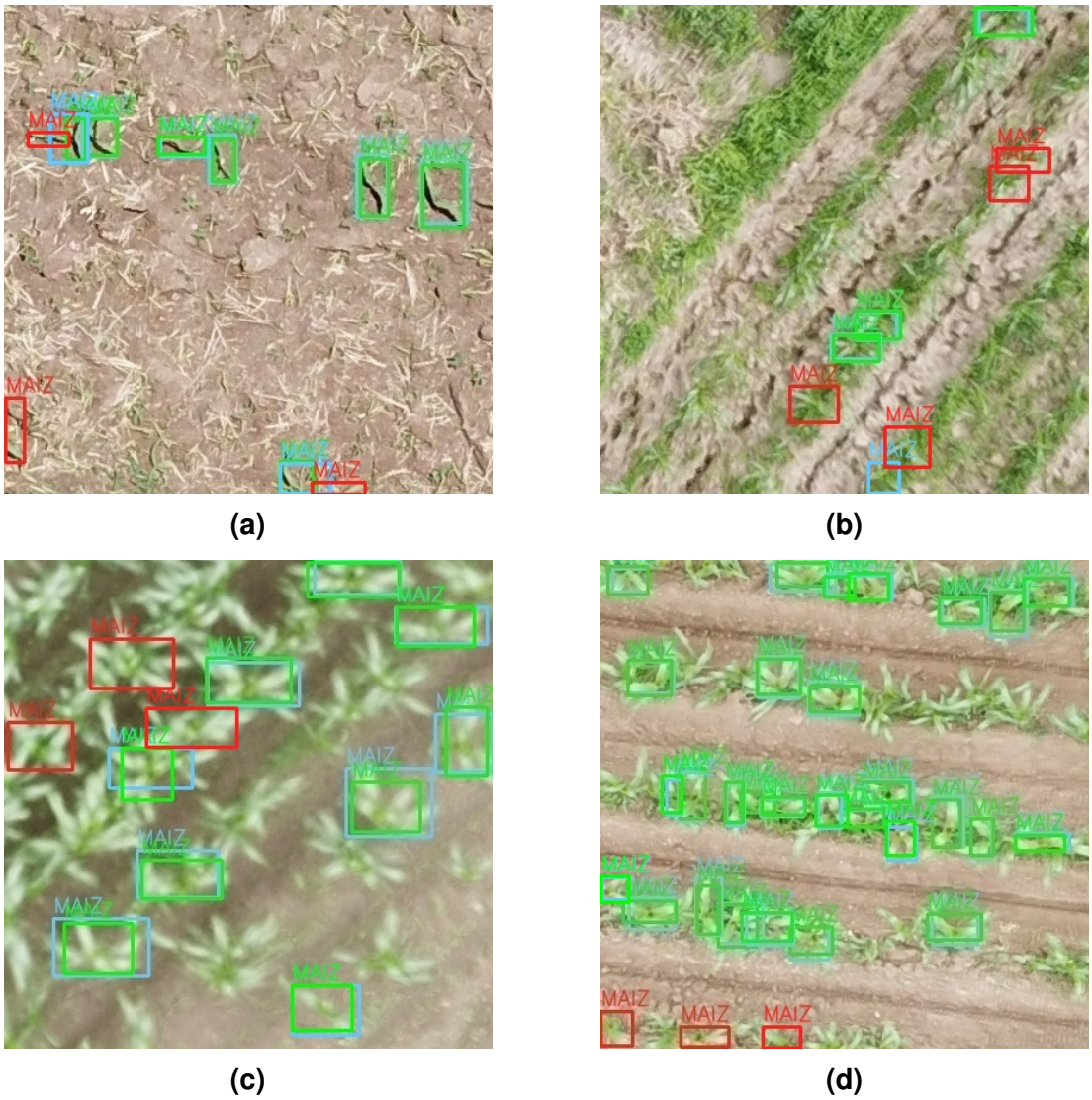


Figura 3.14: *Visualización de las imágenes etiquetadas manualmente (recuadro azul), TP (recuadro verde) y FP (recuadro rojo). Detecciones para etapa vegetativa (**a**) $V3$ con YOLOv4, (**b**) $V7_{1,00}$ con YOLOv5s, (**c**) $V8_{1,00}$ con YOLOv5s y (**d**) $V6_{1,00}$ con YOLOv5s*

3.6. Discusión

El valor de confianza para la evaluación se eligió con la moda. Cuando los modelos alcanzaron el F1-Score máximo, este valor de 0.3 es inferior al de Velumani et al. (2021), quienes reportan una confianza de 0.5 al evaluar la arquitectura Faster-RCNN, indicando mejores resultados en cuanto a clasificación de plantas. Los modelos basados en YOLOv4 son más confiables en términos de clasificación de plantas de maíz.

La mayoría de los trabajos sobre detección de objetos en grandes conjuntos de datos evalúan los modelos CNN en umbrales de IoU superiores a 0.5 (Padilla et al., 2021). Analizando las gráficas en la Figura 3.9 considerar umbrales IoU mayores a 0.5 representa una disminución de la métrica F1, indicando que los modelos pierden precisión en la estimación del tamaño de la planta de maíz. Esto se puede apreciar en la Figura 3.14a, donde se observa que en algunos casos la predicción de la etiqueta no incluye las hojas de la planta y al no superar el umbral IoU de 0.5 se considerarían predicciones FP. Al igual que en Velumani et al. (2021), se evaluó la métrica F1-Score en un umbral IoU de 0.25. Se logró un incremento promedio de 4.92 % para todos los modelos YOLO a partir del umbral IoU 0.50. Para efectos del conteo y detección de plantas no se considera crítica la estimación precisa de las dimensiones de la planta (Velumani et al., 2021). En consecuencia, el valor umbral de IoU de 0.25 y una confianza de 0.3 se utilizaron para explicar mejor el tamaño más pequeño de los cuadros delimitadores detectados y la clasificación de las plantas de maíz, como se hizo en Velumani et al. (2021).

En relación con el conteo de plantas, YOLOv4 tiene mayor cantidad de TP, por lo que se correlaciona mejor con el verdadero número de plantas $R^2 = 0{,}81$ y rRMSE = 39.55 %, seguido del modelo YOLOv5-s con $R^2 = 0{,}78$ y rRMSE = 42.06 %. Aunque hay una alta correlación con el número real de plantas en ambos modelos, de acuerdo con Daubige et al. (2021) aún se considerarían resultados muy pobres al tener valores rRMSE mayores al 20 %.

Para un mejor análisis de los datos, se evaluaron los modelos para cada etapa de crecimiento de la planta y su resolución espacial. Para la evaluación de los modelos en las etapas $V3_{0,33}$ y $V4_{0,33}$, a excepción de YOLOv4 Tiny 3L, el desempeño de los resultados es consistente con lo reportado en la literatura. Se encontraron resultados similares en Daubige et al. (2021), quien reportó 10 % <rRMSE <20 % para las etapas $V3$ y $V4$ bajo condiciones moderadas de maleza. En Gnädinger y Schmidhalter (2017) se reporta correlación $R^2 = 0{,}89$ para las etapas V3 y V5, mientras que en García-Martínez et al. (2020) se reportan $R^2 = 0{,}98$ para la etapa V2.

Para los modelos YOLOv4-tiny y YOLOv4-tiny-3l, los resultados decaen considerablemente desde la etapa $V5_{0,33}$, lo cual es comprensible, ya que reducen la cantidad de capas convolucionales. Además, se rechazó la idea de que YOLOv4-tiny-3l tendría mejores resultados que YOLOv4-tiny al tener una salida más.

Los modelos YOLOv4 y las versiones de YOLOv5 evaluados en las etapas $V5_{0,33}$, $V6_{0,33}$ y $V7_{0,33}$ mantienen los resultados de 70 % <presion <80 % y 85 % <recall <90 % , con los mejores puntajes en $V5_{0,33}$ seguido de $V7_{0,33}$. Debido a que los valores de rMSE para $V6_{0,33}$ y $V7_{0,33}$ superan el 20 % los resultados se consideran muy pobres y pobres para $V5_{0,33}$. Estos resultados son consistentes con la limitación mencionada por Varela et al. (2018), donde las plantas son propensas a la superposición de hojas, lo que reduce el rendimiento general de la arquitectura YOLO evaluada en este trabajo.

Una inspección visual de las detecciones realizadas por cada modelo YOLO, ayudó a comprender que los FP en etapas inferiores a V5 con GSD de 0.33 cm/píxel se deben a detecciones realizadas en los bordes de las imágenes y en casos aislados por confusión con maleza. En estos casos el recuento de FP se puede disminuir al filtrar los resultados con valores de confianza superiores a 0.30. Para las etapas $V6_{0,33}$, $V7_{0,33}$, $V6_{1,00}$ y $V7_{1,00}$ los FP se deben en su mayoría a predicciones realizadas en plantas no etiquetadas. Aunque el etiquetado parcial no se recomiende en tareas abordadas con aprendizaje supervisado, en este caso fue sumamente complicado el etiquetado completo debido a diversos errores en la imagen. Aun así, la robustez del modelo YOLOv5-s para detectar plantas de maíz bajo condiciones de malezas altamente complejas se puede observar en la Figura 3.14b.

Si bien en el trabajo de Velumani et al. (2021) se evaluó el efecto de la resolución espacial en la detección de plantas de maíz, obteniendo mejores resultados con un GSD de 0.3 cm/pixel en estadios entre V3 y V5, en este trabajo se observó que, para estadios superiores a V5, se debe considerar un GSD superior a 0,3 pero inferior a 1,00 cm/píxel porque las imágenes se vuelven difíciles de

interpretar visualmente para el etiquetado.

Finalmente, debido a las características de la cámara montada en el dron utilizado en este trabajo de investigación, la altura de vuelo a la que se obtuvieron mejores resultados fue de 10 m (GSD = 0,33 cm/pixel), lo que hace inviable un despliegue a gran escala debido a la limitada capacidad de adquisición de datos. Se requieren mejores cámaras que permitan la adquisición de imágenes más nítidas con detalles a nivel de planta en alturas de vuelo más altas para obtener mejores resultados al detectar plantas de maíz y hacer que la aplicación sea factible. Otra limitación de este estudio es que no se exploró un rango de GSD para determinar un óptimo para la detección de plantas de maíz en etapas vegetativas por encima de V5.

3.7. Conclusiones

En este trabajo de investigación se creó una base de datos de imágenes aéreas de cultivos de maíz con diferentes niveles de infestación de malezas y distancia de muestreo en tierra. La detección y conteo de plantas de maíz se evaluó utilizando las arquitecturas YOLOv4, YOLOv4-tiny, YOLOv4-tiny-3l, YOLOv5-s, YOLOv5-m y YOLOv5-l. Se demostró que las arquitecturas YOLOv5 y YOLOv4 son robustas para detectar y contar plantas de maíz en etapas inferiores a V5 en imágenes de alta resolución (GSD = 0.33 cm/píxel) incluso en condiciones de infestación de malezas, obteniendo resultados de Pr entre 77 y 85 %, un Rc por encima del 90 % y rRMSE entre el 10 y el 20 %.

Sin embargo, en el caso de etapas posteriores a V5 con GSD de 1.00 cm/pixel, los resultados no fueron favorables, debido a la baja calidad de las imágenes, que ni siquiera permitieron el etiquetado completo de las plantas de maíz. Las imágenes de alta resolución son cruciales para mejorar los resultados en la detección de plantas; por lo tanto, se recomienda determinar un GSD óptimo para la adquisición de imágenes aéreas en etapas posteriores a V5.

También se observó el efecto de considerar diferentes valores de confianza y umbrales de IoU como modelos de detección de evaluación. En este caso,

YOLOv4 tiene niveles de confianza más altos que las versiones de YOLOv5, aunque las versiones de YOLOv5 son más precisas para determinar la ubicación y el tamaño de la planta. Los errores más grandes en los recuentos de plantas se obtuvieron en el caso de las versiones diminutas de YOLOv4 debido al número reducido de capas convolucionales.

Finalmente, para que la detección de plantas sea factible a mayor escala, una dirección para el trabajo futuro sería explorar el uso de arquitecturas de súper resolución acopladas a un detector entrenable de extremo a extremo, resolviendo el problema de adquirir imágenes de baja resolución.

Bibliografía

Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. doi:10.48550/ARXIV.2004.10934. (Vid. pág. 59)

Brewer, K. ., Clulow, A. ., Sibanda, M. ., Gokool, S. ., Naiken, V. . & Mabhaudhi, T. . (2022). Predicting the Chlorophyll Content of Maize over Phenotyping as a Proxy for Crop Health in Smallholder Farming Systems. *Remote Sensing*, *14*(3), 518. doi:10.3390/rs14030518. (Vid. pág. 54)

Daubige, Joudelat, Burger, Comar, de Solan & Baret. (2021). Plant detection and counting from high-resolution RGB images acquired from UAVs: comparison between deep-learning and handcrafted methods with application to maize, sugar beet, and sunflower. *bioRxiv*. doi:10.1101/2021.04.27.441631. (Vid. págs. 51, 54, 63, 73)

Everingham, M., Eslami, S., Van Gool, L., Williams, C. K., Winn, J. & Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, *111*(1), 98-136. (Vid. pág. 62).

Fan, Z. ., Lu, J. ., Gong, M. ., Xie, H. . & Goodman, E. D. (2018). Automatic Tobacco Plant Detection in UAV Images via Deep Neural Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(3), 876-887. doi:10.1109/jstars.2018.2793849. (Vid. pág. 49)

Flores-Cruz, L. A., García-Salazar, J. A., Mora-Flores, J. S. & Pérez-Soto, F. (2014). Producción de maíz (Zea mays L.) en el Estado de Puebla: un enfoque de equilibrio espacial para identificar las zonas productoras más competitivas. *Agricultura, sociedad y desarrollo*, *11*, 223-239. (Vid. pág. 49).

García-Martínez, H. ., Flores-Magdaleno, H. ., Khalil-Gardezi, A. ., Ascencio-Hernández, R. ., Tijerina-Chávez, L. ., Vázquez-Peña, M. A. & Mancilla-

Villa, O. R. (2020). Digital Count of Corn Plants Using Images Taken by Unmanned Aerial Vehicles and Cross Correlation of Templates. *Agronomy*, *10*(4), 469. doi:10.3390/agronomy10040469. (Vid. págs. 50, 73)

Gnädinger, F. . & Schmidhalter, U. . (2017). Digital Counts of Maize Plants by Unmanned Aerial Vehicles (UAVs). *Remote Sensing*, *9*(6), 544. doi:10.3390/rs9060544. (Vid. págs. 50, 73)

Gómez-Ramos, M., Ruíz-Castilla, J. & García-Lamont, F. (2020). Clasificación de plantas de maíz y maleza: Hacia la mejora de la fertilización en México. *Research in Computing Science*, *149*(8), 683-697. (Vid. págs. 50, 51).

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., ... Minh, M. T. (2022). Ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference. Consultado el 26 de octubre de 2022, desde https://doi.org/10.5281/zenodo.6222936. (Vid. pág. 59)

Khaki, S. ., Safaei, N. ., Pham, H. . & Wang, L. . (2022). WheatNet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. *Neurocomputing*, *489*, 78-89. doi:10.1016/j.neucom.2022.03.017. (Vid. pág. 49)

Kitano, B. T., Mendes, C. C. T., Geus, A. R., Oliveira, H. C. & Souza, J. R. (2019). Corn Plant Counting Using Deep Learning and UAV Images. *IEEE Geoscience and Remote Sensing Letters*, 1-5. doi:10.1109/lgrs.2019.2930549. (Vid. pág. 49)

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., ... Dollár, P. (2014). Microsoft COCO: Common Objects in Context. doi:10.48550/ARXIV.1405.0312. (Vid. pág. 57)

Liu, H. ., Sun, H. ., Li, M. . & Iida, M. . (2020). Application of Color Featuring and Deep Learning in Maize Plant Detection. *Remote Sensing*, *12*(14), 2229. doi:10.3390/rs12142229. (Vid. pág. 50)

Messina, G. . & Modica, G. . (2020). Applications of UAV Thermal Imagery in Precision Agriculture: State of the Art and Future Research Outlook. *Remote Sensing*, *12*(9), 1491. doi:10.3390/rs12091491. (Vid. pág. 49)

Nepal, U. . & Eslamiat, H. . (2022). Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs. *Sensors*, *22*(2), 464. doi:10.3390/s22020464. (Vid. pág. 59)

Oh, S. ., Chang, A. ., Ashapure, A. ., Jung, J. ., Dube, N. ., Maeda, M. ., ... Landivar, J. . (2020). Plant Counting of Cotton from UAS Imagery Using Deep Learning-Based Object Detection Framework. *Remote Sensing*, *12*(18), 2981. doi:10.3390/rs12182981. (Vid. pág. 49)

Osco, L. P., dos Santos de Arruda, M. ., Gonçalves, D. N., Dias, A. ., Batistoti, J. ., de Souza, M. ., ... Gonçalves, W. N. (2021). A CNN approach to simultaneously count plants and detect plantation-rows from UAV imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *174*, 1-17. doi:10.1016/j.isprsjprs.2021.01.024. (Vid. págs. 49, 51)

Padilla, R. ., Passos, W. L., Dias, T. L. B., Netto, S. L. & da Silva, E. A. B. (2021). A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. *Electronics*, *10*(3), 279. doi:10.3390/electronics10030279. (Vid. págs. 62, 63, 73)

Panday, U. S., Pratihast, A. K., Aryal, J. . & Kayastha, R. B. (2020). A Review on Drone-Based Data Solutions for Cereal Crops. *Drones*, *4*(3), 41. doi:10.3390/drones4030041. (Vid. pág. 49)

Pang, Y. ., Shi, Y. ., Gao, S. ., Jiang, F. ., Veeranampalayam-Sivakumar, A. N., Thompson, L. ., ... Liu, C. . (2020). Improved crop row detection with deep neural network for early-season maize stand count in UAV imagery. *Computers and Electronics in Agriculture*, *178*, 105766. doi:10.1016/j.compag.2020.105766. (Vid. pág. 50)

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. En *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 779-788). doi:10.1109/CVPR.2016.91. (Vid. pág. 59)

Redmon, J. & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv*. doi:10.48550/ARXIV.1804.02767. (Vid. pág. 59)

Santos, C. F. G. D. & Papa, J. P. (2022). Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks. *ACM Computing Surveys*, *54*(10s), 1-25. doi:10.1145/3510413. (Vid. pág. 58)

Shuai, G. ., Martinez-Feria, R. A., Zhang, J. ., Li, S. ., Price, R. . & Basso, B. . (2019). Capturing Maize Stand Heterogeneity Across Yield-Stability Zones Using Unmanned Aerial Vehicles (UAV). *Sensors*, *19*(20), 4446. doi:10.3390/s19204446. (Vid. pág. 50)

SIAP. (2022). Servicio de Información Agroalimentaria y Pesquera: Anuario estadístico de la producción agrícola. Consultado el 26 de octubre de 2022, desde https://nube.siap.gob.mx/cierreagricola/. (Vid. pág. 49)

Sozzi, M. ., Cantalamessa, S. ., Cogato, A. ., Kayad, A. . & Marinello, F. . (2022). Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms. *Agronomy*, *12*(2), 319. doi:10.3390/agronomy12020319. (Vid. págs. 59, 61)

Tzutalin. (2015). LabelImg. Consultado el 26 de octubre de 2022, desde https://github.com/tzutalin/labelImg. (Vid. pág. 55)

Valente, J. ., Sari, B. ., Kooistra, L. ., Kramer, H. . & Mücher, S. . (2020). Automated crop plant counting from very high-resolution aerial imagery. *Precision Agriculture*, *21*(6), 1366-1384. doi:10.1007/s11119-020-09725-3. (Vid. pág. 49)

Varela, S. ., Dhodda, P. ., Hsu, W. ., Prasad, P. V., Assefa, Y. ., Peralta, N. ., ... Ciampitti, I. . (2018). Early-Season Stand Count Determination in Corn via Integration of Imagery from Unmanned Aerial Systems (UAS) and Supervised Learning Techniques. *Remote Sensing*, *10*(3), 343. doi:10.3390/rs10020343. (Vid. págs. 49, 50, 74)

Velumani, K. ., Lopez-Lozano, R. ., Madec, S. ., Guo, W. ., Gillet, J. ., Comar, A. . & Baret, F. . (2021). Estimates of Maize Plant Density from UAV RGB Images Using Faster-RCNN Detection Model: Impact of the Spatial Resolution. *Plant Phenomics*, *2021*, 1-16. doi:10.34133/2021/9824843. (Vid. págs. 50, 59, 72-74)

Vong, C. N., Conway, L. S., Zhou, J. ., Kitchen, N. R. & Sudduth, K. A. (2021). Early corn stand count of different cropping systems using UAV-imagery

and deep learning. *Computers and Electronics in Agriculture*, *186*, 106214. doi:10.1016/j.compag.2021.106214. (Vid. pág. 50)

Voulodimos, A. ., Doulamis, N. ., Doulamis, A. . & Protopapadakis, E. . (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, *2018*, 1-13. doi:10.1155/2018/7068349. (Vid. pág. 58)

Wang, L. ., Xiang, L. ., Tang, L. . & Jiang, H. . (2021). A Convolutional Neural Network-Based Method for Corn Stand Counting in the Field. *Sensors*, *21*(2), 507. doi:10.3390/s21020507. (Vid. pág. 50)

Wang, Z. ., Wu, Y. ., Yang, L. ., Thirunavukarasu, A. ., Evison, C. . & Zhao, Y. . (2021). Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors*, *21*(10), 3478. doi:10.3390/s21103478. (Vid. pág. 57)

Wenkel, S. ., Alhazmi, K. ., Liiv, T. ., Alrshoud, S. . & Simon, M. . (2021). Confidence Score: The Forgotten Dimension of Object Detection Performance Evaluation. *Sensors*, *21*(13), 4350. doi:10.3390/s21134350. (Vid. pág. 59)

Yang, B. ., Gao, Z. ., Gao, Y. . & Zhu, Y. . (2021). Rapid Detection and Counting of Wheat Ears in the Field Using YOLOv4 with Attention Module. *Agronomy*, *11*(6), 1202. doi:10.3390/agronomy11061202. (Vid. pág. 63)

Apéndices

Apéndice A

Arquitecturas YOLO

A.1. Estructura de la red YOLOv4

Cuadro A.1: *Estructura de la red YOLOv4*

Index	Layer	Filters	Size/Strd(dil)	Input
0	conv	32	3 x 3/ 1	416 x 416 x 3
1	conv	64	3 x 3/ 2	416 x 416 x 32
2	conv	64	1 x 1/ 1	208 x 208 x 64
3	route 1			
4	conv	64	1 x 1/ 1	208 x 208 x 64
5	conv	32	1 x 1/ 1	208 x 208 x 64
6	conv	64	3 x 3/ 1	208 x 208 x 32
7	Shortcut Layer: 4	wt = 0, wn = 0, outputs: 208 x 208 x 64		
8	conv	64	1 x 1/ 1	208 x 208 x 64
9	route 8 2			
10	conv	64	1 x 1/ 1	208 x 208 x 128
11	conv	128	3 x 3/ 2	208 x 208 x 64
12	conv	64	1 x 1/ 1	104 x 104 x 128
13	route 11			
14	conv	64	1 x 1/ 1	104 x 104 x 128
15	conv	64	1 x 1/ 1	104 x 104 x 64
16	conv	64	3 x 3/ 1	104 x 104 x 64
17	Shortcut Layer: 14	wt = 0, wn = 0, outputs: 104 x 104 x 64		
18	conv	64	1 x 1/ 1	104 x 104 x 64

Continua en la pagina siguiente...

Cuadro A.1 – de la pagina anterior.

Index	Layer	Filters	Size/Strd(dil)	Input
19	conv	64	3 x 3/ 1	104 x 104 x 64
20	Shortcut Layer: 17	wt = 0, wn = 0, outputs: 104 x 104 x 64		
21	conv	64	1 x 1/ 1	104 x 104 x 64
22	route 21 12			
23	conv	128	1 x 1/ 1	104 x 104 x 128
24	conv	256	3 x 3/ 2	104 x 104 x 128
25	conv	128	1 x 1/ 1	52 x 52 x 256
26	route 24			
27	conv	128	1 x 1/ 1	52 x 52 x 256
28	conv	128	1 x 1/ 1	52 x 52 x 128
29	conv	128	3 x 3/ 1	52 x 52 x 128
30	Shortcut Layer: 27	wt = 0, wn = 0, outputs: 52 x 52 x 128		
31	conv	128	1 x 1/ 1	52 x 52 x 128
32	conv	128	3 x 3/ 1	52 x 52 x 128
33	Shortcut Layer: 30	wt = 0, wn = 0, outputs: 52 x 52 x 128		
34	conv	128	1 x 1/ 1	52 x 52 x 128
35	conv	128	3 x 3/ 1	52 x 52 x 128
36	Shortcut Layer: 33	wt = 0, wn = 0, outputs: 52 x 52 x 128		
37	conv	128	1 x 1/ 1	52 x 52 x 128
38	conv	128	3 x 3/ 1	52 x 52 x 128
39	Shortcut Layer: 36	wt = 0, wn = 0, outputs: 52 x 52 x 128		
40	conv	128	1 x 1/ 1	52 x 52 x 128
41	conv	128	3 x 3/ 1	52 x 52 x 128
42	Shortcut Layer: 39	wt = 0, wn = 0, outputs: 52 x 52 x 128		
43	conv	128	1 x 1/ 1	52 x 52 x 128
44	conv	128	3 x 3/ 1	52 x 52 x 128
45	Shortcut Layer: 42	wt = 0, wn = 0, outputs: 52 x 52 x 128		
46	conv	128	1 x 1/ 1	52 x 52 x 128
47	conv	128	3 x 3/ 1	52 x 52 x 128
48	Shortcut Layer: 45	wt = 0, wn = 0, outputs: 52 x 52 x 128		
49	conv	128	1 x 1/ 1	52 x 52 x 128

Continua en la pagina siguiente...

Cuadro A.1 – de la pagina anterior.

Index	Layer	Filters	Size/Strd(dil)	Input
50	conv	128	3 x 3/ 1	52 x 52 x 128
51	Shortcut Layer: 48	wt = 0, wn = 0, outputs: 52 x 52 x 128		
52	conv	128	1 x 1/ 1	52 x 52 x 128
53	route 52 25			
54	conv	256	1 x 1/ 1	52 x 52 x 256
55	conv	512	3 x 3/ 2	52 x 52 x 256
56	conv	256	1 x 1/ 1	26 x 26 x 512
57	route 55			
58	conv	256	1 x 1/ 1	26 x 26 x 512
59	conv	256	1 x 1/ 1	26 x 26 x 256
60	conv	256	3 x 3/ 1	26 x 26 x 256
61	Shortcut Layer: 58	wt = 0, wn = 0, outputs: 26 x 26 x 256		
62	conv	256	1 x 1/ 1	26 x 26 x 256
63	conv	256	3 x 3/ 1	26 x 26 x 256
64	Shortcut Layer: 61	wt = 0, wn = 0, outputs: 26 x 26 x 256		
65	conv	256	1 x 1/ 1	26 x 26 x 256
66	conv	256	3 x 3/ 1	26 x 26 x 256
67	Shortcut Layer: 64	wt = 0, wn = 0, outputs: 26 x 26 x 256		
68	conv	256	1 x 1/ 1	26 x 26 x 256
69	conv	256	3 x 3/ 1	26 x 26 x 256
70	Shortcut Layer: 67	wt = 0, wn = 0, outputs: 26 x 26 x 256		
71	conv	256	1 x 1/ 1	26 x 26 x 256
72	conv	256	3 x 3/ 1	26 x 26 x 256
73	Shortcut Layer: 70	wt = 0, wn = 0, outputs: 26 x 26 x 256		
74	conv	256	1 x 1/ 1	26 x 26 x 256
75	conv	256	3 x 3/ 1	26 x 26 x 256
76	Shortcut Layer: 73	wt = 0, wn = 0, outputs: 26 x 26 x 256		
77	conv	256	1 x 1/ 1	26 x 26 x 256
78	conv	256	3 x 3/ 1	26 x 26 x 256
79	Shortcut Layer: 76	wt = 0, wn = 0, outputs: 26 x 26 x 256		
80	conv	256	1 x 1/ 1	26 x 26 x 256

Continua en la pagina siguiente...

Cuadro A.1 – de la pagina anterior.

Index	Layer	Filters	Size/Strd(dil)	Input
81	conv	256	3 x 3/ 1	26 x 26 x 256
82	Shortcut Layer: 79	wt = 0, wn = 0, outputs: 26 x 26 x 256		
83	conv	256	1 x 1/ 1	26 x 26 x 256
84	route 83 56			
85	conv	512	1 x 1/ 1	26 x 26 x 512
86	conv	1024	3 x 3/ 2	26 x 26 x 512
87	conv	512	1 x 1/ 1	13 x 13 x1024
88	route 86			
89	conv	512	1 x 1/ 1	13 x 13 x1024
90	conv	512	1 x 1/ 1	13 x 13 x 512
91	conv	512	3 x 3/ 1	13 x 13 x 512
92	Shortcut Layer: 89	wt = 0, wn = 0, outputs: 13 x 13 x 512		
93	conv	512	1 x 1/ 1	13 x 13 x 512
94	conv	512	3 x 3/ 1	13 x 13 x 512
95	Shortcut Layer: 92	wt = 0, wn = 0, outputs: 13 x 13 x 512		
96	conv	512	1 x 1/ 1	13 x 13 x 512
97	conv	512	3 x 3/ 1	13 x 13 x 512
98	Shortcut Layer: 95	wt = 0, wn = 0, outputs: 13 x 13 x 512		
99	conv	512	1 x 1/ 1	13 x 13 x 512
100	conv	512	3 x 3/ 1	13 x 13 x 512
101	Shortcut Layer: 98	wt = 0, wn = 0, outputs: 13 x 13 x 512		
102	conv	512	1 x 1/ 1	13 x 13 x 512
103	route 102 87			
104	conv	1024	1 x 1/ 1	13 x 13 x1024
105	conv	512	1 x 1/ 1	13 x 13 x1024
106	conv	1024	3 x 3/ 1	13 x 13 x 512
107	conv	512	1 x 1/ 1	13 x 13 x1024
108	max		5x 5/ 1	13 x 13 x 512
109	route 107			
110	max		9x 9/ 1	13 x 13 x 512
111	route 107			

Continua en la pagina siguiente...

Cuadro A.1 – de la pagina anterior.

Index	Layer	Filters	Size/Strd(dil)	Input
112	max		13x13/ 1	13 x 13 x 512
113	route 112 110 108 107			
114	conv	512	1 x 1/ 1	13 x 13 x2048
115	conv	1024	3 x 3/ 1	13 x 13 x 512
116	conv	512	1 x 1/ 1	13 x 13 x1024
117	conv	256	1 x 1/ 1	13 x 13 x 512
118	upsamp		2x	13 x 13 x 256
119	route 85			
120	conv	256	1 x 1/ 1	26 x 26 x 512
121	route 120 118			
122	conv	256	1 x 1/ 1	26 x 26 x 512
123	conv	512	3 x 3/ 1	26 x 26 x 256
124	conv	256	1 x 1/ 1	26 x 26 x 512
125	conv	512	3 x 3/ 1	26 x 26 x 256
126	conv	256	1 x 1/ 1	26 x 26 x 512
127	conv	128	1 x 1/ 1	26 x 26 x 256
128	upsample		4x	26 x 26 x 128
129	route 23			
130	conv	128	1 x 1/ 1	104 x 104 x 128
131	route 130 128			
132	conv	128	1 x 1/ 1	104 x 104 x 256
133	conv	256	3 x 3/ 1	104 x 104 x 128
134	conv	128	1 x 1/ 1	104 x 104 x 256
135	conv	256	3 x 3/ 1	104 x 104 x 128
136	conv	128	1 x 1/ 1	104 x 104 x 256
137	conv	256	3 x 3/ 1	104 x 104 x 128
138	conv	18	1 x 1/ 1	104 x 104 x 256
139	yolo			
	[yolo] params: iou loss: ciou (4), iou_norm: 0.07, obj_norm: 1.00, cls_norm: 1.00, delta_norm: 1.00, scale_x_y: 1.20			
140	route 136			

Continua en la pagina siguiente...

Cuadro A.1 – de la pagina anterior.

Index	Layer	Filters	Size/Strd(dil)	Input
141	conv	256	3 x 3/ 4	104 x 104 x 128
142	route 141 126			
143	conv	256	1 x 1/ 1	26 x 26 x 512
144	conv	512	3 x 3/ 1	26 x 26 x 256
145	conv	256	1 x 1/ 1	26 x 26 x 512
146	conv	512	3 x 3/ 1	26 x 26 x 256
147	conv	256	1 x 1/ 1	26 x 26 x 512
148	conv	512	3 x 3/ 1	26 x 26 x 256
149	conv	18	1 x 1/ 1	26 x 26 x 512
150	yolo			
	[yolo] params: iou loss: ciou (4), iou_norm: 0.07, obj_norm: 1.00, cls_norm: 1.00, delta_norm: 1.00, scale_x_y: 1.10			
151	route 147			
152	conv	512	3 x 3/ 2	26 x 26 x 256
153	route 152 116			
154	conv	512	1 x 1/ 1	13 x 13 x1024
155	conv	1024	3 x 3/ 1	13 x 13 x 512
156	conv	512	1 x 1/ 1	13 x 13 x1024
157	conv	1024	3 x 3/ 1	13 x 13 x 512
158	conv	512	1 x 1/ 1	13 x 13 x1024
159	conv	1024	3 x 3/ 1	13 x 13 x 512
160	conv	18	1 x 1/ 1	13 x 13 x1024
161	yolo			
	[yolo] params: iou loss: ciou (4), iou_norm: 0.07, obj_norm: 1.00, cls_norm: 1.00, delta_norm: 1.00, scale_x_y: 1.05			

A.2. Estructura de la red YOLOv5 versión 6.0/6.1

Cuadro A.2: *Estructura de la red YOLOv5 para V6.0/6.1*

Index	From	n	Model	N Kernels	Kernel	Stride	Padding
0	-1	1	CBM	64	6	2	2
1	-1	1	CBM	128	3	2	1
2	-1	3	C3_1	128			
3	-1	1	CBM	256	3	2	1
4	-1	6	C3_1	256			
5	-1	1	CBM	512	3	2	1
6	-1	9	C3_1	512			
7	-1	1	CBM	1024	3	2	1
8	-1	3	C3_1	1024			
9	-1	1	SPPF	1024			
10	-1	1	CBM	512	1	1	0
11	-1	1	Upsampling x2 , Nearest				
12	-1 , 6	1	Concat				
13	-1	3	C3_0	512			
14	-1	1	CMB	256	1	1	0
15	-1	1	Upsampling x2 , Nearest				
16	-1 , 4	1	Concat				
17	-1	3	C3_0	256			
18	-1	1	CMB	256	3	2	1
19	-1 , 14	1	Concat				
20	-1	3	C3_0	512			
21	-1	1	CMB	512	3	2	1
22	-1 , 10	1	Concat				
23	-1	3	C3_0	1024			
24	17 , 20 , 23		YOLO Detect Anchors 10, 13, 16, 30, 33, 23 30, 61, 62, 45, 59, 119 116, 90, 156, 198, 373, 326				