



UNIVERSIDAD AUTÓNOMA CHAPINGO

POSGRADO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

DETECCIÓN Y CONTEO DE PLANTAS DE AGAVE MEDIANTE APRENDIZAJE PROFUNDO EN IMÁGENES OBTENIDAS POR RPAS

TESIS

Que como requisito parcial para obtener el grado de:

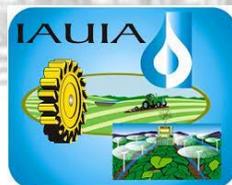
MAESTRO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

Presenta:

JUAN ESPINOZA HERNÁNDEZ

Bajo la supervisión de:

GILBERTO DE JESÚS LÓPEZ CANTEÑS, DR.



APROBADA



Chapingo, Estado de México, septiembre de 2022

**DETECCIÓN Y CONTEO DE PLANTAS DE AGAVE MEDIANTE
APRENDIZAJE PROFUNDO EN IMÁGENES OBTENIDAS POR RPAS**

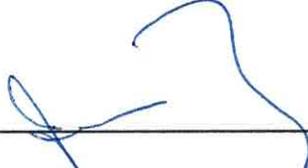
Tesis realizada por el C. Juan Espinoza Hernández bajo la supervisión del Comité Asesor indicado, aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

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

CONTENIDO

LISTA DE CUADROS	vi
LISTA DE FIGURAS	vii
AGRADECIMIENTOS	ix
DATOS BIOGRÁFICOS.....	x
RESUMEN GENERAL.....	xi
GENERAL ABSTRACT.....	xii
1 INTRODUCCIÓN GENERAL.....	13
1.1 Introducción	13
1.2 Objetivos.....	16
1.2.1 Objetivo general	16
1.2.2 Objetivos específicos	16
1.3 Hipótesis.....	16
1.4 Estructura de la tesis	16
1.4.1 Revisión de literatura.....	17
1.4.2 Detección y conteo de plantas de agave mediante aprendizaje profundo en imágenes obtenidas por RPAS.....	17
1.5 Referencias.....	18
2 REVISIÓN DE LITERATURA	20
2.1 El cultivo de agave.....	20
2.2 Agricultura inteligente	21
2.3 Teledetección y RPAS en la agricultura.....	22
2.3.1 Sensores en los RPAS para la agricultura	26
2.4 Generación del ortomosaico	29
2.5 Fenotipado de cultivos.....	30
2.5.1 Métodos para el fenotipado de agave	32
2.6 Aprendizaje profundo.....	33
2.6.1 Redes neuronales convolucionales.....	34

2.6.2	Enfoques de las CNNs	36
2.6.3	Métodos de detección de objetos.....	38
2.6.4	YOLO	39
2.6.5	Conceptos clave.....	42
2.7	Referencias.....	44
3	ARTÍCULO CIENTÍFICO	50
3.1	Resumen	50
3.2	Abstract.....	51
3.3	Introducción	52
3.4	Trabajo relacionado	55
3.4.1	Redes neuronales convolucionales.....	55
3.4.2	Algoritmos de detección de objetos YOLO.....	57
3.5	Materiales y métodos.....	58
3.5.1	Área de estudio	58
3.5.2	Adquisición de imágenes con RPAS y procesamiento fotogramétrico.....	59
3.5.3	División de datos y etiquetado de imágenes	61
3.5.4	YOLOv4 y YOLOv4-tiny	62
3.5.5	Transferencia de aprendizaje.....	63
3.5.6	Hiperparámetros.....	64
3.5.7	Entrenamiento.....	64
3.5.8	Métricas de evaluación.....	65
3.5.9	Detecciones y conteo de plantas	65
3.5.10	Cálculo de distancia entre plantas	66
3.6	Resultados.....	67
3.6.1	Entrenamiento.....	67
3.6.2	Evaluación.....	68
3.6.3	Detecciones y conteo de plantas	69
3.6.4	Distancia entre plantas.....	70

3.7	Discusión	71
3.8	Conclusiones	73
3.9	Referencias.....	74
APÉNDICE A. Planificación del vuelo fotogramétrico y captura de las imágenes del cultivo de agave por medio de RPAS.....		77
APÉNDICE B. Diagrama de flujo en el programa de cómputo Pix4Dmapper para la obtención de los ortomosaicos.....		78
APÉNDICE C. Procedimiento para el etiquetado de la base de datos.		80

LISTA DE CUADROS

CAPÍTULO 2

Cuadro 1. Producción nacional de agave tequilero 2020.....	20
Cuadro 2. Comparación entre tipos de RPAS ala fija y multirroto.....	24
Cuadro 3. Utilización de RPAS de ala fija y multirroto en la agricultura.....	25
Cuadro 4. Sensores utilizados en RPAS en la agricultura para conocer la densidad de plantas.....	28
Cuadro 5. Diferentes métodos y resultados obtenidos en el fenotipado de cultivos.....	31

CAPÍTULO 3

Cuadro 1. Caracterización de los vuelos realizados.....	61
Cuadro 2. Composición de la base de datos.....	62
Cuadro 3. Hiperparámetros utilizados en los algoritmos.....	64
Cuadro 4. Métricas para el conjunto de prueba.....	68
Cuadro 5. Evaluación de los algoritmos.....	69
Cuadro 6. Distancia entre plantas.....	70

LISTA DE FIGURAS

CAPÍTULO 2

Figura 1. RPAS utilizados en la agricultura: a) ala fija, b) multirrotor.....	24
Figura 2. Sensores usados en los RPAS para la agricultura inteligente: a) térmico, b) RGB, c) multiespectral, d) hiperespectral.....	27
Figura 3. Arquitectura general de una CNN.....	35
Figura 4. Componentes típicos de una CNN: a) capa convolucional, b) capa de agrupación (pooling), c) capa completamente conectada.....	35
Figura 5. Diferentes tareas de reconocimiento.....	37
Figura 6. Arquitectura de YOLOv4.....	40
Figura 7. Coordenadas del cuadro delimitador predicho.....	41
Figura 8. Supresión no máxima y la salida de YOLO.....	41

CAPÍTULO 3

Figura 1. Arquitectura general de una CNN.....	55
Figura 2. Componentes típicos de una CNN: a) convolución, b) capa de agrupación, c) capa completamente conectada.....	56
Figura 3. Supresión no máxima y la salida de YOLO.....	58
Figura 4. Localización del campo de agave.....	59
Figura 5. Campo de agave con diferentes condiciones.....	59
Figura 6. Flujo de captura y procesamiento de imágenes.....	60
Figura 7. RPAS DJI Phantom 4 Pro V2.0.....	60

Figura 8. Etiquetas creadas en una imagen.....	61
Figura 9. Arquitectura del algoritmo YOLOv4.....	62
Figura 10. Coordenadas del cuadro delimitador predicho.....	66
Figura 11. Curvas pérdida-iteraciones de los algoritmos para el entrenamiento.....	67
Figura 12. Curvas mAP-iteraciones de los algoritmos durante el entrenamiento.....	68
Figura 13. Detecciones y conteo de plantas de agave con los algoritmos YOLOv4 y YOLOv4-tiny.....	69
Figura 14. Centros de las detecciones de los cuadros delimitadores.....	70

AGRADECIMIENTOS

A la Universidad Autónoma Chapingo y al posgrado en Ingeniería Agrícola y Uso Integral del Agua por conceder la formación académica y los medios para mis estudios de maestría.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada para seguir mi formación académica.

A mi comité asesor, el Dr. Gilberto de Jesús López Canteñs, el Dr. Irineo Lorenzo López Cruz y el Dr. Eugenio Romantchik Kriuchkova que con su sabiduría, conocimiento y apoyo me motivaron a desarrollarme profesionalmente.

A mis compañeros de generación por su amistad y consejos brindados.

A mi familia: padres, hermanos y amigos, por el apoyo que siempre me han brindado motivándome a seguir adelante.

DATOS BIOGRÁFICOS

DATOS PERSONALES

Nombre: Juan Espinoza Hernández
Fecha de nacimiento: 24 de junio de 1993
Lugar de nacimiento: Santa Cruz Itundujia, Oaxaca
CURP: EIHJ930624HOCSRN09
Profesión: Ingeniero Mecánico Agrícola
Cédula profesional: 11897287

DESARROLLO ACADÉMICO

Bachillerato: Preparatoria Agrícola, Universidad Autónoma Chapingo
Licenciatura: Departamento de Ingeniería Mecánica Agrícola, Universidad Autónoma Chapingo

RESUMEN GENERAL

DETECCIÓN Y CONTEO DE PLANTAS DE AGAVE MEDIANTE APRENDIZAJE PROFUNDO EN IMÁGENES OBTENIDAS POR RPAS

Los inventarios de plantas de agave ayudan a monitorear, estimar y planificar la producción. Tradicionalmente el conteo de plantas se ha realizado manualmente, labor que requiere mucho tiempo y es propensa a errores. Actualmente se emplean sistemas de aeronaves pilotadas a distancia (RPAS) para apoyar con información estadística de plantaciones de agave. Por lo que, este trabajo tiene como objetivo detectar y contar plantas de agave en condiciones de campo mediante aprendizaje profundo en imágenes aéreas. Para ello, se capturaron imágenes de alta resolución rojo-verde-azul (RGB) a 40, 60 y 80 m de altura con un RPAS. Se generaron tres ortomosaicos y se usaron para crear la base de datos de plantas de agave. Esta base de datos se utilizó para entrenar dos algoritmos convolucionales: YOLOv4 y YOLOv4-tiny. La evaluación de los algoritmos muestra un valor mean average precision (mAP) de 0.99 para ambos algoritmos y una puntuación F1 de 0.95 para YOLOv4 y de 0.96 para YOLOv4-tiny. Para alturas de 40 y 60 m se obtuvo el 100% de plantas detectadas en ambos algoritmos. Sin embargo, a una altura de 80 m se obtuvo el 96.8% de plantas detectadas con YOLOv4, y el 93.6% con YOLOv4-tiny. La salida de los algoritmos YOLO se utilizó para realizar un cálculo aproximado de la distancia entre plantas detectadas, teniendo un error máximo de 20 cm. Los resultados indican que los algoritmos son robustos, mostrando que es posible detectar y contar agaves en tres alturas de vuelo diferentes. Así mismo, es posible usar la distancia entre plantas de agave calculada para detectar espacios vacíos.

Palabras clave: *agave tequilana Weber*, algoritmos, dron, agricultura de precisión, YOLO

Tesis de Maestría en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo.

Autor: Juan Espinoza Hernández

Director de Tesis: Gilberto de Jesús López Canteñs

GENERAL ABSTRACT

DETECTION AND COUNTING OF AGAVE PLANTS USING DEEP LEARNING IN IMAGES BY RPAS

Agave plant inventories help to monitor, estimate, and plan production. Plant counting has traditionally been done manually, a time-demanding and error-prone task. Remotely piloted aircraft systems (RPAS) are currently used to provide statistical information on agave plantations. Therefore, this work aims to detect and count agave plants under field conditions using deep learning in aerial images. High-resolution red-green-blue (RGB) images were captured at 40, 60, and 80 m height with an RPAS. Three orthomosaics were generated and used to create the agave plant database. This database was used to train two convolutional algorithms: YOLOv4 and YOLOv4-tiny. The algorithms' evaluation shows a mean average precision (mAP) value of 0.99 for both algorithms and an F1 score of 0.95 for YOLOv4 and 0.96 for YOLOv4-tiny. For heights of 40 and 60 m, 100% of plants detected in both algorithms were obtained. However, at a height of 80 m, 96.8% of plants were detected with YOLOv4, and 93.6% with YOLOv4-tiny. The YOLO algorithms' output was used to make an approximate calculation of the distance between detected plants, having a maximum error of 20 cm. The results show that the algorithms are robust, showing that it is possible to detect and count agaves at three different flight heights. In addition, it is possible to use the calculated distance between agave plants to detect empty spaces.

Keywords: *agave tequilana* Weber, algorithms, drone, precision agriculture, YOLO.

1 INTRODUCCIÓN GENERAL

1.1 Introducción

El cultivo de agave azul (*Agave tequilana Weber*) utilizado para la elaboración de la famosa bebida mexicana tequila, se cultiva en la región conocida como Denominación de Origen del Tequila, que incluye a los estados de Jalisco, Michoacán, Tamaulipas, Nayarit y Guanajuato, contribuyendo el estado de Jalisco con la mayor producción (Ceja-Ramírez et al., 2017). Donde, el Consejo Regulador del Tequila A.C. (CRT) es el encargado de inspeccionar, certificar y promover la calidad y el control en lo referente a la cadena de producción agave – tequila.

De acuerdo con los registros, en el año 2020 la producción de agave tequilero superó las 100 mil hectáreas sembradas y 18 mil hectáreas cosechadas, de las cuales más del 70% pertenecen al estado de Jalisco (SIAP, 2020). Dada la importancia de la industria del tequila, el campo de agave ha innovado a través de los años para enfrentar diversos retos, aplicando las herramientas que actualmente ofrece la agricultura inteligente.

La agricultura inteligente, de acuerdo con Radoglou-Grammatikis et al., (2020) persigue los siguientes objetivos: aumentar el rendimiento de los cultivos, mejorar la calidad de los productos agrícolas, hacer un uso más eficiente de agroquímicos, ahorrar energía, proteger el medio ambiente de la contaminación y maximizar los beneficios económicos. Para ello hace uso de tecnologías de posicionamiento global, sensores y más recientemente Sistemas de Aeronaves Pilotadas a Distancia (RPAS) (Paneque-Gálvez et al., 2014), sumado al Internet de las Cosas (IoT), Sistemas de Información Geográfica (SIG), Data Science e Inteligencia Artificial (IA) (Sishodia et al., 2020).

Dentro de la agricultura inteligente, una de las tecnologías más importantes es la teledetección aérea ya que haciendo uso de ella se pueden monitorear cultivos y parámetros de la vegetación mediante imágenes aéreas capturadas en diferentes longitudes de onda (Tsouros et al., 2019 & Sishodia et al., 2020). Principalmente la salud de los cultivos (Radoglou-Grammatikis et al., 2020; Shafi et al., 2019),

sin embargo también se está utilizando para la detección plantas (Flores et al., 2021).

La base en la teledetección aérea son las propias imágenes, las cuales pueden adquirirse mediante diversas plataformas como lo son: satélites, aeronaves tripuladas y RPAS. Sin embargo, recientemente los RPAS han tomado más campo de acción debido a ciertas ventajas que ofrecen frente a las otras plataformas, como lo es su capacidad de ofrecer una mejor resolución espacial y temporal, además de un monitoreo de bajo costo (Calvario et al., 2017). Así mismo, ayudan a identificar anomalías en los cultivos de forma rápida y no destructiva, dando la capacidad de actuar a tiempo, llegando a regiones inaccesibles para otro tipo de plataformas (Yang et al., 2017 & Burdziakowski, 2020).

Para la captura de las imágenes se emplean diferentes tipos de sensores montados en los RPAS, y a través de estos se logra monitorear diferentes características de los cultivos (Sishodia et al., 2020), dando así paso al fenotipado. Entendiéndose como fenotipado a la detección y cuantificación de rasgos en diferentes momentos del ciclo de vida de las plantas, el cual con ayuda de los RPAS y sensores se puede realizar de forma no destructiva mediante el análisis de imágenes (Das Choudhury et al., 2019). Entre los diferentes tipos de sensores empleados en la teledetección agrícola están los RGB, multiespectrales, hiperespectrales y térmicos (Yang et al., 2017), sin embargo para tareas de fenotipado los sensores térmicos no son utilizados, mientras que los RGB son los más empleados, y en su caso, en combinación con otros.

El monitoreo de los cultivos es de vital importancia para mejorar la productividad de una manera eficiente. Existen diferentes métodos basados en visión artificial para el fenotipado de cultivos que resuelven ciertos desafíos, en particular en la detección y conteo de plantas, que es el tema de la presente tesis. Ya que, en el caso del cultivo de agave azul, en relación con los controles de calidad llevados a cabo por el CRT basados en las normas ISO, el elemento estratégico de la cadena productiva del tequila lo constituye la propia planta de agave, lo cual

conlleva a la necesidad de realizar inventarios de estas para así poder estimar rendimientos, además predecir y planificar la producción, teniendo un control sobre el número de plantas establecidas (Calvario et al., 2020). Mediante los inventarios se pueden responder cuestiones como la disponibilidad del agave, etapa vegetativa, ubicación, propietario, estado fitosanitario, entre otras para llevar un mejor manejo y control. Además, la norma oficial mexicana NOM-006-SCFI-2012 menciona que todo agave que es utilizado para la producción del tequila debe estar debidamente registrado ante el CRT y que los productores de *Agave tequilana* Weber variedad azul deben actualizar anualmente su registro de plantaciones y predios, manifestando la condición y/o estado en su inventario correspondiente (DOF, 2012). Dando así un valor agregado a la cadena de producción, elaboración y comercialización del tequila (CRT, 2016).

Debido a que el ciclo del agave es largo (más de 7 años), durante su estancia en campo se expone a diversos factores como heladas, sequias, incendios, plagas y enfermedades. En tanto los procedimientos tradicionales de conteo de plantas normalmente han sido realizados mediante técnicas manuales, recorriendo los campos y contando planta por planta. Las tareas manuales frecuentemente requieren hasta seis meses, dando como resultado de uno a dos conteos como máximo por año. Por otro lado, son propensas a errores humanos, además de que se expone la salud de los trabajadores, aumenta la contaminación y tiene un alto costo económico. Por lo cual existe la necesidad de desarrollar técnicas automatizadas de conteo de agaves, que sean precisas, para agilizar la elaboración de inventarios de una forma no destructiva y segura.

1.2 Objetivos

1.2.1 Objetivo general

Evaluar los algoritmos YOLOv4 y YOLOv4-tiny basados en aprendizaje profundo con el fin de detectar y contar plantas de *Agave tequilana Weber* variedad azul a través de imágenes RGB de alta resolución obtenidas por un sistema de aeronave pilotada a distancia (RPAS).

1.2.2 Objetivos específicos

- Crear una base de datos de agave, a partir de imágenes tomadas por un RPAS, considerando que exista variabilidad de tamaños, presencia de suelo desnudo y maleza.
- Evaluar los algoritmos de detección de objetos YOLOv4 y YOLOv4-tiny respecto al conteo manual de plantas de agave en imágenes obtenidas a diferentes alturas de vuelo.
- Determinar la distancia entre plantas de agave, usando sus ubicaciones en las imágenes, en base a las detecciones obtenidas por los algoritmos YOLO.

1.3 Hipótesis

Es posible detectar y contabilizar, bajo condiciones de campo, las plantas de *Agave tequilana Weber* variedad azul, a través de imágenes RGB de alta resolución obtenidas por un sistema de aeronave pilotada a distancia (RPAS), mediante su procesamiento e implementado los algoritmos de aprendizaje profundo YOLOv4 y YOLOv4-tiny.

1.4 Estructura de la tesis

Este documento se divide en 3 capítulos, los cuales son: 1) introducción general, 2) revisión de literatura y 3) artículo científico, que lleva por nombre “Detección y conteo de plantas de agave mediante aprendizaje profundo en imágenes obtenidas por RPAS”.

1.4.1 Revisión de literatura

En este capítulo se describen los puntos más importantes relacionados con el trabajo de investigación. Primeramente, se describe la importancia de la planta de agave y el problema al cual se le dio solución, es decir al conteo de plantas para la elaboración de inventarios de plantas. Posteriormente se describe la agricultura inteligente, en donde, a través de herramientas como la teledetección, RPAS y diferentes tipos de sensores es posible realizar el fenotipado en diferentes cultivos utilizando distintos métodos. Así mismo, se describen diferentes trabajos relacionados al fenotipado de agave. Además, se presentan temas relacionados con el aprendizaje profundo, específicamente con las redes neuronales convolucionales (CNNs), enfoques de las CNNs y métodos para la detección de objetos. Finalmente se menciona el algoritmo convolucional YOLO y algunos conceptos clave.

1.4.2 Detección y conteo de plantas de agave mediante aprendizaje profundo en imágenes obtenidas por RPAS

En el capítulo tres se incluye el manuscrito del trabajo de investigación para detectar las plantas de agave a través de imágenes RGB obtenidas por RPAS, partiendo de su procesamiento y la creación de una base de datos. Se evalúan dos algoritmos convolucionales de detección de objetos teniendo como salida los cuadros delimitadores de las plantas detectadas por YOLOv4 y YOLOv4-tiny, los cuales también sirven para realizar un cálculo estimado de la distancia entre plantas en el campo.

1.5 Referencias.

- Burdziakowski, P. (2020). A Novel Method for the Deblurring of Photogrammetric Images Using Conditional Generative Adversarial Networks. *Remote Sensing*, 12(16), 2586. <https://doi.org/10.3390/rs12162586>
- Calvario, G., Alarcón, T. E., Dalmau, O., Sierra, B., & Hernández, C. (2020). An Agave Counting Methodology Based on Mathematical Morphology and Images Acquired through Unmanned Aerial Vehicles. *Sensors*, 20(21), 6247. <https://doi.org/10.3390/s20216247>
- Calvario, G., Sierra, B., Alarcón, T. E., Hernández, C., & Dalmau, O. (2017). A Multi-Disciplinary Approach to Remote Sensing through Low-Cost UAVs. *Sensors (Basel, Switzerland)*, 17(6), E1411. <https://doi.org/10.3390/s17061411>
- Ceja Ramírez, R., González Eguiarte, D. R., Ruiz Corral, J. A., Rendón Salcido, L. A., & Flores Garnica, J. G. (2017). Detección de restricciones en la producción de agave azul (Agave tequilana Weber var. Azul) mediante percepción remota. *Terra Latinoamericana*, 35(3), 259-268. <https://doi.org/10.28940/terra.v35i3.252>
- CRT. (2016). Consejo Regulador del Tequila. <https://www.crt.org.mx/index.php/es/>
- Das Choudhury, S., Samal, A., & Awada, T. (2019). Leveraging Image Analysis for High-Throughput Plant Phenotyping. *Frontiers in Plant Science*, 10. <https://www.frontiersin.org/article/10.3389/fpls.2019.00508>
- DOF. (2012). Diario Oficial de la Federación. http://www.dof.gob.mx/nota_detalle.php?codigo=5282165&fecha=13/12/2012
- Flores, D., González-Hernández, I., Lozano, R., Vázquez-Nicolas, J. M., & Hernández Toral, J. L. (2021). Automated Agave Detection and Counting Using a Convolutional Neural Network and Unmanned Aerial Systems. *Drones*, 5(1), 4. <https://doi.org/10.3390/drones5010004>
- Paneque-Gálvez, J., McCall, M. K., Napoletano, B. M., Wich, S. A., & Koh, L. P. (2014). Small Drones for Community-Based Forest Monitoring: An Assessment of Their Feasibility and Potential in Tropical Areas. *Forests*, 5(6), 1481-1507. <https://doi.org/10.3390/f5061481>
- Radoglou-Grammatikis, P., Sarigiannidis, P., Lagkas, T., & Moscholios, I. (2020). A compilation of UAV applications for precision agriculture. *Computer Networks*, 172, 107148. <https://doi.org/10.1016/j.comnet.2020.107148>
- Shafi, U., Mumtaz, R., García-Nieto, J., Hassan, S. A., Zaidi, S. A. R., & Iqbal, N. (2019). Precision Agriculture Techniques and Practices: From Considerations to Applications. *Sensors*, 19(17), 3796. <https://doi.org/10.3390/s19173796>
- SIAP. (2020). Cierre de la producción agrícola 2020. Anuario Estadístico de la Producción Agrícola. <https://nube.siap.gob.mx/cierreagricola/>
- Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sensing*, 12(19), 3136. <https://doi.org/10.3390/rs12193136>

- Tsouros, D. C., Bibi, S., & Sarigiannidis, P. G. (2019). A Review on UAV-Based Applications for Precision Agriculture. *Information*, 10(11), 349. <https://doi.org/10.3390/info10110349>
- Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Zhao, X., Li, Z., Li, H., & Yang, H. (2017). Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives. *Frontiers in Plant Science*, 8. <https://www.frontiersin.org/article/10.3389/fpls.2017.01111>

2 REVISIÓN DE LITERATURA

2.1 El cultivo de agave

El agave azul (*Agave tequilana Weber*) empleado en la elaboración del tequila se cultiva en la región conocida como Denominación de Origen del Tequila, que incluye a los estados mexicanos de Jalisco, Michoacán, Tamaulipas, Nayarit y Guanajuato, contribuyendo el estado de Jalisco con la mayor producción (Ceja-Ramírez et al., 2017) (Cuadro 1).

Cuadro 1. Producción nacional de agave tequilero 2020. Fuente: (SIAP, 2020).

Estados	Superficie (ha)	
	Sembrada	Cosechada
Guanajuato	12393.20	2574.00
Jalisco	73371.82	12765.01
Michoacán	3881.30	423
Nayarit	12042.50	2411.00
Tamaulipas	1539.28	0
Nacional	103228.10	18173.01

Debido a que el elemento estratégico de la cadena productiva del tequila lo constituye la propia planta de agave, se tiene la necesidad de realizar inventarios de estas para así poder estimar rendimientos, además de predecir y planificar la producción, teniendo un control sobre el número de plantas establecidas (Calvario et al., 2020).

Mediante los inventarios se pueden responder cuestiones como la disponibilidad del agave, etapa vegetativa, ubicación, propietario, estado fitosanitario, entre otras, para llevar un mejor manejo y control. Además, la norma oficial mexicana NOM-006-SCFI-2012 menciona que todo agave que es utilizado para la producción del tequila debe estar debidamente registrado ante el Consejo Regulador del Tequila A.C. y que los productores de *Agave tequilana Weber* deben actualizar anualmente su registro de plantaciones y predios, manifestando la condición y/o estado en su inventario correspondiente (DOF, 2012). Dando así un valor agregado a la cadena de producción, elaboración y comercialización de la bebida (CRT, 2016).

Tradicionalmente el conteo de plantas de agave ha sido un proceso largo, costoso y que pone en riesgo la salud de los trabajadores. Este conteo manual es llevado a cabo por grupos de personas que recorren los campos contando plantas una a una, tarea que frecuentemente requiere hasta seis meses, dando como resultado de uno a dos conteos como máximo por año, añadiendo que es propenso a errores humanos. Por lo cual es de importancia el implementar nuevas tecnologías y metodologías automáticas para agilizar el conteo de plantas de agave presentes en las parcelas apoyando en la elaboración de inventarios, dando paso a una agricultura inteligente.

2.2 Agricultura inteligente

Las labores agrícolas del campo tradicionalmente se han realizado basándose en la experiencia de los agricultores adquirida a través de los años. En la actualidad, con la innovación y el empleo de tecnologías emergentes que se adaptan a la agricultura, se puede mejorar la toma de decisiones basándose en el sensado de variables, su almacenamiento y análisis de grandes cantidades de datos. Esta nueva revolución en la agricultura conlleva a la llamada agricultura inteligente (Shafi et al., 2019).

Agricultura inteligente es el término que se utiliza para referirse a una tecnología de producción cuya esencia es el aumento de la eficiencia en la administración de los distintos procesos y operaciones de producción en la agricultura, mediante la aplicación de insumos en cantidades variables, de acuerdo con la variabilidad de las características en un sitio específico.

En base a lo anterior, dentro de la agricultura inteligente es necesario manejar conceptos como la variabilidad espacial y temporal. Ya que, a través de este conocimiento el productor puede dividir la parcela en áreas de manejo, aplicando cantidades variables de insumos, optimizando recursos y aumentando la producción, cumpliendo así con ciertos objetivos (Radoglou-Grammatikis et al., 2020).

Un cultivo, a través de esta tecnología de producción, puede ser mejorado desde diferentes perspectivas como lo son: agronómica, técnica, ambiental, económica,

entre otras. En este sentido, Radoglou-Grammatikis et al., (2020) mencionan que algunos de los principales objetivos que persigue la agricultura inteligente son los siguientes:

- Aumentar el rendimiento de los cultivos.
- Mejorar la calidad de los productos agrícolas.
- Hacer un uso más eficiente de agroquímicos.
- Ahorrar energía.
- Proteger el medio ambiente de la contaminación.
- Maximizar los beneficios económicos.

Para ello, en esta tecnología de producción se emplean herramientas tecnológicas como el posicionamiento global, sensores distribuidos en el área de la parcela y más recientemente también han entrado los sistemas de aeronaves pilotadas a distancia (RPAS) (Paneque-Gálvez et al., 2014), sumado al Internet de las Cosas (IoT), los Sistemas de Información Geográfica (SIG), Data Science y la Inteligencia Artificial (IA) (Sishodia et al., 2020). Con ayuda de estas herramientas se realiza el proceso de recolección de la información, la cual es finalmente plasmada en mapas digitales sobre los cuales se toman decisiones de manejo de los cultivos (Zarco-Tejada et al., 2013).

2.3 Teledetección y RPAS en la agricultura

González et al., (2016) comentan que la confianza que los productores depositan en la agricultura inteligente va en aumento, y cada vez más hacen uso de ella. En consecuencia, al hacer uso de esta tecnología de producción y recopilar datos de las parcelas que posteriormente son analizados conlleva al concepto conocido como teledetección agrícola.

Dentro de las tecnologías mencionadas por Sishodia et al., (2020), una de las más importantes en la agricultura inteligente es la teledetección, ya que a través de ella se puede monitorear diversos cultivos y parámetros de la vegetación mediante imágenes aéreas capturadas en diferentes longitudes de onda del espectro electromagnético (Tsouros et al., 2019).

Durante los últimos años, esta tecnología ha sido utilizada comúnmente para monitorear la salud de los cultivos mediante el análisis de imágenes (Radoglou-Grammatikis et al., 2020; Shafi et al., 2019), sin embargo también se está utilizando para la detección y conteo de plantas (Flores et al., 2021).

Las imágenes aéreas pueden adquirirse mediante diversas plataformas como lo son: satélites, aeronaves tripuladas, y RPAS. No obstante, de las plataformas anteriores la que ha tomado más campo de acción en los últimos años debido a las ventajas que ofrece son los RPAS. Por ejemplo, algunas de las ventajas sobre las otras plataformas es que son capaces de ofrecer una mejor resolución espacial y temporal en las imágenes capturadas. Además, el desarrollo de RPAS ha llevado a la agricultura a otro nivel, ofreciendo mediante la teledetección un monitoreo de bajo costo, ya que los datos se pueden adquirir de forma rápida y sencilla, comparado con los satélites y aeronaves tripuladas (Calvario et al., 2017).

Los RPAS equipados con sensores, tecnologías de posicionamiento y comunicación, actualmente se utilizan en aplicaciones de teledetección en la agricultura para identificar anomalías en los cultivos de forma rápida y no destructiva, dando la capacidad a los productores de actuar a tiempo ya que se pueden utilizar en el momento que el agricultor disponga (Yang et al., 2017). Además los RPAS permiten llegar a regiones a menudo inaccesibles para otro tipo de plataformas (Burdziakowski, 2020).

La teledetección junto con el procesamiento de las imágenes obtenidas por RPAS ha contribuido a la agricultura monitoreando y mejorando el control de los cultivos (Calvario et al., 2020). Con la introducción de los RPAS en la agricultura no es necesario recorrer personalmente toda la parcela para inspeccionarla, ya que con cámaras de alta resolución acopladas a la aeronave se obtiene información georeferenciada (González et al., 2016). Sumado a que estas plataformas tienen la capacidad de volar de manera autónoma mediante un plan de vuelo (Radoglou-Grammatikis et al., 2020).

Hassanalian & Abdelkefi, (2017) mencionan que los diseños de los RPAS son variados, con características y funciones diferentes. Panday et al., (2020) agregan que los RPAS se pueden clasificar según diferentes criterios, como lo es el principio de vuelo, su tamaño, capacidad de carga y rango de vuelo. Sin embargo, el principio de vuelo es el más notorio a simple vista. Dentro de esta clasificación se encuentran RPAS de ala fija y multirrotor (Figura 1) (Shafi et al., 2019).

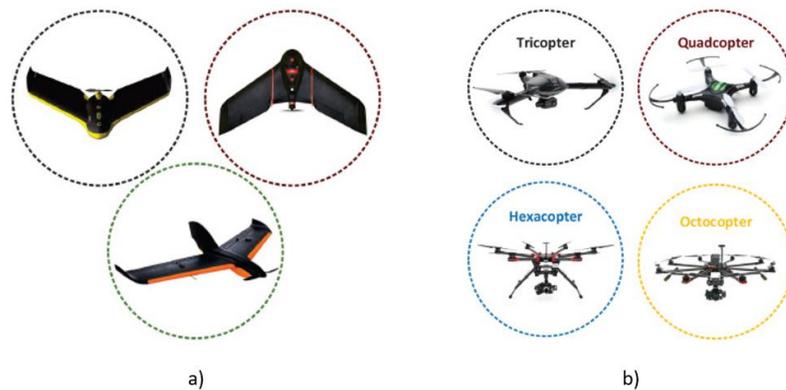


Figura 1. RPAS utilizados en la agricultura: a) ala fija, b) multirrotor. Fuente: (Radoglou-Grammatikis et al., 2020)

En el Cuadro 2 se muestra un resumen con las ventajas y desventajas de los RPAS multirrotor y ala fija, recopiladas por Panday et al., (2020).

Cuadro 2. Comparación entre tipos de RPAS ala fija y multirrotor. Fuente: (Panday et al., 2020).

Tipo	Ventajas	Desventajas
Multirrotor	<ul style="list-style-type: none"> • Mejor accesibilidad, más baratos • Mejor estabilidad • Despegue y aterrizaje vertical • No se requiere pista • Capacidad para mantenerse en un punto • Mayor capacidad de carga útil • Fácil de empezar a volar 	<ul style="list-style-type: none"> • Resistencia y alcance cortos • Velocidad de vuelo más baja • Menor eficiencia energética • Arquitectura compleja y difícil mantenimiento
Ala fija	<ul style="list-style-type: none"> • Arquitectura más simple y mantenimiento más sencillo • Resistencia y largo alcance • Mayor velocidad de vuelo 	<ul style="list-style-type: none"> • Accesibilidad limitada, más caros • Menor resistencia al viento

- Cubre mayor superficie
- Mayor eficiencia energética
- Mayor capacidad para sobrevivir a fallas técnicas debido a la capacidad de deslizamiento
- Dificultades para despegar y aterrizar
- Más difícil de maniobrar
- Más difícil de volar. Se requiere más entrenamiento

La selección del RPAS adecuado depende de las ventajas que ofrece y además de las necesidades del agricultor. En la elección es posible especificar características generales a considerar, como son: la forma de despegue y aterrizaje, resistencia al viento y capacidad de carga, entre otras (Hassanalian & Abdelkefi, 2017; van der Merwe et al., 2020). El tiempo de vuelo se basa en el peso de la carga útil. En general, los sistemas de ala fija logran un tiempo de vuelo más largo, lo que exige cargas útiles livianas. Por el contrario, los RPAS multirrotor alimentados por batería con mayor capacidad de carga útil tienen un tiempo de vuelo reducido (Shafi et al., 2019).

Generalmente la mayoría de los RPAS utilizados en la agricultura son multirrotor, ya que son fáciles de operar, manejan velocidades más lentas, tienen mejor maniobrabilidad y un costo relativamente bajo. En los casos en los que la superficie a volar es considerablemente grande, se prefieren aeronaves de ala fija, que permiten el monitoreo de toda el área en un menor tiempo (Tsouros et al., 2019).

En el Cuadro 3 se muestran algunos ejemplos de trabajos que emplean RPAS de ala fija y multirrotor para tareas de detección de plantas.

Cuadro 3. Utilización de RPAS de ala fija y multirrotor en la agricultura.

RPAS	Modelo	Cultivo	Referencia
Ala fija	senseFly eBee	Cítricos	(Csillik et al., 2018)
	---	Arroz	(Shao et al., 2021)
	senseFly eBee X	Palmeras	(Jintasuttisak et al., 2022)
Multirrotor	DJI S1000	Arroz	(Reza et al., 2017)
	Hexacóptero Atechsys	Trigo	(Jin et al., 2017)
	DJI S1000	Viñedo	(Primicerio et al., 2017)
	3DR X8	Sorgo	(Guo et al., 2018)
	DJI Matrice 600	Colza	(Zhao et al., 2018)
	DJI Phantom 4 Pro	Tabaco	(Fan et al., 2018)
	DJI Phantom 3	Platanar	(Neupane et al., 2019)
	DJI Phantom 4 Pro	Maíz	(Kitano et al., 2019)
	DJI S1000	Arroz	(Wu et al., 2019)

DJI Matrice 600	Cítricos	(Ampatzidis & Partel, 2019)
DJI S900	Papa / lechuga	(Machefer et al., 2020)
DJI Phantom 3 Pro	Espinaca	(Valente et al., 2020)
DJI Matrice 600 Pro	Algodón	(Feng et al., 2020)
3DR Solo / DJI Phantom 4	Maíz	(García-Martínez et al., 2020)
DJI Matrice 100	Olivos	(Sarabia et al., 2020)
DJI Phantom 4	Agave	(Calvario et al., 2020)
DJI Phantom 3	Agave	(Flores et al., 2021)
DJI Phantom 4 Pro V2.0	Abetos	(Emin et al., 2021)
3DR Solo	Platanar	(Aeberli et al., 2021)
DJI Phantom 4 Pro / Mavic Pro	Palmeras	(Ammar et al., 2021)

Flores et al., (2021) comentan que debido a la velocidad a la que viajan los RPAS de ala fija requieren una mayor altura de vuelo. Por el contrario, los RPAS multirrotores pueden volar a bajas alturas que, para aplicaciones de conteo de plantas son convenientes ya que se busca tener una mejor resolución en las imágenes.

2.3.1 Sensores en los RPAS para la agricultura

A través de diferentes tipos de sensores montados en los RPAS se pueden monitorear diversas características de los cultivos. Sishodia et al., (2020) mencionan que, en aplicaciones de teledetección, los sensores presentan ciertas diferencias en resolución espacial, espectral y temporal. En donde la resolución espacial se refiere al tamaño del píxel de la imagen que representa cierta área del suelo, estando determinada por la altura del vuelo, ángulo de visión, distancia focal de la lente y el tamaño de píxel del sensor. La resolución temporal está asociada con la plataforma o RPAS en lugar del sensor en sí, es decir se refiere a la frecuencia con la que se recopila una imagen de la misma área o campo. Y la resolución espectral de un sensor está determinada por el número de bandas del espectro electromagnético que puede captar, además del ancho de estas.

Sin embargo, los sensores utilizados deben cumplir criterios como bajo peso, bajo consumo de energía y tamaño reducido (Yang et al., 2017). Los sensores o cámaras utilizados en teledetección que cumplen con las características anteriores son: sensor de luz visible (RGB), multiespectrales, hiperespectrales y térmicos (Figura 2). No obstante, para tareas de detección, los sensores térmicos no son utilizados.



Figura 2. Sensores usados en los RPAS para la agricultura inteligente: a) térmico, b) RGB, c) multiespectral, d) hiperespectral. Fuente: (Tsouros et al., 2019)

Sensor de luz visible (RGB). Los sensores RGB, conocidos comúnmente como cámaras digitales, son los más utilizados en los RPAS para el fenotipado de cultivos, debido a su fácil procesamiento de datos y bajo peso (Yang et al., 2017). Son de bajo costo y tienen una alta resolución espacial, aunque una resolución espectral relativamente pobre debido al rango del espectro electromagnético que capta (Nijland et al., 2014). Dado lo anterior, no son adecuados para analizar una gran variedad de parámetros de los cultivos que requieren información espectral en el espectro no visible, por tal motivo, estos sensores se usan en conjunto con otros (Tsouros et al., 2019). Sin embargo, con sensores RGB se pueden calcular índices de vegetación, generar modelos digitales de elevación y ortomosaicos de alta resolución (Maes & Steppe, 2019).

Sensor multiespectrales e hiperespectrales. Pueden captar información sobre la absorción y reflexión espectral de la vegetación en varias bandas. La información espectral es útil para calcular índices de vegetación, además de evaluar y monitorear la salud de los cultivos al discriminar partes no saludables en una imagen en base a la absorción y reflectancia de cierta radiación (Tsouros et al., 2019). Los sensores multiespectrales e hiperespectrales constan de un

conjunto de sensores con diferentes lentes, con cada sensor sensible en una región espectral (Maes & Steppe, 2019). Sin embargo, el costo de estos sensores es elevado, al igual que el procesamiento de los datos llega a ser más complejo en comparación con imágenes RGB (Yang et al., 2017). La principal diferencia entre los sensores multiespectrales e hiperespectrales es la cantidad de bandas y su ancho que cada sensor puede captar. Los sensores multiespectrales capturan hasta 15 bandas, mientras que los sensores hiperespectrales capturan cientos de bandas, pero en un ancho más estrecho.

En el Cuadro 4 se menciona parte del progreso de detección y conteo de plantas mediante diferentes tipos de sensores montados en RPAS.

Cuadro 4. Sensores utilizados en RPAS en la agricultura para conocer la densidad de plantas.

Sensor	Modelo	Cultivo	Bibliografía
RGB	Sony Alpha a5100	Arroz	(Reza et al., 2017)
	Nikon Coolpix P7700	Viñedo	(Primicerio et al., 2017)
	Sony ILCE-5100L	Arroz	(Jin et al., 2017)
	DJI FC6310	Tabaco	(Fan et al., 2018)
	Sony Cyber-shot DSC-RX100M3	Sorgo	(Guo et al., 2018)
	Nikon D800	Colza	(Zhao et al., 2018)
	Sony QX-100	Arroz	(Wu et al., 2019)
	DJI FC300S	Platanar	(Neupane et al., 2019)
	DJI FC6310	Maíz	(Kitano et al., 2019)
	CannonS100, Sequoia / DJI FC6310	Maíz	(García-Martínez et al., 2020)
	DJI FC6310	Agave	(Calvario et al., 2020)
	Panasonic GH4	Papa / lechuga	(Machefer et al., 2020)
	DJI FC300X	Espinaca	(Valente et al., 2020)
	DJI FC300S	Agave	(Flores et al., 2021)
	XTU S2	Arroz	(Shao et al., 2021)
	DJI FC6310	Abetos	(Emin et al., 2021)
Multiespectral e hiperespectral	DJI FC6310 / DJI FC220	Palmeras	(Ammar et al., 2021)
	senseFly S.O.D.A.	Palmeras	(Jintasuttisak et al., 2022)
	Parrot Sequoia	Cítricos	(Csillik et al., 2018)
	MicaSense RedEdge-M	Cítricos	(Ampatzidis & Partel, 2019)
	BaySpec 1000	Algodón	(Feng et al., 2020)
MicaSense RedEdge-M	Olivos	(Sarabia et al., 2020)	
Parrot Sequoia	Platanar	(Aeberli et al., 2021)	

Como se puede observar los sensores RGB son más empleados en detección de plantas.

2.4 Generación del ortomosaico

Obtener productos finales como los mapas de prescripción requiere de haber volado el dron y haber realizado un procesamiento a las imágenes capturadas.

El método utilizado para procesar los datos se llama fotogrametría. A través de la fotogrametría se puede reconstruir una zona mediante el análisis de fotografías en dos dimensiones superpuestas capturadas desde diferentes puntos de vista (Kiropoulos et al., 2022), es decir, usando el par estereoscópico, dando paso a la llamada estereoscopia y visión estereoscópica (Santamaría-Peña, 2011), y empleando también el concepto conocido como paralaje (Quirós-Rosado, 2015).

La adquisición de imágenes a baja altitud por medio de RPAS permite la reconstrucción de las parcelas con una alta resolución espacial. Las imágenes se procesan aplicando triangulación aérea y calibrando la orientación interna y externa de las cámaras. Además se utilizan métodos para buscar puntos homólogos en las imágenes superpuestas (Tsouros et al., 2019).

A través del proceso fotogramétrico se obtienen diferentes productos desde la nube de puntos, modelos digitales de elevación hasta el ortomosaico (Quirós-Rosado, 2015). En donde el ortomosaico, también llamado ortofoto, se define como el producto provisto de métrica y escala, a través del cual se pueden realizar mediciones reales de gran precisión. Además, el ortomosaico es una imagen georeferenciada en la que la distorsión ha sido corregida y ortorectificada así la escala de la imagen es uniforme, es decir se ha realizado una corrección de paralaje para que cada objeto esté sobre su base.

Se han desarrollado diversos programas de cómputo que permiten el procesamiento de las imágenes georeferenciadas capturadas. Algunos de los más conocidos son: Pix4Dmapper y Agisoft Metashape. Los cuales son utilizados en investigaciones para la obtención de los diferentes productos como la ortofoto

en la cual se aplican diferentes metodologías para, por ejemplo, detección y conteo de plantas.

2.5 Fenotipado de cultivos

Das Choudhury et al., (2019) mencionan que el fenotipado de plantas basado en imágenes se refiere a la detección y cuantificación de los rasgos de las plantas de forma no destructiva mediante el análisis de imágenes. Al ser un proceso no destructivo permite cuantificar rasgos en diferentes momentos durante el ciclo de vida de las plantas con ayuda de los diferentes tipos de sensores montados en los RPAS. En donde las imágenes RGB se utilizan mayormente para el fenotipado de características estructurales, es decir morfológicas.

Las plantas al ser organismos vivos cambian constantemente, por ejemplo, sus hojas se cruzan presentando desafíos para el fenotipado basado en imágenes, además de otros inconvenientes como cambios de iluminación, el entorno, las oclusiones, las características de la cámara utilizada, entre otras. Por lo cual se hace necesario la utilización de diferentes técnicas para lograr el fenotipado de plantaciones.

Existen diferentes enfoques basados en visión artificial para el fenotipado de cultivos que resuelven ciertos desafíos que se presentan, en particular en la detección y conteo de plantas. Por ejemplo, algunos trabajos hacen uso de operaciones morfológicas (MO) a las imágenes como lo es la segmentación, otros más aplican el método de Máquina de Vectores de Soporte (SVM), índices de vegetación (IV) y método Otsu, transformada de Hough, morfología matemática (MM), correlación cruzada normalizada (NCC) y la implementación de técnicas de aprendizaje profundo (DL), entre otras.

En el Cuadro 5 se observan algunos trabajos recientes referentes a la detección de cultivos utilizando imágenes obtenidas por RPAS, en los cuales se probaron diferentes enfoques. Se puede notar que el aprendizaje profundo ofrece los mejores resultados, por lo que está ganando más terreno en el fenotipado de cultivos, aunque también se puede utilizar en conjunto con otras técnicas.

Cuadro 5. Diferentes métodos y resultados obtenidos en el fenotipado de cultivos.

Cultivo	Método	Resultado	Referencia
Arroz	MO y componentes conectados	Precisión = 87 % Recall = 91% Puntuación F1 = 89%	(Reza et al., 2017)
Viñedo	SVM	Curva característica de operación del receptor (ROC) = 0.95 %	(Primicerio et al., 2017)
Trigo	SVM	Coefficiente de determinación (R^2) = 87 % Raíz del error cuadrado medio (RMSE) = 34.05 plantas/m ²	(Jin et al., 2017)
Sorgo	SVM	Precisión = 82 % Recall = 98 %	(Guo et al., 2018)
Colza	IV y método Otsu	R^2 = 84.5 % y 86.7 % RMSE = 9.79 % y 5.11 %	(Zhao et al., 2018)
Tabaco	DL	Recall = 94 % Precision = 93 %	(Fan et al., 2018)
Cítricos	DL	Precisión = 94.59 % Recall = 97.94 % Puntuación F1 = 96.24 %	(Csillik et al., 2018)
Platanar	DL	Precision (40, 60 y 80 m) = 99, 97 y 98 % Recall (40, 60 y 80 m) = 96, 85 y 75 %	(Neupane et al., 2019)
Maíz	DL	Best residual = 2.6 %	(Kitano et al., 2019)
Arroz	DL	Precisión = 93 % R^2 = 94 %	(Wu et al., 2019)
Cítricos	DL	Precisión = 99.9 % Recall = 99.7 %	(Ampatzidis & Partel, 2019)
Papa y lechuga	DL	Precisión de detección de múltiples objetos (MOTA) (papa) = 78.1 MOTA (lechuga) = 91.8	(Machefer et al., 2020)
Espinaca	IV, método Otsu y DL	Precisión = 95 %	(Valente et al., 2020)
Algodón	IV, transformada de Hough	Precisión = 84.1 %	(Feng et al., 2020)
Maíz	NCC	Precisión media = 99 %	(García-Martínez et al., 2020)
Olivos	MM	Precisión = 99.92 % Recall = 99.67 % Puntuación F1 = 99.75 %	(Sarabia et al., 2020)
Agave	MM	Precisión = 83.09 – 98.06 %	(Calvario et al., 2020)
Agave	DL	Precisión = 95 % Recall = 98 % Puntuación F1 = 96 %	(Flores et al., 2021)
Abetos	DL	Precisión = 96 %	(Emin et al., 2021)
Platanar	DL	Precisión = 79 – 89 % Recall = 92 – 97 % Puntuación F1 = 85 – 93 %	(Aeberli et al., 2021)
Palmeras	DL	Mean average precision (mAP) = 99 %	(Ammar et al., 2021)
Arroz	DL	Precisión = 89.88 %	(Shao et al., 2021)
Palmeras	DL	Precisión = 91 % Recall = 92 % mAP = 92.34 %	(Jintasuttisak et al., 2022)

2.5.1 Métodos para el fenotipado de agave

Se han realizado diversas investigaciones enfocadas en la detección de plantaciones de agave, así como también el conteo de plantas. Por ejemplo, Garnica et al., (2008) presentaron un estudio sobre imágenes obtenidas por el satélite LANDSAT 7, con el objetivo de identificar y ubicar plantaciones de agave. Utilizaron métodos de árbol de regresión y clasificación logrando un 70% de precisión. Dentro de los problemas que enfrentaron está la densidad de plantas en las parcelas, la cual no es homogénea, además de los diferentes tipos de suelo en los que se encuentran establecidas las plantas y otras cubiertas vegetales que se pueden confundir con la planta.

Posteriormente, Ceja-Ramírez et al., (2017) realizaron una investigación para determinar y ubicar la superficie ocupada por agave azul que presentó restricciones en su producción. Utilizaron una imagen adquirida por el satélite LANDSAT 5 en la cual aplicaron interpolación digital a través de un proceso de clasificación supervisado, logrando una precisión del 73%. En ese mismo año, Calvario et al., (2017) utilizaron imágenes RGB obtenidas por RPAS a una altura de 60 m, para monitorear cultivos de agave. Aplicaron técnicas de extracción de capas para separar las plantas de agave de las malezas y otros elementos e hicieron uso de espacios de colores como el creado por la Commission Internationale d'Eclairage CIE L^*a^*b (L - iluminación, a – coordenadas rojo/verde y b – coordenadas amarillo/azul). La precisión que obtuvieron fue del 99% pero no resolvieron los problemas de superposición entre plantas presente en las plantaciones de agave.

En el 2018, Morales-Avenidaño et al., (2018) realizaron una propuesta metodológica para contar plantas de agave, mediante la captura de imágenes RGB por RPAS y un trabajo de gabinete en programas de cómputo comerciales. Calvario et al., (2020) describieron un algoritmo, continuación de Calvario et al., (2017), basado en morfología matemática aplicado a imágenes RGB de alta resolución obtenidas por RPAS, con el objetivo de realizar el conteo de plantas. Observaron que la precisión del algoritmo dependía de la heterogeneidad del

tamaño de las plantas en el campo y la iluminación. La precisión osciló entre 83 a 98%. Recientemente, Flores et al., (2021) realizaron la detección y conteo automático de agave usando una red neuronal convolucional aplicada a imágenes RGB obtenidas por RPAS a una altura de 50 m, logrando una precisión del 96%, pero con un alto coste computacional.

Como se logra observar en la actualidad se pueden capturar fotografías con RPAS para utilizarlas en métodos de identificación de objetos de interés, como lo son las plantas de agave. Siguiendo la tendencia de las investigaciones cada vez se investigan metodologías de conteo automático de plantas más precisas y que permiten analizar grandes volúmenes de información de forma rápida, como los son los algoritmos de aprendizaje profundo. Dentro de esto, en años recientes las redes de aprendizaje profundo, en especial las redes neuronales convolucionales (CNNs) han tomado relevancia, pues han demostrado ser poderosas para tareas de detección y clasificación al extraer información de las imágenes (Prasad et al., 2017). Por ejemplo Ubbens et al., (2018) realizaron un algoritmo de aprendizaje profundo para el conteo de hojas en disposición de roseta, entrenando una CNN. Chen et al., (2017) describieron un proceso de conteo de manzanas, teniendo el reto de los cambios de iluminación y las oclusiones del follaje y frutas cercanas utilizando dos CNNs. Por otra parte, Ampatzidis & Partel, (2019) desarrollaron una técnica para evaluar las características fenotípicas en cultivos de cítricos utilizando imágenes obtenidas por RPAS y aplicando CNNs.

2.6 Aprendizaje profundo

El aprendizaje profundo (DL) abarca un conjunto de métodos dentro de los cuales se encuentran las redes neuronales y una variedad de algoritmos de aprendizaje supervisado y no supervisado, a través de métodos de agrupación, clasificación y regresión (Voulodimos et al., 2018).

El interés por trabajar con métodos de DL cada vez va en aumento ya que se ha demostrado dan mejores resultados que otras técnicas. Además ha sido impulsado por datos recopilados etiquetados de cada vez más alta calidad, y

sumado al desarrollo computacional permitiendo el uso de unidades de procesamiento gráficas (GPU), acelerando los procesos de entrenamiento (Voulodimos et al., 2018). Asimismo, el DL supera la desventaja en la que los algoritmos tradicionales se basan en características diseñadas manualmente (Liu & Wang, 2021).

Con la teledetección se ha producido una amplia recopilación de datos adquirida por RPAS. Las imágenes aéreas capturadas de los cultivos agrícolas en conjunto con técnicas de aprendizaje profundo han ayudado a mejorar el rendimiento en la agricultura inteligente mediante la extracción de características (Osco et al., 2021). Por ejemplo, se han dado grandes avances gracias a la implementación del aprendizaje profundo, en detección de objetos, seguimiento de objetos y segmentación semántica (Voulodimos et al., 2018).

2.6.1 Redes neuronales convolucionales

Las redes neuronales convolucionales (CNNs) se encuentran entre el conjunto de algoritmos de aprendizaje profundo más utilizados por su efectividad en tareas de detección de objetos de interés (Liu & Wang, 2021), contribuyendo significativamente en el reconocimiento de imágenes (Sarvamangala & Kulkarni, 2022). Las CNNs transforman datos de entrada (por ejemplo, imágenes obtenidas por RPAS) en salidas (por ejemplo, detección de plantas). Una CNN funciona como un algoritmo supervisado, utilizando datos de entrada para ser entrenadas, combinando características a través de múltiples operaciones, donde en una capa final se obtiene la predicción deseada (Osco et al., 2021).

Las CNNs se inspiran en el sistema visual animal aprendiendo de forma automática y adaptativa jerarquías de características, desde patrones de bajo a alto nivel. Una CNN consta de tres tipos de capas: convolucionales, de agrupación y completamente conectadas, en donde cada tipo de capa desempeña una cierta función (Voulodimos et al., 2018; Yamashita et al., 2018). Una arquitectura típica consta de repeticiones de capas de convolución y una capa de agrupación, seguidas de una o más capas totalmente conectadas y una

capa de salida (Sultana et al., 2018). La capa de convolución y la capa de agrupación se alternan varias veces (Liu & Wang, 2021) (Figura 3).

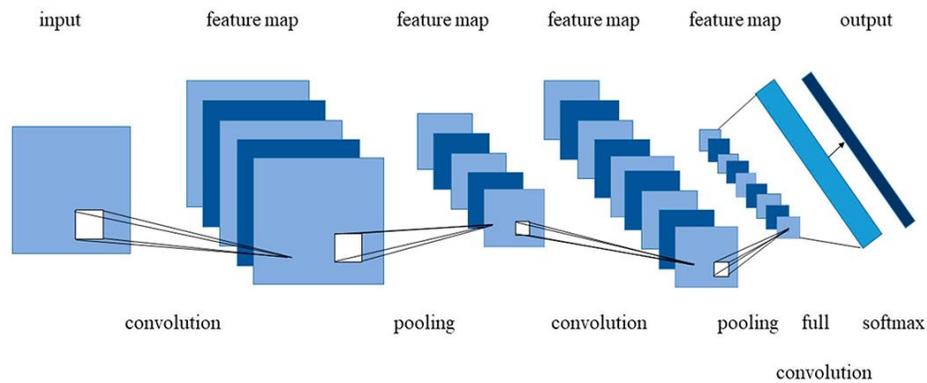


Figura 3. Arquitectura general de una CNN. Fuente: (Zhang et al., 2020)

A continuación, se describe cada una de las capas que conforman una arquitectura típica de una CNN (Figura 4).

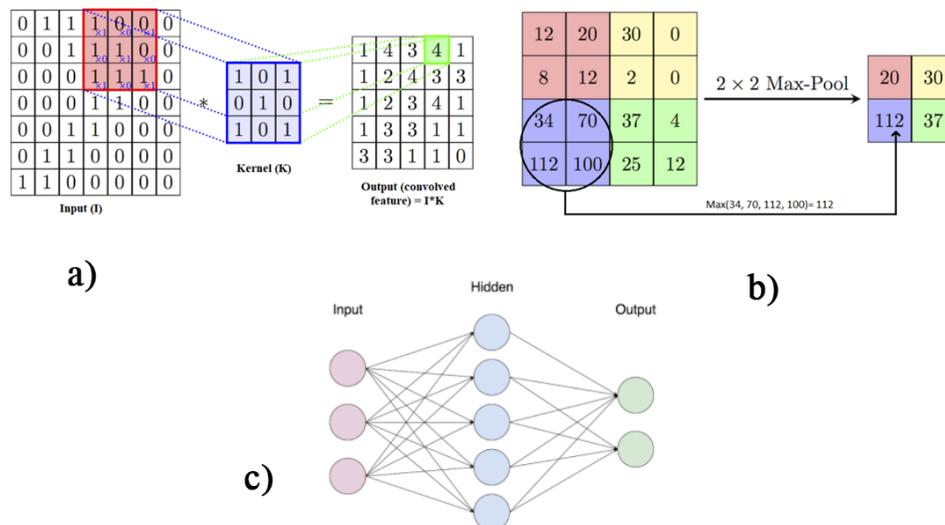


Figura 4. Componentes típicos de una CNN: a) capa convolucional, b) capa de agrupación (pooling), c) capa completamente conectada. Fuente: (Sultana et al., 2018)

Capas convolucionales. Se utilizan filtros o kernels para extraer y generar mapas de características (Figura 4a). Consiste en operaciones lineales y no lineales, operación de convolución y función de activación. En estas capas se aplica una pequeña matriz de números llamada Kernel a la matriz de entrada (Sarvamangala & Kulkarni, 2022).

Capas de agrupación (pooling). Se encarga de reducir las dimensiones de los mapas de características de la entrada (ancho x alto) para la siguiente capa convolucional y disminuye la cantidad de parámetros posteriores. A diferencia de la altura y el ancho, la profundidad de los mapas de características no se ve afectada. También se le conoce como submuestreo o downsampling, ya que la reducción de tamaño conlleva una pérdida de información, que es beneficiosa porque conduce a una menor sobrecarga computacional para las siguientes capas de la red, y también funciona contra el sobreajuste. La agrupación promedio y la agrupación máxima son las más utilizadas (Yamashita et al., 2018). La operación de agrupación toma como entrada una pequeña región de la salida convolucional y la submuestra para producir una única salida (Figura 4b), según el tipo de agrupación seleccionada (Sarvamangala & Kulkarni, 2022).

Capa completamente conectada (FC). Es la última sección de una CNN (Figura 4c), cada nodo tiene conexiones de todas las entradas y todas las conexiones tienen pesos asociados (Sarvamangala & Kulkarni, 2022). La FC convierte los mapas de características de dos dimensiones en un vector de características de una dimensión. Es decir, esta capa toma información de la capa anterior y realiza una operación para generar la salida, por ejemplo, las probabilidades para cada clase en las tareas de clasificación. La capa final normalmente tiene el mismo número de nodos de salida que el número de clases (Yamashita et al., 2018).

El entrenamiento de una CNN es el proceso de encontrar los kernels de las capas de convolución y pesos en las FC que funcionan mejor para una tarea determinada según un conjunto de datos de entrenamiento, es decir que minimizan las diferencias entre las predicciones de salida y las etiquetas de verdad dadas para el entrenamiento (Yamashita et al., 2018).

2.6.2 Enfoques de las CNNs

Wu et al., (2020) mencionan que al emplear algoritmos de DL en datos de teledetección se pueden asistir tareas como clasificación de imágenes, segmentación semántica y de instancias y detección de objetos (Figura 5). A continuación, se describe brevemente cada enfoque, sin embargo, la tarea de

interés en esta investigación es la detección de objetos, concretamente de plantas de agave.

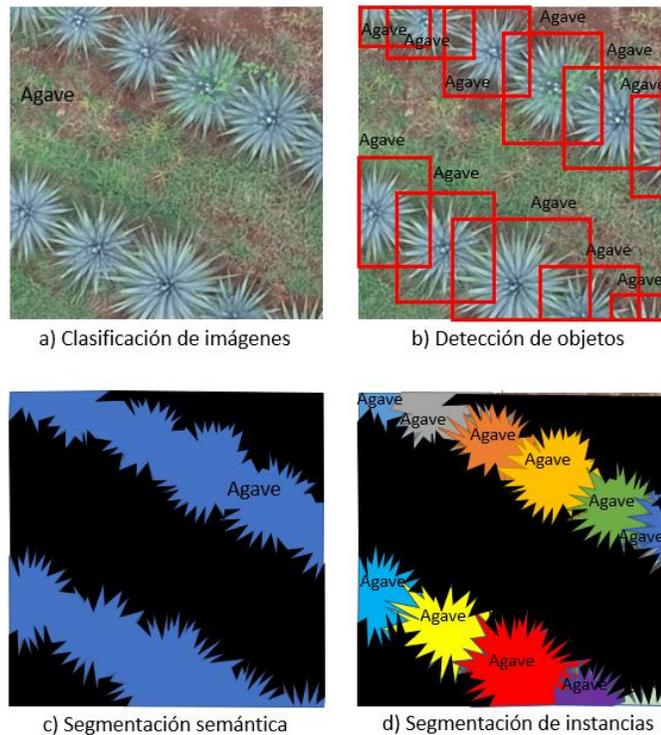


Figura 5. Diferentes tareas de reconocimiento: Fuente: elaboración propia

- La clasificación de imágenes implica asignar una etiqueta de clase a toda una imagen.
- La detección de objetos tiene como objetivo dibujar cuadros delimitadores alrededor de los objetos en una imagen y etiquetar cada uno de ellos de acuerdo con su clase asignada. La detección de objetos puede considerarse una tarea más desafiante ya que requiere ubicar los objetos en la imagen y luego realizar su clasificación.
- En la segmentación semántica se dibujan regiones para distinguir la clase del objeto a nivel de píxel.
- La segmentación de instancias detecta múltiples objetos en máscaras a nivel de píxel y etiqueta cada máscara con una etiqueta de clase. La segmentación de instancias es capaz de individualizar objetos.

En el caso de la detección de plantas de agave la segmentación de instancias se puede utilizar, pero la creación de las máscaras para el entrenamiento requiere mucho más tiempo debido a la forma de la planta en comparación con el etiquetado por recuadros delimitadores para detección de objetos. Además, debido a la superposición entre plantas y la presencia de otras cubiertas vegetales dentro de las plantaciones de agave la tarea de crear máscaras se complica.

2.6.3 Métodos de detección de objetos

Los métodos de detección de objetos se pueden clasificar en dos categorías principales: detectores de una etapa (o métodos basados en regresión) y detectores de dos etapas (o métodos basados en propuestas regionales) (Wu et al., 2020).

- En los detectores de dos etapas se generan propuestas de región (cuadros delimitadores rectangulares candidatos) en el mapa de características. Luego se clasifica cada uno en una etiqueta de clase de objeto con una regresión de cuadro delimitador. Un ejemplo de un detector de dos etapas es el algoritmo Faster-RCNN (Ren et al., 2016).
- Los detectores de una etapa realizan directamente una clasificación y detectan la ubicación de los objetos sin un paso de clasificación de propuesta de región, es decir realizan la predicción para la localización y clasificación de objetos al mismo tiempo. Estos detectores logran una alta velocidad de detección, pero tienden a reducir la precisión de los resultados.

Oscos et al., (2021) mencionan que los métodos basados en la detección de objetos constan de tres componentes: a) backbone, que es responsable de extraer las características en las imágenes; b) neck, utilizado para recopilar las características obtenidas del backbone, y c) head, que realiza la detección y clasificación de las cajas delimitadoras.

No obstante, para aplicar algunos de estos métodos de detección de objetos se requiere de bases de datos etiquetadas (Calvario et al., 2020). Para ello, Flores

et al., (2021) crearon una base de datos de plantas de agave, ya que a la fecha de su investigación no se contaba con una, en la cual implementaron técnicas de aprendizaje profundo. Estos autores agregaron que en el futuro se pueden probar enfoques como los algoritmos convolucionales YOLO, por su versatilidad y robustez, además de que podrían mejorar la velocidad de procesamiento. Dichos algoritmos han sido aplicados con éxito en distintos cultivos, por ejemplo, Ammar et al., (2021), Chowdhury et al., (2022) & Jintasuttisak et al., (2022) aplicaron diversos algoritmos de CNNs recientes para conteo y localización de palmeras, dentro de ellos los algoritmos YOLO. Al igual que Wang et al., (2021) para el conteo de plantas de maíz y así también Lin et al., (2022) detectando plántulas de maní.

2.6.4 YOLO

YOLO (You Only Look Once) es un detector de una etapa que utiliza CNNs para localizar varios objetos en una sola imagen. Lo que significa que además de predecir las clases de los objetos, YOLO también encuentra las ubicaciones de estos objetos en la imagen.

Los algoritmos YOLO han avanzado a través de diferentes versiones implementando mejoras. La primera versión de YOLO (Redmon et al., 2016) fue escrita en un marco de red neuronal de código abierto escrito en C y CUDA llamado Darknet que admite el cálculo de CPU y GPU. Fue el primer algoritmo en combinar la tarea de dibujar cuadros delimitadores e identificar etiquetas de clase. Para YOLOv2 (Redmon & Farhadi, 2016), se incluyó un método para normalizar cada lote de datos mediante Batch Normalization (Ioffe & Szegedy, 2015), así también se incluyeron cajas de anclaje. En YOLOv3 (Redmon & Farhadi, 2018), se agregó una puntuación de objetividad a la predicción del cuadro delimitador, y también se lograron hacer predicciones en objetos más pequeños.

Recientemente, en YOLOv4 (Figura 6) (Bochkovskiy et al., 2020) como Backbone se implementa CSPNet con Darknet53 formando CSPDarknet53, con la cual extrae mapas de características. En el Neck, utilizando Spatial Pyramid

Pooling se recopilan mapas de características generados de diferentes etapas. Y finalmente en Head, la sección de predicción, YOLOv4 utiliza el mismo mecanismo que YOLOv3 detectando el objeto y dibujando el cuadro delimitador, así como también se obtiene la información de las coordenadas bx , by , bw , bh y el valor de confianza.

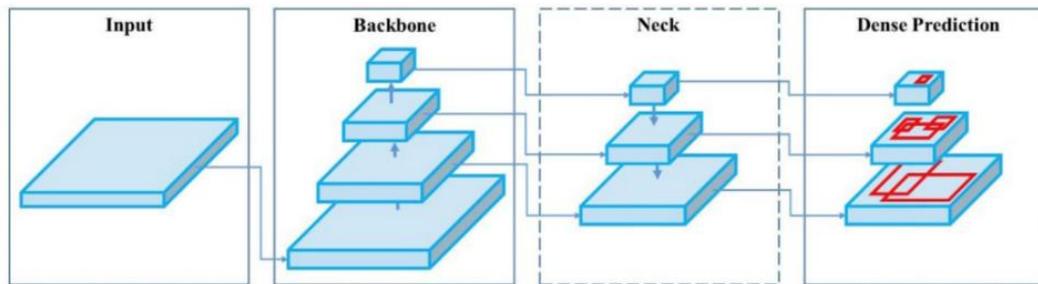


Figura 6. Arquitectura de YOLOv4. Fuente: (Bochkovskiy et al., 2020)

YOLOv4, además de las contribuciones de las versiones anteriores, también implementa nuevas mejoras propias del algoritmo que los autores agregaron (Bochkovskiy et al., 2020). Dentro de ellas están: Mosaic data augmentation, Weighted-Residual-Connections, Cross-Stage-Partial-connections, DropBlock regularization (Ghiasi et al., 2018), Cross mini-Batch Normalization, Self-adversarial-training, CloU-loss y Mish-activation.

Introduciendo CloU (Zheng et al., 2020) durante el entrenamiento como función de pérdida, se mejora la precisión en comparación con YOLOv3. Mientras que la función de activación Mish (Misra, 2020) proporciona mejor precisión, pérdida más baja y suave, y funciona mejor en comparación con los resultados de ReLU y Leaky ReLU.

Resumiendo, con YOLOv4 se hace uso de la función de activación Mish y se mejora la calidad de los mapas de características, que son la entrada de la sección de predicción Head, mejorando los resultados en las detecciones.

Funcionamiento de YOLO. De manera general, YOLO funciona aplicando una única red neuronal a toda la imagen. Divide la imagen de entrada en regiones o celdas, y por cada celda predice la probabilidad de una clase sí hay un objeto y

da los valores bx , by , bh , bw (coordenadas del cuadro delimitador predicho) (Figura 7).

Posteriormente se aplica intersección sobre la unión (IoU) y supresión no máxima para mejorar los resultados, evitando así que varios cuadros delimitadores predichos detecten el mismo objeto (Figura 8).

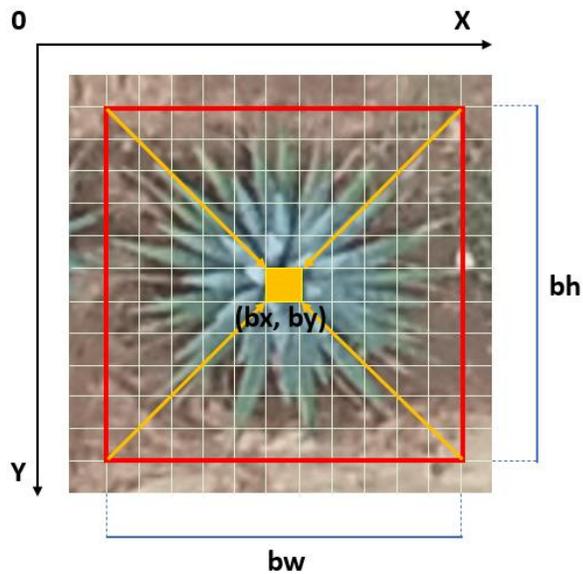


Figura 7. Coordenadas del cuadro delimitador predicho. Fuente: elaboración propia.

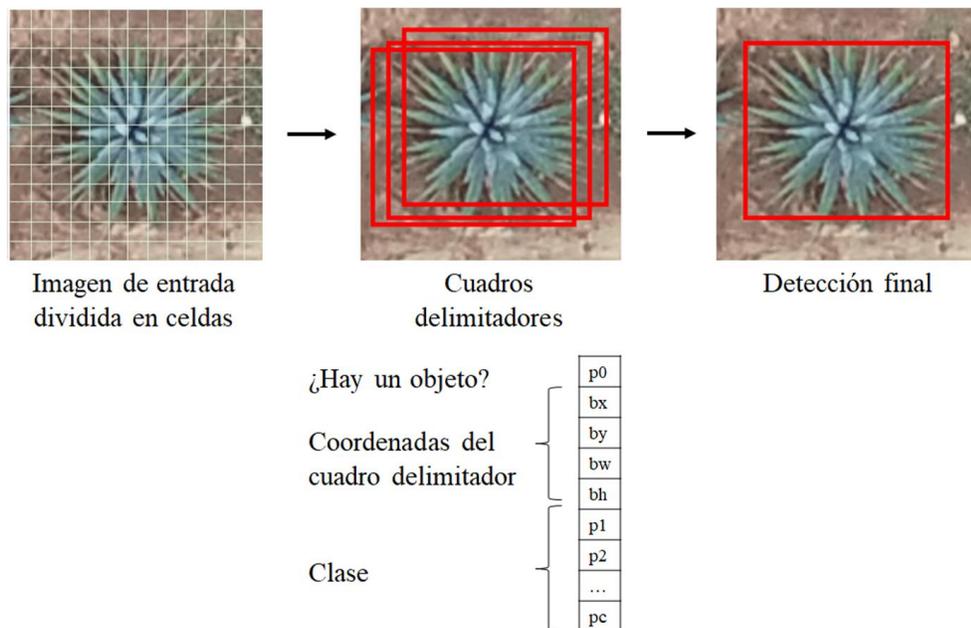


Figura 8. Supresión no máxima y la salida de YOLO. Fuente: elaboración propia.

2.6.5 Conceptos clave

A continuación se describen algunos conceptos importantes a tener presentes dentro de las CNNs (Yamashita et al., 2018):

Parámetro. Representa un valor que cambia a medida que el algoritmo de aprendizaje profundo aprende, y este valor se ve afectado por los hiperparámetros seleccionados.

Hiperparámetro. Son parámetros que debe configurarse antes de comenzar el entrenamiento, ya que controlan el proceso de aprendizaje y determinan los valores de los parámetros del algoritmo con los cuales aprende.

Kernel. Es una cuadrícula de valores que se aplica en las convoluciones. Cada valor se llama peso del kernel. Se asignan números aleatorios que actúan como pesos al comienzo del entrenamiento, que posteriormente se ajustan en cada época de entrenamiento, por lo que el kernel aprende a extraer características (Alzubaidi et al., 2021).

Peso. Es un parámetro fuera de las capas de convolución, es decir, en capas completamente conectadas.

Función de pérdida. Sirve para saber qué tan bien funciona un algoritmo para predecir un resultado. Calcula el error (también llamado pérdida) entre la salida actual y la salida esperada (con las etiquetas de verdad dadas).

Descenso de gradiente. El descenso de gradiente se usa como un algoritmo de optimización que actualiza iterativamente los parámetros que se pueden aprender, es decir, kernels y pesos, de la red para minimizar la pérdida.

Tasa de aprendizaje. Es uno de los hiperparámetros más importantes que se deben establecer antes de que comience el entrenamiento. Es el tamaño de paso de la actualización de los parámetros durante el entrenamiento, es decir, controla el ajuste de los pesos del algoritmo, en relación con el error estimado.

Base de datos. Las bases de datos etiquetadas son los componentes más importantes al aplicar técnicas de aprendizaje profundo, ya que es con la que se

entrenar y prueba un algoritmo. La base de datos normalmente se divide en tres conjuntos: entrenamiento, validación y prueba. Se utiliza un conjunto de entrenamiento para entrenar una red, un conjunto de validación para evaluar el algoritmo durante el proceso de entrenamiento y ajustar los hiperparámetros. El conjunto de prueba se usa solo una vez al final para evaluar el rendimiento del algoritmo.

Sobre ajuste. El sobreajuste se refiere a una situación en la que un algoritmo termina memorizando en lugar de aprender y, por lo tanto, se desempeña peor en un nuevo conjunto de datos. Este es uno de los principales desafíos en el aprendizaje automático, ya que un algoritmo sobreajustado no es generalizable a datos nunca vistos.

Transferencia de aprendizaje. Con esta técnica se transfiere conocimiento de pesos de entrenamiento de un algoritmo de detección llevado a cabo en cierto conjunto de datos a un nuevo conjunto (Huang et al., 2017). Este enfoque mejora la eficiencia durante el entrenamiento, en contraste con inicializar los pesos aleatoriamente (Wang et al., 2022). Es decir, acelera el proceso de aprendizaje y también mejora la capacidad de generalizar del algoritmo (Voulodimos et al., 2018).

Función de activación. Las salidas de una operación lineal, como la convolución, se pasan luego a través de una función de activación no lineal (Yamashita et al., 2018). Se utiliza para garantizar que la representación en el espacio de entrada se asigne a un espacio de salida diferente según los requisitos: Sigmoid, Tan hyperbolic, Rectified linear unit (ReLU), Mish (Sarvamangala & Kulkarni, 2022).

2.7 Referencias

- 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. <https://doi.org/10.3390/rs13112123>
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1), 53. <https://doi.org/10.1186/s40537-021-00444-8>
- 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. <https://doi.org/10.3390/agronomy11081458>
- Ampatzidis, Y., & Partel, V. (2019). UAV-Based High Throughput Phenotyping in Citrus Utilizing Multispectral Imaging and Artificial Intelligence. *Remote Sensing*, 11(4), 410. <https://doi.org/10.3390/rs11040410>
- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv:2004.10934 [cs, eess]*. <http://arxiv.org/abs/2004.10934>
- Burdziakowski, P. (2020). A Novel Method for the Deblurring of Photogrammetric Images Using Conditional Generative Adversarial Networks. *Remote Sensing*, 12(16), 2586. <https://doi.org/10.3390/rs12162586>
- Calvario, G., Alarcón, T. E., Dalmau, O., Sierra, B., & Hernández, C. (2020). An Agave Counting Methodology Based on Mathematical Morphology and Images Acquired through Unmanned Aerial Vehicles. *Sensors*, 20(21), 6247. <https://doi.org/10.3390/s20216247>
- Calvario, G., Sierra, B., Alarcón, T. E., Hernández, C., & Dalmau, O. (2017). A Multi-Disciplinary Approach to Remote Sensing through Low-Cost UAVs. *Sensors (Basel, Switzerland)*, 17(6), E1411. <https://doi.org/10.3390/s17061411>
- Ceja Ramírez, R., González Eguarte, D. R., Ruiz Corral, J. A., Rendón Salcido, L. A., & Flores Garnica, J. G. (2017). Detección de restricciones en la producción de agave azul (Agave tequilana Weber var. Azul) mediante percepción remota. *Terra Latinoamericana*, 35(3), 259-268. <https://doi.org/10.28940/terra.v35i3.252>
- Chen, S. W., Shivakumar, S. S., Dcunha, S., Das, J., Okon, E., Qu, C., Taylor, C. J., & Kumar, V. (2017). Counting Apples and Oranges With Deep Learning: A Data-Driven Approach. *IEEE Robotics and Automation Letters*, 2(2), 781-788. <https://doi.org/10.1109/LRA.2017.2651944>
- Chowdhury, P. N., Shivakumara, P., Nandanwar, L., Samiron, F., Pal, U., & Lu, T. (2022). Oil palm tree counting in drone images. *Pattern Recognition Letters*, 153, 1-9. <https://doi.org/10.1016/j.patrec.2021.11.016>
- CRT. (2016). Consejo Regulador del Tequila. <https://www.crt.org.mx/index.php/es/>
- Csillik, O., Cherbini, J., Johnson, R., Lyons, A., & Kelly, M. (2018). Identification of Citrus Trees from Unmanned Aerial Vehicle Imagery Using

- Convolutional Neural Networks. *Drones*, 2(4), 39. <https://doi.org/10.3390/drones2040039>
- Das Choudhury, S., Samal, A., & Awada, T. (2019). Leveraging Image Analysis for High-Throughput Plant Phenotyping. *Frontiers in Plant Science*, 10. <https://www.frontiersin.org/article/10.3389/fpls.2019.00508>
- DOF. (2012). Diario Oficial de la Federación. http://www.dof.gob.mx/nota_detalle.php?codigo=5282165&fecha=13/12/2012
- Emin, M., Anwar, E., Liu, S., Emin, B., Mamut, M., Abdukeram, A., & Liu, T. (2021). Target Detection-Based Tree Recognition in a Spruce Forest Area with a High Tree Density—Implications for Estimating Tree Numbers. *Sustainability*, 13(6), 3279. <https://doi.org/10.3390/su13063279>
- 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. <https://doi.org/10.1109/JSTARS.2018.2793849>
- Feng, A., Zhou, J., Vories, E., & Sudduth, K. A. (2020). Evaluation of Cotton Emergence Using UAV-Based Narrow-Band Spectral Imagery with Customized Image Alignment and Stitching Algorithms. *Remote Sensing*, 12(11), 1764. <https://doi.org/10.3390/rs12111764>
- Flores, D., González-Hernández, I., Lozano, R., Vázquez-Nicolas, J. M., & Hernández Toral, J. L. (2021). Automated Agave Detection and Counting Using a Convolutional Neural Network and Unmanned Aerial Systems. *Drones*, 5(1), 4. <https://doi.org/10.3390/drones5010004>
- 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. <https://doi.org/10.3390/agronomy10040469>
- Garnica, J. G. F., Reich, R., Zúñiga, E. T., & Bravo, C. A. (2008). Using remote sensing to support different approaches to identify agave (*Agave tequilana* Weber) crops. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37, 4.
- Ghiasi, G., Lin, T.-Y., & Le, Q. V. (2018). DropBlock: A regularization method for convolutional networks. *arXiv:1810.12890 [cs]*. <http://arxiv.org/abs/1810.12890>
- González, A., Amarillo, G., Amarillo, M., & Sarmiento, F. (2016). Drones Aplicados a la Agricultura de Precisión. *Publicaciones e Investigación*, 10, 23-37. <https://doi.org/10.22490/25394088.1585>
- Guo, W., Zheng, B., Potgieter, A. B., Diot, J., Watanabe, K., Noshita, K., Jordan, D. R., Wang, X., Watson, J., Ninomiya, S., & Chapman, S. C. (2018). Aerial Imagery Analysis – Quantifying Appearance and Number of Sorghum Heads for Applications in Breeding and Agronomy. *Frontiers in Plant Science*, 9. <https://www.frontiersin.org/article/10.3389/fpls.2018.01544>
- Hassanalian, M., & Abdelkefi, A. (2017). Classifications, applications, and design challenges of drones: A review. *Progress in Aerospace Sciences*, 91, 99-131. <https://doi.org/10.1016/j.paerosci.2017.04.003>

- Huang, Z., Pan, Z., & Lei, B. (2017). Transfer Learning with Deep Convolutional Neural Network for SAR Target Classification with Limited Labeled Data. *Remote Sensing*, 9(9), 907. <https://doi.org/10.3390/rs9090907>
- Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167 [cs]*. <http://arxiv.org/abs/1502.03167>
- Jin, X., Liu, S., Baret, F., Hemerlé, M., & Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sensing of Environment*, 198, 105-114. <https://doi.org/10.1016/j.rse.2017.06.007>
- Jintasuttisak, T., Edirisinghe, E., & Elbattay, A. (2022). Deep neural network based date palm tree detection in drone imagery. *Computers and Electronics in Agriculture*, 192, 106560. <https://doi.org/10.1016/j.compag.2021.106560>
- Kiropoulos, K., Tsouros, D. C., Dimaraki, F., Triantafyllou, A., Bibi, S., Sarigiannidis, P., & Angelidis, P. (2022). Monitoring Saffron Crops with UAVs. *Telecom*, 3(2), 301-321. <https://doi.org/10.3390/telecom3020017>
- 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. <https://doi.org/10.1109/LGRS.2019.2930549>
- Lin, Y., Chen, T., Liu, S., Cai, Y., Shi, H., Zheng, D., Lan, Y., Yue, X., & Zhang, L. (2022). Quick and accurate monitoring peanut seedlings emergence rate through UAV video and deep learning. *Computers and Electronics in Agriculture*, 197, 106938. <https://doi.org/10.1016/j.compag.2022.106938>
- Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: A review. *Plant Methods*, 17(1), 22. <https://doi.org/10.1186/s13007-021-00722-9>
- 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. <https://doi.org/10.3390/rs12183015>
- Maes, W. H., & Steppe, K. (2019). Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture. *Trends in Plant Science*, 24(2), 152-164. <https://doi.org/10.1016/j.tplants.2018.11.007>
- Misra, D. (2020). Mish: A Self Regularized Non-Monotonic Activation Function. *arXiv:1908.08681 [cs, stat]*. <http://arxiv.org/abs/1908.08681>
- Morales-Avendaño, J. R., Hernández-Cruz, J., López-Cid, I., & Röss, M. (2018). Propuesta Metodológica para Monitoreo de Cultivos de Agave con Vehículos Aéreos no Tripulados (VANT) en Oaxaca, México. *Revista de Ingeniería e Investigación Aplicada UPB/UPTap*, 3, 8.
- Neupane, B., Horanont, T., & Hung, N. D. (2019). Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV). *PLOS ONE*, 14(10), e0223906. <https://doi.org/10.1371/journal.pone.0223906>
- Nijland, W., de Jong, R., de Jong, S. M., Wulder, M. A., Bater, C. W., & Coops, N. C. (2014). Monitoring plant condition and phenology using infrared

- sensitive consumer grade digital cameras. *Agricultural and Forest Meteorology*, 184, 98-106. <https://doi.org/10.1016/j.agrformet.2013.09.007>
- Oscó, L. P., Marcato Junior, J., Marques Ramos, A. P., de Castro Jorge, L. A., Fatholahi, S. N., de Andrade Silva, J., Matsubara, E. T., Pistori, H., Gonçalves, W. N., & Li, J. (2021). A review on deep learning in UAV remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102456. <https://doi.org/10.1016/j.jag.2021.102456>
- 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. <https://doi.org/10.3390/drones4030041>
- Paneque-Gálvez, J., McCall, M. K., Napoletano, B. M., Wich, S. A., & Koh, L. P. (2014). Small Drones for Community-Based Forest Monitoring: An Assessment of Their Feasibility and Potential in Tropical Areas. *Forests*, 5(6), 1481-1507. <https://doi.org/10.3390/f5061481>
- Prasad, M. V. D., Lakshamma, B. J., Chandana, A. H., Komali, K., Manoja, M., Rajesh Kumar, P., Raghava Prasad, C., Inthiyaz, S., & Sasi Kiran, P. (2017). An efficient classification of flower images with convolutional neural networks. *International Journal of Engineering & Technology*, 7(1.1), 384. <https://doi.org/10.14419/ijet.v7i1.1.9857>
- Primicerio, J., Caruso, G., Comba, L., Crisci, A., Gay, P., Guidoni, S., Genesio, L., Ricauda Aimonino, D., & Vaccari, F. P. (2017). Individual plant definition and missing plant characterization in vineyards from high-resolution UAV imagery. *European Journal of Remote Sensing*, 50(1), 179-186. <https://doi.org/10.1080/22797254.2017.1308234>
- Quirós-Rosado, E. (2015). *Introducción a la Fotogrametría y Cartografía aplicadas a la Ingeniería Civil* (1.ª ed.). Universidad de Extremadura.
- Radoglou-Grammatikis, P., Sarigiannidis, P., Lagkas, T., & Moscholios, I. (2020). A compilation of UAV applications for precision agriculture. *Computer Networks*, 172, 107148. <https://doi.org/10.1016/j.comnet.2020.107148>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *arXiv:1506.02640 [cs]*. <http://arxiv.org/abs/1506.02640>
- Redmon, J., & Farhadi, A. (2016). YOLO9000: Better, Faster, Stronger. *arXiv:1612.08242 [cs]*. <http://arxiv.org/abs/1612.08242>
- Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *arXiv:1804.02767 [cs]*. <http://arxiv.org/abs/1804.02767>
- Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv:1506.01497 [cs]*. <http://arxiv.org/abs/1506.01497>
- Reza, M. N., Na, I. S., & Lee, K.-H. (2017). Automatic Counting of Rice Plant Numbers After Transplanting Using Low Altitude UAV Images. *International Journal of Contents*, 13(3), 1-8. <https://doi.org/10.5392/IJoC.2017.13.3.001>
- Santamaría-Peña, J. (2011). *Fundamentos de fotogrametría*. Universidad de La Rioja. <https://dialnet.unirioja.es/servlet/libro?codigo=492591>
- 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. <https://doi.org/10.3390/rs12050748>
- Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: A survey. *Evolutionary Intelligence*, 15(1), 1-22. <https://doi.org/10.1007/s12065-020-00540-3>
- Shafi, U., Mumtaz, R., García-Nieto, J., Hassan, S. A., Zaidi, S. A. R., & Iqbal, N. (2019). Precision Agriculture Techniques and Practices: From Considerations to Applications. *Sensors*, 19(17), 3796. <https://doi.org/10.3390/s19173796>
- Shao, H., Tang, R., Lei, Y., Mu, J., Guan, Y., & Xiang, Y. (2021). Rice Ear Counting Based on Image Segmentation and Establishment of a Dataset. *Plants*, 10(8), 1625. <https://doi.org/10.3390/plants10081625>
- SIAP. (2020). Cierre de la producción agrícola 2020. Anuario Estadístico de la Producción Agrícola. <https://nube.siap.gob.mx/cierreagricola/>
- Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sensing*, 12(19), 3136. <https://doi.org/10.3390/rs12193136>
- Sultana, F., Sufian, A., & Dutta, P. (2018). Advancements in Image Classification using Convolutional Neural Network. *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 122-129. <https://doi.org/10.1109/ICRCICN.2018.8718718>
- Tsouros, D. C., Bibi, S., & Sarigiannidis, P. G. (2019). A Review on UAV-Based Applications for Precision Agriculture. *Information*, 10(11), 349. <https://doi.org/10.3390/info10110349>
- Ubbens, J., Cieslak, M., Prusinkiewicz, P., & Stavness, I. (2018). The use of plant models in deep learning: An application to leaf counting in rosette plants. *Plant Methods*, 14(1), 6. <https://doi.org/10.1186/s13007-018-0273-z>
- Valente, J., Sari, B., Kooistra, L., Kramer, H., & Mùcher, S. (2020). Automated crop plant counting from very high-resolution aerial imagery. *Precision Agriculture*, 21(6), 1366-1384. <https://doi.org/10.1007/s11119-020-09725-3>
- van der Merwe, D., Burchfield, D. R., Witt, T. D., Price, K. P., & Sharda, A. (2020). Chapter One—Drones in agriculture. En D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. 162, pp. 1-30). Academic Press. <https://doi.org/10.1016/bs.agron.2020.03.001>
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 2018, e7068349. <https://doi.org/10.1155/2018/7068349>
- 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. <https://doi.org/10.3390/s21020507>
- Wang, Z., Jin, L., Wang, S., & Xu, H. (2022). Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system. *Postharvest Biology and Technology*, 185, 111808. <https://doi.org/10.1016/j.postharvbio.2021.111808>

- 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. <https://doi.org/10.3390/rs11060691>
- Wu, X., Sahoo, D., & Hoi, S. C. H. (2020). Recent advances in deep learning for object detection. *Neurocomputing*, 396, 39-64. <https://doi.org/10.1016/j.neucom.2020.01.085>
- Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging*, 9(4), 611-629. <https://doi.org/10.1007/s13244-018-0639-9>
- Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Zhao, X., Li, Z., Li, H., & Yang, H. (2017). Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives. *Frontiers in Plant Science*, 8. <https://www.frontiersin.org/article/10.3389/fpls.2017.01111>
- Zarco-Tejada, P. J., Guillén-Climent, M. L., Hernández-Clemente, R., Catalina, A., González, M. R., & Martín, P. (2013). Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV). *Agricultural and Forest Meteorology*, 171-172, 281-294. <https://doi.org/10.1016/j.agrformet.2012.12.013>
- Zhang, Q., Liu, Y., Gong, C., Chen, Y., & Yu, H. (2020). Applications of Deep Learning for Dense Scenes Analysis in Agriculture: A Review. *Sensors*, 20(5), 1520. <https://doi.org/10.3390/s20051520>
- Zhao, B., Zhang, J., Yang, C., Zhou, G., Ding, Y., Shi, Y., Zhang, D., Xie, J., & Liao, Q. (2018). Rapeseed Seedling Stand Counting and Seeding Performance Evaluation at Two Early Growth Stages Based on Unmanned Aerial Vehicle Imagery. *Frontiers in Plant Science*, 9. <https://www.frontiersin.org/article/10.3389/fpls.2018.01362>
- Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07), 12993-13000. <https://doi.org/10.1609/aaai.v34i07.6999>

3 ARTÍCULO CIENTÍFICO

DETECCIÓN Y CONTEO DE PLANTAS DE AGAVE MEDIANTE APRENDIZAJE PROFUNDO EN IMÁGENES OBTENIDAS POR RPAS

Juan Espinoza Hernández¹

Gilberto de Jesús López Canteñs^{1*}

Irineo Lorenzo López Cruz¹

Eugenio Romantchik Kriuchkova¹

¹Universidad Autónoma Chapingo, Posgrado en Ingeniería Agrícola y Uso Integral del Agua. Carretera México-Texcoco km 38.5, Chapingo, México, C. P. 56230, México. Autor de correspondencia: glopezc@chapingo.mx

3.1 Resumen

Las plantas de agave son un elemento estratégico en la cadena de producción del tequila. Los inventarios de agave ayudan a monitorear, estimar y planificar la producción. Este trabajo tiene como objetivo detectar y contar plantas de agave en condiciones de campo mediante aprendizaje profundo en imágenes aéreas. Para ellos, se capturaron imágenes de alta resolución rojo-verde-azul (RGB) a 40, 60 y 80 m de altura con un sistema de aeronave pilotada a distancia (RPAS). Se generaron tres ortomosaicos y usaron para crear la base de datos de plantas de agave. Esta base de datos se utilizó para entrenar dos algoritmos convolucionales: YOLOv4 y YOLOv4-tiny. La evaluación de los algoritmos muestra un valor mean average precision (mAP) de 0.99 para ambos algoritmos y una puntuación F1 de 0.95 para YOLOv4 y de 0.96 para YOLOv4-tiny. Para alturas de 40 y 60 m se obtuvo el 100% de plantas detectadas en ambos algoritmos. Sin embargo, a una altura de 80 m se obtuvo el 96.8% de plantas detectadas con el algoritmo YOLOv4, y el 93.6% con el algoritmo YOLOv4-tiny. La salida de los algoritmos YOLO se utilizó para realizar un cálculo aproximado de la distancia entre plantas detectadas, teniendo un error máximo de 20 cm. Los resultados indican que los algoritmos son robustos, mostrando que es posible detectar y contar agaves en tres alturas de vuelo diferentes. Así mismo, es posible usar la distancia entre plantas de agave calculada para detectar espacios vacíos.

Palabras clave: *agave tequilana Weber*, algoritmos, dron, agricultura de precisión, YOLO

Tesis de Maestría en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo.

Autor: Juan Espinoza Hernández

Director de Tesis: Gilberto de Jesús López Canteñs

3.2 Abstract

Agave plants are a strategic element in the tequila production chain. Agave plant inventories help to monitor, estimate, and plan production. This work aims to detect and count agave plants under field conditions using deep learning in aerial images. To do this, high-resolution red-green-blue (RGB) images were captured at 40, 60, and 80 m height with a Remotely Piloted Aircraft System (RPAS). Three orthomosaics were generated and used to create the agave plant database. This database was used to train two convolutional algorithms: YOLOv4 and YOLOv4-tiny. The algorithms' evaluation shows a mean average precision (mAP) value of 0.99 for both algorithms and an F1 score of 0.95 for YOLOv4 and 0.96 for YOLOv4-tiny. For heights of 40 and 60 m, 100% of plants detected in both algorithms were obtained. However, at a height of 80 m, 96.8% of plants were detected with YOLOv4, and 93.6% with YOLOv4-tiny. The YOLO algorithms' output was used to make an approximate calculation of the distance between detected plants, having a maximum error of 20 cm. The results show that the algorithms are robust, showing that it is possible to detect and count agaves at three different flight heights. In addition, it is possible to use the calculated distance between agave plants to detect empty spaces.

Keywords: agave tequilana Weber, algorithms, drone, precision agriculture, YOLO.

3.3 Introducción

El agave azul (*Agave tequilana Weber*), empleado en la elaboración del tequila se cultiva en la región conocida como Denominación de Origen del Tequila, que incluye a los estados mexicanos de Jalisco, Michoacán, Tamaulipas, Nayarit y Guanajuato, contribuyendo el estado de Jalisco con la mayor producción (Ceja-Ramírez et al., 2017).

Dada la importancia de la industria tequilera, el campo de agave ha venido innovando para enfrentar diversos retos. Ejemplo de esta innovación es el empleo de sistemas de aeronaves pilotadas a distancia (RPAS), conocidos comúnmente como drones, que han entrado en apoyo con información estadística de plantaciones (CRT, 2016).

Es importante señalar que, el elemento estratégico de la cadena productiva del tequila lo constituye la propia planta de agave, lo cual conlleva a la necesidad de realizar inventarios de estas para así poder estimar rendimientos, además de predecir y planificar la producción, teniendo un control sobre el número de plantas establecidas (Calvario et al., 2020). Mediante los inventarios se pueden responder cuestiones como la disponibilidad del agave, etapa vegetativa, ubicación, propietario, estado fitosanitario, entre otras, para llevar un mejor manejo y control. Además, la norma oficial mexicana (NOM-006-SCFI-2012) menciona que todo agave que es utilizado para la producción del tequila debe estar debidamente registrado ante el Consejo Regulador del Tequila A.C. y que los productores de *Agave tequilana Weber* deben actualizar anualmente su registro de plantaciones y predios, manifestando la condición y/o estado en su inventario correspondiente (DOF, 2012). Dando así un valor agregado a la cadena de producción, elaboración y comercialización de la bebida (CRT, 2016).

Tradicionalmente el conteo de plantas de agave ha sido un proceso largo, costoso y que pone en riesgo la salud de los trabajadores. Este conteo manual es llevado a cabo por grupos de personas que recorren los campos contando plantas, tarea que frecuentemente toma hasta seis meses, dando como resultado de uno a dos conteos como máximo por año, añadiendo que es propenso a

errores. En este marco cobra importancia la teledetección aérea, a través del análisis de imágenes adquiridas mediante dispositivos que no tienen contacto físico con el objeto de interés (Lillesand et al., 2015). En ese sentido, en la actualidad se pueden capturar fotografías con RPAS para utilizarlas en métodos de identificación de objetos de interés (Prasad et al., 2017), como lo son los algoritmos de aprendizaje profundo.

Con algoritmos de aprendizaje profundo se pueden analizar grandes volúmenes de información de forma rápida y precisa. En años recientes las redes de aprendizaje profundo, en especial las redes neuronales convolucionales (CNNs) han tomado relevancia. Estas redes son un tipo especial de red neuronal multicapa cuya arquitectura típica se compone de bloques únicos o múltiples de capas de convolución y agrupación, después una o más capas completamente conectadas y una capa de salida (Sultana et al., 2018). En la última década la implementación de aprendizaje profundo en el análisis de imágenes ha sido la mejor opción, pues las CNNs son poderosas para resolver tareas como la detección y clasificación de objetos al extraer información de las imágenes (Prasad et al., 2017). Por ejemplo, Ubbens et al., (2018) realizaron un algoritmo de aprendizaje profundo para el conteo de hojas en disposición de roseta, entrenando una CNN. Chen et al., (2017) describieron un proceso de conteo de manzanas, teniendo el reto de los cambios de iluminación y las oclusiones del follaje y frutas cercanas utilizando dos CNNs. Por otra parte, Ampatzidis & Partel, (2019) desarrollaron una técnica para evaluar las características fenotípicas en cultivos de cítricos utilizando imágenes obtenidas por RPAS y aplicando CNNs. En relación con lo anterior, se nota una tendencia a implementar algoritmos de aprendizaje profundo en imágenes, ya que se ha verificado dan buenos resultados tanto en clasificación como en detección de objetos.

En cuanto a investigaciones enfocadas en la detección y conteo de plantas de agave está el trabajo de Garnica et al., (2008) en el cual presentaron un estudio en el 2008 sobre imágenes obtenidas por el satélite LANDSAT 7, con el objetivo de identificar y ubicar plantaciones de agave. Utilizaron métodos de árbol de

regresión y clasificación logrando un 70% de precisión. Dentro de los problemas que enfrentaron esta la densidad de plantas en las parcelas, la cual no es homogénea, además los diferentes tipos de suelo en los que se encuentran establecidas las plantas, y otras cubiertas vegetales que se pueden confundir con la planta. Más recientemente, Ceja-Ramírez et al., (2017) realizaron una investigación para determinar y ubicar la superficie ocupada por agave azul que presentó restricciones en su producción. Utilizaron una imagen adquirida por el satélite LANDSAT 5 en la cual aplicaron interpolación digital a través de un proceso de clasificación supervisado, logrando una precisión del 73%. En ese mismo año, Calvario et al., (2017) utilizaron imágenes RGB obtenidas por RPAS a una altura de vuelo de 60 m, para monitorear cultivos de agave. Aplicaron técnicas de extracción de capas para separar las plantas de agave de las malezas y otros elementos e hicieron uso de espacios de colores como el CIE. La precisión que obtuvieron fue del 99% pero no resolvieron los problemas de superposición entre plantas presente en las plantaciones de agave.

En el 2018, Morales-Avenidaño et al., (2018) realizaron una propuesta metodológica para contar plantas de agave, mediante la captura de imágenes RGB por RPAS y un trabajo de gabinete en programas de cómputo comerciales. Calvario et al., (2020) describieron un algoritmo, continuación de Calvario et al., (2017), basado en morfología matemática aplicado a imágenes RGB de alta resolución obtenidas por RPAS, con el objetivo de realizar el conteo de plantas. Observaron que la precisión del algoritmo dependía de la heterogeneidad del tamaño de las plantas en el campo y la iluminación. La precisión osciló entre un 83% a 98%. Recientemente Flores et al., (2021) realizaron la detección y conteo automático de agave usando una CNN aplicada a imágenes RGB obtenidas por RPAS a una altura de 50 m, logrando una precisión del 96%, pero con un alto coste computacional.

Se reafirma la tendencia a implementar técnicas de aprendizaje profundo en conjunto con imágenes adquiridas por RPAS. Sin embargo, para aplicar estas técnicas en agave se requiere de bases de datos (Calvario et al., 2020). Para

ello, Flores et al., (2021) crearon una base de datos de plantas de agave, ya que a la fecha de su investigación no se contaba con una, en la cual implementaron las técnicas mencionadas. Los mismos autores Flores et al., (2021), agregaron que en el futuro se pueden probar enfoques como los algoritmos convolucionales YOLO, por su versatilidad y robustez, además de que podrían mejorar la velocidad de procesamiento. Dichos algoritmos han sido aplicados con éxito en distintos cultivos, por ejemplo, Ammar et al., (2021); Chowdhury et al., (2022); Jintasuttisak et al., (2022) aplicaron diversos algoritmos de CNNs recientes para conteo y localización de palmeras, dentro de ellos los algoritmos YOLO. Al igual que en Wang et al., (2021) para el conteo de plantas de maíz y así también Lin et al., (2022) detectando plántulas de maní.

Por lo anterior, se plantea el objetivo de detectar y contar plantas de agave mediante aprendizaje profundo, específicamente con los algoritmos YOLOv4 y YOLOv4-tiny, en imágenes RGB de alta resolución adquiridas por RPAS.

3.4 Trabajo relacionado

3.4.1 Redes neuronales convolucionales

Las CNNs se inspiran en el sistema visual animal aprendiendo de forma automática y adaptativa jerarquías de características, desde patrones de bajo a alto nivel. Una CNN consta de tres tipos de capas: convolucionales, de agrupación y completamente conectadas, en donde cada tipo de capa desempeña cierta función (Voulodimos et al., 2018; Zhang et al., 2020) (Figura 1).

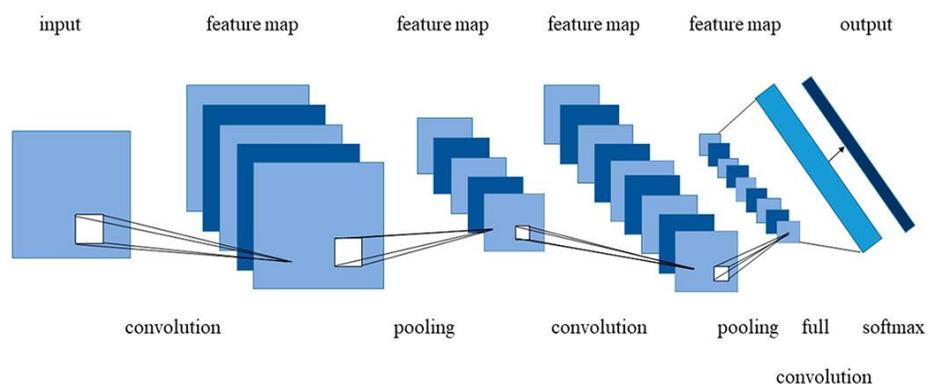


Figura 1. Arquitectura general de una CNN. Fuente: (Zhang et al., 2020).

La arquitectura típica consta de repeticiones de capas de convolución y una capa de agrupación, seguidas de una o más capas completamente conectadas y una capa de salida (Sultana et al., 2018). La capa de convolución y la capa de agrupación se alternan varias veces (Liu & Wang, 2021).

Capa convolucional. Utiliza filtros o kernels, para extraer y generar mapas de características (Figura 2a). Consiste en operaciones lineales y no lineales, operación de convolución y función de activación. En estas capas se aplica una pequeña matriz de números llamada kernel a la matriz de entrada (Sarvamangala & Kulkarni, 2022).

Capa de agrupación (Pooling). Se encarga de reducir las dimensiones de los mapas de características de la entrada (ancho x alto) para la siguiente capa convolucional. La agrupación promedio y la agrupación máxima son las más utilizadas (Yamashita et al., 2018). La operación de agrupación da como resultado la partición de su entrada en un conjunto de parches rectangulares (Figura 2b). Cada parche se reemplaza por un solo valor según el tipo de agrupación aplicada (Sarvamangala & Kulkarni, 2022).

Capa completamente conectada (FC). Es la última sección de una CNN (Figura 2c). Esta capa toma información de la capa anterior y realiza una operación para generar la salida, por ejemplo, las probabilidades para cada clase en las tareas de clasificación. La capa final normalmente tiene el mismo número de nodos de salida que el número de clases.

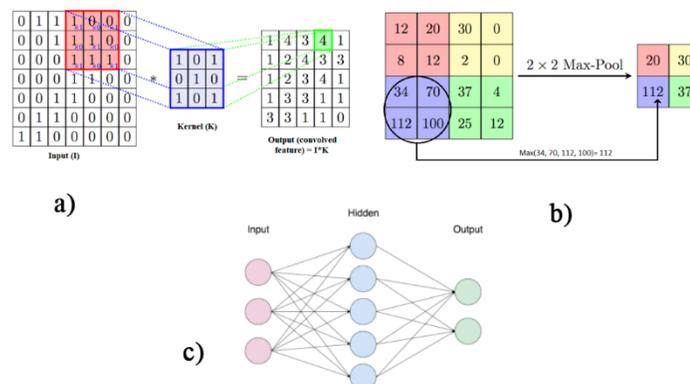


Figura 2. Componentes típicos de una CNN: a) convolución, b) capa de agrupación, c) capa completamente conectada. Fuente: (Sultana et al., 2018)

Entrenar una CNN es el proceso de encontrar los kernels de las capas de convolución y pesos en las FC que funcionan mejor para una tarea determinada en función de un conjunto de datos de entrenamiento, es decir que minimizan las diferencias entre las predicciones de salida y las etiquetas de verdad dadas para el entrenamiento (Yamashita et al., 2018).

3.4.2 Algoritmos de detección de objetos YOLO

YOLO es la forma abreviada de You Only Look Once. Utiliza CNNs para la detección de objetos, pudiendo detectar varios objetos en una sola imagen. Lo que significa que además de predecir las clases de los objetos, YOLO también detecta ubicaciones de estos objetos en la imagen.

Los algoritmos YOLO han avanzado a través de diferentes versiones implementando mejoras. La primera versión de YOLO (Redmon et al., 2016) fue escrita en un marco de red neuronal de código abierto escrito en C y CUDA llamado Darknet, que admite el cálculo de CPU y GPU. Fue el primer algoritmo en combinar la tarea de dibujar cuadros delimitadores e identificar etiquetas de clase. Para YOLOv2 (Redmon & Farhadi, 2016), se incluyó un método para normalizar cada lote de datos mediante Batch Normalization (Ioffe & Szegedy, 2015), así también se incluyeron las cajas de anclaje. En YOLOv3 (Redmon & Farhadi, 2018), se agregó una puntuación de objetividad a la predicción del cuadro delimitador, y también se lograron hacer predicciones en tres niveles separados para mejorar el rendimiento en objetos más pequeños. Recientemente, en el año 2020 fue desarrollado YOLOv4 (Bochkovskiy et al., 2020).

De manera general, YOLO funciona aplicando una única red neuronal a toda la imagen. Divide la imagen de entrada en regiones o celdas, y por cada celda predice la probabilidad de una clase sí hay un objeto y da los valores b_x , b_y , b_h , b_w (coordenadas del cuadro delimitador predicho). Posteriormente se aplica intersección sobre la unión (IoU) y supresión no máxima para mejorar los resultados, evitando así que varios cuadros delimitadores predichos detecten el mismo objeto (Figura 3).

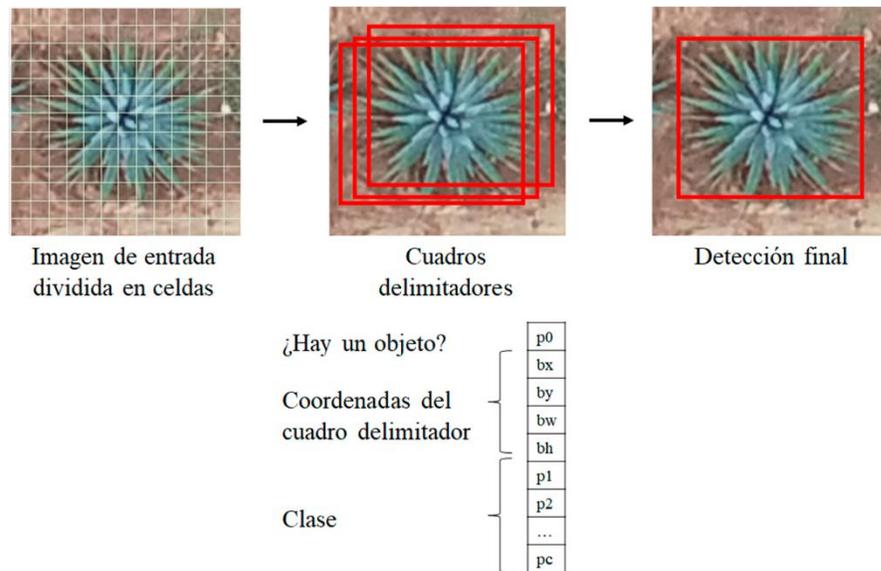


Figura 3. Supresión no máxima y la salida de YOLO.

Cabe mencionar que existen algoritmos de detección de objetos de una etapa y de dos etapas. Los detectores de dos etapas, como Faster R-CNN (Ren et al., 2016), separan las tareas de localización y clasificación de objetos para un cuadro delimitador, mientras que los detectores de una etapa realizan las predicción para la localización y clasificación de objetos al mismo tiempo. Por lo tanto, el algoritmo YOLO es un detector de una etapa. Una diferencia entre los detectores de una y dos etapas es que detectores como YOLO tienen altas velocidades de inferencia (Ameijeiras-Sánchez et al., 2020).

3.5 Materiales y métodos

3.5.1 Área de estudio

El área de estudio consta de una superficie de 4.4 hectáreas de una plantación de agave azul ubicada sobre la carretera Guadalajara-Tepic km 36.2, municipio de El Arenal, Jalisco, México, ubicada geográficamente en las coordenadas 20°47'35.39"N y 103°42'4.27"O (Figura 4).

Se seleccionó esta área de estudio ya que contaba con agave en diferentes etapas fenológicas, desde hijuelos hasta agaves desarrollados (de más de 5 años), dispuestos a una hilera y en doble hilera. Esta parcela se caracteriza por la presencia de maleza y suelo desnudo en la plantación (Figura 5).

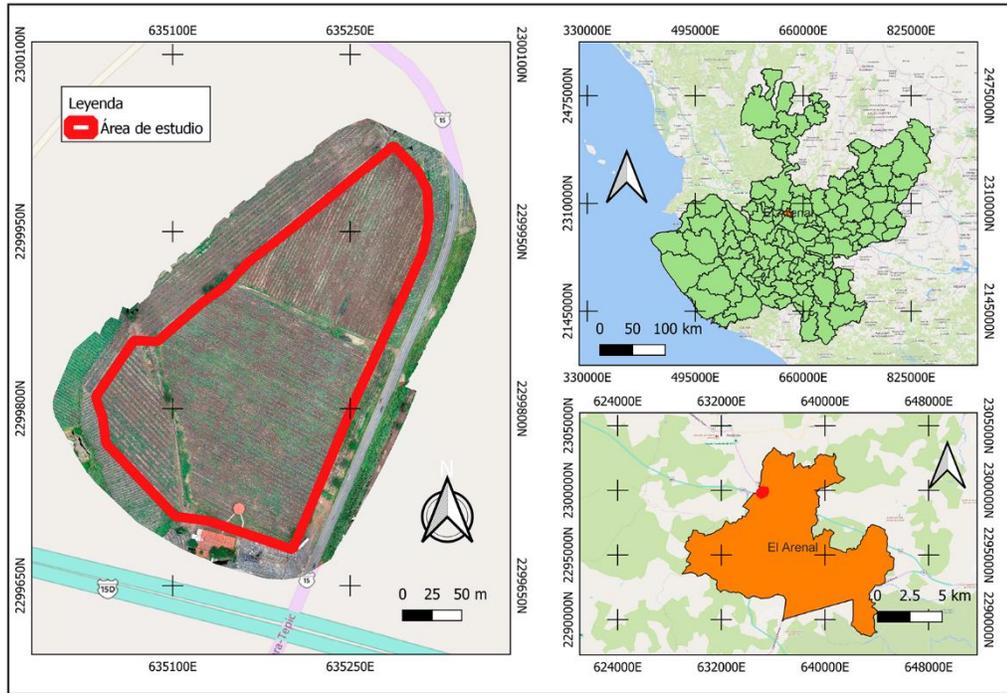


Figura 4. Localización del campo de agave.

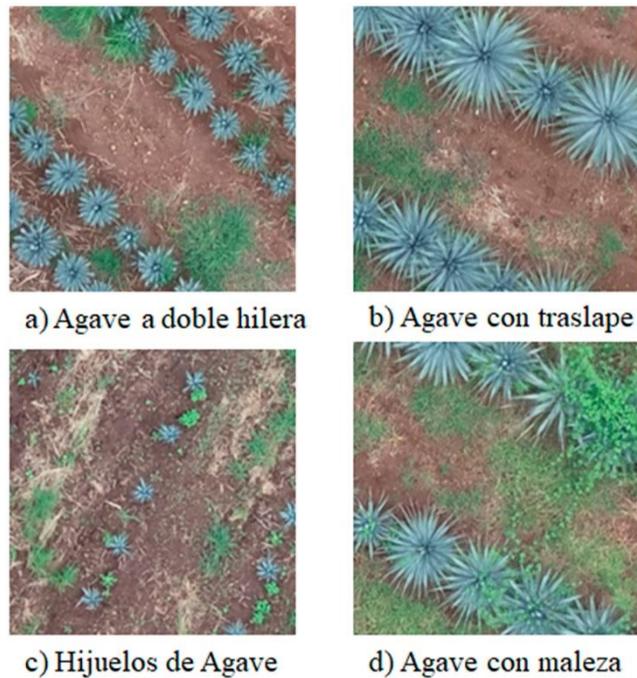


Figura 5. Campo de agave con diferentes condiciones.

3.5.2 Adquisición de imágenes con RPAS y procesamiento fotogramétrico

El proceso de adquisición de imágenes con RPAS y procesamiento fotogramétrico se muestra en la Figura 6.

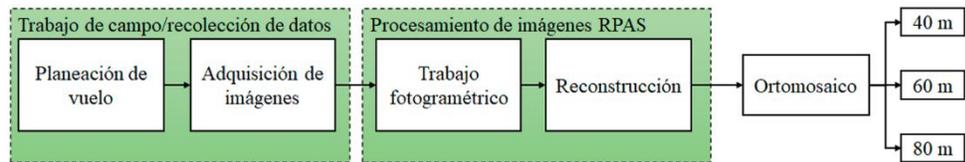


Figura 6. Flujo de captura y procesamiento de imágenes.

Se utilizó un RPAS multirrotor DJI Phantom 4 Pro V2.0 (SZ DJI Technology Co., Shenzhen, Guangdong, China) para adquirir las imágenes de alta resolución (Figura 7). El sistema RPAS estaba equipado con una cámara RGB de 20 MP con distancia focal equivalente a 24 mm.



Figura 7. RPAS DJI Phantom 4 Pro V2.0

El área de estudio se voló el 05/09/2021 entre las 11:00 y las 14:00, hora local, a tres diferentes alturas (40, 60, y 80 m), para tener una base de datos con más variación en cuanto a tamaño de plantas de agave. La temperatura era de 27°C, con nubes dispersas y una velocidad del viento de 5 km/h.

Las misiones de vuelo se planificaron y llevaron a cabo utilizando la aplicación móvil DJI Pilot (SZ DJI Technology Co., Shenzhen, Guangdong, China), para lo cual se configuró el valor ISO a 100, el valor de exposición a 0, la obturación a 1/400 y el traslape tanto lateral como frontal de las imágenes a un 80%. Las imágenes se capturaron desde la vista superior (de manera vertical), es decir con el eje óptico de la cámara perpendicular a la plantación.

Se utilizó el programa de cómputo Pix4Dmapper (Pix4D S. A., Lausanne, Switzerland) para importar las imágenes y crear los ortomosaicos del área de

estudio, realizando la calibración y orientación interna y externa de las cámaras. Al final se obtuvieron tres ortomosaicos RGB, cada uno con su respectivo valor de distancia de muestreo en tierra (GSD) (Cuadro 1).

Cuadro 1. Caracterización de los vuelos realizados.

Altura de vuelo (m)	GSD (cm/píxel)
40	1.10
60	1.64
80	2.19

3.5.3 División de datos y etiquetado de imágenes

Los tres ortomosaicos obtenidos se dividieron en imágenes de tamaño 416x416 píxeles, obteniendo un total de 2403.

El empleo de métodos de aprendizaje supervisado implica indicar al modelo el objeto que debe encontrar en las imágenes y así poder extraer las características. Para ello se llevó a cabo un etiquetado de datos. Se utilizó la herramienta Labellmg (*Labellmg.*, s. f.) para etiquetar las imágenes. Se localizó manualmente la posición de cada planta de agave en cada una de las imágenes y se dibujó un cuadro delimitador asignando una clase, en este caso se nombró Agave a la única clase. Durante el proceso Labellmg guarda para cada imagen etiquetada un archivo con extensión .txt, formato estándar de YOLO, que incluye la clase y las coordenadas normalizadas correspondientes a cada cuadro delimitador de todos los objetos presentes en la imagen, donde cada renglón es una planta etiquetada (Figura 8).

```

0 0.152644 0.094952 0.156250 0.170673
0 0.057692 0.037260 0.110577 0.069712
0 0.462740 0.097356 0.108173 0.093750

```

▲
▲
▲
▲
▲

Clase
X
Y
Ancho
Alto

Figura 8. Etiquetas creadas en una imagen.

Las 2403 imágenes etiquetadas se dividieron aleatoriamente en carpetas separadas. La base de datos quedó dividida en 70% para entrenamiento, 15%

para prueba y 15% para evaluación, respectivamente. El Cuadro 2 muestra la distribución de la base de datos en los diferentes conjuntos.

Cuadro 2. Composición de la base de datos.

Conjunto	Número de imágenes
Entrenamiento	1683
Prueba	360
Evaluación	360
Total	2403

3.5.4 YOLOv4 y YOLOv4-tiny

Uno de los algoritmos utilizado fue YOLOv4 que implementa una detección de extremo a extremo, y que se compone de cuatro partes principales, que son: Input, Backbone, Neck y Head (Dense Prediction) (Figura 9).

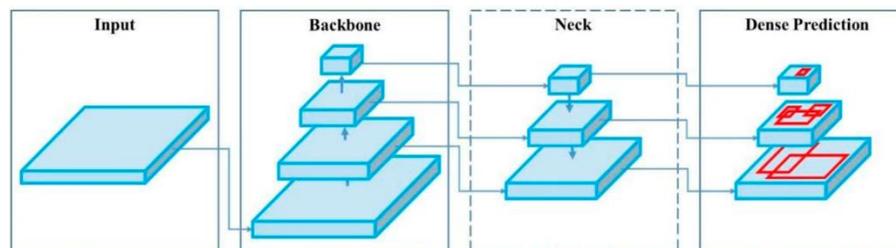


Figura 9. Arquitectura del algoritmo YOLOv4. Fuente: (Bochkovskiy et al., 2020)

Como Backbone YOLOv4 implementa CSPNet con Darknet53 formando CSPDarknet53, con la cual extrae mapas de características. En el Neck, utilizando Spatial Pyramid Pooling se recopilan mapas de características generados de diferentes etapas. Y finalmente en Head, la sección de predicción, el algoritmo YOLOv4 utiliza el mismo mecanismo que YOLOv3 detectando el objeto y dibujando el cuadro delimitador, así como también se obtiene la información de las coordenadas b_x , b_y , b_w , b_h y el valor de confianza, como se mencionó anteriormente.

YOLOv4, además de las contribuciones de las versiones anteriores, también implementa nuevas mejoras propias del algoritmo que los autores Bochkovskiy et al., (2020) agregaron. Dentro de ellas están: Mosaic data augmentation, Weighted-Residual-Connections, Cross-Stage-Partial-connections, DropBlock

regularization (Ghiasi et al., 2018), Cross mini-Batch Normalization, Self-adversarial-training, CloU-loss y Mish-activation.

Introduciendo CloU (Zheng et al., 2020) durante el entrenamiento como función de pérdida, se mejora la precisión en comparación con YOLOv3. Mientras que la función de activación Mish (Misra, 2020) proporciona una mejor precisión, una pérdida más baja y suave, y funciona mejor en comparación con los resultados de ReLU y Leaky ReLU.

En base a lo anterior, se seleccionó el algoritmo convolucional YOLOv4 ya que se hace uso de la función de activación Mish y se mejora la calidad de los mapas de características que se convierten en entrada de la sección de predicción Head, obteniendo mejores resultados.

El otro algoritmo es YOLOv4-tiny, versión comprimida de YOLOv4 diseñada para poder ser entrenada en equipos de cómputo con pocos recursos. Una de las grandes diferencias entre YOLOv4 y YOLOv4-tiny es el tamaño de la red, ya que en la versión tiny su tamaño se reduce, es decir tiene un menor número de capas convolucionales en el Backbone, contando con un total de 29 capas. Además, la red hace detecciones en dos escalas diferentes, en lugar de tres, y por consiguiente el número de cuadros de anclaje para la predicción es menor.

3.5.5 Transferencia de aprendizaje

Para el entrenamiento de los algoritmos se emplea la técnica conocida como transferencia de aprendizaje, ya que con esta se transfiere conocimiento de pesos de entrenamiento de un algoritmo de detección llevado a cabo en cierto conjunto de datos a un nuevo conjunto de datos cuando las tareas de detección originales son diferentes a las de interés (Huang et al., 2017). Este enfoque mejora la eficiencia durante el entrenamiento, en contraste con inicializar los pesos aleatoriamente (Wang et al., 2022).

Para el caso de detección de plantas de agave se emplearon los pesos de YOLOv4 y YOLOv4-tiny previamente entrenados en la base de datos COCO (Lin

et al., 2015), con el fin de disminuir el tiempo de entrenamiento y mejorar el rendimiento de los algoritmos.

3.5.6 Hiperparámetros

La configuración de los hiperparámetros en los algoritmos YOLOv4 y YOLOv4-tiny es de importancia en el rendimiento al momento del entrenamiento (Wang et al., 2022). Ajustarlos facilita el entrenamiento y maximiza el número de detecciones correctas, en este caso de detecciones de plantas de agave, no sólo en los datos de entrenamiento, sino también en los datos de prueba y otras entradas.

En el código fuente del marco Darknet de YOLO se realizaron modificaciones para adecuar los algoritmos al conjunto de datos personalizado. Se fijó el tamaño de la entrada del algoritmo al tamaño de las imágenes de la base de datos, 416x416x3. El algoritmo original fue entrenado con 80 clases por lo que se cambió el número a una clase, "Agave". El número de filtros para extraer características, en las capas convolucionales antes de las capas YOLO, se modificó de 255 a 18 considerando la fórmula: (número de clases + 5) x 3. El número de iteraciones se fijó en 6000. Los pasos se calcularon usando 80% y 90% del número iteraciones, quedando entre 4800 y 5400. En el Cuadro 3 se muestran los hiperparámetros utilizados que se mantuvieron como predeterminados.

Cuadro 3. Hiperparámetros utilizados en los algoritmos.

Algoritmo	Batch	Momentum	Tasa de aprendizaje
YOLOv4	64	0.949	0.001
YOLOv4-tiny	64	0.9	0.00261

3.5.7 Entrenamiento

El entrenamiento de los algoritmos se llevó a cabo utilizando el servicio Google Colaboratory Pro que brinda una GPU para disminuir los tiempos de entrenamiento cuando se trabaja con redes neuronales. En este caso la GPU asignada fue una Tesla P100-PCIE-16GB.

3.5.8 Métricas de evaluación

Para evaluar el rendimiento de los algoritmos implementados se utilizaron las métricas Precision, Recall, F1-score y Mean average precisión (mAP) (Padilla et al., 2021). En donde, Precision es la probabilidad de que un objeto identificado sea una planta de agave, es decir el porcentaje de detecciones correctas. Recall es la capacidad para identificar las plantas. F1-score es el promedio armónico entre Precision y Recall, es utilizada para medir el desempeño general obteniendo un valor alto si tanto Precision como Recall son altos.

A continuación, se muestran las ecuaciones de las métricas consideradas para la evaluación de los algoritmos:

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

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

$$F1\ Score = 2 \left(\frac{Precision * Recall}{Precision + Recall} \right) \quad (3)$$

Donde TP representa los verdaderos positivos (plantas de agave identificadas correctamente), FP representa los falsos positivos (otros objetos identificados como agave que sin embargo no lo son) y FN representa los falsos negativos (agave no identificado).

Mientras que mean average precisión (mAP) resume la precisión de detección promedio en un valor definido de intersección sobre la unión (IoU).

3.5.9 Detecciones y conteo de plantas

Este procedimiento se realizó con una computadora portátil con Windows 10 de 64 bits, procesador AMD Ryzen 5 4600H, con 512 GB de disco de estado sólido, 8 GB de memoria RAM y tarjeta de video NVIDIA GTX 1650 de 4 GB respectivamente.

Cada imagen consta de un cierto número de plantas que se contaron manualmente para compararlas con los resultados experimentales al aplicar los algoritmos YOLOv4 y YOLOv4-tiny con los mejores pesos obtenidos tras los

entrenamientos. Para ello se consideraron todas las plantas dentro de la imagen, incluyendo las plantas completamente visibles y además las plantas de los bordes, en las cuales solo se puede observar una parte de estas.

Para la probabilidad de que un objeto detectado fuera considerado como una planta de agave se utilizó un umbral del 50%, es decir un valor de IoU de 0.5. Si el valor de intersección sobre la unión es superior a 0.5 y una confianza superior al 0.25 se consideró al objeto detectado como una planta de agave.

3.5.10 Cálculo de distancia entre plantas

Para el cálculo de las distancias se utilizaron las propias detecciones (salidas de los algoritmos YOLO), de las cuales se conocen las coordenadas de cada cuadro delimitador dentro de la imagen. A través de estas coordenadas se obtuvo el centro del cuadro delimitador (bx , by), que aproximadamente coincide con el centro de una planta de agave completamente visible en la imagen debido a su forma (Figura 10).

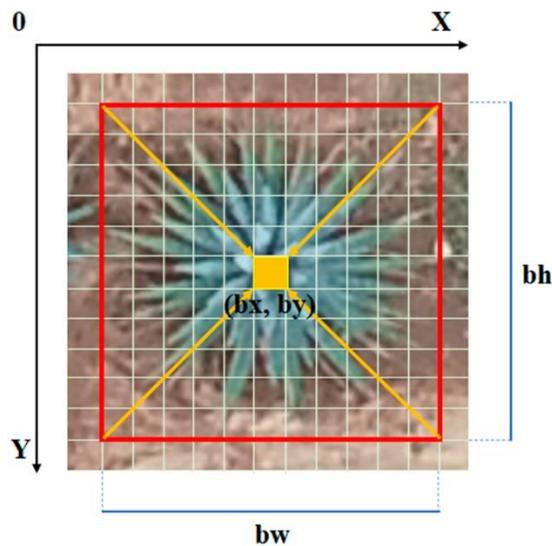


Figura 10. Coordenadas del cuadro delimitador predicho.

A partir de las coordenadas de cada centro de planta se pueden realizar los cálculos de las distancias en metros. Para ello se propuso la siguiente fórmula:

$$d(A, B) = \sqrt{(bx_2 - bx_1)^2 + (by_2 - by_1)^2} \left(\frac{GSD}{100} \right) \quad (4)$$

Donde A y B son los centros de las plantas de agave entre los cuales se calcula la distancia. Los valores bx_1, by_1 son las coordenadas del centro de la planta A, y bx_2, by_2 son las coordenadas del centro de la planta B. El GSD es la distancia de muestreo en tierra dada en cm/píxel, relacionada con la resolución espacial de las imágenes capturadas por el RPAS.

3.6 Resultados

3.6.1 Entrenamiento

En la Figura 11 se observan las curvas de la función de pérdida de los dos algoritmos en cada iteración durante el entrenamiento, mientras que en la Figura 12 se muestra el mAP calculado para el conjunto de prueba durante el entrenamiento.

En ambos algoritmos se observa que la curva de pérdida durante el entrenamiento se vuelve asintota en un rango de 3 a 5 a partir de la iteración número 2500 para el algoritmo YOLOv4. Y en un rango de 0.9 a 1.2 a partir de las 2000 iteraciones para el algoritmo YOLOv4-tiny, aunque alcanza el valor más bajo en el rango de 0.7 a 0.9 a partir de la iteración número 5000.

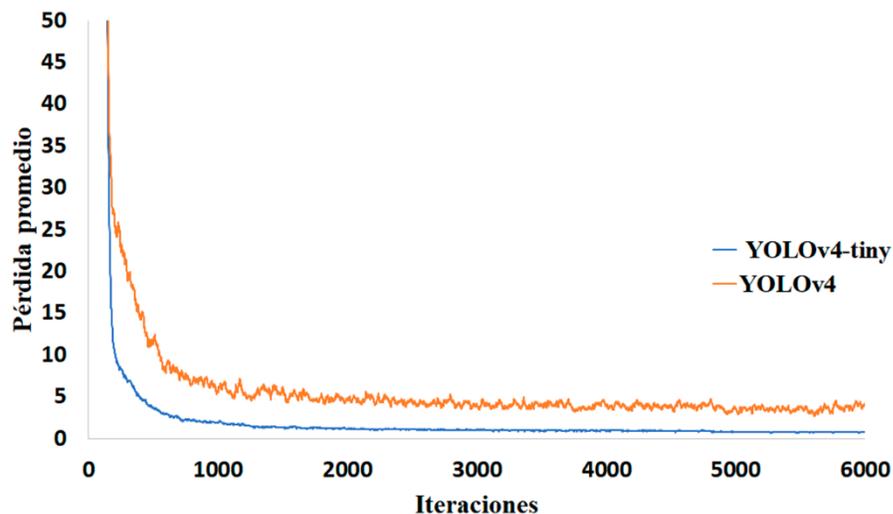


Figura 11. Curvas pérdida-iteraciones de los algoritmos para el entrenamiento.

El valor mAP se mantiene estable por encima del 96%, logrando un valor máximo para el algoritmo YOLOv4 de 99.22% en la iteración 3660, y para YOLOv4-tiny de 99.16% en la iteración 5340.

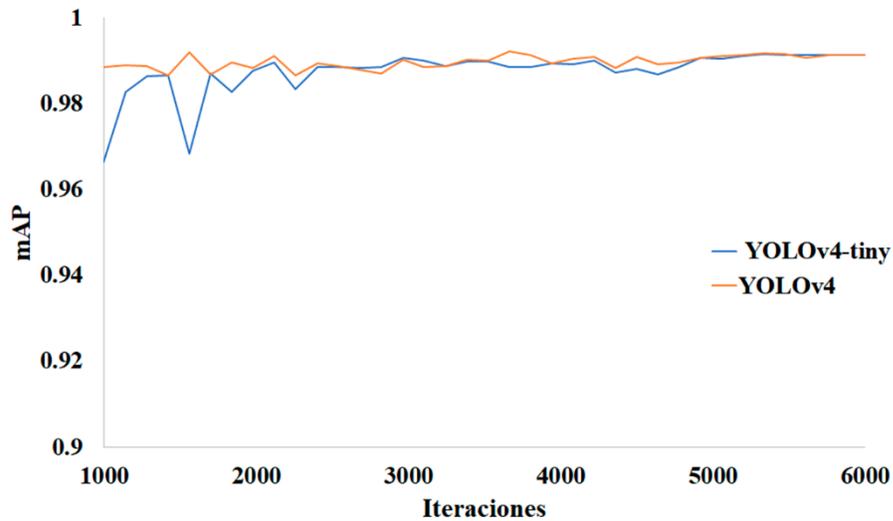


Figura 12. Curvas mAP-iteraciones de los algoritmos durante el entrenamiento. Durante el entrenamiento se calculan las métricas para el conjunto de prueba después de cada época, mostrando cuándo el algoritmo alcanzó el mejor rendimiento. Los resultados mostrados en el Cuadro 4 brindan información sobre el rendimiento de los algoritmos en la detección de plantas de agave para una confianza mayor al 25% y un umbral IoU mayor al 50%. Así también se pueden observar los tiempos que toma cada algoritmo para realizar el entrenamiento, en donde YOLOv4 completo requiere casi 9 veces el tiempo que tarda en entrenar YOLOv4-tiny.

Cuadro 4. Métricas para el conjunto de prueba.

Algoritmo	Tiempo entrenamiento (horas)	Precision	Recall	F1	mAP @0.50
YOLOv4	6.58	0.90	0.99	0.94	99.22
YOLOv4-tiny	0.76	0.92	0.99	0.95	99.16

3.6.2 Evaluación

Posterior al entrenamiento se llevó a cabo la evaluación, utilizando los mejores pesos obtenidos durante el entrenamiento de los algoritmos. Los resultados mostrados en el Cuadro 5 brindan información sobre el rendimiento en el conjunto de evaluación, obteniendo en ambos algoritmos un valor mAP superior al 99%.

Cuadro 5. Evaluación de los algoritmos.

Algoritmo	Precision	Recall	F1	mAP @0.50
YOLOv4	0.90	0.99	0.95	99.40
YOLOv4-tiny	0.92	0.99	0.96	99.28

3.6.3 Detecciones y conteo de plantas

Para evaluar el rendimiento de YOLOv4 y YOLOv4-tiny, se compararon los resultados de las detecciones de plantas de agave en los dos algoritmos, con el conteo manual de las plantas. Del conjunto de datos de evaluación se seleccionaron al azar imágenes, considerando que fueran a diferentes alturas de vuelo, en donde se contaron manualmente las plantas de agave.

La Figura 13 muestra los resultados del conjunto de evaluación para las detecciones. Para las alturas de 40 y 60 m se obtuvo el 100% de plantas detectadas en ambos algoritmos. Sin embargo, a una altura de 80 m para YOLOv4 las detecciones fueron del 96.8% mientras que para YOLOv4-tiny fue de 93.6%.

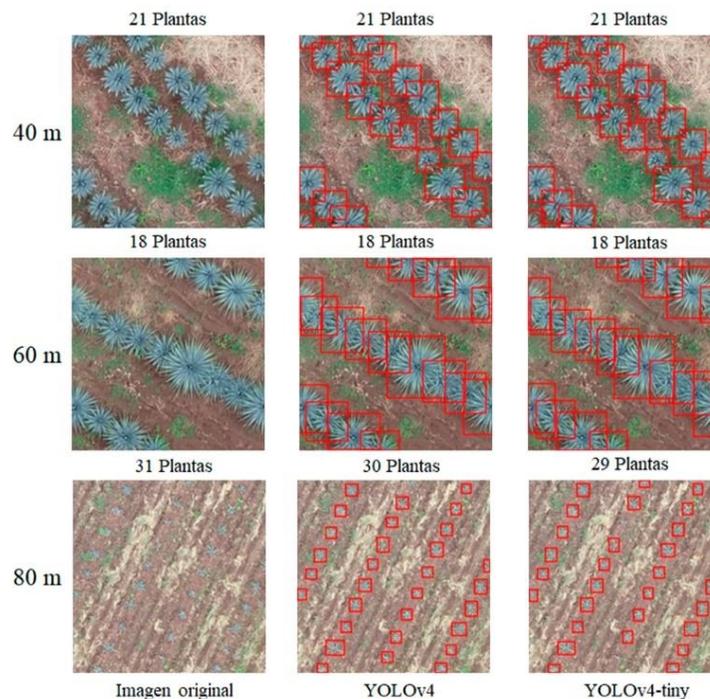


Figura 13. Detecciones y conteo de plantas de agave con los algoritmos YOLOv4 y YOLOv4-tiny.

3.6.4 Distancia entre plantas

Para el cálculo de las distancias entre plantas se obtuvieron los centros de los cuadros delimitadores de las detecciones hechas por YOLO mediante código en Python.

La Figura 14 muestra los centros de las detecciones en una imagen tomada a 40 m de altura. Se asignaron letras en mayúsculas para identificar las plantas en las cuales se realizaron los cálculos.

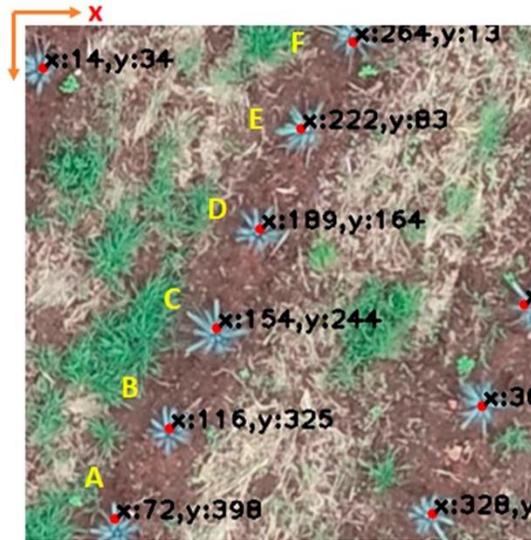


Figura 14. Centros de las detecciones de los cuadros delimitadores.

Los resultados de las distancias entre plantas se muestran en el Cuadro 6. Para el cálculo se usó (4), considerando que el valor del GSD a 40 m fue de 1.10 cm/píxel. Así también se muestran las distancias reales medidas en campo.

Cuadro 6. Distancia entre plantas.

Plantas	Distancia calculada (m)	Distancia real (m)	Error (m)
A – B	0.94	1.14	0.20
B – C	0.98	1.10	0.12
C – D	0.96	1.09	0.13
D – E	0.96	1.13	0.17
E – F	0.90	1.10	0.20

Se observa que la mayor diferencia en las distancias calculadas y reales se da en las plantas A – B y E – F. El error máximo es de 20 cm.

3.7 Discusión

El presente trabajo se llevó a cabo para detectar y contar plantas de agave mediante los algoritmos YOLOv4 en su versión completa y también en su versión reducida, YOLOv4-tiny, en imágenes RGB de alta resolución adquiridas a tres diferentes alturas por un RPAS.

De acuerdo con las Figura 11 y 12, tanto el algoritmo YOLOv4 como el algoritmo YOLOv4-tiny no presentaron problemas de sobre ajuste o ajuste insuficiente durante el entrenamiento.

De los Cuadros 4 y 5 se observa que tanto el valor F1 como el mAP son ligeramente mayores en la evaluación, comparados con la prueba. Esto debido a que los algoritmos YOLOv4 son lo suficientemente robustos para lograr una correcta generalización de las plantas de agave. Uno de los aspectos que puede influir es la regularización DropBlock (Ghiasi et al., 2018), en el entrenamiento del algoritmo YOLOv4, que consiste en ocultar secciones de la imagen al apagar un grupo de neuronas vecinas, obligando a la red a aprender diferentes regiones de cada imagen, haciendo que el algoritmo logre una mejor generalización. Otro aspecto es la división aleatoria de datos, en la cual la proporción de imágenes de “difícil detección” es menor en el subconjunto para evaluación, haciendo que los errores en la detección sean menores incrementando el valor de las métricas. En este contexto se entiende como imágenes de difícil detección a aquellas que cuentan con hijuelos de agave, capturadas a una altura de 80 m.

En la evaluación, Cuadro 5, el valor F1 es bueno para ambos algoritmos, ya que tanto la precisión y recall son altos, cercanos a 1. En Flores et al., (2021) informan que para la detección de plantas de agave obtuvieron un valor F1 de 0.96, mientras que en Calvario et al., (2020), para la misma tarea, obtuvieron un valor F1 que oscilaba en el rango de 0.88 a 0.98. En contraste, Ammar et al., (2021), para la detección de palmeras obtuvieron un F1 de 0.9. Todos estos valores similares al obtenido por los dos algoritmos YOLO en la detección de agave.

Una diferencia con otros trabajos relacionados con el conteo de agaves son las etapas de procesamiento, costo computacional y equipos utilizados. En Flores

et al., (2021) & Calvario et al., (2020) estas etapas fueron superiores comparadas con el presente trabajo, además de que volaron a una sola altura de vuelo, mientras que la base de datos que se presenta fue creada considerando tres diferentes alturas.

Referente a detecciones y conteo de plantas, los dos algoritmos lograron identificar agaves en etapas de desarrollo avanzadas en condiciones reales del cultivo. Es decir, cuando se encuentra maleza como pasto e incluso algunas hierbas más grandes dentro de la propia planta. Así mismo también fueron capaces de detectar hijuelos de agave, lo que comprueba la robustez en los algoritmos entrenados. Se puede notar en la Figura 13 que hubo muy pocas plantas de agave que no fueron detectadas. Esto se dio principalmente al evaluar en imágenes a la mayor altura de vuelo, que fue a 80 m, en donde debido a efectos de la altura la planta se puede mezclar con el entorno. Cabe resaltar que esta situación se da mayormente cuando se trata de detectar hijuelos de agave. Pero si el agave ya está en una etapa de desarrollo avanzada los algoritmos son capaces de detectarlas. Por lo cual, es recomendable volar a 60 m cuando se tienen campos con variedad de tamaños de agaves y se requiere detectar también hijuelos.

En Flores et al., (2021) se reporta el 98% de agaves detectados a una altura de 50 m. Por consiguiente, en base a lo anterior, se puede inferir que los algoritmos YOLOv4 y YOLOv4-tiny pueden detectar en un 100% los agaves a una altura de vuelo similar.

Una de las diferencias en los algoritmos utilizados en el trabajo es que el algoritmo YOLOv4, al ser más grande y contar con más capas convoluciones y tres salidas, toma más tiempo al momento del entrenamiento. Mientras que el algoritmo YOLOv4-tiny al tener menos capas convolucionales y dos salidas, es más rápido, lo que lo hace adecuado inclusive para implementaciones en tiempo real.

Finalmente, es posible aprovechar la salida de los algoritmos YOLO debido a la forma de las detecciones de las plantas de agave que se parecen a

paralelogramos cuyos centros se aproximan a los centroides de las plantas, para realizar un cálculo estimado de la distancia entre estas. Dado que los cuadros delimitadores sólo toman en cuenta la parte visible de la planta, hay un desfase entre el centroide real y el predicho para las plantas localizadas en los bordes de la imagen. Por lo tanto, los cálculos de distancia sólo se pueden considerar precisos en las predicciones de las plantas completamente visibles. Aun así, un error de ± 20 cm en el cálculo de la distancia entre plantas no es significativo. Para que sea significativo la diferencia tendría que ser de más de 0.90 m para considerar que falta una planta, esto debido a la distribución habitual de los agaves en las plantaciones. Por ejemplo, en Calvario et al., (2020) se reporta que la distancia entre plantas de agave es de 1.0 a 1.20 m, mientras que en Gerritsen et al., (2011) se reporta una distancia entre plantas de 1.0 m. Por lo cual la distancia calculada mediante las salidas de los algoritmos YOLO puede ser utilizada como referencia.

3.8 Conclusiones

Se creó una base de datos de plantas de agave que incluyó 2403 imágenes capturadas a diferentes alturas de vuelo: 40, 60 y 80 m. Las imágenes contaban con agaves en diferentes etapas fenológicas dispuestos en una o doble hilera, además de la presencia de maleza y suelo desnudo.

Empleando los algoritmos YOLOv4 y YOLOv4-tiny, se logró la detección y conteo de plantas de agave en etapas de desarrollo avanzadas, considerando las alturas de vuelo mencionadas. Sin embargo, si se requiere detectar hijuelos de agave, la altura de vuelo no debe ser mayor a 60 m.

Se utilizaron las coordenadas de los cuadros delimitadores predichos por la salida de YOLO para realizar el cálculo de la distancia entre plantas. Conocer este dato, sumado al número de plantas, es benéfico en la elaboración de inventarios para determinar en qué partes de las plantaciones faltan agaves.

3.9 Referencias

- Ameijeiras-Sánchez, D., González-Diez, H. R., & Hernández-Heredia, Y. (2020). Revisión de algoritmos de detección y seguimiento de objetos con redes profundas para videovigilancia inteligente. *Revista Cubana de Ciencias Informáticas*, 14(3), 165-195.
- 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. <https://doi.org/10.3390/agronomy11081458>
- Ampatzidis, Y., & Partel, V. (2019). UAV-Based High Throughput Phenotyping in Citrus Utilizing Multispectral Imaging and Artificial Intelligence. *Remote Sensing*, 11(4), 410. <https://doi.org/10.3390/rs11040410>
- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv:2004.10934 [cs, eess]*. <http://arxiv.org/abs/2004.10934>
- Calvario, G., Alarcón, T. E., Dalmau, O., Sierra, B., & Hernández, C. (2020). An Agave Counting Methodology Based on Mathematical Morphology and Images Acquired through Unmanned Aerial Vehicles. *Sensors*, 20(21), 6247. <https://doi.org/10.3390/s20216247>
- Calvario, G., Sierra, B., Alarcón, T. E., Hernández, C., & Dalmau, O. (2017). A Multi-Disciplinary Approach to Remote Sensing through Low-Cost UAVs. *Sensors (Basel, Switzerland)*, 17(6), E1411. <https://doi.org/10.3390/s17061411>
- Ceja Ramírez, R., González Eguiarte, D. R., Ruiz Corral, J. A., Rendón Salcido, L. A., & Flores Garnica, J. G. (2017). Detección de restricciones en la producción de agave azul (Agave tequilana Weber var. Azul) mediante percepción remota. *Terra Latinoamericana*, 35(3), 259-268. <https://doi.org/10.28940/terra.v35i3.252>
- Chen, S. W., Shivakumar, S. S., Dcunha, S., Das, J., Okon, E., Qu, C., Taylor, C. J., & Kumar, V. (2017). Counting Apples and Oranges With Deep Learning: A Data-Driven Approach. *IEEE Robotics and Automation Letters*, 2(2), 781-788. <https://doi.org/10.1109/LRA.2017.2651944>
- Chowdhury, P. N., Shivakumara, P., Nandanwar, L., Samiron, F., Pal, U., & Lu, T. (2022). Oil palm tree counting in drone images. *Pattern Recognition Letters*, 153, 1-9. <https://doi.org/10.1016/j.patrec.2021.11.016>
- CRT. (2016). Consejo Regulador del Tequila. <https://www.crt.org.mx/index.php/es/>
- DOF. (2012). Diario Oficial de la Federación. http://www.dof.gob.mx/nota_detalle.php?codigo=5282165&fecha=13/12/2012
- Flores, D., González-Hernández, I., Lozano, R., Vázquez-Nicolas, J. M., & Hernández Toral, J. L. (2021). Automated Agave Detection and Counting Using a Convolutional Neural Network and Unmanned Aerial Systems. *Drones*, 5(1), 4. <https://doi.org/10.3390/drones5010004>
- Garnica, J. G. F., Reich, R., Zúñiga, E. T., & Bravo, C. A. (2008). Using remote sensing to support different approaches to identify agave (Agave tequilana

- Weber) crops. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37, 4.
- Gerritsen, P. W., Estrella-Chulim, N., Moreno-Hernández, A., Escobedo-Garrido, S., & Bustamante-González, Á. (2011). Prácticas de manejo agronómico para la sustentabilidad: Características y medición en Agave tequilana Weber en la región Sierra Amula, Jalisco. *Tropical and Subtropical Agroecosystems*, 14(1), 159-169.
- Ghiasi, G., Lin, T.-Y., & Le, Q. V. (2018). DropBlock: A regularization method for convolutional networks. *arXiv:1810.12890 [cs]*. <http://arxiv.org/abs/1810.12890>
- Huang, Z., Pan, Z., & Lei, B. (2017). Transfer Learning with Deep Convolutional Neural Network for SAR Target Classification with Limited Labeled Data. *Remote Sensing*, 9(9), 907. <https://doi.org/10.3390/rs9090907>
- Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167 [cs]*. <http://arxiv.org/abs/1502.03167>
- Jintasuttisak, T., Edirisinghe, E., & Elbattay, A. (2022). Deep neural network based date palm tree detection in drone imagery. *Computers and Electronics in Agriculture*, 192, 106560. <https://doi.org/10.1016/j.compag.2021.106560>
- Labellmg. (s. f.). Recuperado 22 de julio de 2021, de <https://github.com/tzutalin/labellmg>
- Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote Sensing and Image Interpretation* (7.^a ed.). John Wiley & Sons.
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2015). Microsoft COCO: Common Objects in Context. *arXiv:1405.0312 [cs]*. <http://arxiv.org/abs/1405.0312>
- Lin, Y., Chen, T., Liu, S., Cai, Y., Shi, H., Zheng, D., Lan, Y., Yue, X., & Zhang, L. (2022). Quick and accurate monitoring peanut seedlings emergence rate through UAV video and deep learning. *Computers and Electronics in Agriculture*, 197, 106938. <https://doi.org/10.1016/j.compag.2022.106938>
- Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: A review. *Plant Methods*, 17(1), 22. <https://doi.org/10.1186/s13007-021-00722-9>
- Misra, D. (2020). Mish: A Self Regularized Non-Monotonic Activation Function. *arXiv:1908.08681 [cs, stat]*. <http://arxiv.org/abs/1908.08681>
- Morales-Avendaño, J. R., Hernández-Cruz, J., López-Cid, I., & Röss, M. (2018). Propuesta Metodológica para Monitoreo de Cultivos de Agave con Vehículos Aéreos no Tripulados (VANT) en Oaxaca, México. *Revista de Ingeniería e Investigación Aplicada UPB/UPTap*, 3, 8.
- 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. <https://doi.org/10.3390/electronics10030279>
- Prasad, M. V. D., Lakshamma, B. J., Chandana, A. H., Komali, K., Manoj, M., Rajesh Kumar, P., Raghava Prasad, C., Inthiyaz, S., & Sasi Kiran, P. (2017). An efficient classification of flower images with convolutional neural

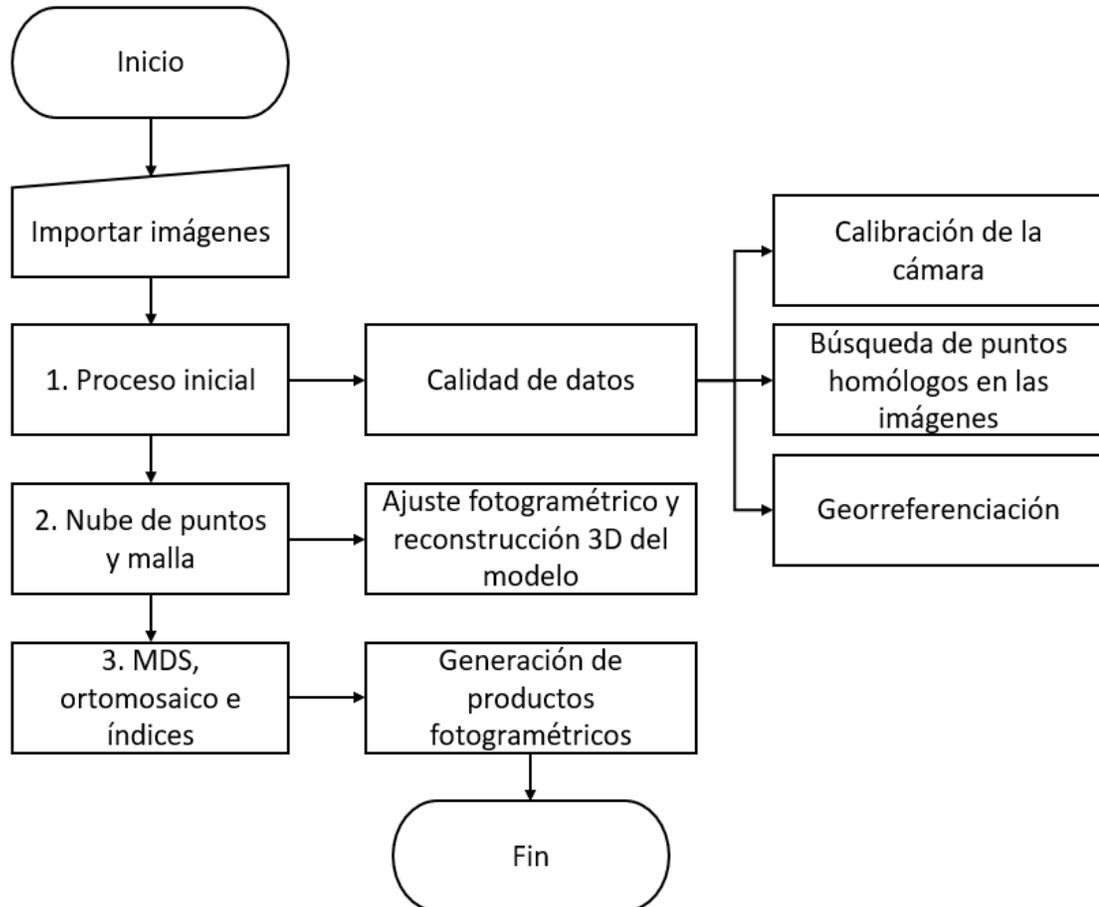
- networks. *International Journal of Engineering & Technology*, 7(1.1), 384. <https://doi.org/10.14419/ijet.v7i1.1.9857>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *arXiv:1506.02640 [cs]*. <http://arxiv.org/abs/1506.02640>
- Redmon, J., & Farhadi, A. (2016). YOLO9000: Better, Faster, Stronger. *arXiv:1612.08242 [cs]*. <http://arxiv.org/abs/1612.08242>
- Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *arXiv:1804.02767 [cs]*. <http://arxiv.org/abs/1804.02767>
- Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv:1506.01497 [cs]*. <http://arxiv.org/abs/1506.01497>
- Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: A survey. *Evolutionary Intelligence*, 15(1), 1-22. <https://doi.org/10.1007/s12065-020-00540-3>
- Sultana, F., Sufian, A., & Dutta, P. (2018). Advancements in Image Classification using Convolutional Neural Network. *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 122-129. <https://doi.org/10.1109/ICRCICN.2018.8718718>
- Ubbens, J., Cieslak, M., Prusinkiewicz, P., & Stavness, I. (2018). The use of plant models in deep learning: An application to leaf counting in rosette plants. *Plant Methods*, 14(1), 6. <https://doi.org/10.1186/s13007-018-0273-z>
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 2018, e7068349. <https://doi.org/10.1155/2018/7068349>
- 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. <https://doi.org/10.3390/s21020507>
- Wang, Z., Jin, L., Wang, S., & Xu, H. (2022). Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system. *Postharvest Biology and Technology*, 185, 111808. <https://doi.org/10.1016/j.postharvbio.2021.111808>
- Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging*, 9(4), 611-629. <https://doi.org/10.1007/s13244-018-0639-9>
- Zhang, Q., Liu, Y., Gong, C., Chen, Y., & Yu, H. (2020). Applications of Deep Learning for Dense Scenes Analysis in Agriculture: A Review. *Sensors*, 20(5), 1520. <https://doi.org/10.3390/s20051520>
- Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07), 12993-13000. <https://doi.org/10.1609/aaai.v34i07.6999>

APÉNDICE A. Planificación del vuelo fotogramétrico y captura de las imágenes del cultivo de agave por medio de RPAS.

La planificación del vuelo es el paso principal para garantizar buenos resultados en los productos fotogramétricos. Para ello se consideraron los siguientes puntos:

- **Terreno, clima y plataforma RPAS.** Se consideró la zona de estudio para contemplar la topografía del terreno, evitando posibles obstáculos, establecer el punto de partida, considerar los tiempos de vuelo y contar con los mapas correspondientes a la zona de vuelo. Se verificaron las condiciones climáticas, probabilidad de lluvia, nubosidad y velocidad del viento altas. En cuanto a la plataforma RPAS, se utilizó un multirrotores DJI Phantom 4 Pro V2.0.
- **App para la planificación.** Se utilizó la aplicación móvil DJI Pilot.
- **Configuración de la cámara.** Los parámetros que se configuraron fueron el valor ISO a 100, el valor de exposición EV a 0, la obturación a 1/400.
- **Altura de vuelo.** Se realizaron vuelos a tres diferentes alturas de vuelo, a 40, 60 y 80 m. Estas alturas se ajustaron en la aplicación para la planificación de cada vuelo fotogramétrico.
- **Distancia de muestreo en tierra (GSD).** Para una altura de vuelo de 40 m el GSD fue de 1.10 cm/píxel, para una altura de vuelo de 60 m fue de 1.64 cm/píxel y para una altura de vuelo de 80 m fue de 2.19 cm/píxel.
- **Porcentaje de recubrimiento estereoscópico.** El traslape configurado en la aplicación móvil fue de 80% tanto lateral como frontal.
- **Velocidad de vuelo del RPAS.** Las imágenes se capturaron a una velocidad de 5 m/s.
- **Inclinación de la cámara.** Los vuelos se planificaron con una inclinación de la cámara del 90%.
- **Calibración de sensores del RPAS.** Se realizó la calibración de IMU y brújula.

APÉNDICE B. Diagrama de flujo en el programa de cómputo Pix4Dmapper para la obtención de los ortomosaicos.



Los ortomosaicos se obtuvieron utilizando el programa de cómputo Pix4Dmapper (Pix4D S. A., Lausanne, Switzerland) siguiendo el siguiente procedimiento:

1. Importación de las imágenes.
 - a. Se creó un nuevo proyecto desde la ventana principal de Pix4Dmapper, asignando un nombre y la ruta de ubicación para guardarlo.
 - b. Posteriormente se seleccionó la carpeta que contiene las imágenes adquiridas mediante el RPAS.
 - c. Añadidas las imágenes, Pix4Dmapper presenta el sistema de coordenadas en el que está ubicado el proyecto. Así mismo, el

programa de cómputo reconoce el modelo de la cámara y parámetros de configuración.

d. La siguiente configuración es referente a la salida deseada, se selecciona una plantilla basada en los datos disponibles y la información requerida. Para este caso como se deseó obtener un ortomosaico, se seleccionó la plantilla 3D Maps.

e. Pix4Dmapper tiene un mapa base en el cual se muestra la geolocalización del proyecto, corroborando que realmente corresponde a la zona volada.

2. Proceso inicial.

a. Consiste en la orientación de las imágenes y una nube de puntos dispersa. Se configura la calidad del producto, forma en la que se adquirieron las imágenes y el algoritmo de calibración.

3. Nube de puntos y malla.

a. Se puede seleccionar la densidad de la nube de puntos y la calidad de malla deseada.

4. En DSM, ortomosaico e índice.

a. Se seleccionan los formatos y calidad de salida de los productos mencionados.

5. En recursos y notificaciones se configura la cantidad de memoria RAM a disponer y también si se cuenta con una tarjeta de video dedicada para acelerar el procesamiento de las imágenes.

6. En este caso no se contó con puntos de apoyo en tierra.

7. Con la configuración de calidad del proyecto, el último paso es obtener y guardar el ortomosaico en formato .tiff

APÉNDICE C. Procedimiento para el etiquetado de la base de datos.

- Los ortomosaicos generados (a 40, 60 y 80 m) en Pix4Dmapper se abrieron en QGIS. En este programa de cómputo se crearon archivos con extensión .shp para recortar las ortofotos en las áreas de interés.
- Mediante código en Python los recortes generados de los ortomosaicos se partieron en imágenes de tamaño 416x416 pixeles.
- Todas las imágenes de tamaño 416x416 pixeles de las tres diferentes alturas de vuelo se guardaron en una misma carpeta
- Se utilizó la herramienta Labellmg para etiquetar las imágenes de agave, dibujando un cuadro delimitador en cada una de las plantas presentes en cada imagen, asignando una única clase llamada “Agave”. Se consideraron tanto las plantas visibles completamente, así como las de los bordes de las imágenes.
- De cada imagen etiquetada se creó un archivo con extensión .txt, que es el formato de YOLO que incluye la clase y las coordenadas normalizadas correspondientes a cada cuadro delimitador de todas las plantas presentes en la imagen, en donde cada renglón es una planta etiquetada.
- Las imágenes etiquetadas junto con sus correspondientes archivos .txt se dividieron aleatoriamente en carpetas separadas. La base de datos quedó dividida en 70% para entrenamiento, 15% para prueba y 15% para evaluación, respectivamente.