



**CENTRO DE INVESTIGACIÓN EN
CIENCIAS DE INFORMACIÓN GEOESPACIAL, A.C,
CentroGeo**

Centro Público de Investigación SECIHTI

Desarrollo de un sistema automatizado de visión computacional
con VANT para la detección y segmentación de afectaciones por
Spodoptera frugiperda en cultivos de maíz en Yucatán

TESIS

Que para obtener el grado de
Maestro en Ciencias de Información Geoespacial

Presenta

Alejandro Carrillo Gómez

Directora de Tesis

Dra. Daniela Alejandra Moctezuma Ochoa

Ciudad de México
2025

CENTRO DE INVESTIGACIÓN EN
CIENCIAS DE INFORMACIÓN GEOESPACIAL, A.C,
CentroGeo

Centro Público de Investigación SECIHTI

Desarrollo de un sistema automatizado de visión computacional
con VANT para la detección y segmentación de afectaciones por
Spodoptera frugiperda en cultivos de maíz en Yucatán

TESIS

Que para obtener el grado de

Maestro en Ciencias de Información Geoespacial

Presenta

Alejandro Carrillo Gómez

Directora de Tesis

**Dra. Daniela Alejandra Moctezuma
Ochoa**

Sinodales

**Dra. Daniela Juanita López Araujo
Dr. Néstor Miguel Cid García
Dr. León Felipe Dozal García**

Ciudad de México, noviembre, 2025

Resumen

Esta tesis aborda el desafío fitosanitario que representa la plaga *Spodoptera frugiperda* (gusano cogollero) en los cultivos de maíz de Yucatán, México. El objetivo principal fue desarrollar y validar un sistema automatizado de monitoreo, basado en visión computacional y aprendizaje profundo, para la detección y segmentación de plantas de maíz y de las afectaciones foliares causadas por esta plaga a partir de imágenes aéreas.

La investigación se fundamenta en la problemática agrícola regional, donde los métodos tradicionales de control de plagas enfrentan limitaciones de escalabilidad y precisión. Para ello, se implementó una metodología que incluyó la adquisición sistemática de imágenes multiescala con vehículos aéreos no tripulados (VANTs) en parcelas de los municipios de Muna y Kantunil. Se generaron y curaron tres conjuntos de datos especializados: parcelas representativas, plantas individuales y plantas afectadas.

El núcleo metodológico consistió en el entrenamiento y evaluación de modelos de segmentación de instancias de la familia YOLO, específicamente YOLOv11. El modelo definitivo (A1), optimizado para identificar la clase "Maíz" y la clase "Afectadas", demostró un alto rendimiento en la detección de cajas delimitadoras, con una precisión del 92.9% y un mAP50 del 94.2% para plantas de maíz. Para las afectaciones por *S. frugiperda*, el sistema alcanzó una precisión del 79.2% y un mAP50 del 71.7%. Para el procesamiento de imágenes de alta resolución se integró el método Slicing Aided Hyper Inference (SAHI), logrando un mapeo espacialmente preciso de la distribución de la plaga en ortomosaicos de parcelas completas.

Se concluye que el sistema propuesto constituye una herramienta eficaz y escalable para el monitoreo objetivo de *S. frugiperda*, con potencial inmediato para apoyar la evaluación estadística de técnicas de control en el marco de la transición agroecológica en Yucatán. El trabajo sienta las bases para un protocolo de monitoreo periódico que puede transformar datos en decisiones agrícolas ágiles y fundamentadas, contribuyendo a la sostenibilidad del cultivo de maíz en la región.

Agradecimientos

Al Posgrado del **Centro de Investigación en Ciencias de información Gepespacial**, A.C, por la oportunidad de la formación académica y profesional en un agradable ambiente de trabajo.

A la **Secretaría de Ciencia, Humanidades, Tecnología e Innovación** por el apoyo y el respaldo académico.

Un especial agradecimiento a mi asesora la Dra. Daniela Alejandra Moctezuma Ochoa por su disciplina, atención y sus comentarios que fueron siempre una excelente guía durante todo el proceso de elaboración del trabajo.

A mis sinodales, por su atención cuidadosa y sus comentarios puntuales, que me guiaron y me permitieron presentar adecuadamente la investigación.

Este trabajo no habría sido posible sin la colaboración y participación de los productores milperos de Muna y Kantunil que amablemente nos permitieron visitar sus parcelas, realizar los muestreos y que nos compartieron sus experiencias como agricultores, lo que les agradecemos profundamente.

Agradezco también a mis compañeros de la maestría por su amistad y apoyo durante estos años y a los amigos que me acompañaron en el camino para llegar aquí.

A mi familia, por su apoyo, comprensión y afecto.

A la Mtra. Sabina E. Forti por apoyarme con consejos prácticos y académicos durante el trabajo y por ser un ejemplo e inspiración.

Índice General

Capítulo 1. Introducción.....	1
1.1. Problemática.....	2
1.2. Hipótesis.....	3
1.3. Objetivo general.....	3
1.3.1 Objetivos específicos.....	3
1.4 Antecedentes y trabajos relacionados.....	3
Capítulo 2. Marco Teórico.....	6
2.1 Situación del cultivo de maíz en Yucatán.....	6
2.2. Características de la plaga de gusano.....	6
2.3. Características de la planta de maíz en etapas sensibles.....	8
2.3.1. Técnicas plaguicidas aplicadas en la zona de estudio.....	8
2.4. Elección de Modelo de aprendizaje profundo.....	9
2.5. Estructura de procesamiento del modelo YOLOv11.....	10
2.5.1. Columna vertebral (Backbone).....	11
2.5.2. Cuello (Neck).....	14
2.5.3. Cabeza (Head).....	16
2.5.4. Hiperparámetros y Optimización.....	16
2.6. Método SAHI.....	16
2.7 Métrica F1.....	17
2.8 Métrica mAP50.....	17
2.8.1. Fundamentos Conceptuales.....	18
2.9 Métrica mAp50-95.....	18
Capítulo 3. Documentación de los sitios de estudio.....	19
3.1. Muna.....	19
3.2. Kantunil.....	23
Capítulo 4. Metodología.....	24
4.1. Levantamientos de 2023 en Muna.....	24
4.2. Levantamiento de imágenes 2024 en Kantunil.....	25
4.2.1 Marcas de referencias en campo.....	25
4.2.2. Variabilidad controlada de altura.....	25
4.2.3. Identificación y Registro manual de Plantas Afectadas.....	26
4.3. Generación de bases de datos.....	28
4.3.1. Conjunto de imágenes 1. Parcelas representativas.....	29
4.3.2. Conjunto de imágenes 2. Plantas de maíz individuales.....	29
4.3.3. Conjunto de imágenes 3. Plantas afectadas por plaga.....	30
4.4. Preparación de los datos.....	31

4.4.1. Preprocesamiento con la plataforma Roboflow.....	31
4.4.2 Sustracción de fondo por espacio de color.....	32
4.4.3. Aumento de datos etiquetados.....	33
4.4.4. Inclusión de imágenes a nivel de tierra.....	33
4.5 Generación de modelos de segmentación.....	34
4.5.1. Modelo P1 (Maíz y Plantas secundarias).....	34
4.5.2. Modelo A2 (Afectaciones de 2 clases).....	35
4.5.3 Modelo A1 (Afectaciones de 1 clase).....	36
4.6 Implementación en Dispositivo Autónomo.....	37
Capítulo 5. Resultados y discusión.....	38
5.1. Descripción General de los Resultados.....	38
5.2 Resultados Modelo P1 (Maíz y Plantas secundarias).....	38
5.3. Resultados el Modelo A2.....	40
5.4. Resultados del Modelo A1.....	45
5.4.1. Evaluación en conjunto de Test.....	48
5.5 Aplicación de método SAHI.....	54
5.6 Ortomosaicos de la parcela.....	55
5.7. Consideraciones para adquisición de imágenes aéreas.....	58
5.8 Resultados con la computadora Raspberry Pi.....	61
Capítulo 6. Conclusiones y trabajo futuro.....	63
Referencias.....	64

Lista de figuras

Figura 2.1. Ciclo biológico de <i>S. frugiperda</i> . Tomado de Hardke et al., (2015).....	7
Figura 2.2. Lesiones provocadas por <i>S. frugiperda</i> a una planta de maíz. Hojas y centro afectados.	8
Figura 2.3. Arquitectura general del modelo YOLOv11, mostrando las 3 etapas principales y los módulos que trabajan con los mapas de características a diferentes escalas. Tomado de Jocher & Qiu, (2024), Ultralytics.....	10
Figura 2.4. Comparativo de velocidad y precisión de YOLOv11 con YOLOv5 y EfficientDet	11
Figura 2.5. Módulo Conv y los métodos que contiene.....	12
Figura 2.6. Comparación de función ReLu con Silu (Izquierda) y comparación de la derivada de SiLu con la Sigmoide (Derecha).....	13
Figura 2.7. Módulo C3K (Derecha). C3K es incorporado n veces dentro del módulo C3K2 (Izquierda)..	13
Figura 2.8. Capa de cuello de botella conformada por módulos Conv con opción de entrada de la parte residual.....	14
Figura 2.9. Módulo de agrupación piramidal SPFF que utiliza 3 capas de Max Pooling secuencial para realizar los mapeos de características a diferentes escalas (derecha). Ejemplo de operación Max Pulling (izquierda)	15
Figura 2.10. Módulo PSA (izquierda) que está contenido dentro del módulo C2PSA (derecha).....	15
Figura 3.1. Ubicación del Municipio de Muna, Yucatán, dentro de la península a la izquierda, en el recuadro rojo se encuentra el municipio y en gris en la parte superior, la ciudad de Mérida. Se muestra un acercamiento satelital al municipio a la derecha.....	20
Figura 3.2 Imagen de Google Maps con los pines rojos que marcan la ubicación de las parcelas 1, 2 y 3 en las que se realizaron los mapeos de Muna.....	21
Figura 3.3 Sesión de Escuela de Campo en la parcela 1 del municipio de Muna.....	22
Figura 3.4 Parcela con monocultivo de maíz en Kantunil.....	23
Figura 4.1. Ejemplo de imagen de 5472 x 3078 px, capturada por medio del VANT a 6 m de altura.....	24
Figura 4.2. Perspectivas desde diferentes alturas (1, 3, 6 y 10 m.), para habilitar seguimiento de los detalles pequeños de cada planta.....	26
Figura 4.3. Registro manual longitudinal al nivel de piso de las afectaciones por <i>S. frugiperda</i>	27
Figura 4.4 Resultado de la distribución de plantas afectadas, evaluadas de forma manual en campo, como registro de referencia. Los colores representan el grado de afectación en las plantas de cada fila de la parcela.....	27
Figura 4.5. Diagrama de flujo de los datos, se ilustran los 3 principales conjuntos de imágenes arriba y los modelos que se generaron a partir de ellos abajo. El conjunto 3 se utilizó para el modelo A1 y A2.....	28

Figura 4.6. Imágenes de 684 x 615 px., representativas de la base de datos de imágenes de cultivos de maíz en Yucatán, que contuvieran al menos una planta de maíz completa.....	30
Imagen 4.7. Imágenes ilustrativas de plantas en la base de datos de afectadas por plaga.....	31
Figura 4.8 Muestra de imágenes a las que se les ha generado sus máscaras de segmentación de clase, utilizando Roboflow. La clase Maíz en amarillo, la clase no maíz en morado, lo no seleccionado oscurecido.....	32
Figura 4.9. Espacios de color RGB y HSV arriba, y plantas de maíz segmentadas mediante técnica de corte de color en espacio HSV, abajo.....	33
Figura 4.10 Segmentación de plantas y de hojas afectadas, en rojo las hojas con huecos y en morado las hojas con restos.....	36
Figura 5.1. Ejemplo del resultado de la segmentación realizada por el Modelo P1 de segmentación de plantas, en imagen de 640 x 640	39
Figura 5.2. Imagen de alta resolución a la que se implementó el modelo P1 de detección de maíz y hierbas secundarias.	40
Figura 5.3. Comparativo de seis imágenes de un batch del etiquetado manual de validación (izquierda) contra las mismas imágenes, pero predichas por el modelo A2 (derecha), que indica el nivel de confianza de cada instancia.	42
Figura 5.4. Comparativo entre 2 imágenes del batch de evaluación arriba y las mismas 2 imágenes, con la segmentación automática realizada por el modelo A2 abajo.	43
Figura 5.5. Matriz de confusión absoluta (arriba) y normalizada (abajo) del modelo A2, con el etiquetado manual en el eje horizontal y en el vertical los generados por el modelo A2.	44
Figura 5.6 Gráfico de puntuación F1 del modelo A2 para cada nivel de corte de confianza por cada clase. En azul grueso se resalta el resultado promedio del modelo.	45
Figura 5.7. Se presentan los resultados de las mejoras en cada una de las 250 épocas del entrenamiento del modelo A1, en las métricas de precisión, exhaustividad y de mAP50 referente a las cajas delimitadoras de detección (Bouding box) del modelo A1.	46
Figura 5.8. Presenta los resultados de las mejoras en cada una de las 250 épocas del entrenamiento del modelo A1, en las métricas de precisión, exhaustividad y de mAP50 referente a las máscaras de segmentación del modelo A1.	47
Figura 5.9. Se presentan 6 de 16 imágenes de un batch generado durante el entrenamiento, a la izquierda el batch con los etiquetados manuales y a la derecha los etiquetados generados automáticamente por el modelo A1.	49
Figura 5.10. Comparativo de 2 imágenes provenientes del etiquetado manual (izquierda), con 2 imágenes provenientes del batch, realizada por el modelo A1 (derecha). Se aprecia un caso 1 (arriba) en el que las secciones etiquetadas coinciden perfectamente con la segmentación del modelo y otro caso 2 en el que difiere un poco (abajo).	50
Figura 5.11. Imagen que ilustra algunos de los motivos principales que explicarían discrepancias entre el etiquetado y la predicción, como la segmentación de la hoja correcta, pero no en su totalidad, o que el modelo lo hace mejor que el manual.	51

Figura 5.12. Matriz de confusión normalizada del modelo A1, con la clase Maíz, la clase Hojas Afectadas, y los pixeles de fondo no clasificados como "Sin_Clase".	52
Figura 5.13. Diagramas de puntuación F1 contra nivel de confianza para las clases Maíz ("Ok" en las imágenes) y la clase de hoja afectada ("X" en las imágenes). Arriba el diagrama de cajas de detección y abajo máscaras de segmentación. En azul más grueso se representa el promedio de ambas clases, que sería el resultado general del modelo A1.	53
Figura 5.14 Arriba: Comparativo de 2 imágenes tomadas a 10 m de altura con 40% y 50% de corte de confianza. Abajo: Capturadas a 6 m de altura, con corte de confianza de 10% y 50%. Procesadas por el modelo A1 y con la aplicación del método SAHI de inferencia por segmentos. Se probaron diferentes cortes de confianza, se ilustran algunos.	55
Figura 5.15. Ortomosaico con 4 imágenes obtenidas a cada 5 m longitudinales sobre la parcela a 6 m de altura, con 60% de corte de confianza.	56
Figura 5.16. Ortomosaico con 4 imágenes obtenidas a cada 5 m longitudinales sobre la parcela a 6 m de altura, con 50% de corte de confianza.	56
Figura 5.17. Segmento efectivo por recorrido de vuelo a 6 m de altura, con 50% de corte de confianza.	57
Figura 5.18. Segmento de ortomosaico de la parcela analizada, resaltando en rojo los casos de hojas afectadas y en morado las plantas de maíz.	57
Figura 5.19. Kit de hardware libre con costo menor que puede ser conseguido fácilmente por internet.	59
Figura 5.20. VANT marca 3DR modelo SOLO con el que se realizaron experimentos. Imagen propia.	60
Figura 5.21. Ejemplo de un plan de vuelo para el VANT SOLO generado con el software libre ArduPilot. Imagen propia.	61
Figura 5.22. Imagen obtenida y procesada en hardware Raspberry pi, en una escena nocturna de un maíz.	62

Lista de tablas

Tabla 3.1. Fechas de recolección de imágenes y características de las parcelas vistas en Muna 2023.

Tabla 5.1 Resultados de la métrica mAP50 por clase de modelo P1, en gris los resultados por clase sobre el set de Prueba.

Tabla 5.2. Resultados del análisis de las imágenes provenientes del conjunto de Test en cuanto a máscaras de segmentación, obtenidos por el modelo A2. En gris los resultados de la métrica mAP50 de máscaras de cada clase.

Tabla 5.3. Presenta los resultados del análisis de las imágenes provenientes del conjunto de Test en cuanto a cajas de detección (Bounding box). En gris los resultados Del modelo A1 de la métrica mAP50 por cada clase.

Tabla 5.4 Presenta los resultados del análisis de las imágenes provenientes del conjunto de Test en cuanto a máscaras de segmentación, por el modelo A1. En gris los resultados de la métrica mAP50 por cada clase.

Acrónimos y abreviaturas

AP – Average Precision

CNN – Convolutional Neural Network

Faster R-CNN – Faster Region-based Convolutional Neural Networks

FN – Falso negativo

FP – Falsos positivos

FPS – Frames Per Second

IoU – Intersection over Union

mAP – Mean Average Precision

MP – Mega pixeles

P – Precisión

px – Pixeles

V2 – Estadio fenológico con 2 hojas externas anilladas

V4 – Estadio fenológico con 4 hojas externas anilladas

VANT – Vehículo Aéreo No Tripulado

VP – Verdaderos positivos

YOLO – You Only Look Once

Capítulo 1. Introducción

El cultivo de maíz en México, particularmente en la región de Yucatán, es un pilar fundamental tanto para la seguridad alimentaria como para la riqueza cultural desde tiempos prehispánicos. Históricamente, el maíz (*Zea mays L.*) ha constituido un componente esencial en la dieta de la población yucateca, proporcionando nutrientes vitales y siendo un ingrediente clave en numerosos platillos regionales. La domesticación del maíz comenzó en Mesoamérica, donde se transformó a partir de una gramínea ancestral conocida como "teosinte", un proceso que se extendió a lo largo de aproximadamente diez mil años (Matsuoka et al., 2002). En la cosmovisión maya, el maíz es considerado un regalo divino, representando la base espiritual de la existencia humana, como se ilustra en el libro tradicional Popol Vuh, que describe a los humanos como "hijos del maíz" (Arellano, 1987).

La producción de maíz en Yucatán ha enfrentado un notable decrecimiento, en parte debido a políticas que no favorecen a los pequeños productores, quienes constituyen la mayoría de los agricultores en la región. Durante el año 2020, se produjeron 100,779.52 toneladas de maíz en Yucatán, lo que representó el 0.6% de la producción nacional (Echánove, 2021).

Uno de los retos más significativos que enfrenta la producción de maíz es la amenaza constante de plagas, siendo el conocido como gusano cogollero (*S. frugiperda*), una de las más destructivas. Esta plaga puede causar pérdidas significativas en la cosecha en regiones afectadas. Su capacidad de adaptación y rápida reproducción complican su control, lo que ha llevado a un aumento en el uso de pesticidas, lo que a su vez afecta a polinizadores, generando preocupaciones tanto económicas como ambientales. Por lo que, la detección temprana de plagas se vuelve crucial para minimizar el daño a los cultivos y optimizar la producción agrícola (Amein & Aziza, 2023).

Las Ciencias de Información Geoespacial ofrecen técnicas y herramientas que hacen posible implementar un sistema de monitoreo continuo de cultivos en momentos clave, que contribuya a la reducción de pérdidas económicas para los agricultores.

En este trabajo se desarrolla e implementa un sistema de monitoreo basado en el uso de un vehículo aéreo no tripulado (VANT) y visión computacional para la detección temprana de afectaciones causadas por *S. frugiperda* en cultivos de maíz. Mediante redes neuronales convolucionales (CNN), se logra identificar individualmente patrones morfológicos clave asociados a la plaga, como perforaciones foliares o residuos de deposiciones, con alta precisión espacial. Esta aproximación indirecta permite cuantificar tres dimensiones críticas del impacto agronómico: 1) Presencia/ausencia de infestación, 2) Distribución espacial de focos de daño, y 3) Grado de afectación por planta (leve/moderado/severa).

El método propuesto genera mapas geo etiquetados que pueden servir como base analítica para evaluar científicamente la eficacia de técnicas de control implementadas en parcelas experimentales. Al correlacionar variables como densidad de plantas afectadas, progresión fenológica (etapas V3-V5) y efectividad de intervenciones (ej: agua de vidrio, trampas de feromonas), el sistema proporciona métricas objetivas para optimizar estrategias de manejo integrado en el contexto agroecológico yucateco

1.1. Problemática

S. Frugiperda es una de las plagas polípagas de América que más causa pérdidas si no se controla oportunamente, ya que ataca a la planta en etapas tempranas, lo que provoca retraso en el desarrollo del cultivo y disminución del rendimiento del grano y forraje al alimentarse de tejido vegetal en las primeras etapas fenológicas del cultivo. Las larvas ocasionan pérdidas en la producción de maíz que van del 10 al 100 %, siendo la etapa más delicada cuando las plantas tienen entre 40 y 60 cm de altura (Banda et al., 1981).

La plaga se ha esparcido en años recientes en África, en donde ya representa pérdidas aproximadas de entre 21 y 53 % de la producción anual en 12 de los países productores (Montezano et al., 2018). Se encuentra en el grupo de las plagas constantes que pueden causar daños económicos anualmente porque está presente durante todo el ciclo de los cultivos (Reséndiz et al., 2016).

En el contexto de la agricultura en Yucatán, los problemas relacionados con la disminución del cultivo de maíz se relaciona directamente con la soberanía alimentaria, siendo el maíz un ingrediente central de la cocina regional (Ayora Diaz, 2017). Las comunidades rurales son particularmente vulnerables a este problema, con tasas de inseguridad alimentaria que pueden alcanzar hasta el 60% en algunas áreas (Carrillo et al., 2023).

Debido a la necesidad de mejorar las técnicas de gestión de los cultivos, la agricultura puede buscar soporte en tecnologías de análisis de imágenes, aprendizaje automático y bases de datos geográficos, de tal manera que el agricultor pueda mantener un seguimiento detallado de la salud de sus cultivos, identificando plagas y enfermedades antes de que constituyan un problema mayor.

Frente a esta situación, se planteó en México en la administración del presidente Andrés Manuel López Obrador apoyar a los pequeños productores de grano, en particular a los que cultivan maíz y frijol, con la meta de reforzar la autosuficiencia alimentaria (Echánove, 2021).

El 28 de noviembre de 2022, el secretario de Agricultura y Desarrollo Rural, Víctor Villalobos Arámbula, expuso que *“el sector primario, al igual que otros, tiene que reconocer que se deben hacer cambios e integrar nuevas tecnologías al sector agrícola, puesto que es necesario alcanzar un campo más productivo, tomando en cuenta la importancia del cuidado del medio ambiente, la salud humana y la protección a la biodiversidad”* (SADER, 2022).

Este proyecto tiene como meta generar una herramienta precisa de análisis aéreo de cultivos que sea capaz de cuantificar y dar seguimiento a afectaciones particulares, propias del tipo de cultivo, en este caso las causadas en cultivos de maíz por la plaga de *S. frugiperda*.

Así, se pretende sentar precedentes del uso de aplicaciones de aprendizaje profundo diseñadas específicamente para necesidades particulares del tipo de agricultura característica del estado de Yucatán. Su relevancia científica radica en el hecho de implementar mecanismos de las ciencias de información geoespacial, como los sistemas de

percepción remota con algoritmos de inferencia en tiempo real, como una herramienta para a resolver problemas actuales de la agricultura mexicana.

1.2. Hipótesis

La aplicación de técnicas de detección de objetos mediante redes neuronales convolucionales y vehículos aéreos no tripulados (VANT) permitirá realizar un monitoreo eficiente y preciso del estado de las plantas sobre los cultivos de maíz, facilitando la detección temprana de las afectaciones causadas por la plaga del gusano de *S. frugiperda* y la toma de medidas que eviten pérdidas de productos y ahorro de insumos.

1.3. Objetivo general

Desarrollar un sistema de monitoreo de plagas mediante herramientas de aprendizaje profundo y VANT, que sea útil para detectar la presencia y distribución de plagas importantes en agricultura, siendo en este trabajo la de gusano cogollero (*S. frugiperda*), en parcelas de maíz en Yucatán.

1.3.1 Objetivos específicos

1. Desarrollar un algoritmo que detecte y segmente plantas de maíz.
2. Desarrollar un algoritmo que clasifique las plantas detectadas como afectadas o no afectadas (plaga de *S. frugiperda*).
3. Implementar un método que permita hacer una georreferenciación precisa de las afectaciones registradas por el sistema aéreo.

1.4 Antecedentes y trabajos relacionados

Uno de los primeros trabajos que utilizan en detección y clasificación de enfermedades en plantas mediante el uso de redes neuronales convolucionales y procesamiento de imágenes es el de Huang et al. (2007), quienes utilizaron procesamiento de imágenes para identificar enfermedades en plántulas de *Phalaenopsis*, generaron un clasificador de 4 clases: 1 para saludable, 2 para pudrición blanda bacteriana, 3 para la mancha marrón bacteriana y 4 para la pudrición negra por *Phytophthora*. Las lesiones se segmentaron mediante un procesamiento que incluye ecualización de histograma por transformación exponencial y se analizaron características de textura por medio de la matriz de coocurrencia de niveles de gris y características de color. Mediante un clasificador por retropropagación, logró una precisión del 89.6% en la clasificación de lesiones y un 97.2% en la detección general.

En años más recientes, Zhang et al. (2018) desarrollaron modelos de redes neuronales convolucionales profundas, específicamente, versiones ajustadas de GoogLeNet y Cifar10,

que alcanzaron precisiones superiores al 98% en la clasificación de 8 diferentes enfermedades foliares del maíz. Sus mejoras incluyeron ajustes en los parámetros, modificaciones en las operaciones de pooling y la incorporación de funciones como ReLU y dropout. El trabajo demuestra que con una resolución adecuada, semejante a la que se implementó en ese estudio, es posible detectar las sutiles afectaciones en las plantas de maíz e incluso clasificarlas por tipo de afectación. El trabajo resalta que la captura rápida de imágenes de alta resolución puede hacer posible el diagnóstico oportuno de afectaciones, elemento clave para una gestión agrícola precisa y sostenible.

Por su parte, el uso de los VANT en agricultura ha demostrado ser una herramienta eficaz para la obtención de datos detallados que facilitan la toma de decisiones en el manejo de cultivos. Santos et al. (2019) hacen una revisión de diferentes formas de utilizar estas aeronaves en agricultura y destacan que los VANT permiten capturar imágenes de alta resolución espacial y temporal, lo que posibilita monitorear aspectos clave como el vigor vegetal, la detección de fallas en la plantación y la identificación temprana de problemas fitosanitarios, incluyendo plagas y enfermedades. Además, destacan que el empleo de índices de vegetación derivados de imágenes multiespectrales o incluso de cámaras convencionales con filtros específicos, ha sido fundamental para evaluar la salud de los cultivos y optimizar la aplicación de agroquímicos, contribuyendo así a una agricultura más sostenible y eficiente.

Como antecedente en la aplicación de VANT para el mejoramiento del cultivo de maíz, Pang et al. (2020) desarrollaron un sistema que utiliza imágenes capturadas con VANT a 5 metros de altura (resolución espacial de 2-3 mm/píxel) para evaluar cultivos de maíz en estadio V4-V5. Su metodología, basada en la red neuronal MaxArea Mask Scoring RCNN (Huang et al. 2019), permitió procesar filas completas de cultivo, en lugar de plantas individuales, alcanzando una precisión del 95.8% en la estimación de la densidad de plantas. Este trabajo es particularmente relevante porque destacan explícitamente la utilidad que tiene la capacidad de cuantificar la cantidad de plantas faltantes para apoyar decisiones de resiembra temprana, proporcionando a los agricultores una herramienta objetiva para evaluar la replantación cuando aún es viable agrónomicamente. Sus resultados establecen un precedente sólido para el uso de drones como tecnología habilitadora en el mejoramiento del cultivo de maíz, demostrando cómo la integración de percepción remota y aprendizaje profundo puede optimizar prácticas de manejo agronómico basadas en datos precisos y oportunos.

Otro trabajo de detección automatizada de *S. frugiperda* mediante VANT y visión computacional, se generó en el estudio de Ishengoma et al. (2021), que constituye un referente metodológico clave. Los autores utilizaron un dron Phantom 4 Pro v2 para capturar imágenes a 5 m de altura, las cuales fueron procesadas mediante recortes sucesivos (de 5472 × 3078, a 450 × 300 y después a 150 × 150 píxeles) y la aplicación del detector de esquinas Shi-Tomasi, para resaltar texturas asociadas a daños por *S. frugiperda*. Tras eliminar imágenes de fondo y balancear el dataset (11,280 imágenes etiquetadas como infectadas/no infectadas), entrenaron arquitecturas CNN como MobileNetV2 e InceptionV3, alcanzando 100% de precisión en la clasificación binaria. Este trabajo demuestra que el preprocesamiento dirigido a características morfológicas sutiles, como agujeros y ventanas

foliares, es determinante para lograr una detección confiable, incluso con equipos de bajo coste. Sus hallazgos validan el potencial de los VANT y los algoritmos de aprendizaje profundo para el monitoreo de plagas en condiciones reales, sentando las bases para desarrollos posteriores en agricultura de precisión.

Algunos estudios previos que evalúan el uso en agricultura de las estructuras de redes neuronales convolucionales de la familia de YOLO para desarrollar sistemas de detección de plagas incluyen: J. Liu y Wang (2020) implementaron una detección temprana de enfermedades en tomates basada en MobileNetv2-YOLOv3. Bhatt et al. (2019) usaron YOLOv3 para detectar plagas en plantas de té, probando diferentes arquitecturas como DarkNet-19, MobileNet, Inception v2, ResNet-101 y DarkNet-53. Chen et al. (2021) desarrollaron una aplicación móvil para detectar especies de cochinillas, cóccidos y diáspidos utilizando YOLOv4. Kuzuhara et al. (2020) crearon una red de propuesta regional para detección de plagas de insectos usando YOLOv3 y un modelo de reidentificación Xception.

Otro estudio particularmente relevante para este trabajo es el de Feng et al. (2022), quienes abordaron específicamente la detección de *S. frugiperda* en cultivos de maíz, mediante detección indirecta por medio de las señales de su alimentación utilizando imágenes RGB de VANT. Su enfoque implementó el uso de diferentes redes neuronales convolucionales adaptando los mecanismos de atención de ResNeSt-50, ResNet-50, EfficientNet y RegNet, para clasificar plantas afectadas directamente desde imágenes aéreas, alcanzando una precisión del 89.39% en su conjunto de imágenes de Test, en condiciones de campo reales. Este trabajo demuestra la viabilidad del uso de VANT con sensores RGB estándar, en lugar de equipos multiespectrales costosos, en la tarea de detección de plagas.

Estos antecedentes demuestran que es posible generar un sistema capaz de recopilar imágenes de maíz en condiciones reales de producción agrícola y generar la detección automática de hojas infectadas por *S. frugiperda*, que proporcione una referencia confiable para la formulación e implementación de medidas de prevención y control.

El presente trabajo obtiene resultados similares a los de Feng et al. (2022), en cuanto a las métricas de precisión en la detección de zonas afectadas por *S. frugiperda* en plantas de maíz, lo cual se discute en el capítulo 5. En este trabajo se propone un método de aplicación conjunta de técnicas en el estado del arte para analizar plantas maíz y que sirva para apoyar los esfuerzos de transición agroecológica de la región.

Capítulo 2. Marco Teórico

2.1 Situación del cultivo de maíz en Yucatán

Tradicionalmente, en la Península de Yucatán las comunidades rurales implementan una estrategia en la que se recurre al cultivo de diversos recursos agrícolas, la cual se maneja principalmente a nivel familiar. Esta estrategia contempla el aprovechamiento de una amplia gama de recursos naturales (Toledo et al., 2003), conocido como milpa, en el que el maíz es el elemento central de la misma. Si bien, el cultivo de maíz se da tradicionalmente en el contexto de la milpa, muchos ejidatarios de la región optan actualmente por producción del monocultivo de maíz u otras especies, rotándolo en algunas ocasiones con otros cultivos de temporada.

Ambas técnicas de cultivo de maíz, milpa y monocultivo, se enfrentan a problemas como a la expansión de los cultivos de soja, que son más redituables, sin embargo, requiere mucho más capital comparativamente para realizar su cultivo, ya que son cultivos que requieren ser tecnificados. Eso ha llevado a muchos ejidatarios a rentar sus parcelas de tierra, que antes destinaban a la milpa, a los productores de soja, lo que ha causado situaciones como deforestación y problemas con las abejas (Echánove, 2021). El cambio climático también es causa de varias pérdidas de las cosechas de temporal en años muy secos. (Bueno et al., 2010).

Otro factor importante son las plagas que invaden los cultivos, principalmente la que es conocida como el gusano cogollero, (*S. frugiperda*). Asimismo, su presencia propicia las condiciones para la aparición de plagas secundarias como la mosca de los estigmas (*Euxesta stigmatias* Loew), que puede causar pérdidas económicas significativas (Camacho-Báez et al., 2012). El control de plagas involucra el uso de agroquímicos, algunos de estos compuestos son altamente tóxicos, tanto para la salud humana como para el medioambiente (Rodríguez et al., 2019), eliminando tanto a los insectos de la plaga, como a los demás insectos, presentes en el cultivo de maíz, como las abejas (*Apis mellifera*) (García-Gutiérrez et al., 2012).

2.2. Características de la plaga de gusano

El gusano cogollero (*Spodoptera frugiperda*) es un insecto del orden Lepidoptera, familia Noctuidae, es reconocido como una de las más importantes plagas de palomas nocturnas del continente americano, y que ya se extiende por el africano. Se tiene registro de que se puede alimentar de 353 especies de plantas, de 76 familias distintas (Montezano et al., 2018), la infestación es común y cuando es alta puede provocar la defoliación completa (Valdez-Torres et al., 2012).

Es una plaga de origen tropical, sin embargo, se puede encontrar en casi todo el continente americano. Cada hembra oviposita en promedio 1000 huevos durante aproximadamente 12 días en los que vive en estado adulto. Los huevos de este gusano están cubiertos con hilos

de seda y se encuentran sobre las hojas, en la parte media de la planta de maíz, preferentemente en el envés o en la zona basal de las mismas, tardan de 3 a 5 días en eclosionar (Murúa y Virla, 2004). Las diferentes etapas de su ciclo de desarrollo se pueden observar en la Fig. 2.1.



Figura 2.1. Ciclo biológico de *S. frugiperda*. Tomado de Hardke et al., (2015).

Para poder completar su desarrollo, las larvas de *S. frugiperda* consumen en promedio 179.7 cm² de superficie foliar de maíz, en un periodo de entre 14 y 22 días (Casmuz et al., 2008). Una infestación no controlada de esta plaga puede llevar a una gran reducción del rendimiento de cultivo al afectar el desarrollo de la planta, lo cual ocasiona pérdidas económicas para los productores.

El ciclo de *S. frugiperda* se completa en un periodo medio de 30 días. En la Figura 2.2 se ilustra una planta afectada, en la cual las larvas más pequeñas ya se han comido entre sí, han comido las hojas y las más grandes han llegado al centro de la planta. Se pueden visualizar algunos de los restos que son indicadores de presencia de *S. frugiperda* en el centro de la hoja lesionada, las hojas con huecos y los restos cafés que evidenciando la pérdida de las hojas centrales, las más importantes para el desarrollo de la planta.

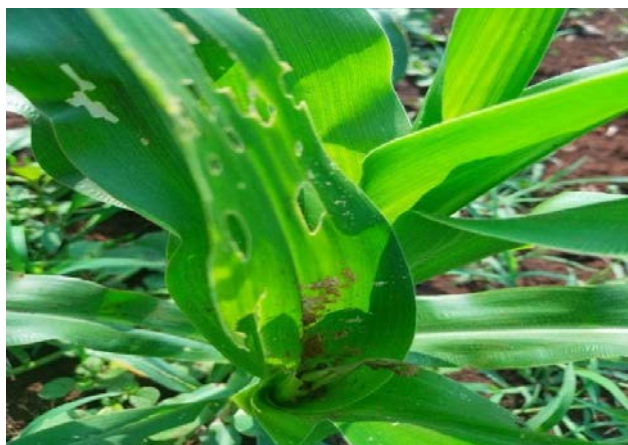


Figura 2.2. Lesiones provocadas por *S. frugiperda* a una planta de maíz. Hojas y centro afectados. Imagen propia.

2.3. Características de la planta de maíz en etapas sensibles

Las etapas fenológicas del maíz se clasifican mediante un sistema que los clasifica como estadios V (vegetativos) y estadios R (reproductivos), una metodología estandarizada que permite monitorear con precisión el desarrollo del cultivo.

Los estadios V se refieren al número de hojas emergidas que han desarrollado un anillo en el tallo y van desde V1 hasta V(n). Donde n equivale al número de hojas anilladas emergidas. Aproximadamente en V5 comienza su etapa de espigamiento y la planta deja de ser susceptible a *S. frugiperda*.

Las plantas de maíz son susceptibles de ser dañadas por *S. frugiperda* solo durante su desarrollo vegetativo, desde el estadio V1, hasta aproximadamente 60 días después en un estadio V5, por lo que es en esta etapa cuando se debe monitorear y tratar la presencia de la plaga (Galindo & Garcés, 2023).

A continuación se mencionan brevemente las técnicas plaguicidas que ya se aplican en las parcelas analizadas

2.3.1. Técnicas plaguicidas aplicadas en la zona de estudio.

Entre los métodos que se estudian para combatir la plaga, se reportan métodos físicos, mecánicos, culturales, biológicos y químicos, los cuales se pueden combinar para mejorar los resultados (Ángel-Ríos et al., 2015). En la región de estudio se exploran técnicas agroecológicas que ayuden en la mitigación de las plagas, conforme al nuevo Plan de Transición Agroecológica, y se estudia la eficacia de esas técnicas y las mejores prácticas en su utilización. Las técnicas que se aplican en la región y que se encuentran en desarrollo y estudio son las siguientes:

-
1. **Aplicación del compuesto conocido como agua de vidrio.** Es una solución alcalina con un pH aproximado de 12, se emplea como fungicida, insecticida y nutriente orgánico; fortalece la respuesta inmunológica vegetal ante el ataque por plagas y/o enfermedades, así como agobios por sequías. El grado de alcalinidad permite deshidratar huevecillos de insectos de cuerpo blando, además de hacer que no sea viable el crecimiento de hongos.
 2. **Trampas de feromonas de confusión sexual.** Es una solución de reciente desarrollo a base de feromonas, que funciona por medio de dispensadores colocados en los sembradíos, los cuales van liberando de manera continua un compuesto químico y forman una "nube" sobre el cultivo que lo protege durante 90 días. Las trampas de feromonas se sitúan a una distancia de 50 m entre cada una.
 3. **Las trampas de bidones con piña y melaza.** Para atrapar palomillas adultas de insectos que, en su fase larvaria (gusanos), atacan a los cultivos de maíz y otras especies. La trampa resulta eficiente para atrapar adultos (palomillas) de la familia Noctuidae tales como *S. frugiperda*. Los bidones se elevan sobre la superficie del cultivo a medida que crece, para que, de esa manera, el aroma se esparza por la superficie.

Otra técnica que se busca implementar es el control biológico, que se sirve de la presencia de depredadores o entomófagos naturales de *S. frugiperda* como catarinas y crisopas, que en su mayoría se alimentan de los huevecillos de este insecto (Hernández et al., 2018).

El presente trabajo pretende proporcionar un método que sea capaz de apoyar los estudios relacionados, mediante recopilación de datos para realizar análisis estadísticos que permitan medir la eficacia de las técnicas implementadas en los diferentes entornos.

2.4. Elección de Modelo de aprendizaje profundo

Existen diversas arquitecturas de modelos de aprendizaje profundo que han sido desarrolladas para tareas de detección y segmentación que pueden ser utilizadas para abordar los presentes desafíos agrícolas. Recientemente, se han desarrollado técnicas que integran detección y segmentación de instancias, mejorando la precisión en la localización y segmentación. La capacidad de YOLO para realizar estas tareas en tiempo real ha consolidado su uso en robótica, vigilancia y vehículos autónomos, donde la rapidez es esencial (Terven, 2023).

Este estudio se enfoca en optimizar la detección de enfermedades en cultivos, requiriendo un modelo que equilibre precisión, velocidad y eficiencia de recursos para procesar grandes volúmenes de imágenes y facilitar su implementación en campo. Tras analizar diversas arquitecturas de red disponibles, su tamaño y capacidades, se seleccionó la arquitectura de red neuronal convolucional YOLOv11n-seg. Esta red es la versión de tamaño nano de los modelos disponibles de la serie YOLOv11, siendo las opciones nano, small, medium, large y Xtra_large. El modelo YOLOv11n-seg cuenta con 2.9 millones de parámetros ajustables, se utilizó el modelo de red especializada en segmentación de instancias.

2.5. Estructura de procesamiento del modelo YOLOv11

YOLOv11 (Jocher & Qiu, 2024) cuenta con diferentes versiones diseñadas para abordar tareas clásicas de visión computarizada como detección de objetos, segmentación de instancias, clasificación de imágenes, estimación de poses y detección de objetos orientados (OBB). También está disponible en diferentes tamaños, dependiendo del número de parámetros (nano, small, medium y large y x-large). En este trabajo se utilizó el modelo nano de segmentación, que es el más apropiado para correr en dispositivos de hardware de baja potencia en campo.

Su arquitectura está dividida en 3 etapas principales: Columna vertebral, Cuello y Cabeza. Cada etapa será descrita a grandes rasgos en esta sección. En la Figura 2.3, es posible visualizar la arquitectura completa del modelo YOLOv11, con sus tres etapas principales y sus módulos internos.

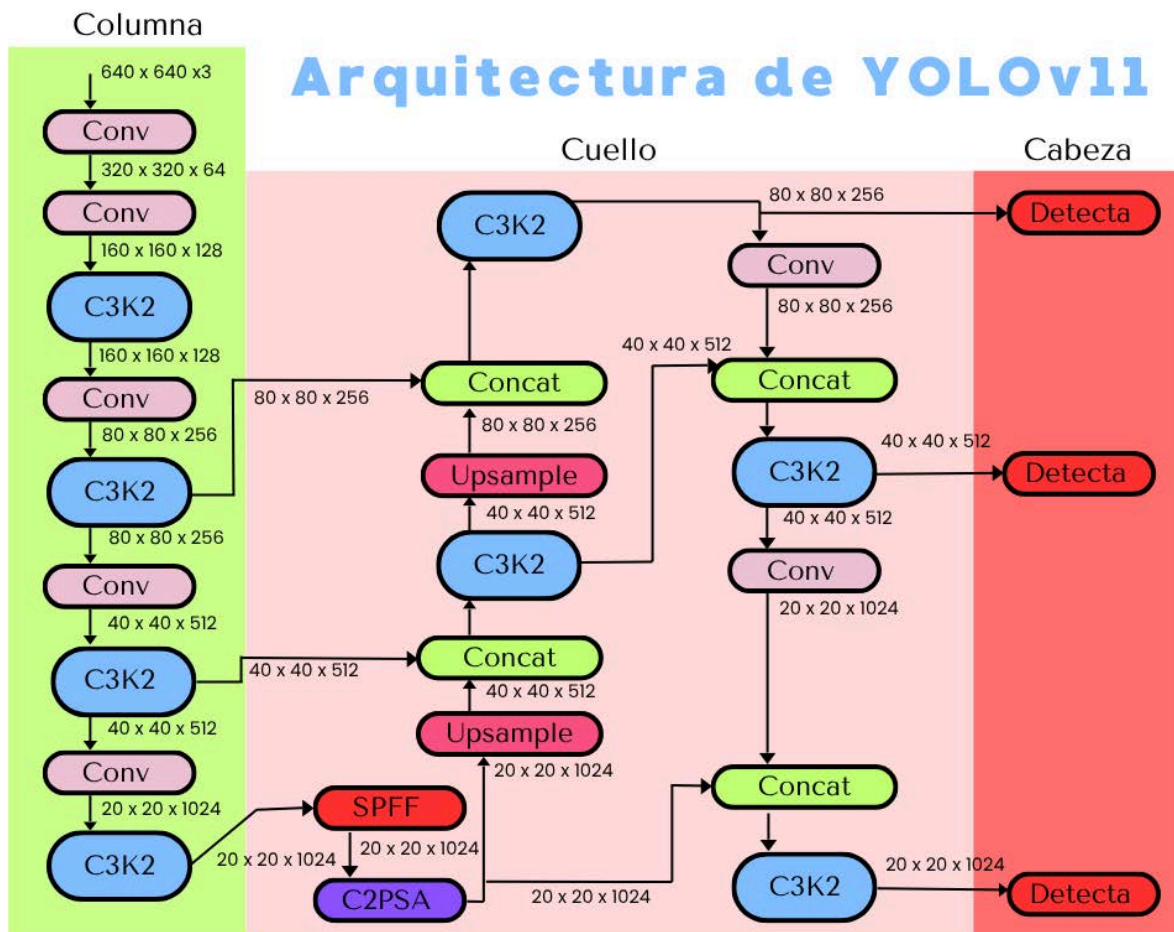


Figura 2.3. Arquitectura general del modelo YOLOv11, mostrando las 3 etapas principales y los módulos que trabajan con los mapas de características a diferentes escalas. Editado de Jocher & Qiu, (2024), Ultralytics.

Esta arquitectura supera a modelos anteriores de la familia YOLO y a otros modelos como EfficientDet, como se muestra en el gráfico comparativo de la precisión contra tiempo de inferencia que se muestra en la Figura 2.4. En el gráfico se comparan YOLOv11 con YOLOv5 y con EfficientDet. La métrica se obtiene de implementar cada modelo sobre la base de datos Common Objects in Context (Lin et al., 2014) de Microsoft y al obtener su magnitud de valor mAP50-95 y latencia. En azul se resaltan los resultados de YOLOv11, en rosa YOLOv5 y en gris EfficientDet (Jocher & Qiu, 2024).

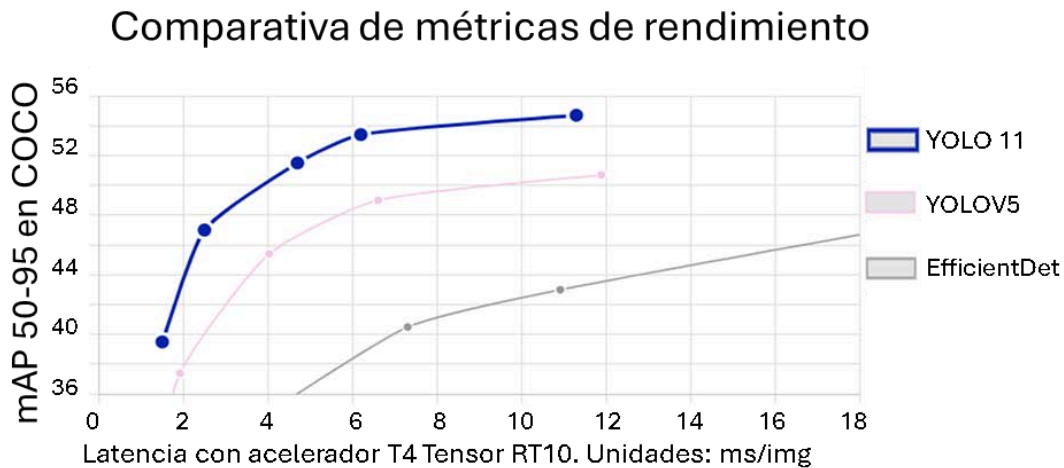


Figura 2.4. Comparativo de velocidad y precisión de YOLOv11 con YOLOv5 y EfficientDet.

Se describen las 3 etapas funcionales: Columna vertebral, Cuello y Cabeza, que conforman a esta arquitectura.

2.5.1. Columna vertebral (Backbone)

La primera etapa es la de columna vertebral y está compuesto de módulos de tipo Conv y C3K2:

1. Módulo Conv

Cada módulo Conv procesa la información comenzando con una capa convolucional 2D, seguida de una capa de normalización por lotes 2D (Batch Normalization 2D) y finalmente con una función de activación SiLu (Sigmoid Linear Unit).

La función SiLU se define como el producto de la función sigmoidea logística, por su entrada. Es decir, para la entrada x , SiLu se calcula mediante la función sigmoidea multiplicada por su entrada:

$$SiLU(x) = x \cdot \sigma(x) \quad (2.1)$$

En donde $\sigma(x)$ es la función sigmoidea logística:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (2.2)$$

. La estructura del módulo Conv se ilustra en la Figura 2.5.

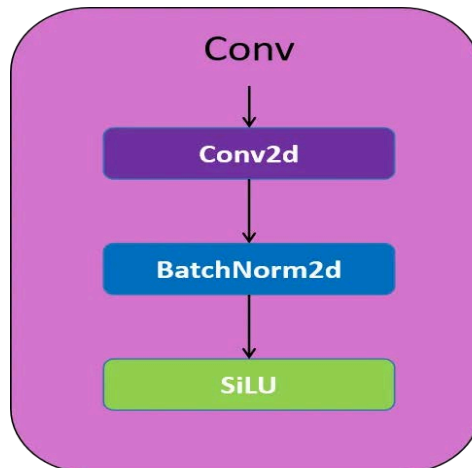


Figura 2.5. Módulo Conv y los métodos que contiene.

La función de activación SiLU presenta dos propiedades clave que la distinguen de otras funciones como ReLU. Para valores $|x| > 4$, su comportamiento converge al de ReLU, manteniendo la eficiencia computacional. Sin embargo, presenta un mínimo global único en $x \approx -1.278$ (valor ≈ -0.278), característica ausente en activaciones convencionales (Figura 2.6, izquierda). Su derivada, definida como:

$$\frac{d}{dx} SiLU(x) = \sigma(x)[1 + x(1 - \sigma(x))] \quad (2.3)$$

actúa como estabilizador implícito durante el entrenamiento. Su perfil suavizado regula automáticamente la magnitud de los gradientes sin saturación, evitando oscilaciones numéricas. Esta propiedad es particularmente valiosa en tareas de segmentación precisa de objetos pequeños, como lesiones foliares en agricultura de precisión, ya que preserva información espacial crítica al mantener gradientes bien condicionados durante la retropropagación, mejorando así la delimitación de características sutiles (Figura 2.6, derecha).

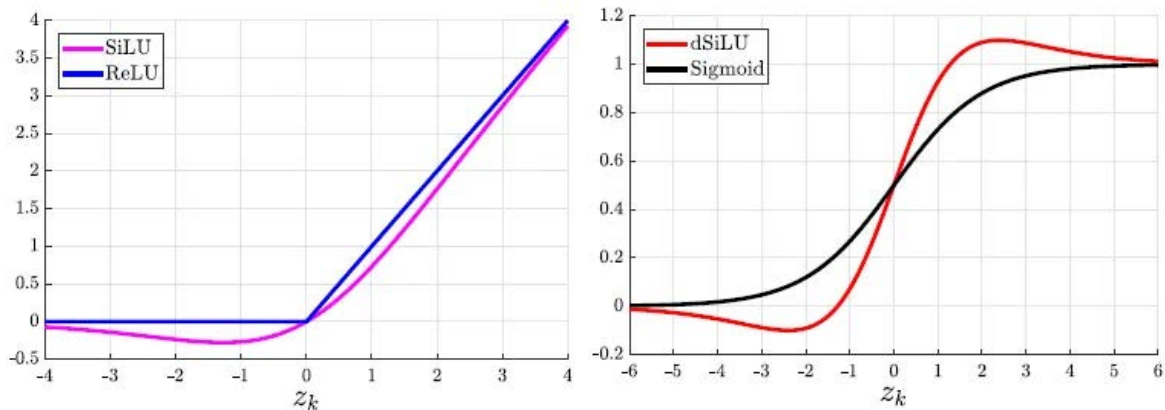


Figura 2.6. Comparación de función ReLU con Silu (Izquierda) y comparación de la derivada de SiLU con la Sigmoide (Derecha).

2. Módulos C3K2

Es una evolución del cuello de botella CSP (Cross Stage Partial) introducido en la versión YOLOv8. El módulo C3K2 optimiza el flujo de información a través de la red, dividiendo el mapa de características y aplicando una serie de convoluciones de núcleo más pequeñas (3x3), que son más rápidas y computacionalmente más económicas que las convoluciones de núcleo más grandes. Comienza con un módulo C3K, en donde la entrada pasa a través de un módulo Conv, seguidos de una serie de 'n' capas de cuello de botella, seguidas de una concatenación y finaliza con el módulo Conv final. Este módulo C3K se ilustra en la Figura 2.7 (derecha).

C3K2 comienza con un módulo convolucional Conv, seguido de una serie de módulos C3K. Al final son concatenados la salida del módulo Conv, con la salida del último módulo C3K, terminando con un módulo de convolución final. El módulo C3K2 se ilustra en la Figura 2.7 (izquierda).

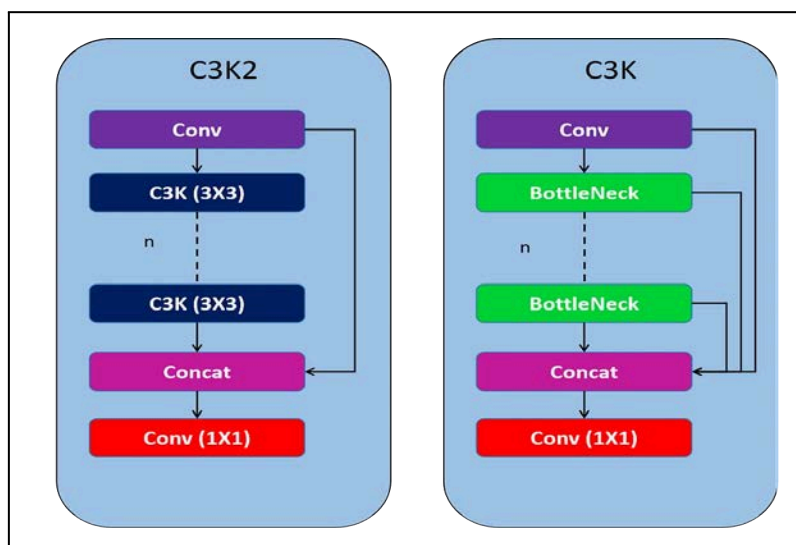


Figura 2.7. Módulo C3K (Derecha). C3K es incorporado n veces dentro del módulo C3K2 (Izquierda).

El módulo de Bottle Neck, dentro de C3k es similar a como funcionan las redes residuales como ResNet, que utilizan funciones residuales para mejorar el desempeño en redes profundas. En este caso se trata de una secuencia de módulos Conv, pero en esta ocasión existe un parámetro de acceso directo que determina si se desea obtener la parte residual o no. Si el acceso directo se establece en falso, no se considerará ningún residuo. El módulo se ilustra en la Figura 2.8.

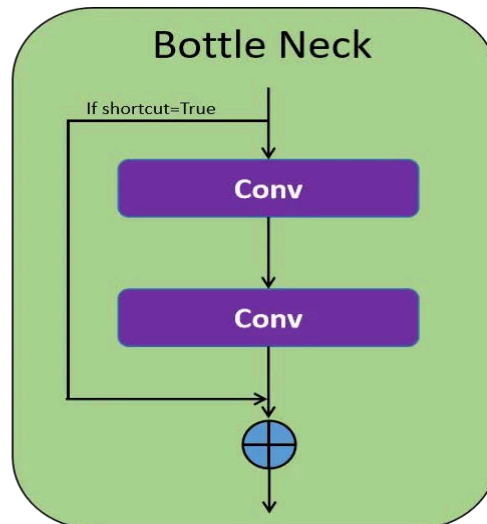


Figura 2.8. Capa de cuello de botella conformada por módulos Conv con opción de entrada de la parte residual.

2.5.2. Cuello (Neck)

El "cuello" de esta arquitectura se encarga de recopilar y fusionar las características extraídas por el "backbone". Esto es crucial porque las características de diferentes niveles del "backbone" capturan información a diferentes escalas. Utiliza en su estructura el mismo tipo de módulos que en el "backbone", pero introduce un módulo SPPF y uno C2PSA:

1. Módulo SPPF

El módulo SPPF (Spatial Pyramid Pooling - Fast) es una técnica clave introducida en YOLOv5, diseñada para realizar agrupación piramidal rápida, mediante max-pooling secuencial para capturar contexto multiescala. El módulo se ilustra en la Figura 2.9, con un ejemplo de la operación de Max Pulling 2D.

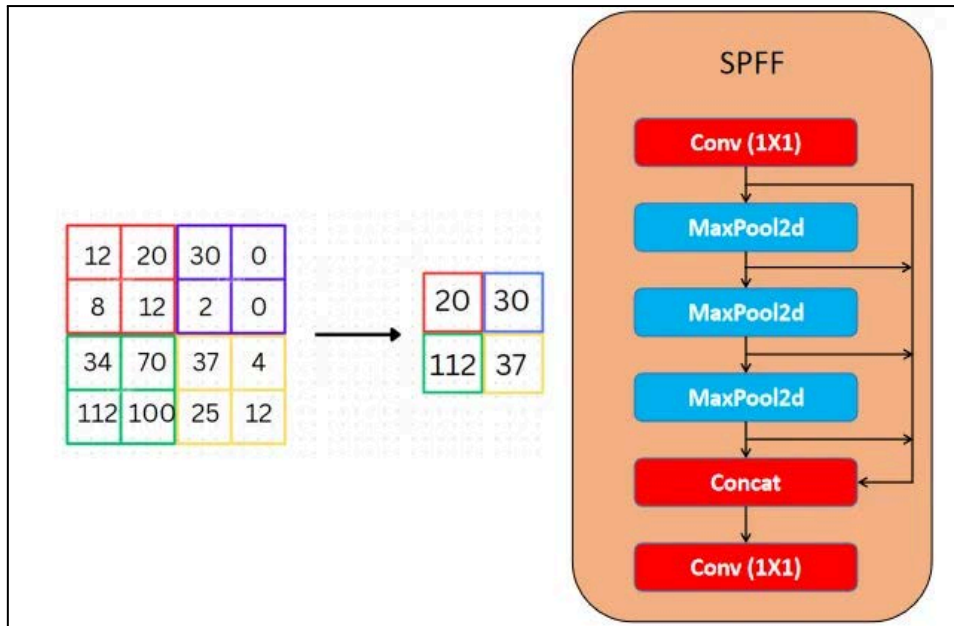


Figura 2.9. Módulo de agrupación piramidal SPFF que utiliza 3 capas de MaxPooling secuencial para realizar los mapeos de características a diferentes escalas (derecha). Ejemplo de operación Max Pulling (izquierda).

2. Módulo C2PSA

El módulo C2PSA (Cross Stage Partial with Spatial Attention) es una de las innovaciones más relevantes de YOLOv11, contiene los mecanismos de atención PSA, que mejoran la focalización del modelo en regiones importantes de la imagen, como objetos pequeños o parcialmente ocluidos, al enfatizar la relevancia espacial en los mapas de características. El módulo C2PSA se ilustra en la Figura 2.10 (derecha), que contiene a los 2 módulos PSA (derecha).

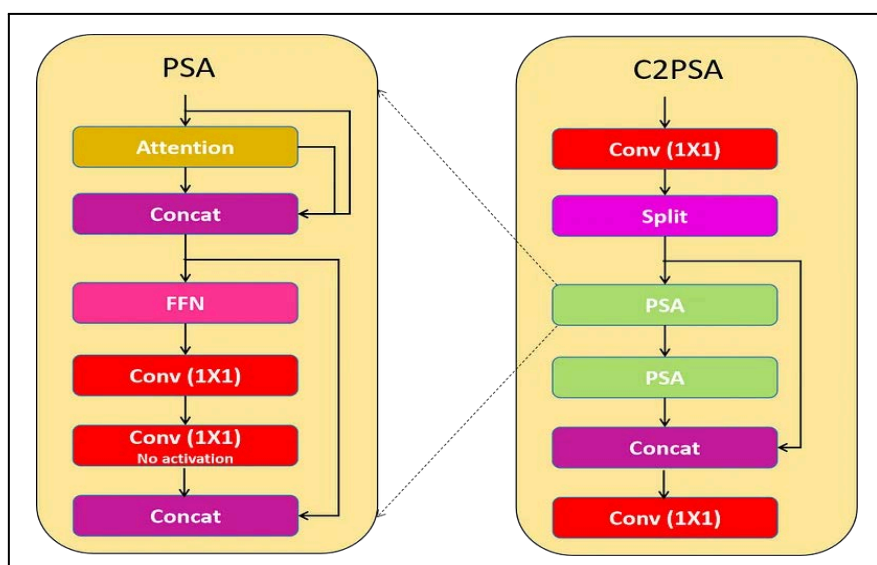


Figura 2.10. Módulo PSA (izquierda) que está contenido dentro del módulo C2PSA (derecha).

El módulo PSA emplea atención espacial para priorizar regiones relevantes (ej: objetos pequeños) y suprimir fondos. Su arquitectura (Fig. 2.16, izquierda) procesa la entrada mediante: 1) Capa de atención, 2) Concatenación triple (entrada+atención+procesado), 3) módulos FFN/Conv, 4) Concatenación final. Esto genera firmas posicionales únicas que diferencian objetos similares en distintas ubicaciones.

2.5.3. Cabeza (Head)

Genera máscaras de segmentación en 3 escalas: baja, mediana y alta resolución, adaptándose a objetos de distintos tamaños. Este enfoque asegura que los objetos pequeños se detecten con mayor detalle, mientras que los objetos más grandes son capturados por características de nivel superior.

2.5.4. Hiperparámetros y Optimización

Los hiperparámetros se seleccionaron siguiendo prácticas establecidas como estándar, para modelos YOLOv11 que ajustaron mediante pruebas preliminares. Se utilizó un tamaño de lote de 16. La tasa de aprendizaje inicial fue $1e-3$, un valor común para el optimizador de descenso de gradiente estocástico con momento. El momentum se fijó en 0.937, lo que suaviza las actualizaciones de pesos y acelera la convergencia al priorizar direcciones de gradiente consistentes.

2.6. Método SAHI

El método SAHI (Slicing Aided Hyper Inference) es una técnica en visión por computadora diseñada para mejorar la detección de objetos pequeños o densamente agrupados en imágenes de alta resolución, como lesiones en plantas (Akyon et al., 2022). Este enfoque aborda la pérdida de información en modelos convencionales al procesar imágenes completas. SAHI se basa en tres etapas clave:

División de la imagen: La imagen original se segmenta en subregiones superpuestas, permitiendo un análisis a resolución adecuada y evitando la reducción de tamaño de objetos pequeños.

Inferencia en parches: Cada parche se procesa individualmente con un modelo de detección, capturando detalles que podrían pasar desapercibidos en un análisis global.

Fusión de resultados: Las detecciones de los parches se consolidan utilizando técnicas como la Supresión de No Máximos (NMS) para eliminar duplicados y generar predicciones precisas.

SAHI es especialmente útil en el contexto agrícola, donde las lesiones suelen ser pequeñas e irregulares, difíciles de identificar debido a la resolución de la imagen o la complejidad del fondo. Al analizar parches a resoluciones nativas, se preserva la morfología de las lesiones y se reducen falsos negativos. Aunque su uso implica un mayor tiempo de inferencia por el procesamiento múltiple, se mantiene dentro de un rango óptimo para aplicaciones como el diagnóstico temprano en cultivos, y funciona con modelos ya entrenados, evitando costos computacionales adicionales.

2.7 Métrica F1

La métrica F1 (F1-score) o medida F, es un indicador de rendimiento utilizado en evaluación de modelos de clasificación que combina la precisión y la exhaustividad (recall) en un único valor. Se define como la media armónica de estas dos métricas:

$$F1 = (2) \left(\frac{\text{Precisión} \times \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}} \right) \quad (2.4)$$

Precisión (P): Proporción de predicciones positivas correctas respecto al total de predicciones positivas:

$$P = \frac{VP}{VP + FP} \quad (2.5)$$

VP: Verdaderos positivos

FP: Falsos positivos

FN: Falsos negativos

Exhaustividad (R): Proporción de casos positivos reales correctamente identificados:

$$R = \frac{VP}{VP + FN} \quad (2.6)$$

2.8 Métrica mAP50

El mAP50 (precisión promedio media con al menos un 50% de superposición) es una métrica estándar en visión computacional para evaluar modelos de detección de objetos. Combina dos aspectos críticos, la capacidad del modelo para identificar correctamente lo que aparece en la imagen o clasificación y la capacidad para localizar con precisión el contorno o máscara de segmentación de dónde está el objeto. Es la métrica que se busca optimizar en este trabajo.

2.8.1. Fundamentos Conceptuales

IoU (Intersección sobre Unión): Es una métrica que indica cuánto se superpone el área predicha por el modelo con el área real del objeto. Se obtiene de calcular un factor:

$$\text{IoU} = \text{Área de Intersección} / \text{Área de Unión} \quad (2.7)$$

Umbral 50: Indica que solamente cuando la superposición IoU es mayor o igual a 50%, se considera una detección válida.

Curva Precisión-Exhaustividad: Para cada clase se varía el umbral de confianza del modelo de 0 a 1. En cada umbral, se calcula Precisión y Exhaustividad. Se traza la curva Precisión-Exhaustividad al recorrer todos los valores de confianza

AP (Precisión Promedio): Es el área bajo la curva Precisión-Exhaustividad para una clase. Sus valores abarcan desde 0 o pésimo y 1 o perfecto.

mAP (Precisión Promedio Media): Promedio de las AP de todas las clases

mAP50: Versión que exige mínimo 50% de superposición (IoU ≥ 0.5)

2.9 Métrica mAp50-95

Es el promedio de mAP calculado sobre 10 umbrales de solapamiento (IoU) desde 0.50 hasta 0.95, en incrementos de 0.05. Para obtener esta métrica se evalúa la robustez del modelo en detección de objetos con distintos niveles de precisión espacial, desde coincidencias bajas de 50% de IoU, hasta coincidencias muy exigentes de 95% de IoU. Es una de las métricas estándar de robustez en modelos de visión computarizada.

Capítulo 3. Documentación de los sitios de estudio

3.1. Muna

En mayo de 2023 comenzaron los levantamientos de imágenes. Debido a que INIFAP tiene un centro de investigación en el municipio de Muna y a que es una zona en la que existe coordinación entre los agricultores locales, se eligió a ese municipio como primer sitio de estudio. Para entrar en contacto con la comunidad agrícola productora de maíz de Yucatán, se asistió a un evento local de intercambio de semillas nativas. Estos son eventos organizados por las comunidades agrícolas para preservar las semillas tradicionales. En el evento se contactó al Coordinador regional del Programa para el Bienestar (PpB) del Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP) y coordinador de las Escuelas de Campo de la región de Yucatán y Quintana Roo, el M. en C. Adrián Javier López Pérez, quien ayudó a coordinar con los agricultores las parcelas en donde se realizarían los mapeos en parcelas de maíz que estuvieran en la etapa adecuada para el estudio.

El municipio de Muna es uno de los 106 municipios del estado de Yucatán, se localiza entre los paralelos 20°24" y 20°35" de latitud norte y los meridianos 89°37" y 89°47", de longitud oeste. Su cabecera municipal es la localidad homónima de Muna. Se localiza a 64 km al sur de la ciudad de Mérida. En la Figura 3.1 se ilustra su posición dentro de la península de Yucatán y un acercamiento satelital al municipio, con una referencia de 20 km. es una comisaría que cuenta con agricultores organizados que participan en las llamadas Escuelas de Campo y que activamente se organizan para compartir sus técnicas y conocimientos de cultivo de maíz y otros productos, por lo que se denomina como "Territorio funcional" en el Programa para el Bienestar y es un territorio que ya se encuentra en coordinación implementando un plan de transición agroecológica.

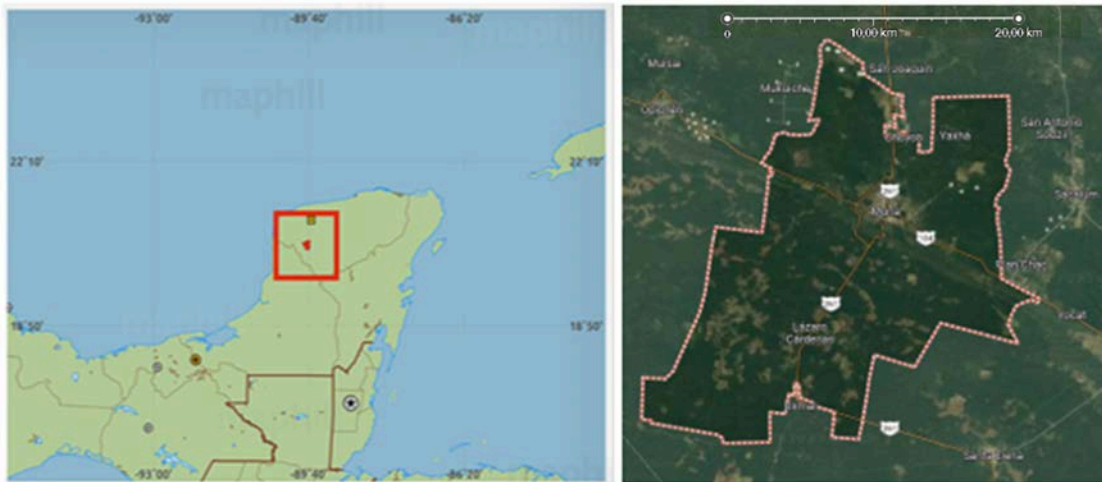


Figura 3.1. Ubicación del Municipio de Muna, Yucatán, dentro de la península a la izquierda, en el recuadro rojo se encuentra el municipio y en gris en la parte superior, la ciudad de Mérida. Se muestra un cercamiento satelital al municipio a la derecha.

Se realizó la primera visita a las parcelas 1 y 2 en conjunto con un técnico de INIFAP para analizar las características de la zona de cultivo y reconocer el entorno y las plantas afectadas el 19 de mayo de 2023. Ese día el técnico del INIFAP impartió una sesión del programa de “Escuela de Campo” en la parcela 1, en donde se compartió conocimiento entre los agricultores con respecto a algunas medidas que se pueden tomar para resolver problemas propios del cultivo de maíz, entre los que se encuentran técnicas para mitigar la aparición de plaga de *S. frugiperda* (como trampas de melaza y de feromonas de confusión sexual), ya que es un problema que suele afectar constantemente a los agricultores a través del daño a sus cultivos de maíz. Ese mismo día se realizó el primer levantamiento de imágenes aéreas en las parcelas 1 y 2 para generar la base de datos.

Se buscó hacer levantamientos en parcelas que tuvieran plantas de maíz en la etapa adecuada, con alturas entre 30 cm y 60 cm. Las parcelas suelen estar divididas en unidades agrícolas de 50 m x 50 m conocidas en la región como “mecates”, mismas que fueron las unidades de muestreo en el estudio.

En la segunda visita, el 8 de junio, se realizó nuevamente levantamiento de imágenes sobre las parcelas 1, 2 y 3.

La ubicación de las parcelas al sur de Muna, cercanas a las instalaciones de INIFAP se muestran en la Figura 3.2, en la parte de arriba se visualiza la cabecera municipal, al centro, los pines rojos que indican la ubicación de las parcelas 1, 2 y el sitio donde se realizó la Escuela de Campo y abajo, la parcela 3 y la comisaría Lázaro Cárdenas, en donde se localizan las instalaciones del INIFAP.

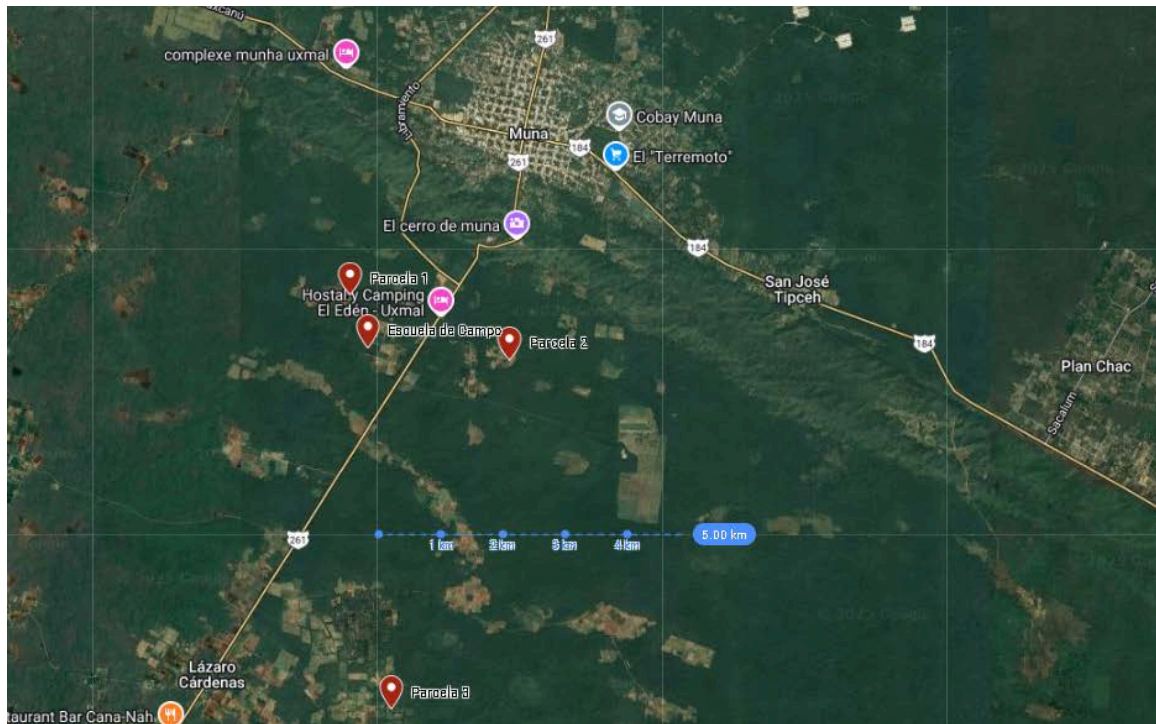


Figura 3.2 Imagen de Google Maps con los pines rojos que marcan la ubicación de las parcelas 1, 2 y 3 en las que se realizaron los mapeos de Muna.

En el municipio de Muna se realizaron diversos muestreos en tres parcelas, sus características principales se describen a continuación:

Parcela 1. Esta parcela tiene un área superficial de 0.214 hectáreas, se encuentra a 3.82 Km al sureste del centro de Muna. Pertenece a un agricultor que ha cultivado maíz en esa parcela durante un periodo aproximado de veinte años, el cual manifiesta que el rendimiento de su cultivo ha ido decreciendo con los años y manifiesta que la tierra está cansada, ya que después de cada cosecha, no se le ha permitido un periodo de “descanso” en el que la tierra recupere sus propiedades naturalmente. El agricultor se había incorporado al sistema de Escuelas de Campo (Figura 3.3), en los dos años anteriores al estudio, y había comenzado a incorporar el agua de vidrio, trampas de feromonas de confusión sexual para desorientar a las polillas de *S. frugiperda* y trampas de bidones con piña y melaza para capturarlas. Se trataba de la segunda vez que cosechaba implementando dichas técnicas. El agricultor había comenzado a implementar métodos para la regeneración del suelo, como aplicación de microorganismos y lixiviados. El sistema de riego utilizado es de riego por goteo. La conductividad de la tierra fue de 1.7 $\mu\text{S} / \text{cm}$ (microSiemens/cm), que se considera dentro del rango ideal para el cultivo del maíz.



Figura 3.3 Sesión de Escuela de Campo en la parcela 1 del municipio de Muna.

Parcela 2. Esta parcela tiene una superficie de 0.121 ha y está situada a 3.58 km al sur de Muna. Pertenece a un agricultor que se ha incorporado al programa de Escuelas de Campo en los seis meses previos al inicio del presente estudio. Su parcela tiene las siguientes características: Realiza fumigación con el insecticida llamado Palgus®, que es un insecticida de amplio espectro para el manejo integrado de larvas, usado en gran variedad de tipos de cultivos. Utiliza también Gro Green, fertilizante en polvo que contiene nitrógeno, fósforo y potasio, complementado con micronutrientes (microelementos). Gro-Green ayuda al desarrollo y crecimiento de los cultivos. También aplica urea. Su sistema de riego es por canales y se realiza una vez por semana.

Parcela 3. Esta parcela pertenece a un agricultor que se ha incorporado al programa de las escuelas de campo un año atrás. Su parcela tiene una superficie de 0.114 ha y está situada a 9.48 km al sur de Muna. Expresa que su situación es precaria y le es difícil conseguir insumos para la mejora de su parcela. Tiene las características siguientes:

Posee trampas de confusión sexual por primera vez con el cultivo analizado. Reporta que utiliza los insecticidas Lorsban®, un insecticida de amplio espectro, particularmente efectivo en el control larvas de lepidópteros, en las etapas desde V1 a V6 y el insecticida Palgus® para etapas desde V6 hasta V15. Para combatir los pastos y controlar las malezas aplica dicloruro de paraquat, sustancia que degrada el tejido vegetal y causa la muerte de la planta. Se utiliza para eliminar vegetación secundaria y es preciso evitar el contacto con la planta de maíz. Se hace una recopilación de las características generales de las parcelas analizadas en Muna en 2023 y se describe en la Tabla 3.1, en la que se indica la fecha de visita, etapa fenológica V de las plantas y algunas de las técnicas aplicadas en cada parcela. Esa información puede ser de utilidad para consideraciones en el estudio de prácticas agroecológicas en posteriores estudios.

Sitio de muestreo	Fecha	Etapas maíz	Imágenes colectadas	Trampa de conf. sexual	Fertilizante	Pesticidas
Parcela 1	19/05/23	V5	58	Si	Si	Si
Parcela 2	19/05/23	V1	60	No	Si	Si
Parcela 2	08/06/23	V3 y V8	88	No	Si	Si
Parcela 3	14/09/23	V4	117	Si	Si	No
Parcela 1	14/09/23	V2 y V8	95	Si	Si	Si

Tabla 3.1. Fechas de recolección de imágenes y características de las parcelas visitadas en Muna 2023.

3.2. Kantunil

En octubre de 2024 se realizaron nuevos levantamientos de imágenes en otra zona de cultivo de la región, en el municipio de Kantunil, localizado a 75 km al este de la ciudad de Mérida, entre los paralelos 20° 44" y 20° 52" de latitud norte y los meridianos 88° 55" y 89° 06" de longitud oeste.

En Kantunil hay comunidades agrícolas que se organizan para preservar las diferentes razas de semilla criolla. En la comunidad se realizan esfuerzos por mantener bancos de germoplasma, donde conservan varias razas diferentes de maíces criollos, de los que actualmente solo se cultivan activamente tres.

En Kantunil hay una serie de parcelas familiares en las que se ha incorporado un sistema de riego que hace posible la siembra en diferentes épocas del año sin depender de la temporada de lluvias para contar con el agua para el riego. La Figura 3.4 muestra una imagen de la visita a las parcelas familiares de Kantunil.

Se ha logrado tener contacto cercano con esta comunidad debido a la participación en ferias de la semilla y otras actividades relacionadas con los sistemas agro productivos, por lo cual se aprovechó para adquirir imágenes para el presente estudio.



Figura 3.4 Parcela con monocultivo de maíz en Kantunil.

Capítulo 4. Metodología

Los apartados de 4.1 y 4.2 refiere a la metodología que se utilizó durante los levantamientos de las imágenes de las plantas de maíz en las parcelas y las consideraciones relevantes tomadas en cuenta. Las secciones 4.3 a 4.7 refieren a la metodología del procesamiento de los datos.

4.1. Levantamientos de 2023 en Muna

La recolección de imágenes de la zona de cultivo se llevó a cabo mediante el uso de un VANT marca DJI modelo Mavic s2, que está equipado de fábrica con una cámara RGB marca Hasselblad de 20 Mpíxeles (MP).

En los primeros cuatro muestreos en la región de Muna se capturaron al rededor de 100 imágenes por parcela, con una resolución de 5472 x 3078 píxeles, aplicando un traslapo de 30% entre imágenes. Las imágenes fueron capturadas a alturas de 6 y 7 metros sobre el suelo sobre parcelas que cumplían con que las plantas estuvieran en las etapas fenológicas apropiadas, desde V2 hasta V4, cuando son más susceptibles a daños importantes. A 6 m de altura, cada imagen cubrió una superficie de suelo de 12 m x 6.7 m, en las que pueden ser capturadas hasta aproximadamente 60 plantas por fotograma. Se ilustra un ejemplo de las imágenes recopiladas en la Figura 4.1.



Figura 4.1. Ejemplo de imagen de 5472 x 3078 px, capturada por medio del VANT a 6 m de altura.

Se realizaron capturas en diferentes parcelas, con el objetivo de que las imágenes capturadas captaran la variabilidad de contextos en los que se puede encontrar un cultivo de maíz en Yucatán, ya que en la región se suele combinar el cultivo de maíz con el de otras especies, además de que varía la cantidad de hierbas secundarias según el tipo de suelo o costumbres del agricultor. Algunos agricultores realizan monocultivos y otros sistemas de milpa, pero comparten espacios en regiones comunes.

Como resultado de los hallazgos obtenidos en estos primeros levantamientos, se determinó que para futuros mapeos sería fundamental incorporar tres consideraciones metodológicas más: (1) contar con un sistema de referencias en campo que facilitara la correlación espacial precisa, (2) una mayor variabilidad controlada en las alturas de vuelo para capturar diferentes escalas de detalle, (3) realizar un registro preciso a nivel de tierra de las plantas afectadas, con el registro de su posición y grado de afectación.

Estas mejoras respondían a la necesidad de obtener datos más finos para el análisis computacional, particularmente ante el desafío que representaba el pequeño tamaño de las afectaciones en las hojas y la sutileza de las diferencias entre plantas sanas y afectadas.

4.2. Levantamiento de imágenes 2024 en Kantunil

Se implementaron las nuevas consideraciones metodológicas mencionadas anteriormente en el levantamiento realizado en el municipio de Kantunil.

4.2.1 Marcas de referencias en campo

Se generaron marcas rectangulares cuadrículadas de 24 x 35 cm, de color rojo y blanco, para ser colocadas en el suelo de la parcela de estudio, de modo que resaltarán lo más posible desde las alturas. Se registraron las coordenadas UTM de los puntos de referencia colocados en el suelo utilizando un GPS submétrico (precisión ≤ 30 cm). Estas marcas cuadrículadas de rojo y blanco, serán la referencia sobre las cuales comenzar los vuelos a diferentes alturas.

Además de las cuadrículas de referencia, se colocó a lo largo de la parcela una cuerda de 20 m, con intercambios de color entre blanco y negro a cada metro, para servir como referencia métrica en las anotaciones de campo y en las imágenes capturadas.

4.2.2. Variabilidad controlada de altura

Se utilizó un método para generar una base de datos de imágenes a múltiples alturas, que permitiera analizar desde diferentes perspectivas durante los procesos de desarrollo y evaluación de los modelos computacionales realizados.

Mediante un VANT, se ejecutaron vuelos verticales sobre las cuadrículas rojiblancas georreferenciadas. La secuencia de captura sistemática incluyó fotogramas tomados a alturas incrementales de 1 a 10 metros, con intervalos de 1 metro.

El proceso fue repetido cada 5 metros longitudinales sobre la sección de estudio. En la Figura 4.2 se ilustra un ejemplo de un mismo punto en el suelo, visto desde 1, 3, 6 y 10 metros de altura. Ese tipo de captura a cada metro hizo posible verificar detalladamente el

estado de cada planta y posteriormente, probar el desempeño de los modelos, probando a diferentes alturas.

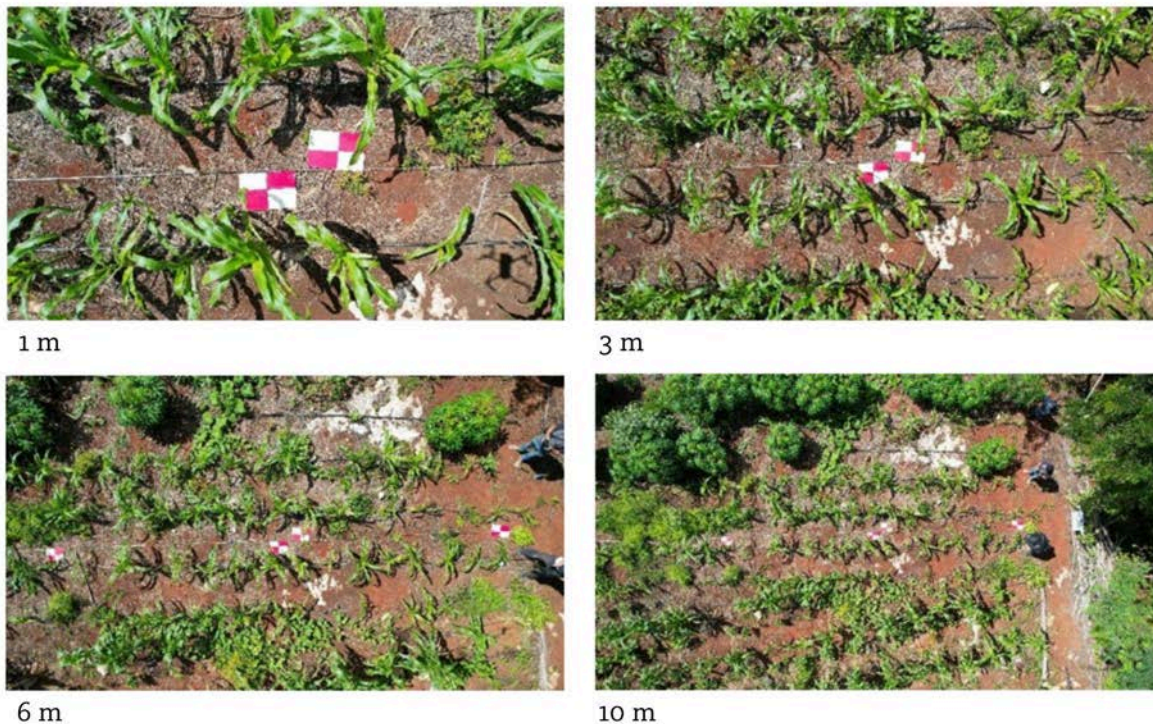


Figura 4.2. Perspectivas desde diferentes alturas (1, 3, 6 y 10 m.), para habilitar seguimiento de los detalles pequeños de cada planta.

4.2.3. Identificación y Registro manual de Plantas Afectadas

Posterior al levantamiento de imágenes, se procedió a la identificación de las plantas afectadas por *S. frugiperda* manualmente a lo largo de todas las filas del cultivo fotografiado. Se registró la localización de cada planta afectada con relación a las marcas de cuadrados rojo y blanco de referencia en tierra

Se asignó un nombre a cada fila desde A, hasta E y se anotó la fila y posición de cada planta afectada. Se evaluó el estado de la afectación utilizando una escala de calificación del 1 al 3, donde 1 indica una afectación leve, 2 una media y 3 una afectación severa. En la Figura 4.3 se ilustra un ejemplo de una planta clasificada en campo como con afectaciones de nivel 2, ya que presentaba pérdida de superficie foliar y restos en su superficie, pero aún no representaba una pérdida total de las hojas centrales.



Figura 4.3. Registro manual longitudinal al nivel de piso de las afectaciones por *S. frugiperda*.

Este registro permitió verificar individualmente cada planta y así conocer y registrar la distribución espacial de las plantas con infestaciones y su severidad, facilitando así el análisis posterior en las imágenes.

Se realizó posteriormente un marcado digital sobre una imagen aérea para revelar la distribución espacial de las plantas afectadas, obtenidas de la evaluación manual a nivel de suelo.

Esta imagen y las equivalentes del registro manual terrestre se usaron posteriormente para probar el desempeño del modelo a diferentes alturas, comparando la predicción realizada por este, con el registro obtenido de esta parcela de referencia. El resultado de una imagen se ilustra en la Figura 4.4, en la que las plantas registradas como afectadas levemente se representan en blanco y con el número 1, las de nivel 2 en naranja y las de daño máximo, de grado 3 con rojo.



Figura 4.4 Resultado de la distribución de plantas afectadas, evaluadas de forma manual en campo, como registro de referencia. Los colores representan el grado de afectación en las plantas de cada fila de la parcela.

4.3. Generación de bases de datos

El procesamiento de los datos comenzó con la generación de los conjuntos de imágenes apropiados para realizar cada etapa del proceso. Se consideran 3 principales conjuntos de imágenes. Uno, de las mejores imágenes de alta resolución, otro de plantas de maíz y el último, de plantas de maíz afectadas por el gusano.

Se utilizaron los conjuntos de imágenes para generar 3 diferentes modelos de segmentación, cada uno con un enfoque particular, con el objetivo de explorar los alcances del método YOLOv11 y conocer mejor su desempeño en diferentes tareas referentes a aplicaciones agrícolas.

Cada enfoque incluía una clase para plantas de maíz, la diferencia entre ellos fueron las otras clases, en la que el primer enfoque considera diferenciar las plantas de maiz de las hierbas secundarias, el segundo enfoque a dos diferentes tipos de afectaciones y el tercer enfoque usa una sola clase para determinar las hojas afectadas.

El flujo de los datos seguido para la generación de un modelo final se representa en la Figura 4.5, comenzando con la totalidad de imágenes capturadas, avanzando a la derecha en tres etapas. Se muestran las principales características de los conjuntos de imágenes y los modelos generados.

Se puede apreciar que se utilizó el Conjunto de imágenes 2 para generar el modelo P1 y el Conjunto 3 se utilizó para generar los modelos A1 y A2.

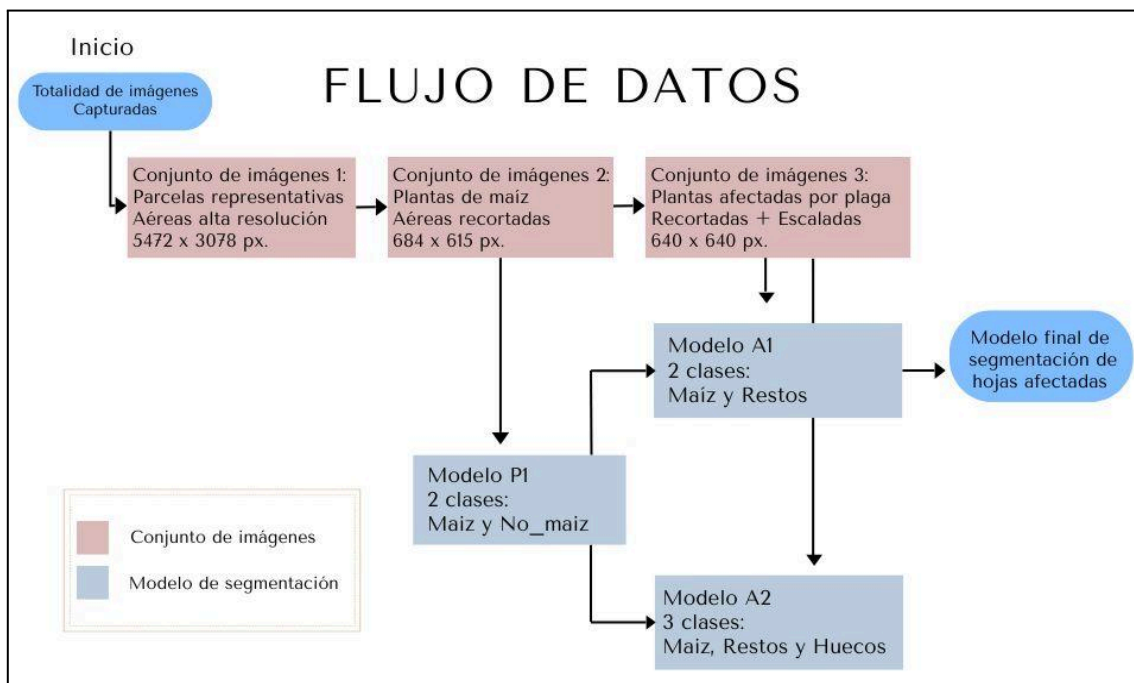


Figura 4.5. Diagrama de flujo de los datos, se ilustran los 3 principales conjuntos de imágenes arriba y los modelos que se generaron a partir de ellos abajo. El conjunto 3 se utilizó para el modelo A1 y A2.

4.3.1. Conjunto de imágenes 1. Parcelas representativas

El conjunto es una selección de las imágenes aéreas más apropiadas para el estudio. Una vez que se contaba con la totalidad de imágenes de alta resolución tomadas con el VANT sobre todas las parcelas y alturas, se seleccionaron las que correspondían a cultivos en las etapas fenológicas entre V2 y V4 (plantas de entre 2 y 4 hojas exteriores anilladas), que miden entre 20 y 60 cm de altura, ya que cuando las plantas llegan a aproximadamente V5 cambia la morfología de las plantas y dejan de ser tan susceptibles a la plaga.

Una vez seleccionadas las imágenes de plantas en etapas fenológicas apropiadas, se seleccionaron, de entre ellas, aquellas que tuvieran una gran densidad de plantas de maíz para ser las representativas de cada parcela y se descartaron las imágenes con zonas sin plantas de maíz.

Respecto a las alturas variables, se seleccionaron las imágenes capturadas a 6 y 7 metros de altura. Esas alturas permiten una resolución suficientemente clara de las zonas afectadas, con una resolución espacial aproximada de 0.219 cm/px, lo que significa que cada píxel de la imagen representará un cuadrado de aproximadamente 0.219 cm x 0.219 cm en el nivel de suelo, lo cual permite visualizar las zonas dañadas con una resolución apropiada.

4.3.2. Conjunto de imágenes 2. Plantas de maíz individuales

Se creó este conjunto de imágenes con el objetivo de generar con él el modelo de segmentación de instancias, que obtenga la vista superior de las plantas de maíz en estados fenológico entre V2 y V4.

Para generar este conjunto se subdividió cada imagen de alta resolución de 5472 x 3078, proveniente del Conjunto de imágenes 1, en 40 sub-imágenes más pequeñas, de modo que cada sección resultante fuera de un tamaño lo más cercano a 640 x 640, ya que es el tamaño al que se escalan las imágenes para su uso en el entrenamiento e inferencias con el modelo YOLOv11. Las medidas obtenidas al dividir en 40 sub-imágenes de alta resolución resultaron en 684 x 615 px, que es muy cercano al tamaño ideal y el efecto del escalamiento es imperceptible.

Después de dividir la imagen de alta resolución en 40 sub-imágenes de menor tamaño, muchas sub-imágenes quedan sin plantas de maíz o solo con trozos de ellas, por lo que se seleccionaron solo las imágenes en las que se visualizara la mayor parte de una planta de maíz. Un ejemplo de las imágenes de ese conjunto se ilustra en la Figura 4.6.



Figura 4.6. Imágenes de 684 x 615 px., representativas de la base de datos de imágenes de cultivos de maíz en Yucatán, que contuvieran al menos una planta de maíz completa.

Con estos criterios se generó una base de datos de 668 imágenes que contienen al menos una planta de maíz completa. La base de datos de imágenes cumple las características necesarias para el proceso de etiquetado de las imágenes que se describe en la siguiente sección.

4.3.3. Conjunto de imágenes 3. Plantas afectadas por plaga

Del conjunto de imágenes 2 se seleccionaron únicamente las imágenes de maíces que efectivamente contuviesen hojas afectadas por la plaga.

Ya que la proporción total de plantas afectadas era bajo en las muestras, se utilizó además los registros de las afectaciones recopilados en los levantamientos verticales para seleccionar las plantas que habían sido registradas como afectadas y se agregaron a la base de datos.

Para contar con la mayor cantidad posible de plantas afectadas con huecos y restos, se decidió incluir las imágenes afectadas del registro fotográfico tomadas a nivel de suelo. Se incluyeron las que coincidieran con una perspectiva y proporción similar a las obtenidas desde las imágenes aéreas. Para que fueran equivalentes a las imágenes aéreas, se transformaron a un tamaño de 640 x 640 px. Se reunieron 707 imágenes originales con plantas afectadas. Ejemplos de plantas en esta base de datos se pueden ver en la Figura 4.7, en donde se ha hecho un ligero acercamiento a la planta afectada, solamente para fines ilustrativos.



Figura 4.7. Imágenes ilustrativas de plantas en la base de datos de afectadas por plaga.

4.4. Preparación de los datos

4.4.1. Preprocesamiento con la plataforma Roboflow

Para la mayor parte del preprocesamiento se empleó la plataforma en línea Roboflow (Roboflow, 2024) con el objetivo de llevar a cabo escalamientos, segmentación y manejo de etiquetados de los objetos de interés. Esta plataforma proporciona funcionalidades significativas que optimizan el proceso de anotación. Específicamente, la capacidad de generar cuadros delimitadores permite al usuario definir regiones de interés, tras lo cual el software ejecuta un algoritmo de segmentación automática basado en el modelo de segmentación SAM (Kirillov et al., 2023) dentro de dichos límites que genera una máscara de segmentación aceptable del objeto de interés. Si bien los polígonos resultantes de esta segmentación automática requieren en general una revisión y modificación manual para asegurar una alta precisión, representan un avance considerable en la eficiencia del etiquetado, especialmente en conjuntos de datos de gran escala. Adicionalmente, la plataforma facilita la edición de las máscaras de segmentación mediante la adición o sustracción intuitiva de áreas. Un clic fuera de la máscara de segmentación, activa la inclusión de la región seleccionada, mientras que un clic dentro de la máscara permite la eliminación selectiva de zona que estaban incluidas. También es posible editar individualmente cada punto del polígono de la máscara de segmentación. Esta funcionalidad de ajuste manual resulta crucial para refinar la calidad de las anotaciones. Un ejemplo ilustrativo del resultado obtenido en las máscaras de segmentación generadas con la plataforma se presenta en la Figura 4.8, que contiene 3 imágenes resultantes del proceso de segmentación. Para mejor claridad, la clase *Maiz* segmentada ha sido delimitada por un contorno color amarillo y la clase *No_Maiz* está delimitada por un contorno morado. Las superficies no etiquetadas aparecen ligeramente opacadas.



Figura 4.8 Muestra de imágenes a las que se les ha generado sus máscaras de segmentación de clase, utilizando Roboflow. La clase Maíz en amarillo, la clase no maíz en morado, lo no seleccionado oscurecido.

4.4.2 Sustracción de fondo por espacio de color

Se generaron algunas imágenes especiales con el uso de Python y paqueterías como OpenCV, para ser usadas en los entrenamientos, y así facilitara la segmentación del contorno de las plantas y las partes delgadas de las hojas. Esto se realizó por medio de convertir las imágenes RGB a espacio de color en el que los valores representan color, saturación y opacidad HSV (hue, saturation and value, por sus siglas en inglés), espacio de color en el que los rangos de color se organizan en coordenadas radiales, lo que permite seleccionar por medio de umbrales de segmentación los valores de color de 0 a 179 y saturación y opacidad de 0 a 255. con la elección correcta de rangos de color, saturación y brillo, es posible segmentar solo las plantas de maíz. Mediante este procedimiento se removió el fondo y otras malezas.

Esta segmentación HSV es útil para la extracción efectiva de las características de la planta y es respaldada por trabajos que discuten diversas técnicas de sustracción de fondo y segmentación que permiten mejorar la detección de objetos en imágenes complejas (Romero et al., 2019) y pueden aumentar significativamente la precisión de los modelos de detección de objetos en aplicaciones agrícolas (Garcia, 2023).

El resultado de dicho procedimiento se muestra en la Figura 4.9. Las imágenes generadas mediante este procedimiento se agregaron al Conjunto de datos 2, para su uso en el modelo detector de plantas P1.

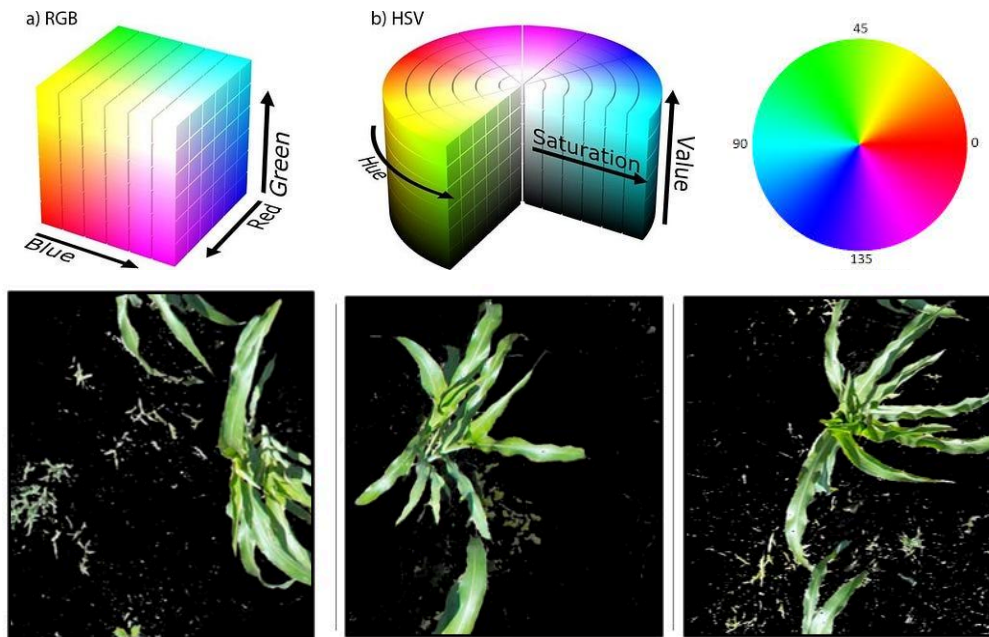


Figura 4.9. Espacios de color RGB y HSV arriba, y plantas de maíz segmentadas mediante técnica de corte de color en espacio HSV, abajo.

4.4.3. Aumento de datos etiquetados

Es preferible contar con más ejemplos de plantas en diferentes condiciones para mejorar los resultados del modelo, pero hay que considerar los tiempos que toma realizar los mapeos y etiquetados, verificar su precisión, además de la disponibilidad de los datos, por lo que un método común es realizar un aumento de datos mediante ligeras modificaciones de las imágenes originales, que aproveche la generación artificial de variaciones en las imágenes para tener un conjunto de datos más robusto, que contenga a las originales más las diferentes variaciones generadas.

Se realizó el proceso de aumento de datos, considerando las siguientes variaciones a las imágenes originales:

- Rotación de 90° , -90° y 180°
- Espejeo horizontal y vertical
- Variación en la saturación del color de $+30\%$ y -30%
- Variación del brillo $+15\%$ y -15%

4.4.4. Inclusión de imágenes a nivel de tierra

Se incluyeron también en la base de datos de afectaciones, las imágenes provenientes de muestras a nivel de tierra que contuviesen plantas afectadas y que coincidieran con el encuadre de los segmentos obtenidos de las fotografías aéreas. Estas fueron también escaladas a 640×640 px.

En todas las imágenes se fijó una orientación para relacionarlas con la orientación de las máscaras de segmentación que se generarán posteriormente.

4.5 Generación de modelos de segmentación

El desarrollo de modelos de visión computacional constituye un proceso iterativo que exige la evaluación sistemática de múltiples parámetros operativos, como puede ser el tamaño de la imagen, la distancia al objeto observado o la resolución de la cámara. Es necesario también analizar los resultados cuantitativos y cualitativos para determinar qué acciones se tomarán con la finalidad de optimizar el modelo.

Debido a que es posible realizar una adaptación de un sistema de detección y segmentación a diferentes tipos de cultivos, se desarrollarán brevemente los resultados de los pasos preliminares P1 y A2, a manera de documentación del proceso de generación de un modelo para un cultivo específico y sus ajustes. Se comentarán los resultados obtenidos y las consideraciones que se tomaron para alcanzar el modelo optimizado A1.

El entrenamiento de los modelos se llevó a cabo en el entorno Google Colab, una plataforma para ejecutar código basada en la web que ofrece acceso gratuito a recursos computacionales, ideal para experimentación ágil y colaborativa. Esta elección permitió aprovechar las GPUs Tesla T4, cuyas capacidades de aceleración mediante CUDA (Compute Unified Device Architecture) y 16 GB de memoria para el entrenamiento de redes neuronales profundas. La GPU T4, diseñada para cargas de trabajo de inferencia y entrenamiento moderado, es adecuada para entrenamiento de modelos tipo YOLOv11.

4.5.1. Modelo P1 (Maíz y Plantas secundarias)

Se desarrolló inicialmente el modelo de segmentación de instancias, orientado a identificar a las plantas de maíz y toda la maleza y plantas secundarias dentro de una sola clase de plantas secundarias (clase *No_Maiz*), para este trabajo se denominó Modelo P1. Para generar este modelo se utilizó el Conjunto de imágenes 2, de plantas de maíz. El objetivo principal fue probar las capacidades del modelo YOLOv11 para generar un modelo segmentador de plantas que facilitara el posterior etiquetado de la base de datos de las afectaciones por plagas y también probar la capacidad de distinguir hierbas secundarias, que también es información útil para la agricultura, ya que en monocultivos se intenta eliminar toda la maleza secundaria, por lo que la información puede servir para optimizar esos procesos.

Clase *Maiz*.

Para la segmentación de plantas de maíz se siguieron los siguientes criterios.

- Las plantas de maíz se suelen sembrar en pares, pero puede haber 3 plantas o plantas separadas, por lo que se consideró para la segmentación a cada conjunto como un solo objeto, a excepción de cuando la separación entre plantas dentro del grupo fuese evidente.

-
- Se procuró incluir a todas las hojas de la copa. En perspectivas laterales no se incluyó el tronco de la planta ni las primeras hojas en la máscara de segmentación.
 - Se ajustaron las máscaras de segmentación para evitar incluir en el área a los espacios internos entre hojas que capturasen partes del fondo. Se encaminó a la máscara de segmentación a lo largo de las hojas para romper los espacios internos que incluyeran áreas que no pertenecieran a la planta.
 - No se etiquetaron pedazos de hojas aisladas. Debido a que la base de datos contenía imágenes con al menos un conjunto completo de plantas, no se etiquetaron como maíz a las hojas sueltas o incompletas debido al corte de la cuadrícula de la imagen de alta definición, con el objetivo de que durante el entrenamiento, el modelo lograra ser capaz de diferenciar las plantas completas y contarlas como objetos enteros.

Clase *No_Maiz*

Esta clase consistía en todas las hierbas secundarias que no fuesen maíz, su segmentación se realizó por zonas generales debido a la dificultad de segmentar todas las hierbas secundarias de manera muy precisa. Posteriormente, se determinó eliminar esa clase del modelo final, debido a que no es muy relevante la segmentación de hierbas secundarias para cumplir con los objetivos específicos del presente trabajo, además de que mantener la clase activa implica costo computacional y temporal adicional. Sin embargo, se presentan los resultados alcanzados en esa clase debido a que se obtuvieron resultados interesantes, que tienen valor potencial en proyectos agrícolas.

4.5.2. Modelo A2 (Afectaciones de 2 clases)

Posteriormente, se utilizó al Modelo que diferencia entre tipos de plantas P1 como herramienta para realizar el etiquetado durante el desarrollo de un nuevo modelo, esta vez enfocado en las pequeñas afectaciones causadas por la plaga de *S. frugiperda*. El nuevo modelo consta de la clase *Maiz* y 2 clases de afectaciones: *Huecos* y *Restos*. Este modelo se denominó A2

Para generar el Modelo A2, detector de plantas y lesiones, se utilizó al Conjunto de imágenes 3 de plantas afectadas, obtenida de dividir las imágenes aéreas. Se verificaron las anotaciones de la clase maíz y posteriormente se agregaron y anotaron las dos nuevas clases (Huecos y Restos) las cuales son efectos visibles de afectaciones en las plantas, indicadores de presencia de la plaga de *S. frugiperda*.

En ese punto, el conjunto de datos contenía 278 imágenes que presentaban al menos una planta o grupo de plantas completo dentro de la imagen.

Una vez que se contó con la base de datos de plantas de maíz segmentadas y etiquetadas ya con las nuevas clases, se realizó una distribución aleatoria de las imágenes en tres grupos. Uno con el 70% de las imágenes para el proceso de entrenamiento, resultando en

195 imágenes, otro con el 10%, equivalente a 28 imágenes de prueba y otro con el 20%, equivalente a 55 imágenes para el proceso de validación.

Las 195 imágenes de entrenamiento se aumentaron a 585 imágenes mediante las técnicas de aumento de datos presentadas en la subsección 4.4.3.

Segmentación de Afectaciones de 2 clases

Se realizó un etiquetado y segmentación manual de las imágenes del dataset de plantas afectadas por plaga, en las que se presentasen evidencia de los restos que genera *S. frugiperda* cuando se alimenta de las hojas de la planta de maíz.

La clase *Restos* corresponde a manchas que quedan en la hoja cuando esta es comida por el gusano y son pequeños pedazos de apariencia blanca alrededor del centro de la planta. En otras ocasiones puede ser café, correspondiente a restos biológicos del gusano sobre las hojas. Ambos tipos de residuos se clasificaron dentro de la clase *Restos*.

La clase *Huecos* corresponde a hojas que han tenido un suficiente grado de afectación, tal que las mordidas han provocado huecos o cortes, es un indicador de daño físico avanzado en la planta. Es algo que puede no apreciarse muy bien en las imágenes, ya que si el hueco ha causado la pérdida parcial de la hoja, no es posible distinguir un hueco, sino los bordes mordidos de la hoja.

Se consideró que esas dos clases de afectaciones serían indicadores suficientes para detectar afectaciones con restos físicos causadas por *S. frugiperda* en plantas de maíz. En la Figura 4.10 se puede visualizar un ejemplo con tres imágenes, en las cuales se han segmentado las tres clases, *Maiz* en amarillo, *Huecos* en rojo y *Restos* en morado.



Figura 4.10 Segmentación de plantas y de hojas afectadas, en rojo las hojas con huecos y en morado las hojas con restos.

4.5.3 Modelo A1 (Afectaciones de 1 clase)

Para este Modelo A1 se cambió el enfoque y en lugar de tratar de diferenciar si la afectación es de tipo hueco o restos, se buscó detectar correctamente la mayor cantidad de hojas afectadas, sin distinción del tipo específico de afectación, ya que se detectó que no es posible diferenciar entre dos clases, si las dos afectaciones ocurren en la misma hoja. Se determinó unificar las clases *Restos* y *Huecos* en una sola clase, que representa a todas las hojas afectadas. Por lo que el modelo conserva una clase para las plantas de maíz completas y otra para las hojas afectadas por *S. frugiperda*. Para generar esa clase se fusionaron todos los polígonos de segmentación de *Restos* y *Huecos* en el set de entrenamiento y se les dio una nueva etiqueta única de “*Afectadas*”, con la cual se realizó el entrenamiento del nuevo modelo para un solo tipo de afectaciones, el cual se denominó A1.

4.6 Implementación en Dispositivo Autónomo

Para la fase de despliegue en entornos agrícolas, se exploró con el uso del el modelo A1 de segmentación en un dispositivo de computación perimetral: una Raspberry Pi 4 con 4 GB de RAM. Esta selección respondió a requisitos críticos de portabilidad, bajo consumo energético y autonomía operativa en campo. Se estableció una conexión remota mediante una plataforma web (Raspberry Pi Connect), que permitió el acceso al escritorio Linux para configurar y supervisar el uso del sistema sin necesidad de interacción física directa.

Durante la configuración del entorno, se generó y activó un entorno virtual Python para garantizar el aislamiento de dependencias y la reproducibilidad del flujo de trabajo. Posteriormente, se instaló la librería Ultralytics NCNN (Jocher & Qiu, 2024), esencial para convertir el modelo entrenado (en formato PyTorch .pt) al formato optimizado NCNN. Esta conversión fue fundamental para maximizar la eficiencia en la arquitectura ARM del procesador Broadcom BCM2711 de la Raspberry Pi, dado que NCNN está específicamente diseñado para acelerar inferencias en dispositivos embebidos mediante operaciones vectorizadas.

La transformación del modelo se ejecutó mediante un comando bash. La exportación preservó todas las capacidades de segmentación del modelo original mientras adaptó su estructura computacional a las restricciones hardware del dispositivo.

Finalmente, se ejecutó el modelo en la computadora Raspberry Pi mediante un script Python, se procesaron los fotogramas capturados por el módulo de videocámara de 8Mp de Raspberry Pi.

Capítulo 5. Resultados y discusión

5.1. Descripción General de los Resultados

En esta sección se presentan los mejores resultados obtenidos con cada aproximación, esto es, la aproximación del Modelo P1 con sus 2 clases de plantas, la del Modelo A2 con la clase de maíz y dos clases de afectaciones y la del modelo A1, de maíz y una sola clase para las afectaciones. Cada aproximación tuvo resultados interesantes. De la aproximación con el modelo P1 se resaltan solo las más sobresalientes, ya que fue un paso intermedio, pero reveló resultados interesantes. Del enfoque con el modelo A2 se hace un análisis más completo para poder ser comparado con la A1. La aproximación con el modelo A1 fue la definitiva y es la que se presenta más detallada. La métrica que se busca optimizar es la del mAP50 (mean Average Precision at 50% IoU) para máscaras de segmentación. Se hace la comparación del mAP50 de segmentación, con el mAP50 de la caja de detección, ya que si se logra generar cajas de detección precisas, también es un buen acierto.

Como herramienta para evaluar el desempeño de los modelos A1 y A2, se utilizó la matriz de confusión absoluta y su versión normalizada por predicciones, en donde cada columna suma 100%. Para el análisis, el fondo (background) se incluyó como clase predicha, con el fin de cuantificar errores (objetos no detectados o predicciones en zonas vacías), pero no como clase real de objetos anotados manualmente.

5.2 Resultados Modelo P1 (Maíz y Plantas secundarias)

Las métricas globales del modelo P1 de solo plantas, evaluadas en segmentación de instancias, en el conjunto de validación durante el entrenamiento, reflejan que se alcanzó un rendimiento general de 59% en mAP50, con una precisión de 69.0% y una exhaustividad (recall) de 59.7%.

Analizando los resultados de mAP50 por categoría, esta vez en el conjunto de test, con imágenes que nunca “vio” el modelo durante el entrenamiento, se puede apreciar que fue capaz de identificar correctamente las plantas de maíz con un mAP50 de 75, con una precisión moderada en la clase No-maíz de 36% de mAP50. Los resultados se reúnen en la Tabla 5.1, resaltando en gris los valores de la métrica mAP50 por clase del conjunto de Test para las máscaras de segmentación, que es la métrica que se buscó optimizar.

mAP50	Set Validación	Set Test
General	59%	56%
Maiz	81%	75%
No_maiz	36%	36%

Tabla 5.1 Resultados de la métrica mAP50 por clase de modelo P1, en gris los resultados por clase sobre el set de prueba.

Resultados del modelo P1:

Clase Maiz: Una puntuación de 75% de mAP50 en el conjunto de prueba, indica una capacidad razonable para un modelo de segmentación de plantas de maíz en condiciones y entornos diversos.

Clase No_maiz: La segmentación de vegetación ajena al maíz presentó limitaciones, alcanzando un 36% de mAP50. Este bajo rendimiento se atribuye a la heterogeneidad morfológica y a la complejidad de los patrones visuales en esta clase, que dificultan su diferenciación precisa frente al entorno, además de que el etiquetado manual de esa clase no fue realizado de manera muy exhaustiva.

En la Figura 5.1 se visualizan 3 imágenes con segmentaciones realizadas por el modelo P1, en la plataforma de Roboflow, se aprecia con máscara amarilla las hojas de maíz y con máscara morada, la segmentación de hierbas secundarias.



Figura 5.1. Ejemplo del resultado de la segmentación realizada por el Modelo P1 de segmentación de plantas, en imagen de 640 x 640.

Con base en las observaciones de los resultados, se determinó eliminar la clase *No_maiz* del modelo final. La medida se tomó con la finalidad de optimizar recursos computacionales, enfocándose en mejorar la segmentación de plantas de maíz y los daños por plagas, ya que identificar todas las hierbas secundarias en la clase *No_maiz* no es necesaria para la tarea principal de identificación de afectaciones por plaga.

Sin embargo, se presentan los resultados obtenidos, ya que muestran un resultado aceptable en la segmentación de hierbas secundarias en parcelas con sistemas de monocultivo, que tienen menor proporción de estas y son más fáciles de identificar.

Conocer la cantidad y ubicación de la hierba secundaria representa también información relevante para la agricultura, ya que se intenta evitar la maleza secundaria con insumos agrícolas, por lo que conocer su cantidad y distribución puede ser tomada en cuenta en trabajos posteriores.

Los resultados de la segmentación de las plantas de la clase *Maiz*, es aceptable y se concluye que puede ser usado como herramienta de segmentación posterior.

Se determinó que en las siguientes iteraciones se etiquetarán también las puntas aisladas incompletas de hojas de maíz truncadas, lo que no se hizo en este modelo y causó pérdida de las puntas en la máscara de segmentación de algunas plantas. Un ejemplo de una imagen de alta resolución tomada a 6 m de altura y procesada con el método SAHI se puede apreciar en la Figura 5.2, en la que se puede apreciar las zonas con hierba secundaria en rojo, segmentadas de una manera aceptable para análisis agrícolas. También se aprecia una correcta segmentación de los maíces en morado.

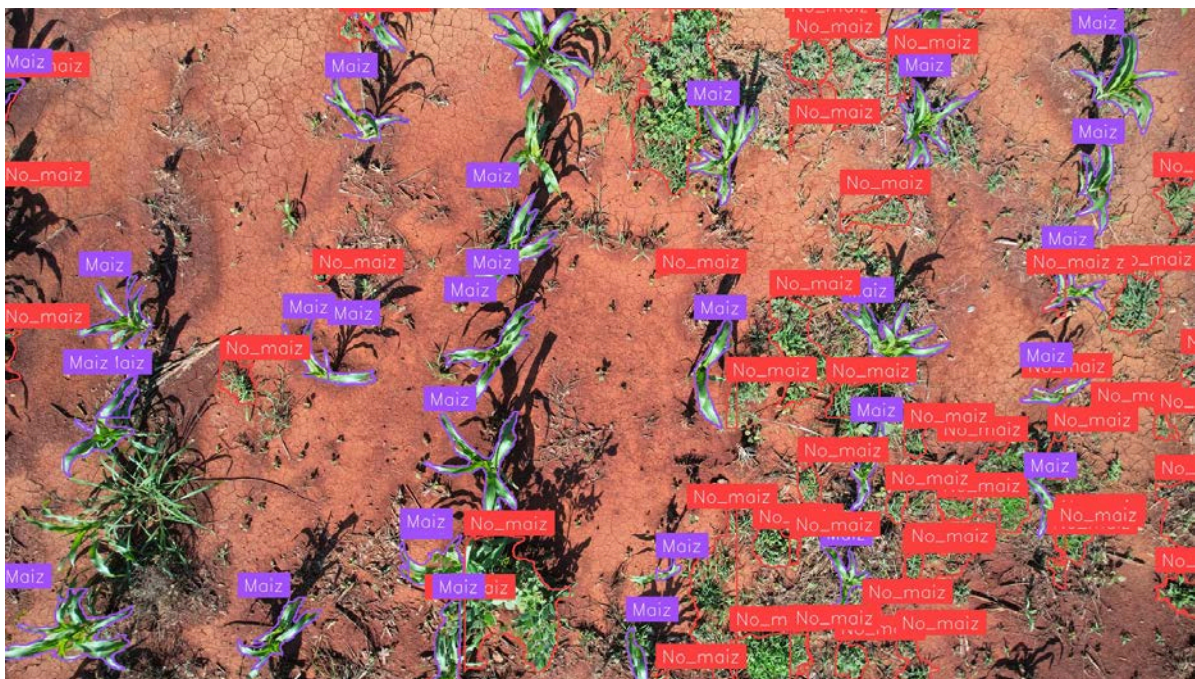


Figura 5.2. Imagen de alta resolución a la que se implementó el modelo P1 de detección de maíz y de hierbas secundarias.

5.3. Resultados el Modelo A2

Cajas de detección y máscaras de segmentación de A2

El enfoque del Modelo A2, se identificó nuevamente una clase para plantas de maíz y se destinaron 2 clases para las afectaciones, mostró un desempeño variable en las

predicciones entre las clases analizadas. La clase "Maíz" obtuvo un rendimiento destacado, con un mAP50 de 89.8% en cajas de detección y de 71.7% para máscaras de segmentación (en gris), lo que refleja una precisión aceptable en la identificación de estructuras vegetales de las plantas de maíz.

Por el contrario, las clases asociadas a lesiones, "Huecos" y "Restos", presentaron resultados limitados, con valores de mAP50 de 21.1% y 31.4% respectivamente para cajas de detección. Para máscaras de segmentación, la puntuación de mAP50 fue de 21.1% para Huecos y de 29% para Restos (marcados en gris).

Los resultados de mAP50 y mAP50-95 de cajas de detección y máscaras de segmentación se puede apreciar en la Tabla 5.2. Se resalta en gris la métrica principal de referencia, el mAP50 de máscaras de segmentación, como referencia principal del desempeño del modelo A2.

Clase	Caja mAP50	mAP50-95 Caja	mAP50 Mask	mAP50-95 (M)
Global	47.4%	32.5%	40.6%	20.1%
Maíz	89.8%	67.3%	71.7%	33.4%
Huecos	21.1%	10.9%	21.1%	9.93%
Restos	31.4%	19.2%	29.0%	17.0%

Tabla 5.2. Resultados del análisis de las imágenes provenientes del conjunto de Test en cuanto a máscaras de segmentación, obtenidos por el modelo A2. En gris los resultados de la métrica mAP50 de máscaras de cada clase.

Esta discrepancia en los resultados podría atribuirse a factores como el desbalance en el conjunto de datos, con menor representación de ejemplos de lesiones, la complejidad morfológica de las anomalías causadas por la plaga (tamaño reducido, formas irregulares o similitud con el fondo), y la separación de los objetos al procesar imágenes divididas, que afecta la recuperación de objetos pequeños.

Para visualizar cualitativamente cómo se puede mejorar el desempeño del modelo A2, se revisaron los resultados de los batches procesados para seleccionar la mayor cantidad de ejemplos de fallas del modelo. En la Figura 5.3 se visualizan 6 de las 16 imágenes de un batch de evaluación. Se muestran únicamente 6 imágenes del batch para ilustrar algunos ejemplos sobresalientes de diferencias entre el etiquetado y el resultado de la predicción, a un tamaño apropiado. En el lado izquierdo se pueden visualizar las imágenes con etiquetado manual y del lado derecho, las mismas imágenes, pero esta vez etiquetas automáticamente por el modelo A2 en la validación.

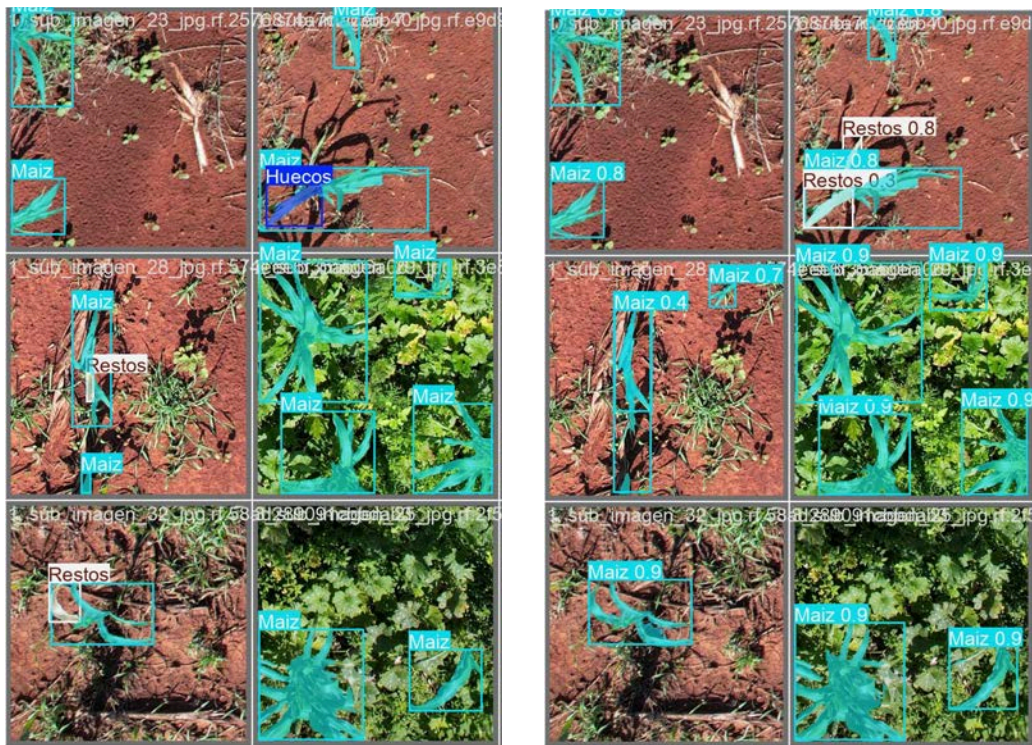


Figura 5.3. Comparativo de seis imágenes de un batch del etiquetado manual de validación (izquierda) contra las mismas imágenes, pero predichas por el modelo A2 (derecha), que indica el nivel de confianza de cada instancia.

La localización, tamaño y patrón de afectación de hojas clasificadas como *Restos* mostraron demasiada similitud con las hojas que fueron etiquetadas como de la clase *Huecos*.

Se identificaron casos donde la hoja detectada difería de la etiqueta asignada, o donde la parte afectada específica no coincidía. Esto sugiere que las diferencias morfológicas sutiles inherentes a estas clases, y la complejidad perceptual para distinguirlas consistentemente, pueden introducir ruido tanto en el etiquetado humano como en la clasificación automática. En la Figura 5.4 se ilustra un caso en el que se detectaron los errores mencionados. Arriba se encuentran los etiquetados manuales y abajo los realizados por el modelo A2. Se puede apreciar que en el etiquetado realizado por el modelo, aparece el nivel de confianza de la detección en la etiqueta de clase. En las imágenes seleccionadas se aprecia qué plantas afectadas pueden ser calificadas como afectadas, pero en hojas diferentes o con la clase equivocada.

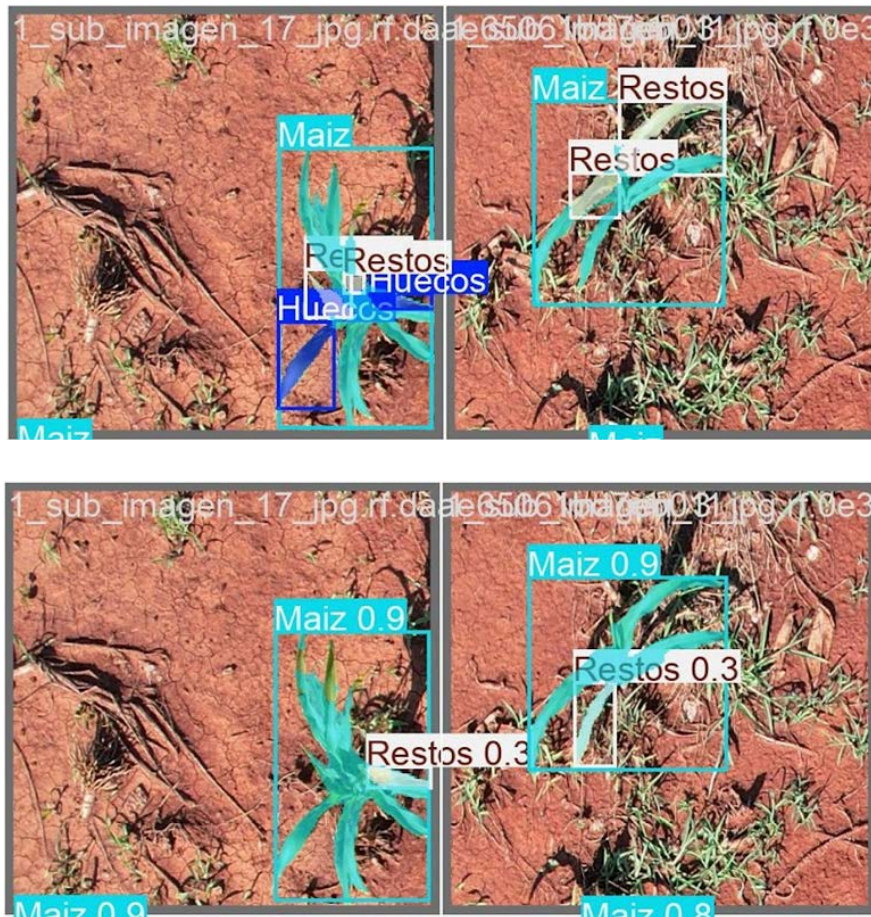


Figura 5.4. Comparativo entre 2 imágenes del batch de evaluación arriba y las mismas 2 imágenes, con la segmentación automática realizada por el modelo A2 abajo.

Se generó la matriz de confusión y la matriz de confusión normalizada del modelo A2, en donde se puede analizar la distribución de los aciertos y fallos del modelo (Figura 5.5). Cada columna representa la totalidad de objetos de cada clase etiquetados manualmente, distribuidos por las clases en la que los clasificó el modelo A2, por lo que la suma de cada columna es igual a 1. Se puede constatar cómo solo la clase Maíz obtiene buenos resultados. Se incluyen en la matriz de confusión los pixeles de fondo, que no fueron clasificados específicamente, aparecen como “Sin Clase” para fines de tener un análisis más completo.

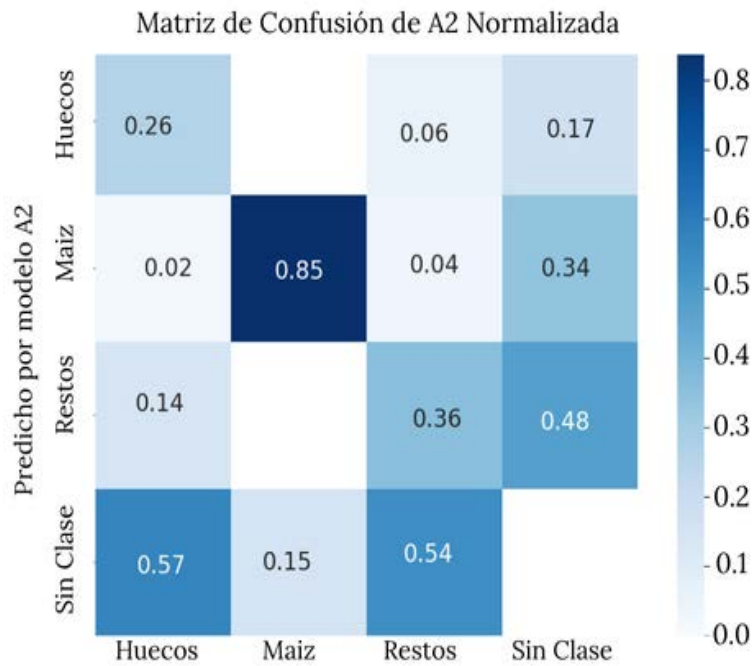


Figura 5.5. Matriz de confusión absoluta (arriba) y normalizada (abajo) del modelo A2, con el etiquetado manual en el eje horizontal y en el vertical los generados por el modelo A2.

El gráfico de las puntuaciones F1, contra confianza del modelo A2 para máscaras de segmentación, se puede observar en la Figura 5.6, donde se aprecia la curva de la clase maíz en naranja, las clases *Restos* y *Huecos* en verde y azul y el promedio del modelo A2 en azul más grueso. Este enfoque no logró hacer una correcta distinción entre los tipos de afectación presente en las hojas afectadas. En el gráfico se puede apreciar que el sistema tiene su mejor cantidad de aciertos en distinguir entre el tipo de afectación en torno al 40%, pero no son puntuaciones suficientemente acertadas.

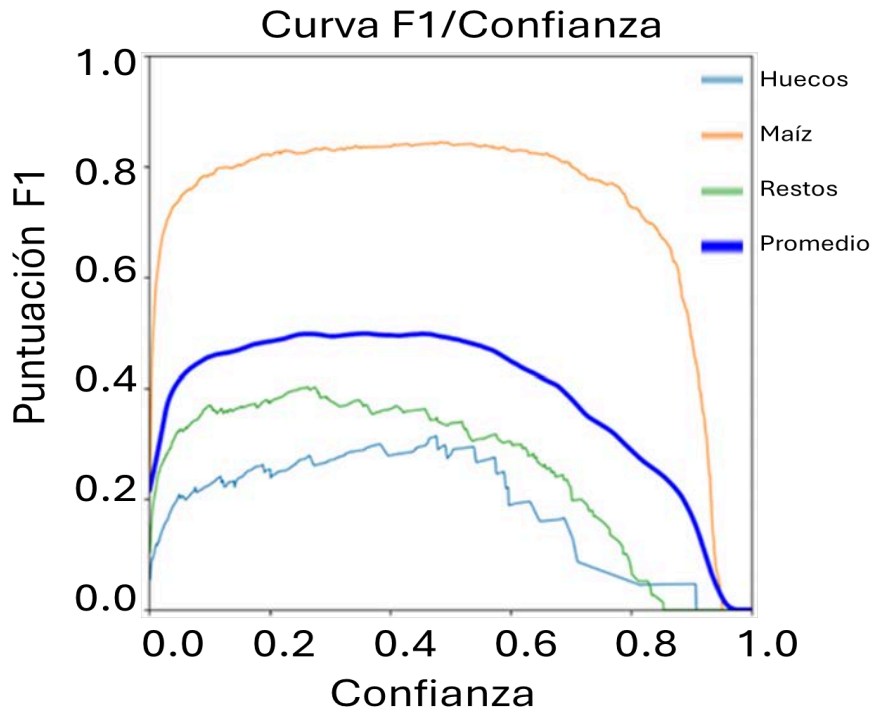


Figura 5.6. Gráfico de puntuación F1 del modelo A2 para cada nivel de corte de confianza por cada clase. En azul grueso se resalta el resultado promedio del modelo.

Para un refinamiento que mantuviera dos clases de afectaciones, sería necesario incrementar muestras de *Huecos*, ya que fue una clase que apareció alrededor de un cuarto de veces con respecto a la clase *Restos*.

Una vez que se constataron las limitaciones para poder distinguir entre hojas afectadas por diferentes características, se determinó utilizar solamente una clase para las hojas afectadas, sin distinción del tipo de afectación, ya que, ambas clases indican presencia de plaga. Ese fue el enfoque para el modelo A1.

5.4. Resultados del Modelo A1

Para el entrenamiento, se tomó como punto de partida la arquitectura y los pesos del modelo YOLOv11n-seg, que fue entrenado durante 250 épocas en 3 horas, utilizando una GPU Tesla T4. La arquitectura, compuesta por 203 capas convolucionales y 2.842.998 parámetros, demostró una eficiencia computacional notable durante la inferencia, procesando cada imagen en un promedio de 10.4 ms. Este tiempo se desglosa en 0.3 ms para preprocesamiento, 3.9 ms para la inferencia central y 6.1 ms para el postprocesamiento, evidenciando su potencial para aplicaciones en tiempo real.

El nombre de las clases se redujo en la visualización de las imágenes, con la finalidad de poder desplegar el nombre de la clase en la imagen sin ocupar mucho espacio y que

aparezca como un pequeño indicador. Las plantas de maíz completas y saludables, que se etiquetaron como "Ok" . Por su parte, la clase de las hojas dañadas por la plaga se etiquetaron como "X".

El conjunto de entrenamiento constó de 1,871 imágenes de 640x640 píxeles, las cuales contenían 3,267 instancias de plantas de maíz sanas de y 2,920 instancias de hojas con daños. La distribución fue de 643 imágenes de entrenamiento (aumentadas a 1,533), 152 de validación y 199 de prueba.

Cajas de detección

En la Figura 5.7 se muestran los resultados de las métricas de precisión, exhaustividad y mAP50 para las cajas de detección durante las 250 épocas de entrenamiento. Se puede apreciar que se encontraban cercanos a un punto de convergencia.

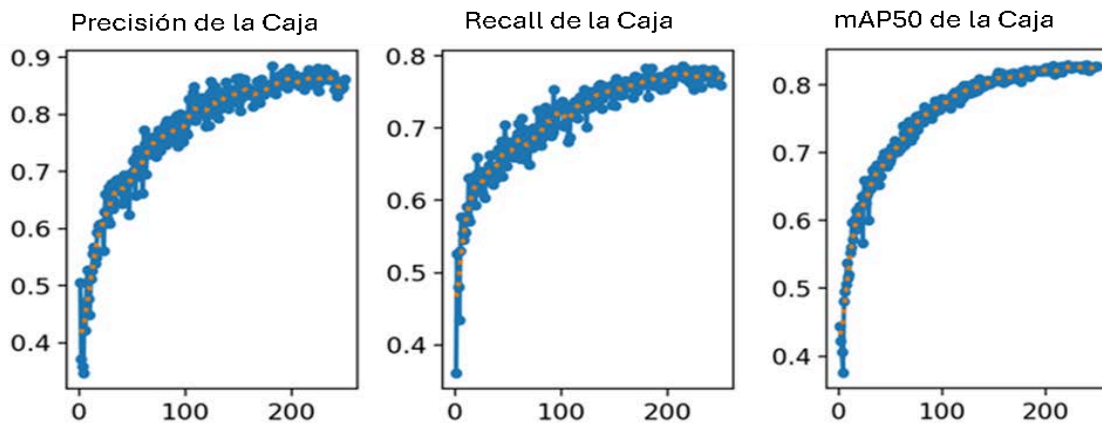


Figura 5.7. Se presentan los resultados de las mejoras en cada una de las 250 épocas del entrenamiento del modelo A1, en las métricas de precisión, exhaustividad y de mAP50 referente a las cajas delimitadoras de detección (Bouding box) del modelo A1.

En la Tabla 5.3 se muestran los resultados desglosados por clase de las métricas obtenidas del set de validación, de 199 imágenes para la Precisión, Exhaustividad, mAP50 y mAP50-95 para las cajas de detección. En la Tabla se resalta en gris los resultados de mAP50 de cada clase del modelo A1, ya que es la métrica de referencia a optimizar.

Clase	Imágenes	Instancias	Prec. B	Recall Box	mAP50	mAP50-95
Gral. A1	199	979	0.86	0.775	0.83	0.685
Maíz	199	435	0.929	0.9	0.942	0.814
Afectada	199	544	0.792	0.65	0.717	0.556

Tabla 5.3. Presenta los resultados del análisis de las imágenes provenientes del conjunto de Test en cuanto a cajas de detección (Bounding box). En gris los resultados del modelo A1 de la métrica mAP50 por cada clase.

Los resultados por clase para las cajas de detección muestran que este modelo ha alcanzado un mAP50 muy alto, de 0.943 en la tarea de detección de plantas de maíz y a pesar de todas las complejidades morfológicas de las hojas afectadas, estas lograron un mAP50 de 0.717. Es un resultado que se puede considerar bueno para un entorno agrícola complejo.

Máscaras de segmentación

Los resultados de las métricas de precisión, exhaustividad y mAP50 para las máscaras de segmentación con el modelo A1 se muestran en la Figura 5.8. Se puede apreciar que estaban también cerca de un punto de convergencia.

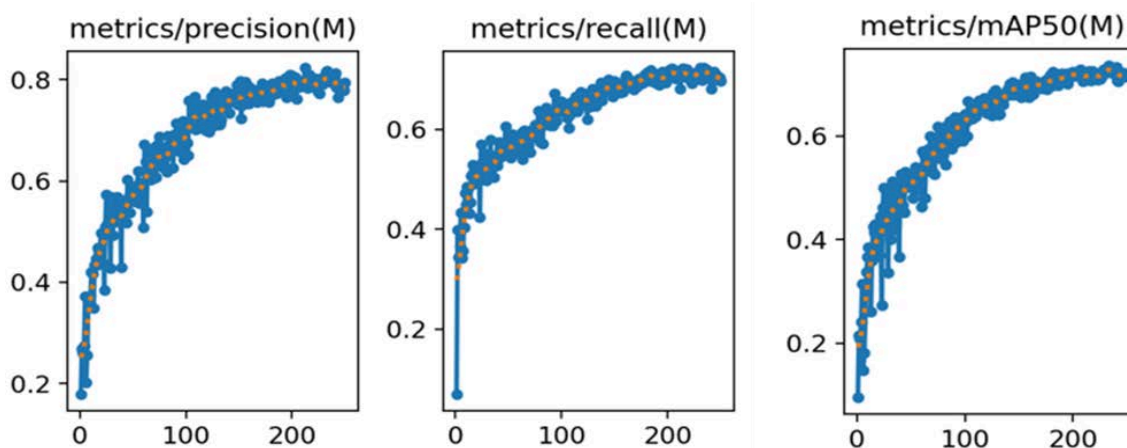


Figura 5.8. Presenta los resultados de las mejoras en cada una de las 250 épocas del entrenamiento del modelo A1, en las métricas de precisión, exhaustividad y de mAP50 referente a las máscaras de segmentación del modelo A1.

Los resultados obtenidos de las métricas para las máscaras de segmentación por clase se pueden apreciar en la Tabla 5.4, en la que se resalta en gris la métrica principal de mAP50, en este caso, para las máscaras de segmentación.

Clase	Imágenes	Instancias	Prec. Masc.	Rec.Masc	Masc. (mAP50)	M (mAP50-95)
Gral. A1	199	979	0.804	0.714	0.733	0.436
Maíz	199	435	0.828	0.784	0.777	0.391
Afectada	199	544	0.781	0.643	0.70	0.481

Tabla 5.4 Presenta los resultados del análisis de las imágenes provenientes del conjunto de Test en cuanto a máscaras de segmentación, por el modelo A1. En gris los resultados de la métrica mAP50 por cada clase.

El análisis de la exhaustividad en la tarea de segmentación reflejó una disparidad equivalente entre clases. La exhaustividad general para las máscaras de segmentación fue del 71.4%, pero con una marcada diferencia, con 78.4% para plantas de maíz completas (etiquetadas como "Ok") frente a 64.3% para hojas dañadas (etiquetadas como "X"). Esta brecha de 14 puntos porcentuales se puede deber a que es más sencillo para el modelo A1 detectar y segmentar la geometría de las plantas completas, que las áreas específicas de daño foliar, donde la delimitación precisa de las pequeñas y diferentes zonas afectadas resulta más compleja.

5.4.1. Evaluación en conjunto de Test

En la evaluación sobre el set de Test (199 imágenes con 979 instancias detectadas), el modelo A1 mostró un desempeño diferenciado entre las dos clases objetivo.

Plantas Completas

Para la detección de plantas completas (Anotada en las imágenes como "Ok" con, 435 instancias en el conjunto de Test), los resultados fueron alentadores, un recall del 78.4% confirma una capacidad apropiada para identificar la mayoría de las plantas presentes, respaldada por una precisión del 82.8% que indica un mínimo de falsos positivos. La métrica de precisión media en localización (mAP50) de plantas completas alcanzó un 77.7%, subrayando la capacidad del modelo de localizar plantas de maíz enteras con precisión espacial en todos los entornos probados.

Hojas Afectadas

Por contraste, la detección de hojas con daños por plagas (Anotada como "X" en las imágenes, con 544 instancias) presentó valores más moderados. Si bien la precisión se mantuvo en un nivel aceptable de 78.1%, asegurando que la mayoría de las detecciones positivas fueran correctas, el recall del 64.3% revela una limitación, aproximadamente un tercio de las hojas afectadas en el set de validación no fueron identificadas.

En la Figura 5.9 se seleccionaron 6 imágenes de un batch etiquetado manual (a la izquierda) usado durante el entrenamiento para la validación del modelo A1 y las mismas imágenes, pero con el etiquetado realizado automáticamente (a la derecha) por el modelo A1 en su proceso de validación, anotando también el nivel de confianza asignado.

Etiquetado manual

Etiquetado por modelo A1

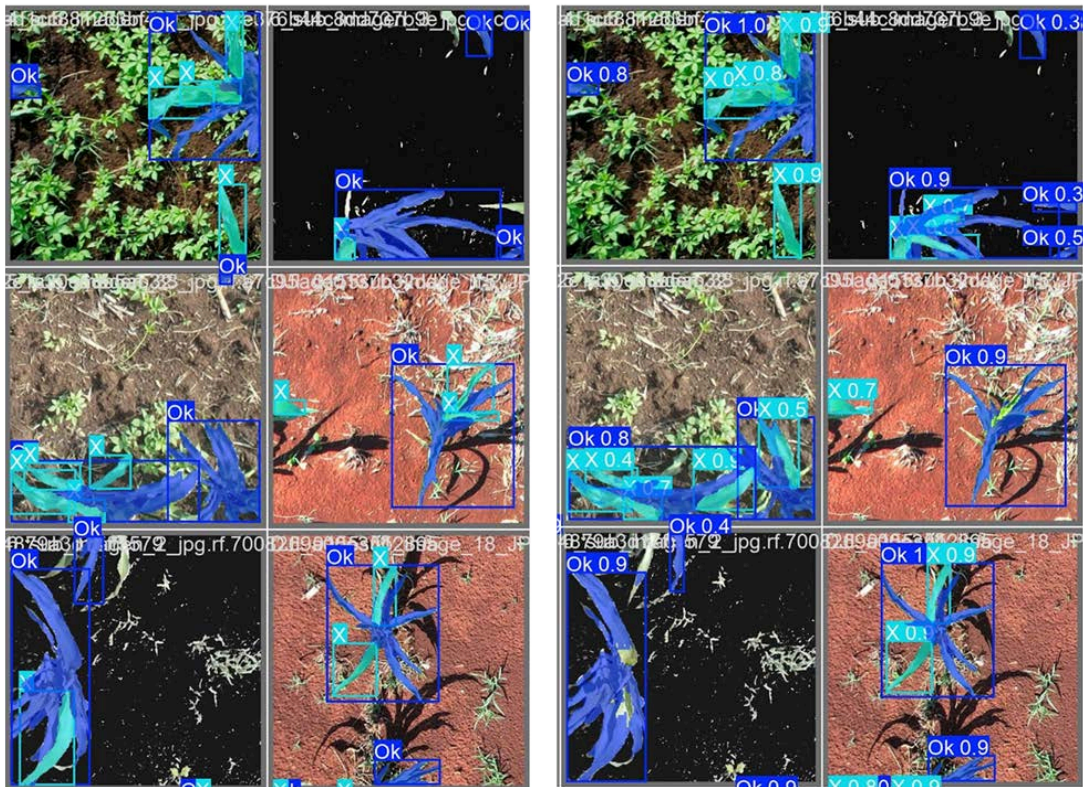


Figura 5.9. Se presentan 6 de 16 imágenes de un batch generado durante el entrenamiento, a la izquierda el batch con los etiquetados manuales y a la derecha los etiquetados generados automáticamente por el modelo A1.

En la Figura 5.9 se aprecia cómo el modelo A1 logra segmentar correctamente a las plantas en las instancias de la clase maíz (etiquetadas como “OK”), coincidiendo casi totalmente en su caja de detección con las generadas manualmente. Se puede constatar la alta correspondencia entre los resultados de la clase de hojas afectadas (etiquetadas como “X”).

Se observa que hay imágenes en las que el sistema detecta problemas en la hoja y la segmenta perfectamente, como en el etiquetado manual que se muestra en la Figura 5.10 en el caso 1 (arriba a la izquierda el manual y arriba derecha el del modelo A1).

El modelo A1 hace buena segmentación en la mayoría de las hojas afectadas, sin embargo, se detectaron algunos casos (como el Caso 2 de la Figura 5.10) en los que mezcla hojas diferentes en la misma selección o detecta hojas que no estaban etiquetadas en el etiquetado manual (Figura 5.10 abajo, Caso 2. La planta del lado derecho).

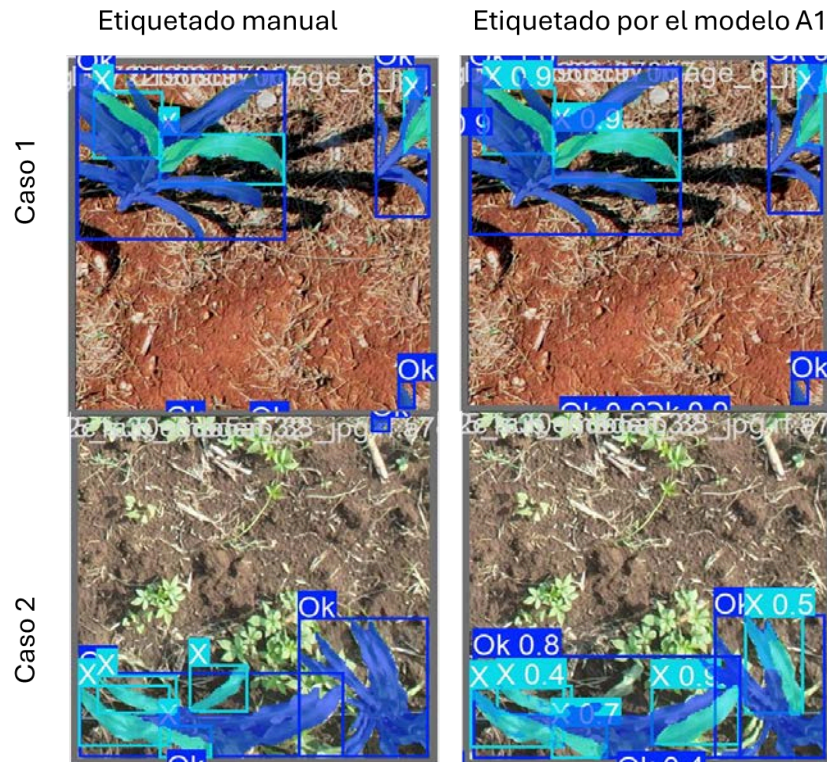


Figura 5.10. Comparativo de 2 imágenes provenientes del etiquetado manual (izquierda), con 2 imágenes provenientes del batch, realizada por el modelo A1 (derecha). Se aprecia un caso 1 (arriba) en el que las secciones etiquetadas coinciden perfectamente con la segmentación del modelo y otro caso 2 en el que difiere un poco (abajo).

Se pueden comprender algunas de las principales limitaciones técnicas para segmentar, relacionadas con el tipo de objetos que se analizan en este trabajo con ayuda de la Figura 5.11, en donde se compara un ejemplo de una imagen proveniente del batch, en su versión segmentada manualmente contra la segmentada por el modelo A1. Se puede apreciar que en cada imagen hay 2 maíces en cajas de detección azul, uno grande y uno pequeño. En el grande se puede apreciar que el modelo A1 hizo un mejor trabajo que el etiquetado manual en la tarea de evitar espacios internos, pero el manual lo hizo mejor en alcanzar todas las puntas de la planta de maíz. Se puede apreciar también que el modelo A1 seleccionó 2 hojas como una sola en la parte inferior del maíz grande. En la planta pequeña se puede ver que el etiquetado manual omitió etiquetar algunas hojas, pero el modelo A1 las recuperó.

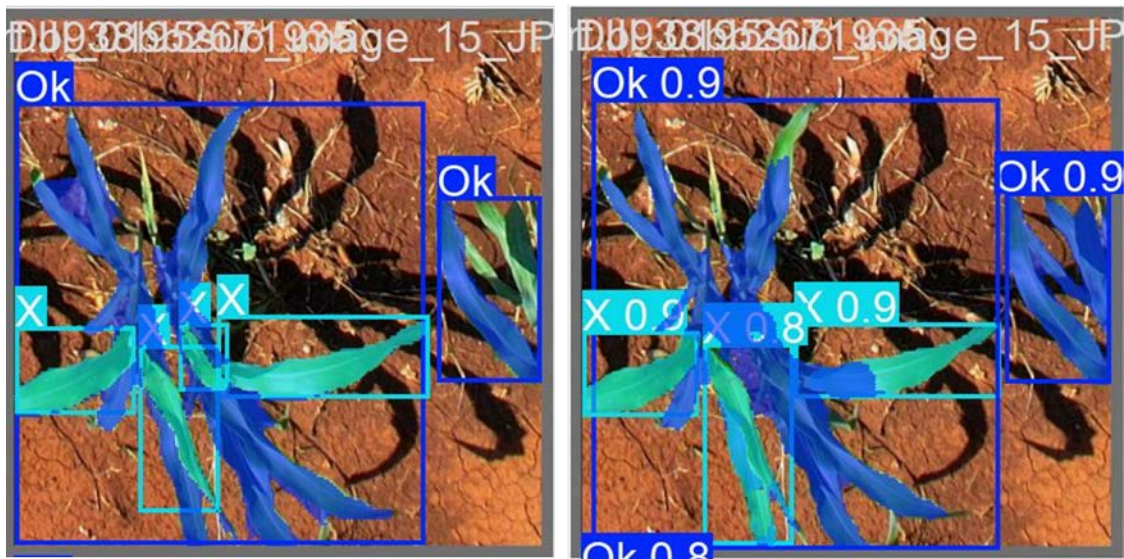


Figura 5.11. Imagen que ilustra algunos de los motivos principales que explicarían discrepancias entre el etiquetado y la predicción, como la segmentación de la hoja correcta, pero no en su totalidad, o que el modelo lo hace mejor que el manual.

El modelo logró detectar correctamente en esta imagen de ejemplo, 3 de 4 hojas afectadas, faltando una hoja más pequeña. Aunque las cajas son prácticamente las mismas, existe cierta diferencia entre las máscaras de segmentación que puede ser introducida tanto por ligeras variaciones en los criterios en el etiquetado al calificar una hoja como afectada, tanto por la forma en la que el modelo A1 resuelve cómo continúa la trayectoria de cada hoja afectada.

Es posible apreciar cómo fueron clasificadas las instancias generadas mediante el etiquetado manual y cómo las clasificó el modelo A1 mediante el análisis de su matriz de confusión. En la Figura 5.12 se ilustran la matriz de confusión normalizada.

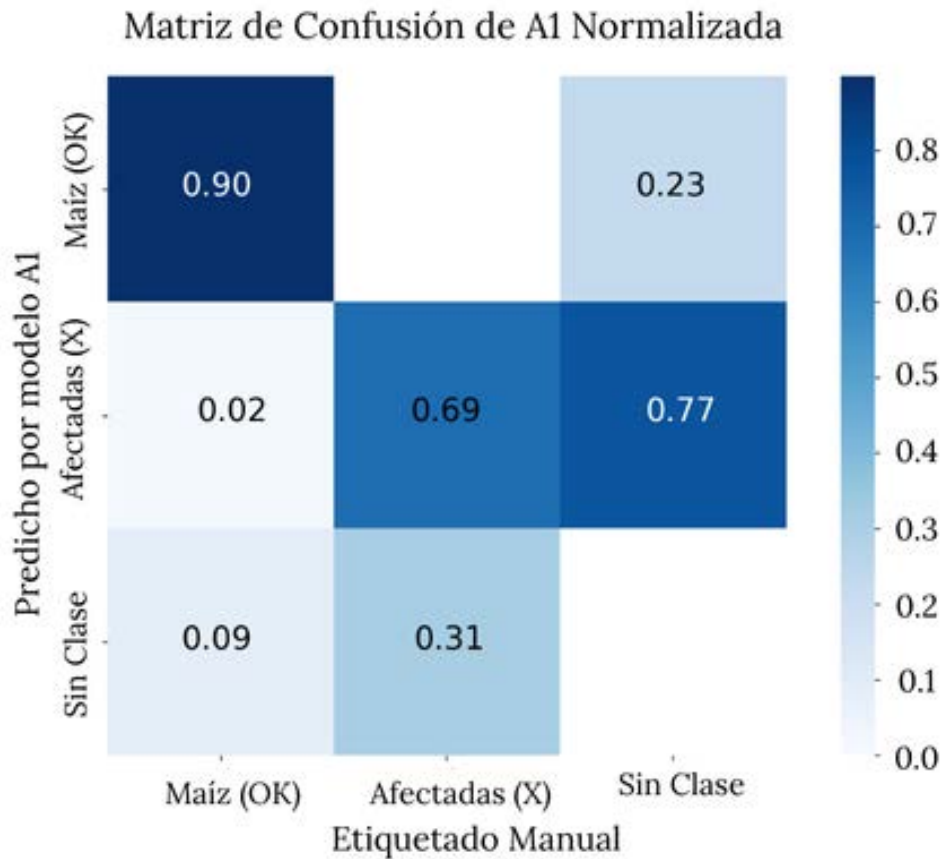


Figura 5.12. Matriz de confusión normalizada del modelo A1, con la clase *Maíz*, la clase *Hojas Afectadas*, y los píxeles de fondo no clasificados como "*Sin_Clase*".

El sistema identificó correctamente a 90% de las 295 apariciones de "Maíz" en las imágenes correspondientes al set de prueba, que contenía imágenes de cultivos de maíz de los diferentes entornos en Yucatán. La clase de "Afectadas" detectó el 69% de los 544 casos etiquetados correctas, y el 31% no fueron detectadas.

En la Figura 5.13 se aprecia el diagrama de las puntuaciones F1 por cada corte de nivel de confianza para cajas de detección (arriba) y máscaras de segmentación (abajo).

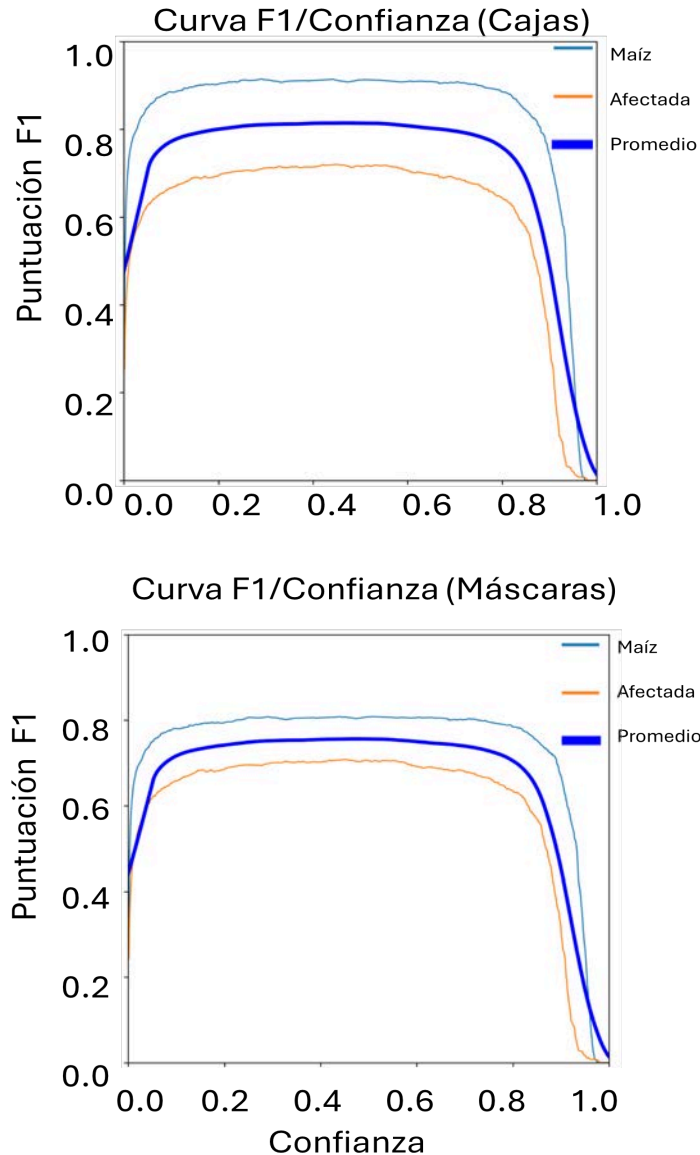


Figura 5.13. Diagramas de puntuación F1 contra nivel de confianza para las clases *Maíz* (“Ok” en las imágenes) y la clase de *hoja afectada* (“X” en las imágenes). Arriba el diagrama de cajas de detección y abajo máscaras de segmentación. En azul más grueso se representa el promedio de ambas clases, que sería el resultado general del modelo A1.

Haciendo una comparativa con los resultados en el trabajo de Feng et al. (2022), que alcanzaron su mejor resultado general de 96% de precisión con ResNeSt-50, en este trabajo se alcanzó 86% en la misma métrica de precisión de caja de detección con YOLOv11n, como se puede constatar en la Tabla 5.3, lo cual mejora el resultado de 83.3% que obtuvo Feng con ResNet50, al 85.66% que obtuvo con EfficientNet y al 85% que obtuvo con RegNet.

En conclusión, si bien los resultados aún pueden ser optimizados realizando nuevos entrenamientos de modelos, los resultados validan al sistema como una herramienta eficaz para la detección y segmentación de plantas de maíz, cumpliendo con los requisitos

buscados de velocidad y precisión. En la detección de afectaciones, si bien la precisión operativa de las cajas de detección (79.5%) es aceptable, la exhaustividad alcanzada señala una oportunidad para realizar optimizaciones futuras. El equilibrio entre velocidad (3.9 ms en inferencia) y precisión general lo posiciona como una herramienta capaz de detectar periódicamente la presencia de *S. frugiperda* y su distribución estadística.

5.5 Aplicación de método SAHI

Para superar las limitaciones asociadas al procesamiento de imágenes de alta resolución obtenidas mediante el VANT, de dimensiones 5472 × 3078 píxeles, se implementó el método Slicing Aided Hyper Inference (SAHI). Esta estrategia permitió implementar el modelo de segmentación, originalmente entrenado con imágenes de 640 × 640 píxeles, al análisis de superficies agrícolas, sin comprometer su precisión por pérdida de resolución.

Cada imagen de alta resolución capturada por el VANT se dividió sistemáticamente en 40 segmentos solapados de 640 × 640 píxeles, dimensiones óptimas para la arquitectura del modelo YOLO11n-seg. Este enfoque garantizó que todas las regiones de interés, incluidas plantas completas y afectaciones foliares, fuesen procesadas en fragmentos compatibles con la entrada del modelo.

Sobre cada sub-imagen se aplicó el modelo de segmentación optimizado, utilizando un umbral de confianza mínimo del 40% para validar las detecciones. Este filtro aseguró que solo las predicciones con suficiente certeza, tanto para plantas sanas (clase "Ok") como para hojas dañadas (clase "X"), fuesen consideradas en el análisis posterior. La inferencia en los 40 segmentos por imagen demostró ser computacionalmente eficiente, manteniendo un tiempo medio de procesamiento de 10.4 ms por fragmento gracias a la optimización previa del modelo y al uso de aceleración por GPU.

Se realizaron pruebas para una verificar la altura apropiada de funcionamiento del modelo A1 en combinación con el método SAHI. Con el uso de la base de datos de imágenes aéreas que varían desde 1 a 10 m de altura, se probó el sistema a diferentes alturas. El resultado más cercano al etiquetado de referencia registrado en tierra se ha logrado a 6 m de altura y con un corte de 50% de confianza. En la Figura 5.14 se visualizan algunos resultados obtenidos en las pruebas de funcionamiento a diferentes alturas y cortes de confianza. Con la máxima altura de 10 m se logró mucha precisión, pero poca exhaustividad en la clase de hojas afectadas. Bajando y probando metro a metro, se encontró que a 6 m de altura se logra alcanzar un buen nivel de exhaustividad respecto de las hojas afectadas y se puede realizar un corte de confianza de 50%, con el cual, se obtuvo el mejor parecido con el registro realizado manualmente en campo a nivel de suelo.

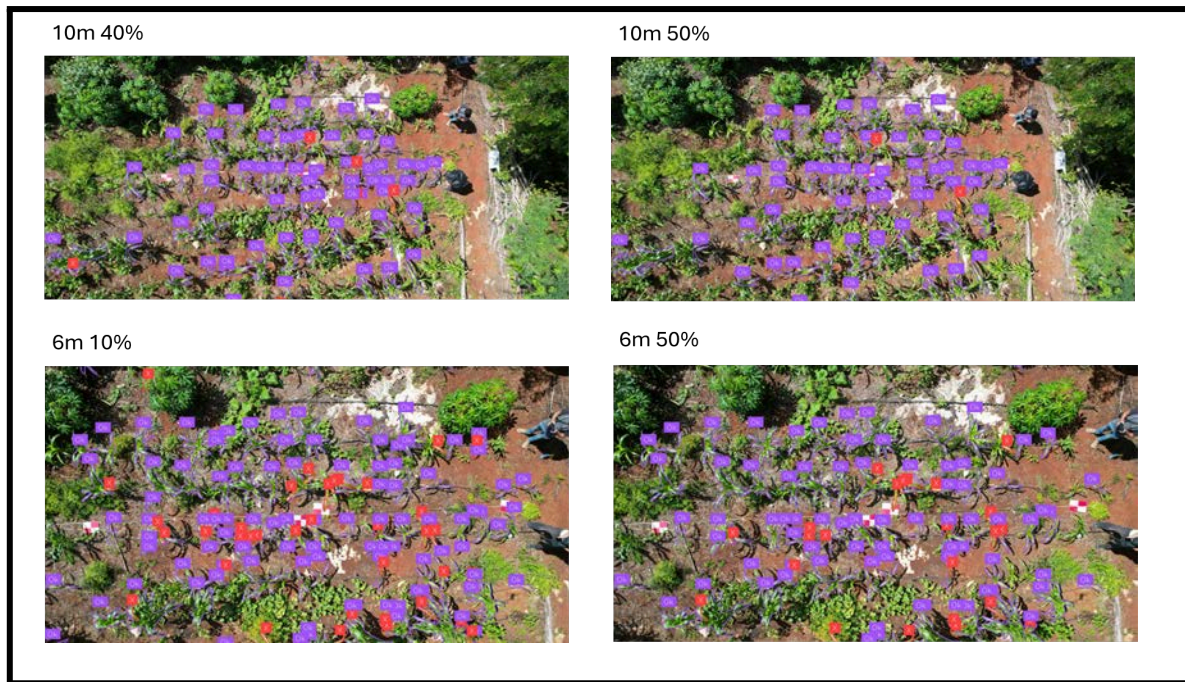


Figura 5.14 Arriba: Comparativo de 2 imágenes tomadas a 10 m de altura con 40% y 50% de corte de confianza. Abajo: Capturadas a 6 m de altura, con corte de confianza de 10% y 50%. Procesadas por el modelo A1 y con la aplicación del método SAHI de inferencia por segmentos. Se probaron diferentes cortes de confianza, se ilustran algunos

5.6 Ortomosaicos de la parcela

Se generaron ortomosaicos uniendo las imágenes de alta resolución analizadas con el modelo A1 y SAHI, a 6m de altura. Para ello se utilizó OpenCV y los puntos de control medidos con GPS submétrico, tomados en campo y con el VANT, utilizando el sistema de coordenadas: EPSG:32616 (WGS 84/UTM zona 16N).

Los ortomosaicos fueron de la sección de la parcela de referencia, se utilizó la base de datos de alturas para verificar la precisión en la distribución espacial de las afectaciones a diferentes alturas y al variar el corte de confianza del modelo. En la Figura 5.15 se aprecia el ortomosaico de la sección de referencia a 6 m de altura y con 60% de corte de confianza.

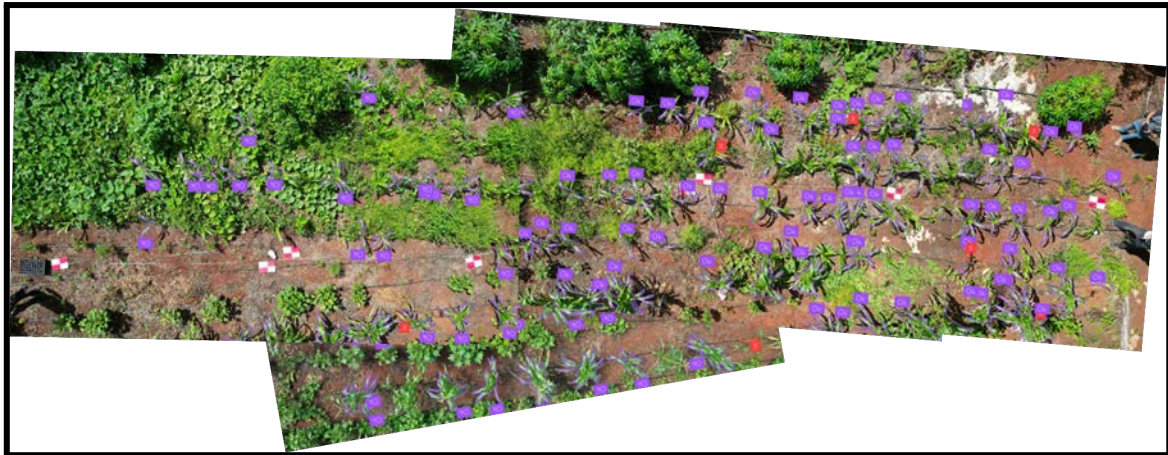


Figura 5.15. Ortomosaico con 4 imágenes obtenidas a cada 5 m longitudinales sobre la parcela a 6 m de altura, con 60% de corte de confianza.

Se puede apreciar que se logra realizar correctamente las detecciones de plantas de maíz, pero faltan varias de las hojas afectadas, por lo que se bajó el umbral de confianza de 60 a 50%. Probando los umbrales se constató que el de 50% es el que mejor coincidía con el registro terrestre. El ortomosaico generado con ese umbral se muestra en la Figura 5.16.



Figura 5.16. Ortomosaico con 4 imágenes obtenidas a cada 5 m longitudinales sobre la parcela a 6 m de altura, con 50% de corte de confianza.

El ortomosaico revela la mejor coincidencia con el registro manual. Si se extrae una sección efectiva que se recuperaría por sección, lo que se puede tomar como referencia de área efectiva del mapeo y considerar a partir de ella los entrelapamientos entre secciones que serían necesarias considerar en un mapeo a gran escala. En este caso, se estima que la sección transversal efectiva es de 6 metros y se ilustra en la Figura 5.17.



Figura 5.17. Segmento efectivo por recorrido de vuelo a 6 m de altura, con 50% de corte de confianza.

Con las métricas obtenidas con la parcela de referencia de Kantunil, como método de calibración, se realizaron inferencias sobre los registros de imágenes de alta resolución levantados en las parcelas de Muna, capturadas anteriormente y a la misma altura. Con ellas se generó un ortomosaico de la parcela. Se presenta la mitad del ortomosaico de la parcela 1 de Muna en la Figura 5.18.

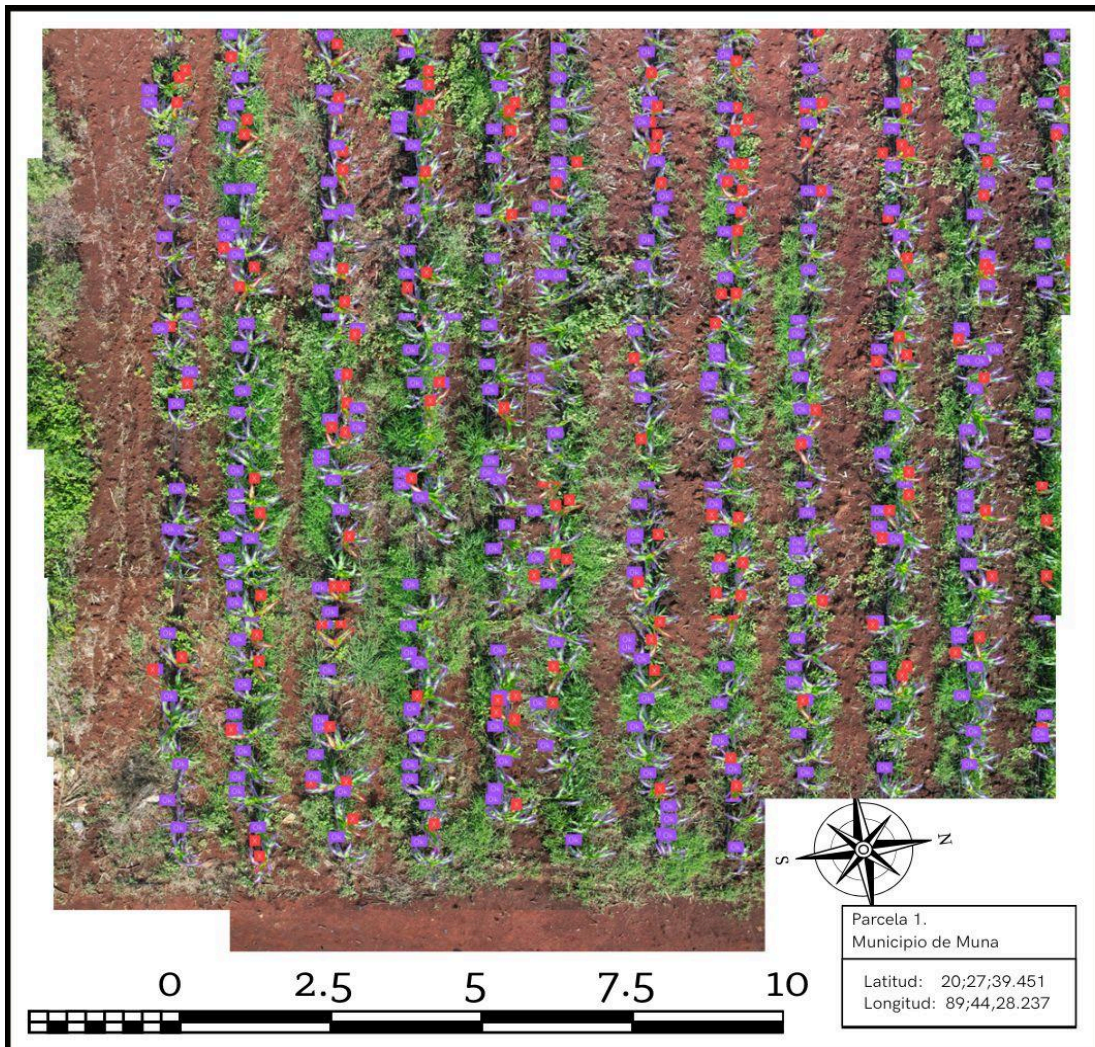


Figura 5.18. Segmento de ortomosaico de la parcela analizada, resaltando en rojo los casos de hojas afectadas y en morado las plantas de maíz.

Este ortomosaico refleja la capacidad del sistema para ser implementado como herramienta en el estudio estadístico de los métodos empleados para combatir las plagas.

Este sistema se puede implementar mediante seguimiento semanal del cultivo durante las etapas clave, desde V1 hasta V5, por un periodo aproximado de 1.5 meses, que permita actuar para mitigar plagas activas y cuantificar estadísticamente la efectividad de las medidas de mitigación aplicadas. Al ser capaz de cubrir parcelas de riego y de ciclo anual, se puede estandarizar un protocolo de evaluación-acción-recolección de datos en intervalos de 7 días, generando métricas objetivas que serían más complicadas de obtener con métodos tradicionales.

En Yucatán, esta herramienta puede apoyar directamente en el trabajo de técnicos de campo que actualmente realizan estudios estadísticos manuales para evaluar técnicas agrícolas. Ya se tiene contacto con cooperativas locales, con las que se compartirán los resultados y se podría implementar el sistema de monitoreo conjunto entre el VANT y los técnicos, potenciando sus esfuerzos contra plagas y en favor de la transición agroecológica. Esta sinergia tecnología-conocimiento ancestral fortalece la resiliencia del maíz nativo y transforma datos en decisiones ágiles para una agricultura sostenible.

5.7. Consideraciones para adquisición de imágenes aéreas

Para el monitoreo agrícola mediante fotogrametría aérea, se evaluaron dos enfoques tecnológicos diferenciados por su costo y accesibilidad.

1. Sistemas Comerciales Integrados:

Se empleó un vehículo aéreo no tripulado (VANT) comercial (modelo DJI Mavic Air 2s) equipado con una cámara integrada de alta resolución (48 MP). Si bien esta solución ofrece operación simplificada y permite la programación precisa de rutas de vuelo mediante aplicaciones propietarias, garantizando puntos de captura y alturas constantes para series temporales, su costo representa una barrera significativa para la escalabilidad en contextos agrícolas con presupuestos limitados.

2. Plataforma de Código Abierto y Bajo Costo:

Como alternativa económicamente más accesible, se exploró el uso de una plataforma modular basada en hardware abierto. La configuración propuesta consta de: Un chasis de cuadricóptero genérico (modelo F450), un controlador de vuelo Pixhawk (estándar en la comunidad de código abierto), una cámara de 64 MP desarrollada por Arducam para Raspberry Pi 4.

El costo estimado de un sistema de código abierto es considerablemente menor (aproximadamente 8000 MXN para el VANT y 2000 MXN para la cámara) que el de los sistemas comerciales, aumentando su viabilidad financiera para proyectos agrícolas extensivos. Adicionalmente, esta solución ofrece mayor flexibilidad técnica. La cámara Arducam permite la multiplexación de hasta cuatro sensores de 64 MP, lo que es mayor resolución que el modelo DJI y además habilita configuraciones multispectrales mediante el uso de filtros ópticos específicos. Esto facilitaría la captura de datos en bandas seleccionadas del espectro visible e infrarrojo cercano (NIR), esenciales para generar índices de vegetación y mapas de vigor con mayor profundidad analítica. El contenido de los kits genéricos de código abierto para ensamblar un VANT se ilustra en la Figura 5.19, que incluye un marco modelo F450 de 4 rotores, controlador de vuelo 3DR Pixhawk®, con GPS, motores, transmisor RC, hélices y baterías. La ruta de vuelo y puntos de fotografía pueden ser programados con el software de código libre Mission Planner y cargados en el controlador de vuelo Pixhawk.



Figura 5.19. Kit de hardware libre con costo menor que puede ser conseguido fácilmente por internet.

Ambos sistemas permiten almacenar las coordenadas GPS de todos los puntos que debe recorrer el VANT durante el vuelo para realizar el levantamiento en cada parcela y generar los ortomosaicos con precisión. Al implementar esa funcionalidad en el sistema, se alcanza una resolución temporal en la que todas las imágenes pueden ser tomadas desde los mismos puntos sobre la parcela, lo que incrementa la precisión del análisis y habilita la posibilidad de replicar automáticamente la toma de muestras para cada parcela.

Durante el desarrollo de este trabajo se realizaron pruebas con los sistemas de recorridos automatizados para generar ortomosaicos y probar sus capacidades. Para probar los sistemas de código libre se implementaron los recorridos utilizando el software de vuelo ArduPilot, utilizando un VANT modelo SOLO de la marca ya discontinuada 3DR, que mantiene soporte por parte de la comunidad de software libre.



Figura 5.20. VANT marca 3DR modelo SOLO con el que se realizaron experimentos. Imagen propia.

Debido a que el VANT SOLO utiliza piezas de código libre, como la controladora de vuelo Pixhawk, en este caso modelo "Cube", es capaz de implementar los planes de vuelo automáticos con las herramientas libres de ArduPilot. Los VANT de código abierto permite realizar el mismo tipo de operaciones por un menor costo, además de que presentan la ventaja de que es posible hacer todos los ajustes necesarios, tanto a las aeronaves como al sistema de vuelo requieren de mayor conocimiento técnico del funcionamiento, pero se considera que ese nivel de complejidad es manejable.

En la Figura 5.21 se muestra un ejemplo de un plan de vuelo para los recorridos generados mediante ArduPilot, que fue implementado con el VANT SOLO. En la imagen se destacan en verde todos los puntos del recorrido desde donde se determina que se va a tomar una fotografía y en amarillo el recorrido y la dirección del VANT. Los marcadores rojos determinan el perímetro.



Figura 5.21. Ejemplo de un plan de vuelo para el VANT SOLO generado con el software libre ArduPilot. Imagen propia.

Estos recorridos programados son fundamentales para un proceso que busque generar mapas de afectaciones y detectar problemas periódicamente y guiar intervenciones localizadas y racionalizadas, por lo que los recorridos planificados se toman en cuenta como parte del sistema de detección en este trabajo.

5.8 Resultados con la computadora Raspberry Pi

El sistema alcanzó una tasa de procesamiento de 1.3 fotogramas por segundo (FPS), evidenciando viabilidad operativa en regímenes cuasi en tiempo real para aplicaciones de monitoreo agrícola. Sin embargo, en el proyecto actual las inferencias se necesitarían realizar sobre las 40 sub imágenes de 640x640 provenientes de las imágenes de alta resolución, que no puede ser procesado en tiempo real por el procesador del dispositivo Raspberry Pi. Un resultado del procesamiento, realizado en tiempo real, directamente desde la computadora Raspberry Pi, se puede ver en la Figura 5.20. Siendo una escena nocturna de una planta de maíz con afectación en las hojas superiores.



Figura 5.22. Imagen obtenida y procesada en hardware Raspberry pi, en una escena nocturna de un maíz.

El desempeño, logrado con recursos limitados (4 GB de RAM sin aceleración GPU externa), valida la eficiencia del formato NCNN y posiciona a la Raspberry Pi como una plataforma viable para inferencias con dispositivos de borde, en escenarios de precisión agrícola con requerimientos de baja latencia, pero no es suficientemente rápido aún para hacerlo escalable para inferencias en tiempo real en grandes extensiones de terreno.

Por ese motivo se cambió el enfoque de procesamiento, de tal forma que se destine toda la potencia de la computadora Raspberry Pi para la adquisición de imágenes de alta resolución y que los procesos de inferencia y postprocesamiento se realicen después en dispositivos especializados en procesos paralelizables.

Capítulo 6. Conclusiones y trabajo futuro

El sistema desarrollado en este trabajo representa un alcance significativo en monitoreo de superficies agrícolas, que puede tener aplicaciones inmediatas en el campo Yucateco.

Los resultados demuestran que nuestro enfoque basado en YOLOv11 y SAHI permite detectar y segmentar con precisión tanto plantas de maíz como afectaciones por *S. frugiperda*, permitiendo realizar mapeos de superficies para analizar el estado de cada planta, cumpliendo así los objetivos planteados inicialmente. La arquitectura propuesta ha mostrado capacidad de detección en condiciones reales de campo, procesando imágenes aéreas con variabilidad de iluminación, escala y complejidad de fondo.

El sistema alcanzó una métrica de precisión de detección de caja delimitadora de plantas de maíz de 92.9%, con un mAP50 de 94.2% y exhaustividad de 78.4%.

Para las afectaciones por *S. frugiperda*, alcanza una precisión de detección de caja delimitadora de 79.2%, con un mAP50 de 71.7% y exhaustividad de 64.3%, demostrando viabilidad operativa en condiciones reales aunque mejorable en precisión espacial

Las métricas del modelo A1, promediando el resultado de sus 2 clases para cajas delimitadoras, obtienen una precisión de 86%, mAP50 de 83% y exhaustividad de 77.5%

En cuanto a los resultados del modelo A1 para máscaras de segmentación, los resultados generales fueron de precisión de 80%, mAP50 de 73.3% y exhaustividad de 73.3%

La implementación en Raspberry Pi (1 fps) evidenció limitaciones para procesamiento en tiempo real, llevando al diseño de una arquitectura híbrida: adquisición local con transmisión a servicios en nube (Google Colab) donde se alcanzan 96 fps, permitiendo análisis de 1 ha en <15 minutos.

Tomando en cuenta la duración de la carga de la batería se estima que de podría mapear aproximadamente una hectárea por cada carga de batería. También es posible contar con varias baterías para hacer recorridos más extensos en una misma jornada. El tiempo de procesamiento se mantiene dentro del rango de minuto para realizar las inferencias en las imágenes de alta resolución, ya con el método SAHI. Después se realiza independientemente la generación del ortomosaico.

La transición agroecológica en Yucatán encontrará en esta tecnología una herramienta muy útil. Al ser capaz de cuantificar el impacto en los cultivos de implementar técnicas como trampas de feromonas o agua de vidrio, el sistema proporcionará evidencia empírica para optimizar su aplicación. Futuras colaboraciones con el INIFAP podrían permitir validar escalas de daño específicas para variedades locales de maíz, mientras que la integración con sensores multispectrales añadiría capas de análisis fisiológico. Este trabajo sienta las bases para una nueva generación de herramientas agrícolas centradas en sostenibilidad y accesibilidad.

Como conclusión final, esta investigación no solo demuestra la viabilidad técnica de la solución propuesta, sino su potencial transformador para la agricultura regional. El sistema está preparado para despliegue en cooperativas agrícolas, donde su implementación piloto durante el próximo ciclo de cultivo puede generar métricas de impacto real. Los desafíos técnicos superados y las lecciones aprendidas constituyen un valioso aporte al campo, abriendo camino para aplicaciones en otros cultivos estratégicos locales, como pitahaya o chile habanero.

Referencias

1. Akyon, F. C., Altinuc, S. O., & Temizel, A. (2022). Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection. arXiv preprint arXiv:2202.06934.
2. Amein, A., & Aziza, A. (2023). A review of the control strategies of fall armyworm (*Spodoptera frugiperda*) in maize. *Journal of Agricultural and Environmental Sciences*, 20(1), 1-15.
3. Ángel-Ríos M.D., Pérez-Salgado J., Morales J.F. 2015. Toxicidad de extractos vegetales y hongos entomopatógenos en el gusano cogollero *Spodoptera frugiperda* J.E. Smith (Lepidoptera: Noctuidae), del maíz en el Estado de Guerrero. *Entomología Mexicana* 2:260-265.
4. Ayora Diaz, S. I. (2017). Translocalidad, globalización y regionalismo: Cómo entender la gastronomía regional yucateca. *Anales de Antropología*, 51(2), 96-105. <https://doi.org/10.1016/j.antro.2017.03.004>
5. Banda JF, Enkerlin D, De Alba G, Garza L. 1981.Importancia de *Heliothis zea* (Boddie) y determinación del umbral económico, distribución matemática y muestreo secuencial de *Spodoptera frugiperda* en maíz criollo. *Fitófilo*, 85:101-118.
6. Bhatt, Prakruti & Sarangi, Sanat & Pappula, Srinivasu. (2019). Detection of diseases and pests on images captured in uncontrolled conditions from tea plantations. 33. 10.1117/12.2518868.
7. Bueno, J., Orellana, R., & Pacheco, J. (2010). Atlas: Escenarios de Cambio Climático en la Península de Yucatán. Centro de Investigación Científica de Yucatán, A.C. (CICY).
8. Camacho-Báez, J. R., García- Gutiérrez, C., Mundo-Ocampo, M., Armenta-Bojorquez, A. D., Nava-Pérez, E., Valenzuela-Hernández, J. I., & González-Guitrón, U. (2012). ENEMIGOS NATURALES DE LAS MOSCAS DE LOS ESTIGMAS DEL MAÍZ: *Euxesta stigmatias* (Loew), *Chaetopsis massyla* (Walker) y *Eumecosomyia nubila* (Wiedemann) EN GUASAVE SINALOA, MÉXICO. *Ra Ximhai*, 8(3b), 71-77.
9. Carrillo, et al. (2023). Mercados locales y seguridad alimentaria. El caso de las regiones Milpera y Puuc en Yucatán. *Nóesis revista de ciencias sociales y humanidades*. <https://doi.org/10.20983/noesis.2023.2.4>
10. Casmuz, A., Juárez, M. L., Socías, M. G., Murúa, M. G., Prieto, S., Medina, S., Willink, E., & Gastaminza, G. (2010). Revisión de los hospederos del gusano cogollero del maíz, *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Revista de la Sociedad Entomológica Argentina*, 69(3-4), 209-231.
11. Dewi, S. (2023). The role of technology in sustainable agriculture: A review. *Journal of Sustainable Agriculture*, 43(2), 1-15.
12. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Housby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

-
13. Echanove, Flavia. (2021). La nueva política agrícola en México y los pequeños productores de maíz de la península de Yucatán, 2019-2021. *Papeles de Geografía*, 67. 66-79. [10.6018/geografia.476931](https://doi.org/10.6018/geografia.476931).
 14. Feng, J., Sun, Y., Zhang, K., Zhao, Y., Ren, Y., Chen, Y., Zhuang, H., & Chen, S. (2022). Autonomous detection of *Spodoptera frugiperda* by feeding symptoms directly from UAV RGB imagery. *Applied Sciences*, 12(5), Article 2592. <https://doi.org/10.3390/app12052592>
 15. Galindo-Gutiérrez, N. F., & Garcés-Gómez, Y. A. (2023). Effects of Silicon Application on Yield, Spectral Index, and Fall Armyworm (*Spodoptera frugiperda*) Infestation on Maize (*Zea mays*) Crop. *AgriEngineering*, 5(4), 2112-2122. <https://doi.org/10.3390/agriengineering5040129>
 16. García, C. (2023). Parámetros de caracterización en zonas urbanas para detectar residuos sólidos mediante visión artificial. *Investigación e Innovación en Ingenierías*, 11(2), 49-59. <https://doi.org/10.17081/invinno.11.2.6690>
 17. García-Gutiérrez C., González-Maldonado M.B., Cortez-Mondaca E. 2012. Uso de enemigos naturales y biorracionales para el control de plagas de maíz. *Ra Ximhai* 8:57-70.
 18. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: surpassing human-level performance on imagenet classification. *arXiv*. <https://doi.org/10.1109/iccv.2015.123>
 19. He, K., Zhang, X., Ren, S., & Sun, J., (2016) "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016. 770-778, doi: 10.1109/CVPR.2016.90.
 20. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* 2961-2969.
 21. Hernández-Trejo A, Osorio-Hernández E., López-Santillán J. A., Ríos-Velasco C., Varela-Fuentes S. E., Rodríguez-Herrera R. (2018) Insectos benéficos asociados al control del gusano cogollero (*Spodoptera frugiperda*) en el cultivo de maíz (*Zea mays* L.).
 22. Huang, K.-Y. (2007). Application of artificial neural network for detecting *Phalaenopsis* seedling diseases using color and texture features. *Computers and Electronics in Agriculture*, 57(1), 3–11. <https://doi.org/10.1016/j.compag.2007.01.015>
 23. Huang, Z., Huang, L., Gong, Y., Huang, C., & Wang, X. (2019). Mask scoring r-cnn. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 6409-6418.
 24. Ishengoma, F. S., Rai, I. A., & Said, R. N. (2021). Identification of maize leaves infected by fall armyworms using UAV-based imagery and convolutional neural networks. *Computers and Electronics in Agriculture*, 184, 106-124. <https://doi.org/10.1016/j.compag.2021.106124>
 25. Jocher, G., & Qiu, J. (2024). Ultralytics YOLO11 (Versión 11.0.0) [Software de computación]. Ultralytics. <https://github.com/ultralytics/ultralytics>
 26. Kuzuhara, H., Takimoto, H., Sato, Y., & Kanagawa, A. (2020). Insect Pest Detection and Identification Method Based on Deep Learning for Realizing a Pest Control System. 2020 59th Annual Conference of the Society of Instrument and Control

27. Lin, T. Y. et al. (2014). Microsoft COCO: Common Objects in Context. arXiv:1405.0312.
28. Liu, J., & Wang, X. (2020). Early recognition of tomato gray leaf spot disease based on MobileNetv2-YOLOv3 model. *Plant Methods*, 16(1), 1-16. <https://doi.org/10.1186/s13007-020-00624-2>
29. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012-10022.
30. Murúa, G., & Virla, E. (2004). Population parameters of *Spodoptera Frugiperda* (Smith) (Lep.: Noctuidae) fed on corn and two predominant grasses in Tucuman (Argentina). *Acta Zoológica Mexicana (nueva serie)*, 20(1), 199-210.
31. Matsuoka, Y., et al. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences*, 99(9), 6080-6084.
32. Montezano D.G, Specht A., D.R. Sosa-Gómez, Roque-Specht V.F., Sousa-Silva J.C., Paula-Moraes S.V., Peterson J.A. and Hunt T.E. "Host Plants of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in the Americas," *African Entomology* 26(2), 286-300, (1 September 2018). <https://doi.org/10.4001/003.026.0286>
33. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection (arXiv:1506.02640). arXiv. <http://arxiv.org/abs/1506.02640>
34. Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 7263–7271. <https://doi.org/10.1109/CVPR.2017.690>
35. Reséndiz R.Z., López S.J.A., Osorio H.E., Estrada D.B, Pecina M.J.A., Mendoza C.M.C., Reyes M.A. (2016). Importancia de la resistencia del maíz nativo al ataque de larvas de lepidópteros. *Temas de Ciencia y Tecnología* 20:3-14.
36. Reyes-Cruz, J. L., Sánchez-Trujillo, M. G., & Mejía-Ramírez, R. (2019). Influencia de la formación universitaria en las actitudes emprendedoras. *Revista CEA*, 5(10), 117–133. <https://doi.org/10.22430/24223182.1240>
37. Roboflow. (2024). Roboflow: Computer vision tools and datasets [Plataforma web]. Recuperado de <https://roboflow.com>
38. Rodríguez, A. G. P., Castro, T. V. M., Cetz, J., & López, R. Q. (2019). Uso de agroquímicos cancerígenos en la región agrícola de Yucatán, México.
39. Romero, R., Fernández, L., Riverón, E., & Hernández, J. (2019). Detección y rastreo de peatones empleando vectores de características de histogramas de gradientes orientados y patrones binarios locales en una máquina de soporte vectorial. *Research in Computing Science*, 148(10), 81-94. <https://doi.org/10.13053/rcs-148-10-7>
40. Santos, L. M., Ferraz, G. A. Silva, Barbosa, B. D. S., & Andrade, A. D. (2019). Use of remotely piloted aircraft in precision agriculture: a review / Utilización de sistemas de aeronaves no tripuladas en agricultura de precisión: una revisión. *DYNA*, 86(210), 284-291.

-
41. Terven, J. (2023). A comprehensive review of yolo architectures in computer vision: from yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction*, 5(4), 1680-1716. <https://doi.org/10.3390/make5040083>
 42. Toledo, V. M., B. Ortiz-Espejel, L. Cortés, P. Moguel, and M. D. J. Ordoñez. 2003. The multiple use of tropical forests by indigenous peoples in Mexico: a case of adaptive management. *Conservation Ecology* 7(3): 9. [online] URL: <http://www.consecol.org/vol7/iss3/art9/>
 43. Valdez-Torres J.B., Soto-Landeros F., Osuna-Enciso T., Báez-Sañudo A.M. 2012. Modelos de predicción fenológica para maíz blanco (*Zea mays* L.) y gusano cogollero (*Spodoptera frugiperda* J.E. Smith). *Agrociencia* 46:399-410
 44. Zhang, X., Qiao, Y., Meng, F., Fan, C., & Zhang, M. (2018). Identification of maize leaf diseases using improved deep convolutional neural networks. *IEEE Access*, 6, 30370–30377. <https://doi.org/10.1109/ACCESS.2018.2844405>