

Article

An Automated Information Processing Framework for UAV-Based Detection and Spatial Mapping of Crop Damage Using Deep Learning

Alejandro Carrillo-Gómez ¹, Daniela Moctezuma ^{1,*} and Enrique Camacho-Pérez ^{2,*}

¹ Centro de Investigación en Ciencias de Información Geoespacial, Mexico City 14240, Mexico; acarrillo@centrogeo.edu.mx

² Facultad de Ingeniería, Universidad Autónoma de Yucatán, Merida 97000, Mexico

* Correspondence: dmoctezuma@centrogeo.edu.mx (D.M.); enrique.camacho@correo.uady.mx (E.C.-P.)

Abstract

The early detection and spatial characterization of crop damage are critical for improving decision-making in precision agriculture, particularly in regions where traditional monitoring methods are limited in scalability and objectivity. This study presents an integrated information processing framework that couples UAV-based image acquisition, instance segmentation, slicing-aided inference of large orthomosaics, and georeferenced spatial analysis into a single reproducible pipeline for the detection and mapping of crop damage. The framework is applied to maize cultivated under traditional milpa systems in Yucatán, Mexico, a region characterized by intercropping, irregular plant spacing, and complex backgrounds rarely represented in mainstream agricultural deep learning benchmarks. High-resolution RGB images were systematically acquired over maize fields in Yucatán, Mexico, and curated into specialized datasets representing parcels, individual plants, and damaged vegetation. Instance segmentation models based on the YOLOv11 architecture were trained and evaluated to extract visual information related to crop condition, while the Slicing-Aided Hyper Inference (SAHI) method was integrated to enable efficient processing of large orthomosaic images. The proposed framework achieved high performance in detecting maize plants, with a precision of 92.9% and an mAP50 of 94.2%, and demonstrated reliable identification of damage patterns associated with *Spodoptera frugiperda*, reaching a precision of 79.2% and an mAP50 of 71.7%. The resulting georeferenced outputs provide spatially explicit information that supports quantitative analysis of crop health and damage distribution. The results indicate that the proposed framework constitutes a scalable and reproducible approach for UAV-based visual information extraction, with potential applicability to broader agricultural monitoring and data-driven decision support systems.



Academic Editors: Xin Lu, Jianhua Yang, Xiaoxia Li, Dehao Wu and Wei Wang

Received: 15 March 2026

Revised: 21 May 2026

Accepted: 22 May 2026

Published: 27 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: UAV imagery; deep learning; precision agriculture; crop damage detection; maize monitoring; *Spodoptera frugiperda*; instance segmentation; spatial mapping; edge computing

1. Introduction

Agricultural production systems are increasingly adopting digital technologies to improve crop monitoring, resource management, and decision-making processes. Advances in remote sensing, unmanned aerial vehicles (UAVs), and artificial intelligence have enabled the development of high-resolution monitoring systems capable of capturing

detailed information about crop conditions at the field scale [1–3]. Within this broader context of digital agriculture, UAV platforms have emerged as a particularly attractive sensing solution because they provide flexible and cost-effective alternatives to satellite imagery, offering centimeter-level spatial resolution and rapid data acquisition that are highly valuable for precision agriculture applications. Recent studies have demonstrated the potential of UAV imagery combined with deep learning techniques for crop monitoring, disease detection, and yield assessment [4–6]. Convolutional neural networks (CNNs) and related architectures have shown strong performance in agricultural image analysis tasks, including plant counting, crop classification, and disease recognition [7,8]. These methods enable automated interpretation of aerial imagery, significantly reducing the need for labor-intensive field inspections.

Among the crops that benefit most from these high-resolution monitoring capabilities, maize (*Zea mays*) plays a particularly critical role in global food security and agricultural economies. Early detection of crop stress and pest damage is essential to prevent yield losses and to support timely intervention strategies. UAV-based remote sensing has been increasingly explored for maize monitoring, including disease detection, crop assessment, and structural damage analysis [9–11]. High-resolution RGB imagery collected by UAVs allows detailed visualization of plant morphology and canopy structure, enabling automated detection of abnormalities at the plant level.

Within the spectrum of biotic threats to maize that can be monitored from UAV imagery, the fall armyworm (*Spodoptera frugiperda*) is one of the most damaging pests affecting maize production worldwide. This pest causes severe foliar damage during early vegetative stages, particularly between the V2 and V4 growth stages, and has rapidly spread across multiple regions including the Americas, Africa, and Asia. Early detection of infestation symptoms is crucial for effective pest management, yet manual scouting methods are time-consuming and often impractical at large spatial scales. Consequently, automated detection approaches based on aerial imagery and machine learning are increasingly being investigated for monitoring pest-induced crop damage.

To address pest detection from aerial imagery, recent advances in deep learning-based object detection have substantially improved the capability to identify plant stress and damage patterns from aerial imagery. In particular, models from the YOLO family have gained considerable attention due to their balance between detection accuracy and computational efficiency. YOLO-based frameworks have been successfully applied to various agricultural monitoring tasks using UAV imagery [11–13]. Transformer-enhanced architectures have also been proposed to improve small-object detection in complex aerial scenes [14]. Despite these advances, practical deployment in agricultural environments remains challenging due to factors such as variable illumination conditions, heterogeneous backgrounds, and limited availability of annotated datasets [2,15].

Another limitation arises from the complexity of damage representation in crop monitoring systems. Many studies attempt to classify multiple categories of damage symptoms, which can introduce ambiguity due to visual similarity between lesion types. In practice, this may reduce detection reliability and increase model complexity. Simplified representations that focus on detecting the presence of damage rather than detailed symptom classification may provide more robust and operationally practical solutions, particularly when processing large UAV datasets.

In addition, large UAV images often require specialized preprocessing strategies to enable efficient analysis. Techniques such as image tiling and slicing have been proposed to address the challenges associated with detecting small objects in high-resolution aerial imagery. These approaches can significantly improve detection accuracy while maintaining computational feasibility for real-world deployment scenarios.

Within this growing field, several recent studies have specifically targeted UAV-based monitoring of maize. RGB-based approaches combined with YOLO architectures have been applied to maize plant detection and counting, as well as to the detection of phenological structures such as tassels under realistic field conditions [11,16]. Transformer-based semantic segmentation methods have also been used for corn-field delineation [10], while the fusion of RGB and LiDAR data has recently been explored for crop damage quantification in maize [17]. Multispectral UAV imagery combined with vegetation indices and CNNs has been used for early detection of maize leaf diseases [9]. Specific applications targeting *Spodoptera frugiperda* are still scarce; the most directly comparable work [18] addresses the autonomous detection of fall armyworm feeding symptoms from UAV-RGB imagery through patch-level CNN classification. Despite this growing body of work, the joint detection of maize plants and *S. frugiperda* damage at the instance level, combined with slicing-aided inference of high-resolution orthomosaics, georeferenced spatial aggregation, and feasibility assessment on low-cost edge hardware, remains underexplored.

Although individual components such as YOLO-based detectors and slicing-aided inference (SAHI) are now well-established techniques, their integration into an operational pipeline that simultaneously addresses (i) high-resolution orthomosaic processing, (ii) georeferenced spatial aggregation of detections into infestation indicators, and (iii) feasibility assessment on low-cost edge hardware has rarely been reported for maize and, to the best of our knowledge, has not been previously addressed for *Spodoptera frugiperda* damage detection under traditional milpa cultivation systems. The novelty of this work therefore does not lie in the underlying detection algorithm, but in the empirically validated integration of these components and in the methodological and operational findings that emerge from this integration.

The integration of the Slicing-Aided Hyper Inference (SAHI) strategy is fundamental for preserving the diagnostic integrity of high-resolution UAV imagery while maintaining computational feasibility for large-scene analysis. Standard inference pipelines typically resize aerial images to the native model input size (640×640 pixels), which can suppress fine-scale morphological features and reduce the detectability of millimetric lesions associated with *S. frugiperda*. By partitioning large UAV images into smaller overlapping patches, SAHI preserves the original pixel density and spatial detail required for small-object detection, improves the identification of fragmented damage patterns, and enables scalable processing of large orthomosaic regions through sequential patch-based inference and automatic merging of overlapping detections.

This study addresses these challenges by developing and evaluating a UAV-based deep learning framework for automated detection of maize plants and foliar damage associated with *Spodoptera frugiperda*. The proposed approach integrates object detection models, preprocessing strategies for high-resolution imagery, and spatial analysis techniques for identifying infestation patterns at the field scale.

The main contributions of this study are summarized as follows:

- An end-to-end UAV monitoring pipeline that integrates SAM-assisted annotation, HSV-based preprocessing, YOLOv11 instance segmentation, SAHI tiled inference for large orthomosaics, kernel-density-based spatial aggregation of damage, and edge deployment, all instantiated and evaluated on the same maize/*S. frugiperda* case study.
- An empirical comparison between multi-class and unified damage representations that quantifies, for the first time in this domain, the robustness gain obtained by collapsing visually similar lesion subclasses (mAP50 of 71.7% for the unified class versus 21–31% for the per-subclass formulation), providing a transferable design recommendation for UAV-based damage detection systems.

- Region-specific datasets acquired over traditional milpa systems in Yucatán, Mexico, including parcel-level, individual-plant, and infested-plant datasets curated under heterogeneous backgrounds (intercropping, weed presence, exposed soil) that are underrepresented in existing agricultural deep learning benchmarks.
- A georeferenced spatial mapping methodology that converts plant-level detections into parcel-level agronomic indicators (damage rate, damage density, infestation hotspots via kernel density estimation), bridging the gap between object detection outputs and decision support outputs.
- A quantitative feasibility study of edge deployment on a Raspberry Pi 4 using the NCNN inference framework (≈ 1.3 FPS), providing concrete evidence regarding the operational limits of low-cost on-device monitoring for smallholder farming contexts.

The results demonstrate that simplified damage representations combined with optimized UAV-based detection pipelines can improve robustness while maintaining computational feasibility for near-real-time agricultural monitoring.

2. Materials and Methods

2.1. Study Area and Field Campaigns

Field data were collected in maize (*Zea mays*) cultivation areas located in the state of Yucatán, Mexico. Two field campaigns were conducted during different agricultural cycles in order to capture environmental variability and improve dataset robustness.

The first exploratory campaign was carried out in 2023 in the municipality of Muna. This region is characterized by traditional milpa agricultural systems, where maize is frequently cultivated together with other species and secondary vegetation. Such heterogeneity generates complex visual backgrounds that represent realistic conditions for computer vision models.

A second campaign was conducted in 2024 in the municipality of Kantunil, located approximately 75 km east of Mérida. In this region, maize is cultivated in family-managed plots that frequently maintain native seed varieties and controlled irrigation systems. These conditions provided complementary scenarios to the initial surveys, enabling image acquisition under diverse agronomic conditions including variations in plant density, weed presence, soil background, and cultivation practices.

2.2. Aerial Image Acquisition

Aerial imagery was collected using a DJI Mavic Air 2S unmanned aerial vehicle (SZ DJI Technology Co., Ltd., Shenzhen, China) equipped with a 20 MP Hasselblad RGB camera with a 1-inch CMOS sensor.

During the 2023 surveys in Muna, approximately 100 images were captured per plot with a spatial resolution of 5472×3078 pixels. Image acquisition was performed with a longitudinal and lateral overlap of 30% to ensure adequate spatial coverage of the plots.

Flights were conducted at altitudes of 6 and 7 m above ground level (AGL). At a flight altitude of 6 m, each frame covered an approximate ground area of 12×6.7 m, capturing up to 60 maize plants per image. The selected plots corresponded to maize plants in phenological stages V2 to V4, which represent the period of highest susceptibility to damage caused by *Spodoptera frugiperda*.

An example of a UAV image captured during the field surveys is presented in Figure 1.



Figure 1. Example of UAV image acquired at 6 m altitude during the field surveys in Muna. Each frame covers approximately 12×6.7 meters and may contain up to 60 maize plants.

2.3. Methodological Improvements in the 2024 Campaign

Based on the experience obtained during the exploratory surveys, several methodological improvements were implemented in the 2024 campaign conducted in Kantunil to increase spatial accuracy and improve dataset quality.

First, Ground Control Points (GCPs) were installed across the experimental plots. These markers consisted of high-contrast red and white grid targets measuring 24×35 cm. Their geographic coordinates were recorded using a sub-metric GPS receiver with an estimated positional precision of approximately 30 cm.

Second, a metric reference element was incorporated into the field measurements. A 20 m graduated rope with alternating black and white segments at one-meter intervals was deployed within the plots to provide a reliable reference for scale calibration.

Third, controlled multi-altitude image acquisition was implemented. Vertical flights were conducted over the GCP markers at heights ranging from 1 m to 10 m in 1 m increments. These measurements were repeated every 5 m along the sampling transects, generating a multi-resolution dataset suitable for evaluating the influence of Ground Sample Distance (GSD) on model performance.

2.4. Hierarchical Dataset Construction and Experimental Roles

The data processing workflow was structured into three specialized datasets, each fulfilling a distinct role in the development and validation of the proposed segmentation framework.

Dataset 1 (Environmental Context Dataset): This dataset consists of aerial frames selected to represent the environmental variability of the Yucatán agricultural landscape, including variations in soil background, illumination conditions, and secondary vegetation. Its primary role is to provide realistic field-scale context for evaluating model robustness under heterogeneous agricultural conditions.

Dataset 2 (Instance Segmentation Training Dataset): High-resolution UAV imagery from Dataset 1 was subdivided into tiles of 684×615 pixels to facilitate efficient deep learning training while preserving plant morphology and spatial detail. This dataset, containing 668 images with complete maize instances, was specifically designed for training the PS1 segmentation model.

Dataset 3 (Diagnostic Damage Dataset): Since *S. frugiperda* damage is naturally sparse in large-scale UAV surveys, a curated dataset containing 707 images of affected plants was constructed to reduce class imbalance and concentrate diagnostic visual patterns such as leaf perforations and frass accumulation. This dataset was used for training the DD1 and DD2 damage detection models, providing the fine-scale features required for reliable foliar damage segmentation.

To address the challenge of detecting fine foliar features within high-resolution UAV imagery, the Slicing-Aided Hyper Inference (SAHI) method was employed. The original UAV images (5472×3078 pixels) were divided into 66 overlapping patches of 640×640 pixels. This patch size corresponds to the native input resolution of the YOLOv11 architecture, avoiding automatic resizing operations that could degrade the visibility of small damage features. An overlap ratio of 20% was applied between adjacent slices to prevent object truncation at patch boundaries and ensure accurate merging of detections during post-processing.

A summary of the datasets, their resolutions, and their corresponding experimental roles is presented in Table 1.

Table 1. Summary of the datasets used in this study.

Dataset	Images	Resolution	Purpose
Environmental Context Dataset	~100	5472×3078	Field-scale contextual variability
Instance Segmentation Training Dataset	668	684×615	PS1 model training
Diagnostic Damage Dataset	707	variable	DD1/DD2 model training
Test Dataset	199	variable	Independent evaluation

2.4.1. Dataset Split

The annotated dataset was divided into training, validation, and testing subsets following a standard protocol commonly used in computer vision experiments. Approximately 70% of the images were used for training, 20% for validation, and the remaining 10% were reserved for independent testing. The split was performed randomly while preserving class balance across the subsets. The independent test dataset used in the experiments contains 199 images with a total of 979 annotated instances.

2.4.2. Annotation Process

Image annotation was performed using the Roboflow platform through a semi-automated segmentation workflow. Bounding boxes were first manually defined to delimit regions of interest corresponding to maize plants and surrounding vegetation. Subsequently, the Segment Anything Model (SAM) was applied to generate segmentation masks, which were manually refined to ensure annotation accuracy.

For the plant segmentation dataset, two semantic classes were defined:

- Maize plants;
- Non-maize vegetation.

The resulting annotations consist of polygon-based segmentation masks describing the spatial extent of plant structures within each image. Examples of the resulting segmentation masks under different field conditions are shown in Figure 2.



Figure 2. Examples of segmentation masks generated during the annotation process. (a) Crop scene with abundant green secondary weeds. (b) Polyculture system of maize with squash plants. (c) Maize monoculture with the presence of grasses. Maize plant segmentation masks are highlighted in yellow, while non-maize vegetation is outlined in purple using general clustering approximations.

2.5. Image Preprocessing

RGB images were converted into the HSV (Hue, Saturation, Value) color space using the OpenCV library as a preprocessing step prior to CNN training. Empirical threshold ranges were defined to isolate vegetation regions: Hue (32–73), Saturation (18–255), and Value (79–255).

These parameters reduced the presence of non-target background elements such as soil and shadows, facilitating the manual annotation of segmentation masks and improving the visual separation between maize plants and surrounding vegetation under heterogeneous field conditions.

2.6. Data Augmentation

To increase the diversity of the training dataset and reduce the risk of overfitting, several augmentation techniques were applied to the annotated images. The dataset was expanded approximately threefold using the following transformations:

- Rotations of 90°, 180°, and 270°;
- Horizontal and vertical flips;
- Saturation variations up to $\pm 30\%$;
- Brightness variations up to $\pm 15\%$.

These transformations simulate variations in illumination conditions, camera orientation, and field perspectives commonly encountered during UAV surveys.

2.7. Model Training

Three segmentation models based on the YOLOv11 architecture were developed and evaluated. YOLOv11 was selected as the core segmentation architecture due to its balance between computational efficiency and fine-scale feature extraction capability, which is particularly important for detecting small foliar damage patterns in high-resolution UAV imagery. Compared with earlier YOLO variants such as YOLOv8, YOLOv11 incorporates optimized feature extraction modules and improved parameter efficiency, enhancing its ability to preserve spatial detail during segmentation tasks. Although YOLOv10 provides lower inference latency through its NMS-free design, its prediction strategy is less straightforward to integrate with the Slicing-Aided Hyper Inference (SAHI) framework, which relies on post-processing and merging of overlapping detections across image patches. In contrast, YOLOv11 provides stable compatibility with tiled inference workflows commonly required for large UAV imagery. Transformer-based detectors such as DETR [19] and RT-DETR variants [20] have demonstrated strong performance in global-context modeling through self-attention mechanisms; however, these architectures typically involve

higher computational requirements when processing large collections of high-resolution UAV image tiles [20,21].

Model training was performed in the Google Colab environment using NVIDIA Tesla T4 GPUs with 16 GB of VRAM. Figure 3 illustrates the general architecture of the YOLOv11n-seg model employed in this study, including the backbone, neck, and detection head components used for instance segmentation. The experiments were conducted using the YOLOv11n-seg configuration, corresponding to the nano version of the architecture with approximately 2.59 million parameters. Model training was performed using an input resolution of 640×640 pixels, a batch size of 16, and 250 training epochs. Optimization was carried out using the AdamW optimizer with a learning rate of 0.001667 and momentum of 0.9. To improve generalization and robustness under heterogeneous field conditions, augmentation techniques including Mosaic and Albumentations-based transformations such as Blur, ToGray, and CLAHE were applied during training. All experiments were executed using CUDA 12.8 support.

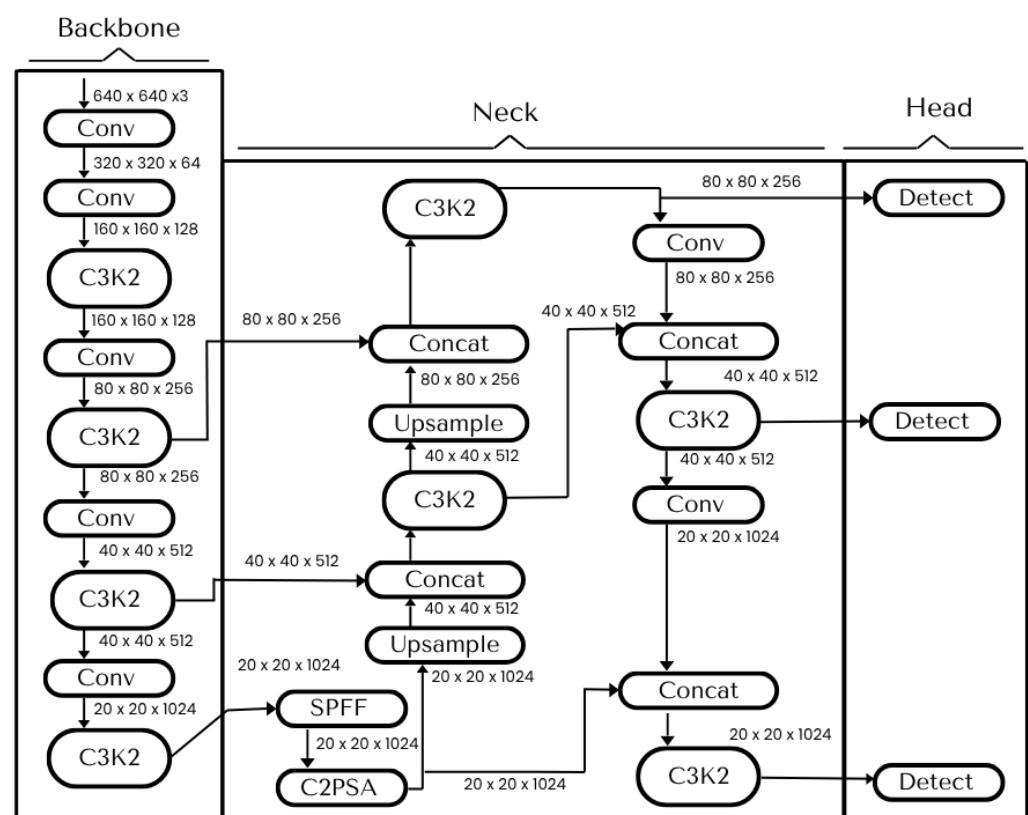


Figure 3. General architecture of the YOLOv11n-seg model used in this study, including the backbone, neck, and detection head components employed for instance segmentation.

The models were designed iteratively in order to evaluate different classification strategies:

- Model PS1 (Plant Segmentation): Distinguishes maize plants from secondary vegetation.
- Model DD1 (Damage Classification): Separates two types of feeding damage: leaf perforations and frass deposits.
- Model DD2 (Unified Damage Model): Combines all damage manifestations into a single class labeled Affected. This model was adopted as the final configuration after observing that overlapping damage patterns reduced the robustness of the multiclass approach.

All three models (PS1, DD1, DD2) share the same YOLOv11 instance segmentation architecture and training configuration. The differences between them are entirely data-centric and arise from the annotation scheme of their respective training datasets, rather than from any modification of the network design or training hyperparameters.

2.8. Evaluation Metrics

The performance of the proposed models was evaluated using standard metrics commonly adopted in object detection and instance segmentation tasks.

The Intersection over Union (IoU) metric measures the overlap between the predicted bounding box or segmentation mask and the corresponding ground truth annotation:

$$IoU = \frac{Area(B_p \cap B_{gt})}{Area(B_p \cup B_{gt})} \quad (1)$$

where B_p represents the predicted region and B_{gt} denotes the ground truth annotation.

Also, the precision and recall metrics were used to evaluate our model; they are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

where TP represents true positives, FP represents false positives, and FN denotes false negatives.

Furthermore, the mean Average Precision was used. It summarizes the precision–recall relationship of the detection model. In this sense, the prediction is considered a True Positive if the IoU exceeds a specified threshold. In this case, we used two evaluation configurations:

- mAP50: Average Precision computed at an IoU threshold of 0.5.
- mAP50-95: Average Precision averaged over IoU thresholds from 0.50 to 0.95 with increments of 0.05 following the COCO evaluation protocol [22].

2.9. Spatial Damage Mapping and Infestation Metrics

To complement the object detection and segmentation results, a spatial analysis of crop damage distribution was performed at the plot level. High-resolution UAV imagery was used to generate georeferenced orthomosaics of the monitored plots.

After inference, the segmentation masks predicted by the trained models were projected onto the orthomosaic. Each detected maize plant was considered an individual spatial observation. Plants containing at least one detected damaged leaf were labeled as affected plants.

The damage rate was defined as:

$$Damage\ Rate = \frac{N_{affected\ plants}}{N_{total\ plants}} \quad (4)$$

Additionally, infestation density was estimated as:

$$Damage\ Density = \frac{N_{affected\ plants}}{A} \quad (5)$$

where A represents the mapped area expressed in square meters.

The resulting spatial detections were aggregated to generate infestation maps, enabling the identification of clusters of affected plants and potential infestation hotspots.

2.10. Edge Deployment

To evaluate the feasibility of field deployment, the optimized model was converted to the NCNN inference format and deployed on a Raspberry Pi 4 equipped with 4 GB of RAM. The system used an 8 MP camera module connected to the Raspberry Pi for image acquisition. The software environment was configured in Python (version 3.10) to ensure compatibility between the inference engine and the embedded hardware.

Experimental tests showed that the system achieved approximately 1.3 frames per second during inference using 640×640 pixel inputs, demonstrating the feasibility of lightweight edge-based monitoring in agricultural environments.

Figure 4 summarizes the complete workflow implemented in this study.

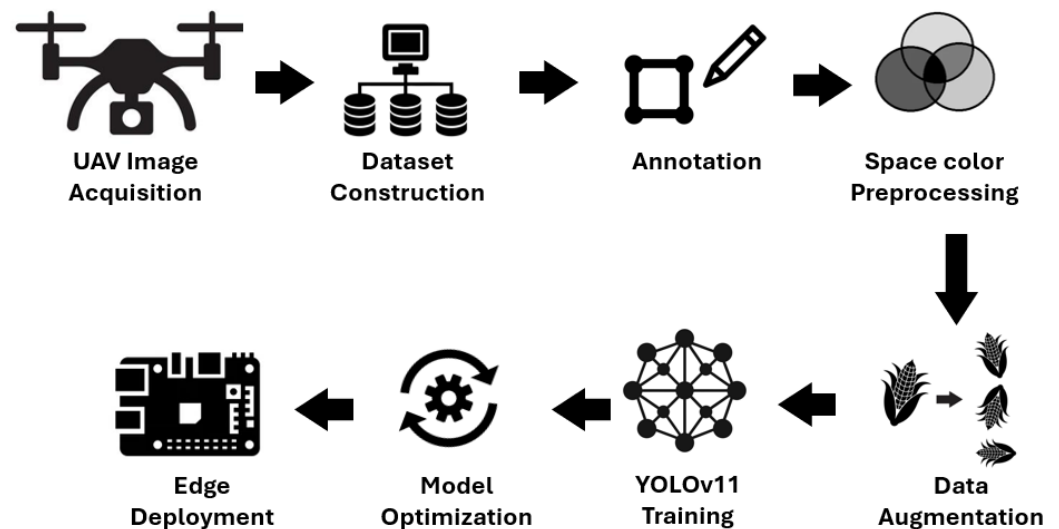


Figure 4. Overview of the methodological workflow proposed in this study. The pipeline includes UAV-based image acquisition, dataset construction, semi-automated annotation using SAM, preprocessing, data augmentation, YOLOv11 training, spatial analysis, and edge deployment.

While the proposed framework achieved an inference speed of approximately 1.3 FPS on the Raspberry Pi 4 using 640×640 pixel patches, this performance remains insufficient for true real-time UAV-based monitoring of full-resolution aerial imagery. A fundamental trade-off exists between spatial resolution and inference speed in precision agriculture applications.

Reliable detection of small foliar damage associated with *Spodoptera frugiperda* requires ultra-high-resolution UAV imagery (5472×3078 pixels) to preserve fine-scale morphological details. However, executing the complete SAHI pipeline on images of this size requires generating overlapping patches, performing multiple inference operations, and merging detections during post-processing, resulting in substantial computational latency on low-power embedded hardware such as the Raspberry Pi 4.

Consequently, the proposed framework follows a decoupled acquisition–processing strategy, where UAV and edge devices are dedicated to rapid image acquisition, while computationally intensive inference tasks are performed asynchronously on higher-performance workstations or cloud-based systems.

This strategy preserves the spatial resolution necessary for early pest detection while maintaining operational feasibility under realistic agricultural field conditions. Future advances in embedded AI accelerators and lightweight deep learning architectures may further improve real-time edge deployment capabilities.

3. Results

This section presents the performance of the three model configurations developed in this study: PS1 for plant segmentation, DD1 for multi-class damage detection, and DD2 for unified damage detection. Results are reported for both bounding-box detection and instance segmentation using mAP50, mAP50–95, precision, and recall.

Overall, the experiments show that reducing the complexity of the damage classification task improved model robustness. In particular, the unified damage formulation adopted in DD2 produced more stable and accurate results than the multi-class alternative.

3.1. Plant Segmentation Performance: Model PS1

PS1 was designed to identify maize plants and heterogeneous vegetation backgrounds in UAV images. During validation, the model achieved an overall segmentation mAP50 of 59%, with a precision of 69.0% and a recall of 59.7%. On the independent test set, the maize class achieved a segmentation mAP50 of 75%, whereas non-maize vegetation reached 36%.

The lower performance observed for non-maize vegetation is attributable to the large morphological variability of weeds and secondary plants, which makes this class less visually consistent than maize under field conditions.

Representative outputs produced by PS1 are shown in Figure 5.

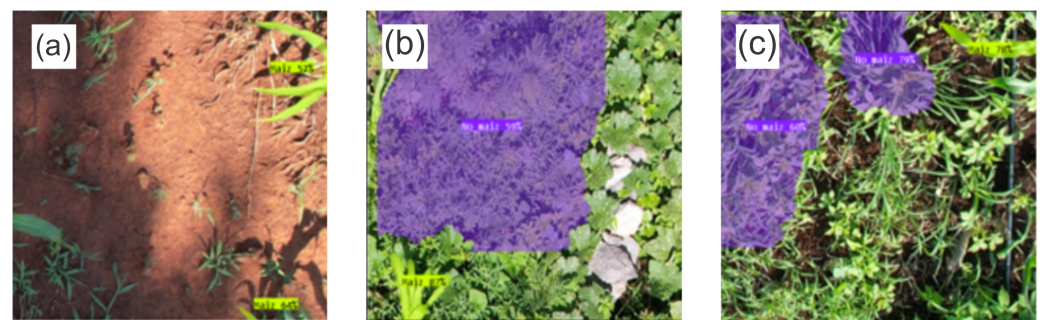


Figure 5. Representative outputs of model PS1 for plant segmentation under heterogeneous field backgrounds. (a) The model correctly segments the maize plant in a scenario without the presence of secondary vegetation. (b) The model performs accurate maize segmentation in the presence of dense secondary vegetation. (c) The model maintains correct maize segmentation under complex weed conditions; however, the segmentation of secondary vegetation is less accurate due to the complexity of the weed structures.

3.2. Multi-Class Damage Detection: Model DD1

DD1 was developed to detect maize plants together with two visible damage categories associated with *S. frugiperda*: holes and residues. The quantitative results obtained on the test set are summarized in Table 2.

The model achieved an overall mAP50 of 47.4% for bounding boxes and 40.6% for masks. The maize class maintained strong performance, with a box mAP50 of 89.8% and a mask mAP50 of 71.7%. In contrast, the damage-related classes were substantially more difficult to detect. The holes class obtained 21.1% mAP50 for both boxes and masks, whereas the residues class reached 31.4% for boxes and 29.0% for masks.

These results indicate that separating visually similar lesion types from UAV imagery is challenging, particularly when the damaged regions occupy only a small portion of the leaf surface and appear under variable illumination and background conditions.

Table 2. Bounding-box detection performance of model DD1 on the independent test set.

Class	Images	Instances	Precision	Recall	mAP50	mAP50–95
Overall	148	571	0.450	0.443	0.474	0.325
Maize	148	281	0.774	0.890	0.898	0.673
Residues	148	231	0.440	0.370	0.314	0.192
Holes	148	59	0.136	0.068	0.211	0.109

Figure 6 presents representative prediction results generated by DD1 on images from the independent test dataset.

**Figure 6.** Representative inference results obtained with model DD1 on images from the independent test dataset.

3.3. Unified Damage Detection: Model DD2

As noted in Section 2.7, DD1 and DD2 share the same YOLOv11 instance segmentation architecture and training configuration; the only difference between the two models lies in the annotation scheme of their training sets. DD1 was trained as a three-class model (*Maize*, *Holes*, *Residues*), whereas DD2 was trained as a two-class model in which both damage symptoms were merged into a single Affected class to remove the boundary ambiguity introduced by their visual overlap in field imagery. The performance gain reported below is therefore attributable to the simplified labeling protocol rather than to any change in network design.

Because DD1 showed limited ability to separate visually similar lesion types, a unified damage representation was introduced in DD2 by merging all damage manifestations into a single class labeled Affected. This simplified formulation reduced class ambiguity and improved detection stability.

Evaluation was performed on an independent test set containing 199 images and 979 annotated instances. The quantitative results are summarized in Tables 3 and 4.

For bounding-box detection, the model achieved an overall precision of 0.860, recall of 0.775, and mAP50 of 0.830. At the class level, maize reached a box mAP50 of 0.942, whereas the affected class achieved 0.717. For instance segmentation, the overall precision, recall, and mAP50 were 0.804, 0.714, and 0.733, respectively. At the class level, mask mAP50 reached 0.777 for maize and 0.700 for affected leaves.

These results confirm that the unified representation improved damage-related performance while preserving strong detection accuracy for maize plants.

Table 3. Bounding-box detection performance of model DD2 on the independent test set.

Class	Images	Instances	Precision	Recall	mAP50	mAP50–95
Overall	199	979	0.860	0.775	0.830	0.685
Maize	199	435	0.929	0.900	0.942	0.814
Affected	199	544	0.792	0.650	0.717	0.556

Table 4. Instance segmentation performance of model DD2 on the independent test set.

Class	Images	Instances	Precision	Recall	mAP50	mAP50–95
Overall	199	979	0.804	0.714	0.733	0.436
Maize	199	435	0.828	0.784	0.777	0.391
Affected	199	544	0.781	0.643	0.700	0.481

Representative inference results obtained with DD2 on the independent test dataset are shown in Figure 7.

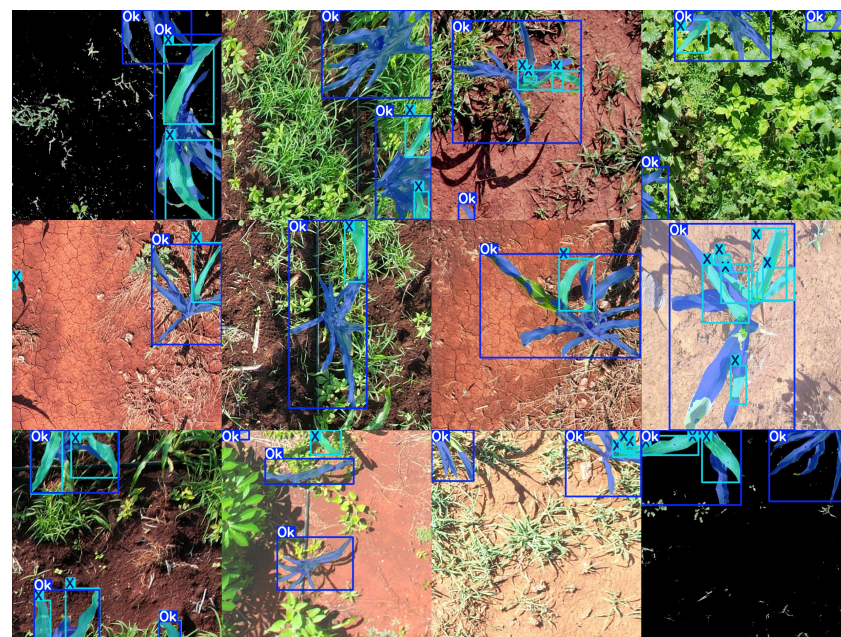


Figure 7. Representative inference results generated by model DD2 on the independent test dataset, illustrating the detection of maize plants and affected regions under heterogeneous field conditions.

To further examine threshold-dependent behavior, the F1 score was analyzed as a function of the confidence threshold. Figure 8 shows that DD2 exhibits a more stable confidence–F1 profile than DD1 across the evaluated threshold range.

3.4. Confusion Matrix and Detection Behavior

The normalized confusion matrix for DD2 is shown in Figure 9. Approximately 90% of maize instances were correctly classified, confirming the robustness of maize detection under heterogeneous field conditions.

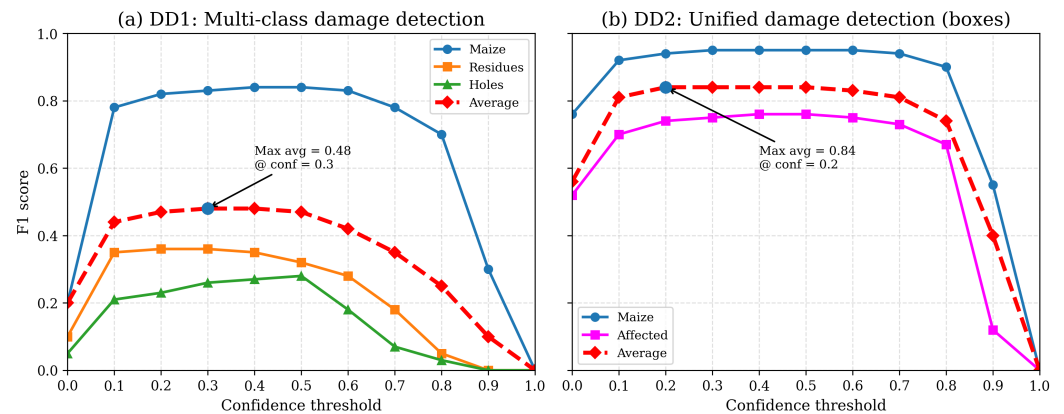


Figure 8. F1 score as a function of the confidence threshold for the evaluated damage detection models: (a) DD1 and (b) DD2.

For the affected class, the detection rate was approximately 69%. Most errors corresponded to missed detections rather than confusion with maize, indicating that the main remaining limitation lies in identifying small, fragmented, or partially occluded damage regions.

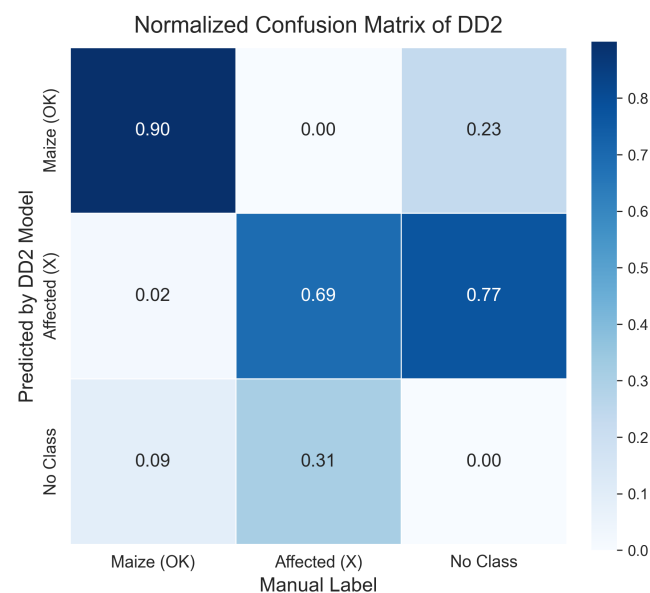


Figure 9. Normalized confusion matrix of model DD2 on the independent test set.

3.5. Spatial Mapping of Crop Damage

Beyond object-level detection, the proposed framework was used to reconstruct the spatial distribution of crop damage at parcel scale. By projecting plant detections and affected leaves onto georeferenced orthomosaics, the system enabled visualization of infestation patterns across the monitored field, as illustrated in Figure 10a.

The resulting maps revealed that affected plants were not uniformly distributed, but instead formed localized clusters, suggesting the presence of infestation hotspots. Figure 10b presents the corresponding Kernel Density Estimation (KDE) heatmap, where the high-density regions highlight potential intervention areas. The best qualitative correspondence with the manual field registry was obtained using imagery acquired at 6 m altitude and a confidence threshold of 50%.

These spatial outputs support the estimation of parcel-level indicators such as the proportion of affected plants and damage density, which are relevant for precision agriculture and localized pest management.

3.6. Feasibility Analysis of Edge Deployment

To assess the feasibility of running the proposed framework on low-cost embedded hardware, the trained YOLOv11 model was converted to the NCNN inference format and deployed on a Raspberry Pi 4 (4 GB RAM). On 8 MP images acquired from the on-board camera module, the system reached an inference speed of approximately 1.3 FPS, which is sufficient for sparse, on-demand monitoring tasks but not for real-time analysis of high-resolution UAV orthomosaics. A clear bottleneck was identified when integrating the SAHI tiled-inference strategy: the combined cost of slicing 20 MP images, running multiple inferences per frame, and reconstructing the segmentation masks exceeded the parallel-processing capacity of the ARM-based architecture, resulting in latencies that are impractical for full-scene mapping in the field.

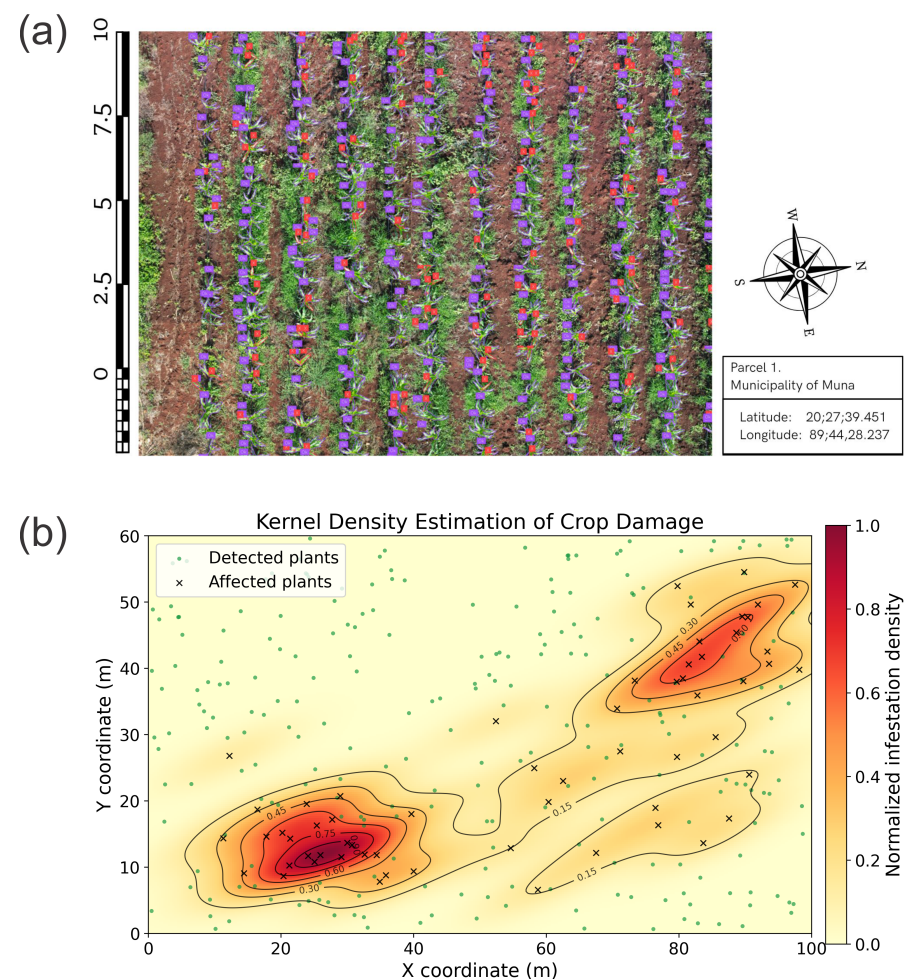


Figure 10. Spatial mapping of *Spodoptera frugiperda* damage in a monitored maize parcel. (a) Orthomosaic-based visualization of the segmentation results. Purple contours and bounding boxes represent detected Maize Plant instances, while red contours indicate detected Affected regions associated with foliar damage. The spatial scale is expressed in meters. (b) Kernel Density Estimation (KDE) heatmap derived from the spatial distribution of affected detections. The color gradient from yellow (low density) to red (high density) highlights localized infestation hotspots and potential intervention areas.

Consistent with this finding, the operational workflow proposed in this study is intentionally hybrid: high-resolution data acquisition is performed using the UAV platform, while SAHI-based inference is delegated to higher-capacity workstations or cloud environments. The edge benchmark reported here is therefore intended as a feasibility reference for projects with lower spatial-resolution or lower temporal-frequency requirements, rather than as evidence of fully on-device real-time deployment. Practical paths to improve on-device performance include the use of lighter YOLO variants, and the integration of dedicated hardware accelerators, which represent natural next steps for closer-to-real-time on-device inference.

4. Discussion

The results obtained in this study demonstrate the feasibility of combining UAV imagery with deep learning-based detection and segmentation models for monitoring maize crops and identifying foliar damage associated with *Spodoptera frugiperda*. The proposed framework successfully detected maize plants and damaged areas under real field conditions using high-resolution RGB imagery acquired at low UAV flight altitudes.

To contextualize the obtained results with respect to recent research in UAV-based agricultural monitoring, Table 5 summarizes representative studies that applied deep learning methods for crop damage detection or disease monitoring using aerial imagery. It is important to clarify the nature of the contribution of this work relative to its individual building blocks. YOLO-family instance segmentation models and slicing-aided inference (SAHI) are well-established techniques, and the present study does not aim to advance their underlying algorithmic design. Rather, the contribution lies in their joint instantiation within an operational pipeline applied to a previously unaddressed problem setting, fall armyworm damage detection on maize cultivated under traditional milpa systems, and in the empirical findings that emerge from this integration. Specifically, the comparison between multi-class and unified damage representations, the quantitative characterization of edge inference performance on a Raspberry Pi 4, and the conversion of detections into parcel-level infestation indicators via kernel density estimation are outcomes that cannot be obtained from any of the individual components in isolation, and that are directly informative for the design of UAV-based monitoring systems in similar agronomic contexts. These studies demonstrate the growing use of UAV platforms combined with computer vision techniques for high-resolution crop monitoring.

Table 5. Comparison of representative UAV-based crop monitoring studies using deep learning approaches.

Study	Crop	Sensor Type	Method	Reported Performance
Feng et al. (2022) [18]	Maize	RGB UAV	CNN patch classification (ResNeSt50)	89.4% accuracy (patch-level)
Zhang et al. (2023) [23]	Wheat	RGB UAV	CNN (Deeplabv3+)	Accuracy ~ 90%
Lu et al. (2024) [11]	Maize	RGB UAV	YOLOv5 plant detection	mAP50 > 0.85 (plant-level)
Chen et al. (2024) [16]	Maize	RGB UAV	RESAM-YOLOv8n (tassel detection)	mAP50 = 95.7%
Martins et al. (2024) [10]	Maize	RGB UAV	Semantic segmentation (SegFormer)	mIoU > 0.81
Dobosz et al. (2025) [17]	Maize	RGB + LiDAR UAV	DL segmentation + DSM analysis	92.9% accuracy (RGB DL)
This study	Maize	RGB UAV	YOLO-based segmentation	Competitive mAP ₅₀ performance

The comparison presented in Table 5 highlights the increasing adoption of UAV-based deep learning approaches for high-resolution agricultural monitoring. Recent studies have explored a wide variety of architectures, including CNN-based classifiers, YOLO detection frameworks, semantic segmentation models, and multimodal sensing strategies

combining RGB and multispectral imagery. Although direct quantitative comparison between studies remains challenging due to differences in crops, datasets, annotation protocols, spatial resolutions, and evaluation metrics, the proposed framework achieves competitive performance while relying exclusively on RGB UAV imagery and a lightweight YOLO-based segmentation pipeline. These results reinforce the feasibility of scalable UAV-based crop damage monitoring using computationally efficient deep learning architectures under heterogeneous field conditions.

Overall, the performance obtained in this study is comparable to that reported in recent UAV-based crop monitoring systems. Several studies have demonstrated that deep learning models trained on high-resolution aerial imagery can achieve detection accuracies above 85–90% for crop disease or damage identification tasks [9,23]. Similarly, UAV-based segmentation approaches have reported Intersection over Union (IoU) values above 0.80 for damaged crop mapping [24]. Although the present work relies exclusively on RGB imagery, the obtained detection performance falls within the range reported by recent studies using both RGB and multispectral sensors.

Recent UAV-based detection studies have increasingly explored architectural enhancements aimed at improving small-object recognition in complex aerial scenes, which is directly relevant to the detection of *S. frugiperda* lesions. Multi-scale feature fusion combined with attention mechanisms has been shown to outperform standard YOLO variants when small targets dominate the scene [25], while shallow detection heads and tile-based inference over UAV orthomosaics have produced similar improvements in instance segmentation tasks [26]. Although these works target different application domains, they collectively support the methodological choices adopted in this study: high-resolution feature maps and slicing-aided inference are particularly beneficial when targets are small or weakly contrasted, which is consistent with the difficulty observed for fragmented and partially occluded damage regions in our experiments.

The most important methodological finding of this work is that simplifying the damage classification scheme significantly improves model performance. The multi-class damage detection configuration (Model DD1) attempted to differentiate between several types of foliar lesions caused by pest activity. However, this configuration produced lower segmentation stability and higher confusion between visually similar damage categories. In contrast, the simplified model (Model DD2), which merged all damage types into a single class, produced more robust detection results and improved segmentation accuracy.

These results suggest that, for UAV imagery with limited spatial resolution and complex agricultural backgrounds, simplifying the damage taxonomy may improve detection reliability. When multiple damage classes exhibit similar visual patterns, forcing the model to distinguish between subtle lesion types can increase classification noise and reduce overall detection performance. A unified damage representation can therefore provide a more stable solution for large-scale crop monitoring pipelines where reliable damage detection is more important than detailed lesion categorization.

The transition from a multi-symptom classification Holes and Residues to a unified diagnostic class Affected leaves was driven by the need to resolve the semantic and spatial overlap between the two original damage classes. In field conditions, *S. frugiperda* damage typically manifests as a combination of both symptoms on a single leaf. By targeting the entire affected leaf as the detection unit, the model avoids the low-precision boundaries associated with distinguishing overlapping lesion patterns and gains a more stable geometric target for instance segmentation. The unification also has practical advantages beyond detection robustness: by providing instance-level localization of affected leaves, the framework supplies the inputs required to compute parcel-level damage

indicators which describe the spatial prevalence of infestation in terms that are more directly actionable for precision pest management than symptom-level classification alone.

This observation is consistent with findings reported in previous UAV-based crop monitoring studies. Environmental variability, illumination changes, soil background heterogeneity, and canopy structure can significantly affect the visual appearance of crop symptoms in aerial imagery [27]. Under such conditions, deep learning models may struggle to consistently differentiate between visually similar stress categories. Simplified classification strategies or hierarchical detection approaches may therefore provide better robustness in real agricultural environments.

Variable illumination conditions, heterogeneous soil backgrounds, and the presence of secondary vegetation represent important challenges for UAV-based agricultural monitoring under real field conditions. To improve robustness against this environmental variability, the proposed framework incorporated datasets collected across heterogeneous agricultural scenarios, including variations in illumination, soil reflectance, plant density, and surrounding vegetation. Additional robustness was introduced through HSV-based preprocessing and augmentation strategies involving brightness and saturation variations, geometric transformations, and image flipping operations. These procedures increase the model's tolerance to visual variability commonly encountered during UAV-based image acquisition. Future work may further improve environmental robustness through adaptive illumination normalization techniques, domain adaptation strategies, and the integration of multispectral information capable of providing complementary spectral features under challenging field conditions.

The results also reinforce the increasing importance of UAV platforms as sensing tools for precision agriculture. Compared with satellite remote sensing systems, UAVs provide significantly higher spatial resolution, enabling plant-level observation of crop conditions. Recent reviews on UAV applications in agriculture highlight their effectiveness for pest detection, crop stress monitoring, and damage mapping in precision agriculture systems [3]. High-resolution UAV imagery allows localized damage patterns to be identified at early stages, which can support targeted pest management strategies and reduce unnecessary pesticide applications.

Another relevant aspect of the proposed framework is the use of RGB imagery as the primary sensing modality. Although multispectral and hyperspectral sensors can improve early stress detection by capturing vegetation indices and spectral signatures, these systems typically involve higher acquisition costs and more complex processing pipelines. Several studies have demonstrated improved disease detection performance when multispectral data are combined with deep learning models [9]. However, RGB-based monitoring systems remain attractive for scalable agricultural monitoring because they rely on widely available commercial UAV platforms and simpler data acquisition workflows. Recent systematic reviews [28,29] confirm that UAV-based deep learning for crop disease and pest monitoring has rapidly expanded since 2019, with RGB sensors dominating practical applications because of their availability, low cost, and high spatial resolution. Multispectral and hyperspectral sensors are reported to improve early or visually subtle stress detection, but at the cost of more complex acquisition and processing pipelines. The framework presented in this study remains aligned with the dominant RGB-based trend identified in these reviews and complements it by providing instance-level outputs and georeferenced infestation indicators that are still rare in the surveyed literature.

While multispectral and hyperspectral sensors may enable earlier detection of physiological stress before visible tissue damage becomes apparent, these technologies typically involve higher acquisition costs, increased processing complexity, and more specialized calibration procedures. In contrast, the proposed framework was intentionally

designed to prioritize accessibility, scalability, and cost-effectiveness through the use of standard RGB imagery acquired with commercially available UAV platforms. Although RGB-based monitoring primarily relies on visible morphological symptoms, the obtained results demonstrate that high-resolution RGB imagery can still provide highly informative spatial and structural features for pest detection under real agricultural conditions. Future research will explore the integration of multispectral sensing to complement the current RGB-based approach with enhanced early physiological stress detection capabilities.

The reproducibility of the proposed framework is supported by the use of a standardized UAV acquisition protocol, including explicitly defined flight altitudes, image overlap, and acquisition conditions using a commercially available UAV platform. In addition, the manuscript describes the complete dataset construction and annotation workflow, the HSV-based preprocessing strategy, and the YOLOv11 + SAHI training and inference pipeline, enabling replication using standard computational resources.

Scalability is supported by the evaluation performed under heterogeneous field conditions using an independent test dataset containing 199 UAV images acquired under varying illumination, soil background, and vegetation density conditions. Furthermore, the integration of slicing-based processing through SAHI enables the analysis of large orthomosaic imagery, while the successful deployment on a low-cost Raspberry Pi 4 demonstrates the feasibility of extending the framework toward scalable and cost-effective agricultural monitoring scenarios.

Recent developments in UAV-based crop monitoring have also explored a variety of deep learning architectures, including transformer-enhanced object detection models and hybrid frameworks integrating structural and spectral information [10]. These approaches aim to improve the detection of small or subtle crop damage patterns in aerial imagery. Nevertheless, increasing model complexity may introduce additional computational requirements, which can limit deployment in edge computing environments or real-time monitoring applications.

Despite the promising results obtained in this work, several limitations should be acknowledged. First, the dataset used in this study was collected within a limited geographic region and under specific environmental conditions. Previous studies have shown that models trained on single-region datasets may exhibit reduced generalization when applied to different agricultural environments [27]. Future work should therefore incorporate multi-location and multi-season datasets to improve model robustness.

While the proposed dataset captures the environmental variability of the Yucatán agricultural region, broader deployment of the framework requires addressing domain shifts associated with variations in soil reflectance, illumination conditions, vegetation structure, and crop morphology across geographic regions. Transfer learning and domain adaptation strategies may help mitigate these effects by enabling the adaptation of pre-trained models to new agricultural environments using smaller locally annotated datasets. In this context, the pre-trained DD2 model could support AI-assisted annotation and pseudo-label generation on local UAV imagery, reducing manual labeling requirements during dataset adaptation. Fine-tuning the model on these curated datasets may facilitate efficient adaptation to new regions while preserving previously learned damage-related representations. Furthermore, given the polyphagous nature of *S. frugiperda*, the framework may also be extended to related crops such as sorghum and rice by retraining the host-plant detection component while maintaining the learned representations associated with foliar damage detection. Across future transfer-learning applications, maintaining a consistent ground sample distance (GSD) threshold remains critical to preserve the spatial scale required for reliable fine-scale lesion detection.

Second, the current framework focuses primarily on binary damage detection rather than detailed damage severity estimation. Recent research has emphasized the importance of incorporating severity analysis into crop monitoring systems to support agronomic decision-making and yield impact assessment. Integrating damage severity estimation or temporal monitoring of infestation progression represents an important direction for future research.

Beyond binary detection, the proposed framework may be extended toward quantitative severity assessment using the instance segmentation masks generated by the DD2 model. By analyzing the spatial distribution and relative proportion of affected regions within the maize canopy, the methodology could estimate severity-related indicators associated with pest pressure intensity and canopy stress.

One possible strategy involves quantifying the density of affected detections relative to the physical area of the host canopy, enabling normalized estimation of infestation pressure independently of plant size. Another complementary indicator may be derived from the proportion of canopy area associated with affected regions, providing an estimate of the spatial extent of visible stress symptoms caused by *S. frugiperda*. Because the proposed framework segments entire affected leaf regions rather than isolated microscopic lesions, these measurements may serve as practical indicators of canopy areas under active pest stress.

An additional extension of the framework involves incorporating multi-temporal UAV surveys throughout the crop phenological cycle. By tracking the temporal evolution of these severity-related indicators across sequential image acquisitions, the framework could support the modeling of infestation progression dynamics and contribute to predictive pest monitoring and localized management strategies.

Another limitation relates to the exclusive use of RGB imagery. While RGB cameras provide a cost-effective sensing solution, multispectral and hyperspectral sensors can capture additional spectral information that may improve early detection of plant stress before visible symptoms appear. Future research should explore multi-sensor UAV platforms combining RGB and spectral data to improve detection capabilities.

Future work should also investigate lightweight deep learning architectures optimized for edge deployment. Real-time crop monitoring using onboard inference on UAVs or embedded devices could significantly improve operational efficiency in large-scale agricultural monitoring systems.

Beyond detection, recent work has explored UAV-based intervention protocols for *S. frugiperda* control, including the optimization of spray parameters and flight altitude for precise insecticide deposition in the maize whorl [30,31]. The framework proposed in the present study can be naturally integrated with such intervention protocols to support a complete monitoring–decision–intervention workflow: instance-level damage detection and parcel-level infestation mapping provide the spatial inputs required to trigger and target site-specific UAV spraying, thereby enabling closed-loop precision pest management within the same UAV platform.

Overall, the findings of this study contribute to the growing body of research demonstrating the potential of UAV-based deep learning systems for high-resolution crop monitoring. The proposed methodology provides a practical and scalable framework for detecting maize plants and identifying pest-related damage under real field conditions, supporting the development of precision agriculture technologies capable of improving crop monitoring and pest management strategies.

5. Conclusions

This study presented an integrated information processing framework for the detection and spatial mapping of *Spodoptera frugiperda*-induced damage on maize cultivated under traditional milpa systems. The contribution of the work resides not in the individual algorithmic components—YOLO-based instance segmentation and slicing-aided inference are by themselves well-established—but in their empirical integration with region-specific dataset construction, georeferenced infestation mapping, and edge-deployment feasibility analysis, all instantiated and validated on the same operational case study.

Experimental results demonstrated that the proposed framework is capable of accurately detecting maize plants and identifying foliar damage associated with *Spodoptera frugiperda* under real field conditions. Among the evaluated model configurations, the unified damage representation (DD2) provided the most robust performance, confirming that simplifying the damage taxonomy can improve detection reliability when processing UAV imagery with heterogeneous backgrounds and limited spatial resolution.

In addition to object-level detection, the integration of georeferenced outputs enabled the reconstruction of spatial damage patterns within agricultural parcels. The resulting density maps revealed localized infestation clusters, highlighting the potential of the framework to support spatially explicit pest monitoring and precision agriculture decision-making.

The deployment experiments conducted on a Raspberry Pi device also demonstrated the feasibility of low-cost edge-based inference, suggesting that hybrid architectures combining edge data acquisition with remote processing may represent a practical solution for operational monitoring systems.

Future research should focus on expanding the dataset to multiple geographic regions and growing seasons, incorporating multi-sensor UAV platforms that combine RGB and spectral data, and developing lightweight deep learning models optimized for real-time edge deployment. These developments could further improve the scalability and operational applicability of UAV-based crop monitoring systems.

Author Contributions: Conceptualization, D.M.; methodology, E.C.-P.; software, E.C.-P.; validation, A.C.-G. and E.C.-P.; formal analysis, E.C.-P.; investigation, A.C.-G.; resources, D.M.; data curation, A.C.-G. and E.C.-P.; writing—original draft preparation, E.C.-P.; writing—review and editing, D.M.; visualization, E.C.-P.; supervision, D.M.; project administration, D.M.; funding acquisition, D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and analyzed during this study are not publicly available due to ongoing research activities but are available from the corresponding author upon reasonable request.

Acknowledgments: Alejandro Carrillo-Gómez acknowledges the scholarship provided by the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Government of Mexico, which supported his master's studies and the development of this research work.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

UAV	Unmanned Aerial Vehicle
CNN	Convolutional Neural Network
IoU	Intersection over Union
mAP	Mean Average Precision
RGB	Red–Green–Blue
SAHI	Slicing-Aided Hyper Inference
FPS	Frames Per Second
KDE	Kernel Density Estimation
PS1	Plant Segmentation Model 1
DD1	Damage Detection Model 1 (Multi-class)
DD2	Damage Detection Model 2 (Unified Damage)

References

1. Yang, C. Remote Sensing and Precision Agriculture Technologies for Crop Disease Detection and Management with a Practical Application Example. *Engineering* **2020**, *6*, 528–532. [\[CrossRef\]](#)
2. Wu, B.; Zhang, M.; Zeng, H.; Tian, F.; Potgieter, A.B.; Qin, X.; Yan, N.; Chang, S.; Zhao, Y.; Dong, Q.; et al. Challenges and opportunities in remote sensing-based crop monitoring: A review. *Natl. Sci. Rev.* **2023**, *10*, nwac290. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Velusamy, P.; Rajendran, S.; Mahendran, R.K.; Naseer, S.; Shafiq, M.; Choi, J.G. Unmanned Aerial Vehicles (UAV) in Precision Agriculture: Applications and Challenges. *Energies* **2022**, *15*, 217. [\[CrossRef\]](#)
4. Osco, L.P.; Marcato Junior, J.; Marques Ramos, A.P.; de Castro Jorge, L.A.; Fatholahi, S.N.; de Andrade Silva, J.; Matsubara, E.T.; Pistori, H.; Gonçalves, W.N.; Li, J. A review on deep learning in UAV remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102456. [\[CrossRef\]](#)
5. Kouadio, L.; El Jarroudi, M.; Belabess, Z.; Laasli, S.E.; Roni, M.Z.K.; Amine, I.D.I.; Mokhtari, N.; Mokrini, F.; Junk, J.; Lahlali, R. A Review on UAV-Based Applications for Plant Disease Detection and Monitoring. *Remote Sens.* **2023**, *15*, 4273. [\[CrossRef\]](#)
6. Domingues, T.; Brandão, T.; Ferreira, J.C. Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey. *Agriculture* **2022**, *12*, 1350. [\[CrossRef\]](#)
7. Khan, A.; Sohail, A.; Zahoor, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [\[CrossRef\]](#)
8. Dolatabadian, A.; Neik, T.X.; Danilevicz, M.F.; Upadhyaya, S.R.; Batley, J.; Edwards, D. Image-based crop disease detection using machine learning. *Plant Pathol.* **2025**, *74*, 18–38. [\[CrossRef\]](#)
9. Logavito, G.; Horanont, T.; Thapa, A.; Intarat, K.; Wuttiwong, K.O. Field-scale detection of Bacterial Leaf Blight in rice based on UAV multispectral imaging and deep learning frameworks. *PLoS ONE* **2025**, *20*, e0314535. [\[CrossRef\]](#)
10. Martins, J.A.C.; Hisano Higuti, A.Y.; Pellegrin, A.O.; Juliano, R.S.; de Araujo, A.M.; Pellegrin, L.A.; Liesenberg, V.; Ramos, A.P.M.; Gonçalves, W.N.; Sant’Ana, D.A.; et al. Assessment of UAV-Based Deep Learning for Corn Crop Analysis in Midwest Brazil. *Agriculture* **2024**, *14*, 2029. [\[CrossRef\]](#)
11. Lu, C.; Nnadozie, E.; Camenzind, M.P.; Hu, Y.; Yu, K. Maize plant detection using UAV-based RGB imaging and YOLOv5. *Front. Plant Sci.* **2024**, *14*, 1274813. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Hamzenejadi, M.H.; Mohseni, H. Fine-tuned YOLOv5 for real-time vehicle detection in UAV imagery: Architectural improvements and performance boost. *Expert Syst. Appl.* **2023**, *231*, 120845. [\[CrossRef\]](#)
13. Zhao, M.; Wang, D.; Yan, Q.; Li, Z.; Liu, X. UAV-Multispectral Based Maize Lodging Stress Assessment with Machine and Deep Learning Methods. *Agriculture* **2025**, *15*, 36. [\[CrossRef\]](#)
14. Miri Rekavandi, A.; Rashidi, S.; Boussaid, F.; Hoefs, S.; Akbas, E.; Bennamoun, M. Transformers in Small Object Detection: A Benchmark and Survey of State-of-the-Art. *ACM Comput. Surv.* **2025**, *58*, 64. [\[CrossRef\]](#)
15. Ahmed, W.A.; Yan, D.; Hamed, J.O.; Olatoyinbo, S.F. Advances in UAV-based deep learning for cassava disease monitoring and detection: A comprehensive review of models, imaging techniques, and agricultural applications. *Smart Agric. Technol.* **2025**, *12*, 101400. [\[CrossRef\]](#)
16. Chen, J.; Fu, Y.; Guo, Y.; Xu, Y.; Zhang, X.; Hao, F. An improved deep learning approach for detection of maize tassels using UAV-based RGB images. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *130*, 103922. [\[CrossRef\]](#)
17. Dobosz, B.; Gozdowski, D.; Koronczok, J.; Žukovskis, J.; Wójcik-Gront, E. Detection of Crop Damage in Maize Using Red–Green–Blue Imagery and LiDAR Data Acquired Using an Unmanned Aerial Vehicle. *Agronomy* **2025**, *15*, 238. [\[CrossRef\]](#)
18. Feng, J.; Sun, Y.; Zhang, K.; Zhao, Y.; Ren, Y.; Chen, Y.; Zhuang, H.; Chen, S. Autonomous Detection of Spodoptera frugiperda by Feeding Symptoms Directly from UAV RGB Imagery. *Appl. Sci.* **2022**, *12*, 2592. [\[CrossRef\]](#)

19. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In *Proceedings of the Computer Vision—ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer: Cham, Switzerland, 2020; pp. 213–229. [\[CrossRef\]](#)
20. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. Dets beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 16–22 June 2024; pp. 16965–16974. [\[CrossRef\]](#)
21. Zhou, Y.; Wei, Y. UAV-DETR: An Enhanced RT-DETR Architecture for Efficient Small Object Detection in UAV Imagery. *Sensors* **2025**, *25*, 4582. [\[CrossRef\]](#)
22. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Proceedings of the Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2014; pp. 740–755. [\[CrossRef\]](#)
23. Zhang, K.; Zhang, R.; Yang, Z.; Deng, J.; Abdullah, A.; Zhou, C.; Lv, X.; Wang, R.; Ma, Z. Efficient Wheat Lodging Detection Using UAV Remote Sensing Images and an Innovative Multi-Branch Classification Framework. *Remote Sens.* **2023**, *15*, 4572. [\[CrossRef\]](#)
24. Aszkowski, P.; Kraft, M.; Drapikowski, P.; Pieczyński, D. Estimation of corn crop damage caused by wildlife in UAV images. *Precis. Agric.* **2024**, *25*, 2505–2530. [\[CrossRef\]](#)
25. Khujamatov, H.; Muksimova, S.; Abdullaev, M.; Cho, J.; Jeon, H.S. Advanced Insect Detection Network for UAV-Based Biodiversity Monitoring. *Remote Sens.* **2025**, *17*, 962. [\[CrossRef\]](#)
26. Bai, Z.; Ji, L.; Tang, H.; Qiu, J.; Kang, S.; Liu, C.; Bian, Z. Rapid Detection and Segmentation of Landslide Hazards in Loess Tableland Areas Using Deep Learning: A Case Study of the 2023 Jishishan Ms 6.2 Earthquake in Gansu, China. *Remote Sens.* **2025**, *17*, 2667. [\[CrossRef\]](#)
27. Manoj, H.M.; Shanthi, D.L.; Lakshmi, B.N.; Archana, K.J.; Venkata, Naga, Jyothi, E.; Archana, K. AI-driven drone technology and computer vision for early detection of crop disease in large agricultural areas. *Sci. Rep.* **2026**, *16*, 2479. [\[CrossRef\]](#)
28. Radočaj, D.; Radočaj, P.; Plaščak, I.; Jurišić, M. Evolution of Deep Learning Approaches in UAV-Based Crop Leaf Disease Detection: A Web of Science Review. *Appl. Sci.* **2025**, *15*, 778. [\[CrossRef\]](#)
29. Zhu, H.; Lin, C.; Liu, G.; Wang, D.; Qin, S.; Li, A.; Xu, J.L.; He, Y. Intelligent agriculture: Deep learning in UAV-based remote sensing imagery for crop diseases and pests detection. *Front. Plant Sci.* **2024**, *15*, 1435016. [\[CrossRef\]](#)
30. Liu, Y.; Liang, X.; Wu, C.; An, X.; Wu, M.; Zhao, Z.; Li, Z.; Chen, Q. Optimization of spray operation parameters of unmanned aerial vehicle confers adequate levels of control of fall armyworm (*Spodoptera frugiperda*). *Front. Plant Sci.* **2025**, *16*, 1581367. [\[CrossRef\]](#)
31. Reyna-Bowen, J.L.; Murillo, L.; Cedeño, H.; Vera Montenegro, L.; Delgado-Moreira, M.I.; Velásquez Cedeño, S. Optimizing the Flight Altitude of the Agras T40 Drone for Controlling *Spodoptera frugiperda* in Maize (*Zea mays* L.) Plantations. *Agroindustrial Sci.* **2025**, *15*, 355–361. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.