agriculture

MDPI

Article

Assessment of UAV-Based Deep Learning for Corn Crop Analysis in Midwest Brazil

José Augusto Correa Martins [1,*], Alberto Yoshiriki Hisano Higuti [1], Aiesca Oliveira Pellegrin [2], Raquel Soares Juliano [2], Adriana Mello de Araújo [2], Luiz Alberto Pellegrin [2], Veraldo Liesenberg [3], Ana Paula Marques Ramos [4,5,*], Wesley Nunes Gonçalves [1], Diego André Sant'Ana [6], Hemerson Pistori [1,7] and José Marcato Junior [1]

1 Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Cidade Universitária, Av. Costa e Silva, Pioneiros 79070-900, MS, Brazil; alberto.yoshiriki@ufms.br(A.Y.H.H.); wesley.goncalves@ufms.br (W.N.G.); pistori@ucdb.br (H.P.); jose.marcato@ufms.br (J.M.J.)

2 Embrapa Pantanal, Rua 21 de Setembro, 1880, Corumbá, MS, 79320-900, Brazil; aiesca.pellegrin@embrapa.br (A.O.P.); raquel.juliano@embrapa.br (R.S.J.); adriana.araujo@embrapa.br (A.M.d.A.); luiz.pellegrin@embrapa.br (L.A.P.)

3 Forest Engineering Department, Santa Catarina State University, Avenida Luiz de Camões 2090, Lages 88520-000, SC, Brazil; veraldo.liesenberg@udesc.br

4 Environment and Regional Development Program, University of Western São Paulo, Rodovia Raposo Tavares, km 572, Bairro Limoeiro 19067-175, SP, Brazil

5 Agronomy Program, University of Western São Paulo, Rodovia Raposo Tavares, km 572, Bairro Limoeiro 19067-175, SP, Brazil

6 Instituto Federal de Mato Grosso do Sul, Campus Aquidauana, Street Amelia Arima, 222, Aquidauana 79200-000, MS, Brazil diego.santana@ifms.edu.br

7 Computer Science Department, Universidade Católica Dom Bosco, Av. Tamandaré, 6000, Campo Grande 79117-010, MS, Brazil

* Correspondence: jose.a@ufms.br (J.A.C.M.); whiteyouzeph@gmail.com or anaramos@unoeste.br (A.P.M.R.); Tel.: +55-67-996363737 (J.A.C.M.)

Citation: Martins, J.A.C.; Hisano Higuti, A.Y.; Pellegrin, A.O.; Juliano, R.S.; de Araújo, 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. https://doi.org/10.3390/agriculture14112029

Academic Editor: Gniewko Niedbała

Received: 27 September 2024
Revised: 24 October 2024
Accepted: 28 October 2024
Published: 11 November 2024

Abstract: Crop segmentation, the process of identifying and delineating agricultural fields or specific crops within an image, plays a crucial role in precision agriculture, enabling farmers and public managers to make informed decisions regarding crop health, yield estimation, and resource allocation in Midwest Brazil. The crops (corn) in this region are being damaged by wild pigs and other diseases. For the quantification of corn fields, this paper applies novel computer-vision techniques and a new dataset of corn imagery composed of 1416 256 $\times$ 256 images and corresponding labels. We flew nine drone missions and classified wild pig damage in ten orthomosaics in different stages of growth using semi-automatic digitizing and deep-learning techniques. The period of crop-development analysis will range from early sprouting to the start of the drying phase. The objective of segmentation is to transform or simplify the representation of an image, making it more meaningful and easier to interpret. For the objective class, corn achieved an IoU of 77.92%, and for background 83.25%, using DeepLabV3+ architecture, 78.81% for corn, and 83.73% for background using SegFormer architecture. For the objective class, the accuracy metrics were achieved at 86.88% and for background 91.41% using DeepLabV3+, 88.14% for the objective, and 91.15% for background using SegFormer.

Keywords: crop segmentation; drones; precision agriculture; semantic segmentation

1. Introduction

The process of crop segmentation, which entails the identification and delineation of agricultural fields or specific crops within an image, is a key factor for precision agriculture. It can empower farmers and managers with the data necessary to optimize crop health, yield predictions, and resource distribution decisions [1,2]. In the agricultural sector, the use of image-processing and computer-vision technologies has expanded significantly due to the decrease in equipment costs, the rise in computational capabilities, and a growing interest in methods of non-destructive food evaluation [3].

For the segmentation research field, the use of Convolutional Neural Networks (CNN) is a hot topic [2,4,5]. CNNs are a type of deep-learning model specifically designed for processing structured grid data, such as images. They are composed of layers that apply convolution operations, which, through back-propagation mechanisms, adaptively learn spatial hierarchies of features from input data. CNNs are useful because of their prowess in capturing spatial features within images if compared to other methods such as Random Forest and Support Vector Machines [6]. Nevertheless, the advent of transformer-based architectures—distinguished by their capacity to effectively model long-range dependencies in data sequences—has unveiled novel prospects for enhancing image-segmentation methodologies [7,8]. In light of the evolving landscape of precision agriculture and the necessity of better resource management, there is a demand for methodologies that are more accurate and efficient in extracting quantitative values about crops in the field. We can refer to some important information that can be obtained with semantic-segmentation methodologies, such as growth monitoring, yield estimation, and disease identification [2,9].

Corn is an important international commercialized crop that is set to become the most widely grown and traded crop in the coming decade [10,11]. Corn represents a cornerstone of Brazilian agriculture and thrives across the Midwest and Southern regions, mirroring the prominence of soybeans in the nation's agricultural portfolio [12]. Enhancing corn productivity hinges on methodical phase monitoring and the eradication of weeds and disease infestations, particularly during the crop's nascent stages, where the risk of production loss peaks.

Recently, transformer network architectures have made an important improvement in the computer-vision arena [13–15]. Unlike Convolutional Neural Networks (CNN), which focus on local feature extraction through convolutions, transformers excel at capturing global dependencies within data. This allows them to understand relationships across the entire image, leading to more comprehensive feature representations [13]. This article is dedicated to exploring how transformer network architectures and, for comparison, other CNNs can be applied to the semantic segmentation of crops in the selected study area of corn fields in Midwest Brazil. This research is being made on a temporally diverse dataset, meaning that it can be used in many stages of the development of corn.

Corn cultivation faces significant development challenges, particularly from wild pigs (Sus scrofa), which cause substantial damage by uprooting plants, consuming seeds, and trampling crops [16]. These invasive animals not only reduce crop yields but also spread diseases to both crops and livestock, leading to economic losses for farmers [17,18]. Addressing this issue is crucial for the sustainability of corn farming, especially in regions where wild pigs are prevalent and pose a recurring threat [19]. Additionally, research by Roda and Roda [20] highlights the broader ecological impacts of wild pigs, including a 45% decline in ground-nesting bird populations due to boar-foraging activities.

The integration of deep learning, UAV imagery, and agricultural monitoring forms an emerging research area focused on addressing agriculture's challenges through technology. Highlighted in studies such as [21], the research achieved an accuracy of more than 80% accuracy for crop yields, thus creating an efficient damage-identification workflow.

Recent advancements in deep learning have significantly enhanced the capabilities of semantic segmentation for a great variety of applications. Studies such as those by Chen et al. (2018) on DeepLabV3+ [22] and Xie et al. (2021) on SegFormer [13] propose robust deep-learning architectures and demonstrate the potential of these architectures to achieve state-of-the-art results on image-segmentation tasks. These kinds of models became known for their efficiency in handling high-resolution images and their effectiveness in extracting and processing spatial features. They have been pivotal in advancing monitoring techniques for many different objectives [23].

The deployment of UAVs for precision agriculture has been explored extensively in the literature. UAVs offer a unique vantage point, providing high-resolution imagery that is crucial for detailed crop analysis. Works such as [4,24–26] highlight the advantages of UAV-based imagery over traditional satellite images, including the ability to capture data at higher spatial and temporal resolutions, which is essential for tracking crop development and assessing damage [27].

The issue of wildlife causing damage to crops is a concern for agricultural communities worldwide. Research such as [28–30] sheds light on the extent of damage, emphasizing the need for effective management strategies. These studies explore various mitigation techniques, ranging from physical barriers to the application of deterrents, and suggest the potential for integrating UAV imagery and deep-learning models for early detection and assessment of wildlife-induced damages [23].

Despite extensive research in each area, a gap remains in integrating deep-learning models, UAV technology, and strategies to address negative impacts, particularly in crop monitoring. Our approach contributes to precision agriculture by offering adaptable methodologies for various crops and regions, advancing automated agricultural monitoring. Additionally, it provides a foundation for future research in semantic segmentation, enhancing efficiency in agricultural practices.

2. Materials and Methods

This section will present a succinct description of the techniques used in this research. For illustration, we present Figure 1. This figure shows an example step-by-step process for using remote-sensing data and deep learning to monitor and analyze the environment.

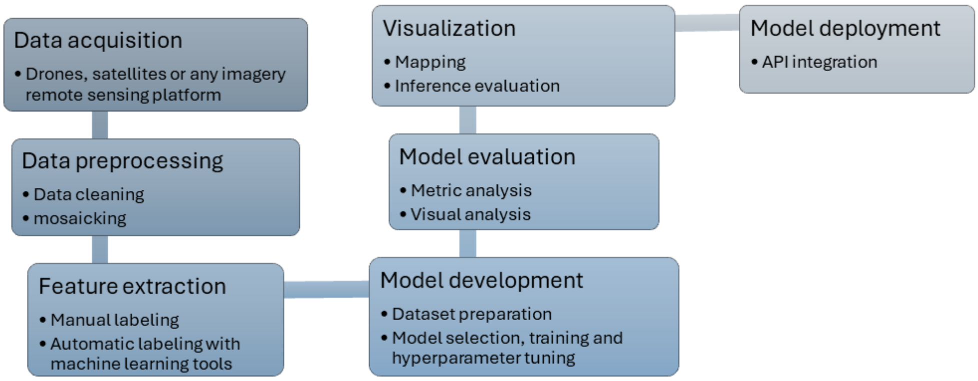


Figure 1. Fluxogram presenting the phases of the workflow.

The first part of the data acquisition was conducted with a UAV in the sites presented in Figure 2. In the preprocessing phase, the images were made into a mosaic using Metashape software [31], version 1.5.1 build 7618 (64 bit). Then, in the feature-extraction phase, the labeling was made with ilastik [32]. Ilastik is an interactive tool that makes it possible to interactively annotate datasets of images. This tool has a machine-learning engine running in the background that facilitates the laborious task of annotation. Ilastik with a human operator was used to segment the images of this research to create the labels. The following process of visualization and model evaluation and deployment was done using Qgis https://qgis.org, programming language Python https://www.python.org, and hosting providers such as GitHub https://github.com.

2.1. Study Area and Image Description

This research is conducted within the Cerrado Biome of the state of Mato Grosso do Sul, situated in the central-western region of Brazil, for the year 2023. Renowned for its significant agricultural potential, this area benefits from fertile soil, favorable climate, and abundant freshwater resources. Mato Grosso do Sul encompasses a diverse range of ecosystems, including savannas, tropical forests, and wetlands. The cerrado is considered to be an optimal area for the production of food, making it a hot spot for the heavily agricultural national economy. For our analysis, we utilized a collection of Unmanned Aerial Vehicle (UAV) mosaics covering corn plantations within the biome in a location inside the Mato Grosso do Sul state. The locations of the mosaics are illustrated in Figure 2.

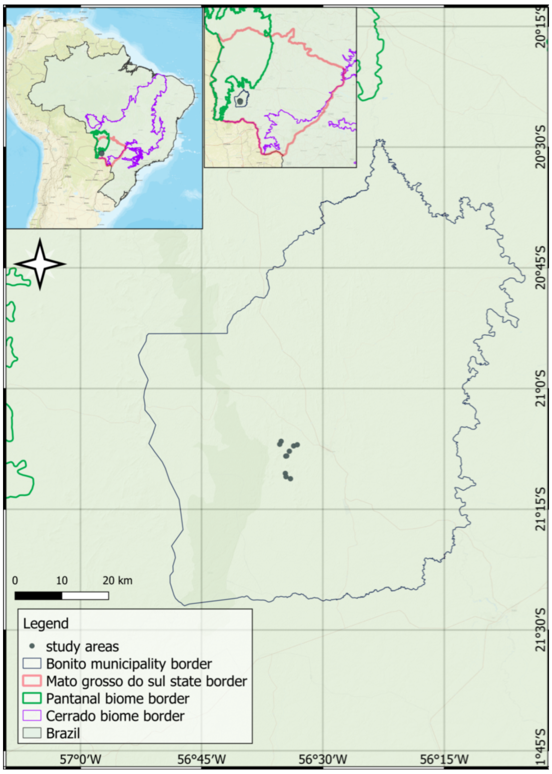


Figure 2. Location map of the study area and mosaics used for the research.

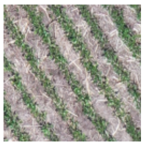
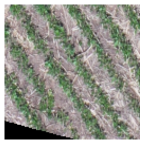
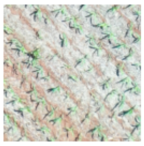
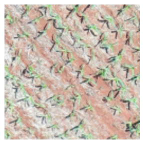
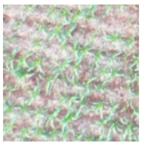
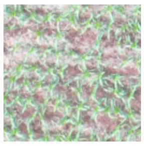
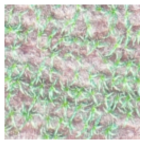
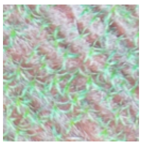
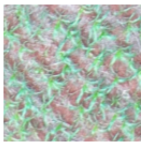
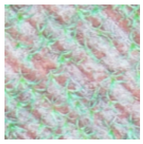
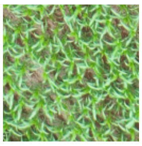
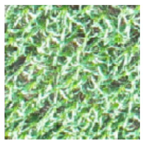
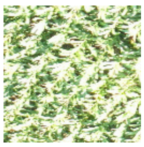
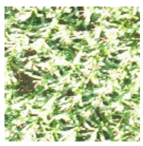
The UAV used for this research was a MAVIC 2 pro. The Mavic 2 Pro camera features a 1-inch CMOS sensor with 20 MP, an adjustable aperture from f/2.8 to f/11, and a 28 mm equivalent focal length. The camera offers an ISO range of 100–12,800 for photos. The average height of the UAV flight for this study was 40 m. Further specifications can be found at the link: https://www.dji.com/br/mavic-2/info, access on 1 July 2024.

Altitude and flight speed significantly impact image resolution. Flying closer to the target improves resolution, but higher speeds generally result in less overlap between images, leading to a mosaic with fewer data points. Balancing these factors is crucial for achieving high-quality imagery while ensuring sufficient overlap for constructing a detailed mosaic, which is key for accurate segmentation and analysis in this study.

With the data collected by the UAV, nine orthomosaics were created. The data were gathered across multiple growing seasons, offering a temporal depth that can enrich the understanding of crop dynamics in the region. This temporal component allows for the observation of crop-growth stages, yield variations, and the impact of environmental factors on agricultural productivity and the health of the environment. Corn growth can be monitored using a variety of scales, one of which is the BBCH scale [33,34]. The abbreviation BBCH derives from the names of the originally participating stakeholders: "Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie". The BBCH scale is utilized to identify the phenological development stages of plants [33,35]. BBCH scales have been established for various crop species, assigning identical codes to similar growth stages across different plants. The scale ranges from 0 Germination to 9 Senescence, which is the last stage of aging.

We present more technical data about the dataset in Table 1. Ground-sampling distance (GSD) represents the distance between the centers of two adjacent pixels. The area is represented in hectares (ha). BBCH is the corresponding BBCH scale and Image Sample is a presentation of what image is being represented at the chosen scale. We can see that we do not have scale mark of 9, and we have a different area of study for each scale. This will certainly pose an impact on the training phase of the model while the network is trying to identify the target object and will have more data on one kind of object and less on the other since an under-sampling was not performed.

Table 1. ID, GSD, Area, growth stages, and image samples of the data used in the study.

Image	GSD (cm)	Area (ha)	BBCH	Image Sample
1	2.10	1.03	1	  
2	2.60	1.71	2	  
3	3.00	1.99	3	  
4	3.00	1.98	5	  
5	2.00	1.32	4	  
6	2.55	1.63	5	  
7	3.00	1.90	5	  
8	1.85	1.09	7	  
9	1.91	1.03	8	  

2.2. Experimental Setup and Data Processing

2.2.1. Dataset

The process of obtaining annotations for the class of interest used in the dataset is detailed in the "Workflow" subsection. Based on this, the 10 orthomosaics of corn fields mentioned in the subsection "Study Area and Image Description" were cropped in patches of 256 × 256 pixels each. These cropped images constituted the dataset for this study, which was divided into training, validation, and test sets with proportions of 70%, 15%, and 15%, respectively. It is important to mention that images lacking any areas with the relevant class were removed from the dataset. Consequently, the dataset was composed of a training set with 992 image patches, a validation set with 212 image patches, and a test set with 212 image patches, each measuring 256 × 256 pixels.

2.2.2. Deep-Learning Architectures

Fully Convolutional Networks (FCN) are a type of Convolutional Neural Network (CNN) designed for pixel-wise prediction tasks such as semantic segmentation. Unlike traditional CNNs that end with fully connected layers, FCNs use only convolutional layers, allowing them to handle input images of arbitrary size. FCNs replace the fully connected layers with convolutional layers, enabling end-to-end learning and inference. The architecture involves a series of downsampling (pooling) layers to extract features, followed by upsampling layers to produce a dense prediction map that matches the input image size [36].

DeepLabV3+ is an advanced semantic-segmentation model that builds on the DeepLab family of models. It incorporates several improvements, such as the Atrous Spatial Pyramid Pooling (ASPP) module, which uses atrous (dilated) convolutions at multiple scales to capture multi-scale context, and the encoder–decoder structure. The encoder captures rich semantic information, while the decoder refines the segmentation results, especially around object boundaries. DeepLabV3+ achieves high performance on segmentation benchmarks by effectively balancing accuracy and computational efficiency [22].

SegFormer is a recent model designed for semantic segmentation that utilizes a transformer-based architecture. Unlike traditional CNN-based models, SegFormer employs transformers to capture long-range dependencies and global context effectively. It combines hierarchical features from multiple layers of the transformer encoder, ensuring both high-resolution details and semantic richness. SegFormer is known for its efficient design and strong performance across various benchmarks, demonstrating the potential of transformers in dense prediction tasks [13].

2.3. Hyperparameters

Various training strategies were implemented, which included modifications to the dataset and adjustments to each model's parameters. The goal was to evaluate and compare the effectiveness and advancements of the models in relation to these changes. In the SegFormer model, the Mix Transformer-B5 (MiT-B5) served as the backbone, and SegformerHead was set as the head, whereas the DeepLabv3+ model employed ResnetV1c as its backbone. Furthermore, the DepthwiseSeparableASPPHead was designated as the head for the DeepLabv3+ model, and the FCNHead was set as the auxiliary head. Lastly, the FCN model also used ResnetV1c as the backbone and FCNHead as the model's head.

Additionally, the training process utilized the AdamW optimizer [37] for SegFormer and DeepLabv3+ models and SGD optimizer [38] for the FCN model. The learning rate was kept consistent across all training variations, set at 6.0×10^{-4} for the SegFormer and 1×10^{-2} for the alternative models. All the models underwent training for a total of 40k iterations. Furthermore, all models were trained using CrossEntropy Loss, with the weight set to 1.0 for SegFormer and DeepLabV3+ and a weight of 0.4 for FCN throughout all training sessions. The model's performance was evaluated using the metrics accuracy and Intersection over Union (IoU) ensuring a comprehensive analysis, even for complex and imbalanced datasets.

2.4. Processing and Repository

The experiments were conducted on the Google Colaboratory (Colab) platform with Intel(R) Xeon(R) CPU @ 2.20 GHz, 16 GB memory and A100 GPU. We used Pytorch (Pytorch) and the semantic-segmentation toolbox MMsegmentation [39] for the model design. For future reproductions of similar procedures, a repository with the code configuration was created https://github.com/Jose-Augusto-C-M/crop_image_deeplearning_segmentation, access on 1 August 2024.

3. Results

3.1. Quantitative Evaluation

This section presents the quantitative results of the study, using the cited metrics to evaluate the performance of the proposed network architectures in segmenting crops at different developmental stages. The results for the inference on the test set are in Table 2. The table presents the median of the Intersection over Union (IoU) and Accuracy (Acc) metrics for three different semantic segmentation networks—Deeplabv3+, SegFormer, and FCN—across two classes—corn and background. mIoU is the mean IoU for the training checkpoints and mAcc is the mean accuracy for the training checkpoints. For each Network, we have 40,000 epochs, and the checkpoints are at 5000 epochs.

Table 2. Quantitative results for the selected networks. Blue colors represent the biggest values, and red colors are the worst.

Network	**fcn**		**DeepLabV3+**		**Segformer**	
Metric\Metric	**Corn**	**Back**	**Corn**	**Back**	**Corn**	**Back**
mIoU	65.14	73.27	80.62	83.88	81.72	84.73
mAcc	77.93	85.37	89.12	91.36	90.09	91.64

As shown in Table 2, Deeplabv3+ demonstrates strong performance for both the corn and background classes, achieving an mIoU of 80.62 for corn and 83.88 for background. The accuracy is particularly high for the background class, reaching 89.12%. SegFormer, on the other hand, exhibits the highest mIoU and mAcc for the corn class at 78.81, indicating an edge over Deeplabv3+ in this segmentation task. FCN lags behind the other two networks in both IoU and accuracy for both classes. Overall, SegFormer outperforms the other networks in terms of mIoU and mAcc for both classes, making it the best choice for segmentation tasks in this context. Deeplabv3+ is a strong performer, especially for the background class, where it achieves good accuracy. FCN demonstrates the weakest performance among the three networks, indicating it may not be the best choice for this particular segmentation task.

The loss graph is a visual representation of how the model's loss function evolves throughout training. The loss function quantifies how well or poorly the model's predictions match the actual ground truth labels. During training, the goal is to minimize this loss, leading to a model that makes more accurate predictions. Figure 3 presents the loss values of the analyzed networks during the training phase. FCN had a higher minimum and exhibited more fluctuations, which might suggest sensitivity to the training process. DeepLabV3+ shows a stable loss towards the end but with a slightly higher baseline, suggesting it is still learning or possibly encountering minor challenges in fine-tuning. SegFormer shows a lower final loss, indicating a potentially better fit than the other two. Several techniques were used to improve model performance, such as dropout to prevent overfitting, data augmentation methods like RandomResize, RandomCrop, and RandomFlip, and tuning of hyperparameters, including the learning rate and the AdamW optimizer. The CrossEntropy loss function was applied to emphasize penalties on target classes during training.

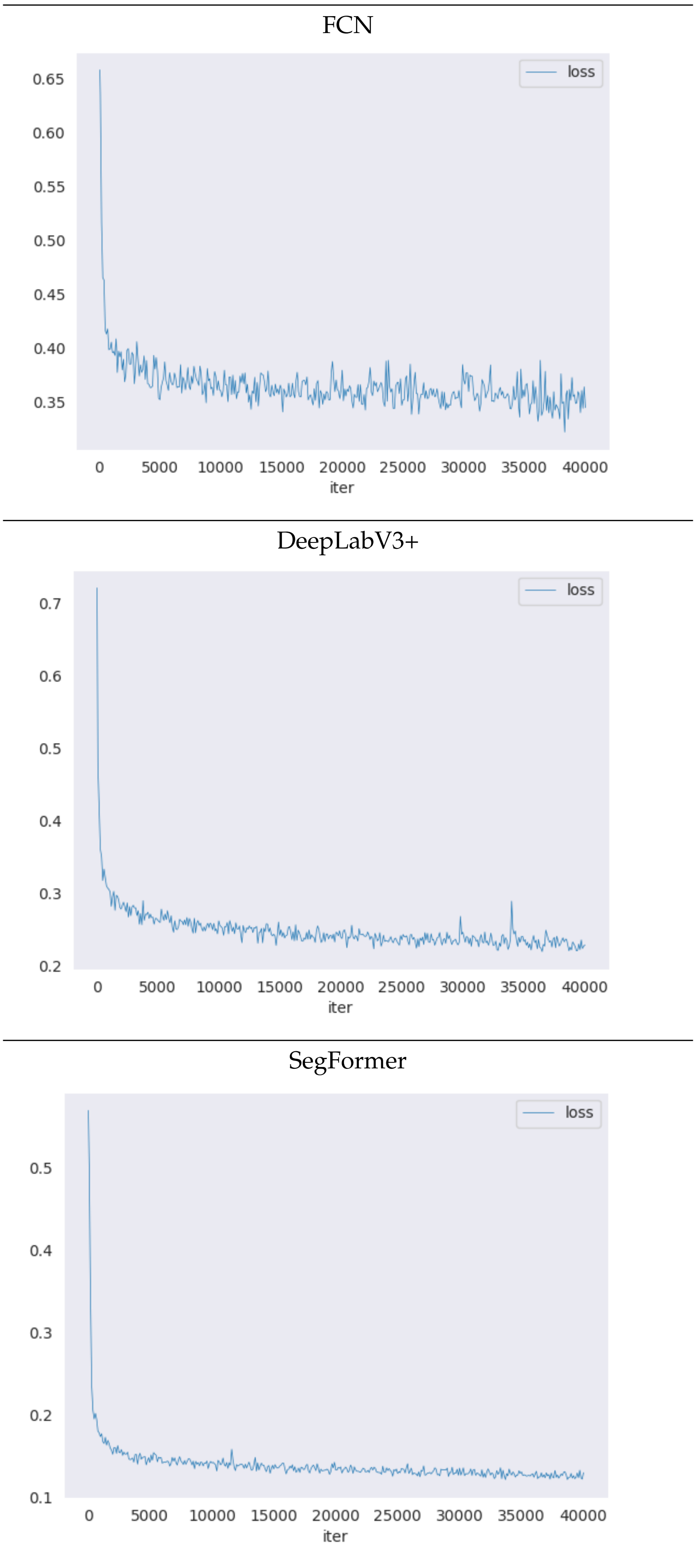


Figure 3. The training losses of the pattern recognition CNN plotted as a function of iteration for 40,000 epochs.

For a more profound analysis of the network training on our target class, a two-factor ANOVA without replication assesses the impact of epochs and different network architectures on the Accuracy (Acc) metric in Table 3. The results show that while the "Rows" factor, representing the combination of epochs and architectures, does not have a statistically significant effect (F = 1.0256, p = 0.4556), the "Columns" factor, representing the different network architectures, has a highly significant impact (F = 425.3365, $p = 2.92 \times 10^{-13}$). The very low p-value for network architectures suggests that the choice of network is crucial in influencing the accuracy outcomes, whereas variations in epochs within this range appear to have less effect.

Table 3. Two-factor ANOVA test of the accuracy metric for the different networks and epochs.

Epochs	**Metric**		**Count**	**Sum**	**Mean**	**Variance**
5000.00	Acc		3	254.77	84.9233	43.4802
10,000.00	Acc		3	257.99	85.9967	35.2816
15,000.00	Acc		3	259.34	86.4467	42.2014
20,000.00	Acc		3	257.14	85.7133	47.8814
25,000.00	Acc		3	258.91	86.3033	42.7900
30,000.00	Acc		3	255.12	85.0400	64.8673
35,000.00	Acc		3	256.91	85.6367	49.1480
40,000.00	Acc		3	256.9287	85.6429	45.5073
	fcn		8	623.4561	77.9320	1.2618
	DeepLabV3+		8	712.93	89.1163	0.5110
	segformer		8	720.7227	90.0903	0.8246
Source of Variation	SS	df	MS	F	p-Value	F crit
Rows	6.1634	7	0.8805	1.0256	0.4556	2.7642
Columns	730.2960	2	365.1480	425.3365	2.92×10^{-13}	3.7389
Error	12.0189	14	0.8585			
Total	748.4783	23				

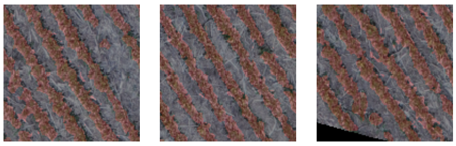
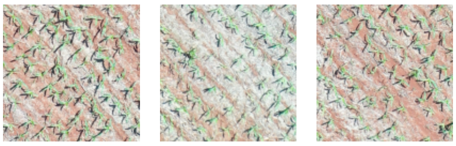
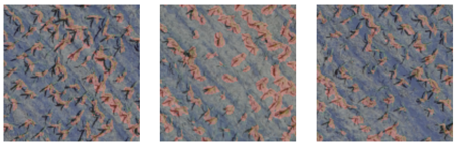
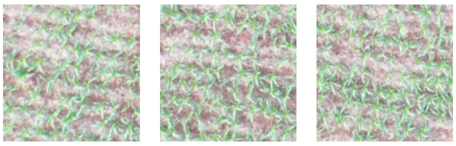
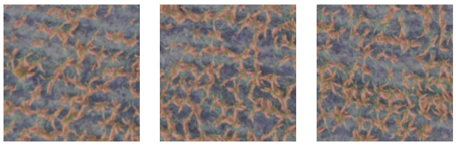
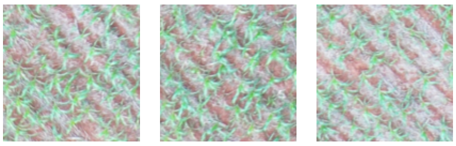
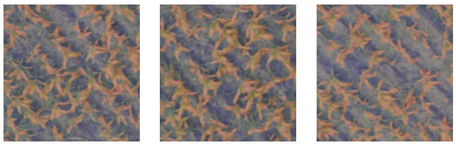
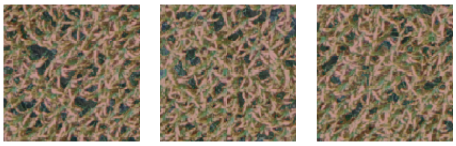
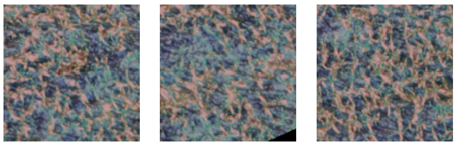
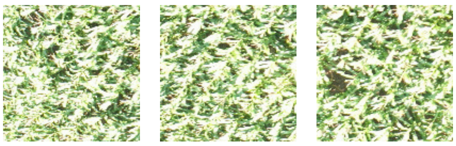
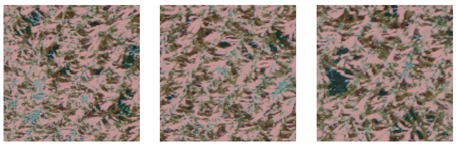
3.2. Network Inference Presentation

In the context of using deep learning for the semantic segmentation of corn crops, the inference process refers to the application of a trained deep-learning model to new, unseen data to make the desired predictions.

Table 4 presents the inference results using the SegFormer architecture. The networks effectively segment the dataset, though some conflicts occur at the borders of segmented regions. These issues likely stem from the inherent complexity and ambiguity at the boundaries, where the model may struggle to accurately distinguish between adjacent regions with similar features. Additionally, as the corn matures, the network faces increasing challenges in accurately segmenting the target class. This difficulty could be due to several factors, with the most significant being a lack of diverse training data for the model to learn from and potential inaccuracies during the annotation process.

Addressing these border conflicts is crucial for improving the overall segmentation accuracy and ensuring more precise and reliable results across the entire dataset. Further refinement of the model, such as enhancing the edge-detection capabilities or incorporating additional contextual information, could help mitigate these issues and enhance the quality of the segmentation.

Table 4. Image sample and inference results. The target class is presented with a reddish color.

Image	Image Sample	Inference Result
1		
2		
3		
4		
5		
6		
7		
8		
9		

3.3. Training Dataset Size Variation Experiment

In an experiment on semantic segmentation, the training dataset size was varied systematically to observe the impact on model performance. The experiment involved using 80%, 70%, 60%, and 50% of the total images in the dataset for training, with the remaining portions reserved for validation and testing. As the training set size decreased, the model had access to fewer labeled examples, which likely affected its ability to generalize. This reduction in training data can lead to poorer feature extraction and less precise segmentation boundaries, as the model has less information to learn from. Table 5 shows the results of this dataset variation on the experiment.

Table 5. Performance of the semantic-segmentation model across varying training set sizes.

Training Test Size Variation	80%		70%		60%		50%	
Class	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc
Corn	80.87	90.2	81.6	89.47	81.64	89.76	80.84	89.52
Background	85.95	91.88	86.58	93.1	85.8	92.46	86.1	92.45

The results from the table show that varying the size of the training dataset had minimal impact on the performance of the semantic-segmentation model for both the "Corn" class and the "Background" class. The Intersection over Union (IoU) and accuracy metrics remain relatively stable across the different training set sizes, indicating that reducing the training data did not significantly degrade the model's performance. This suggests that the model is resilient to variations in the size of the training set, as its performance remains nearly unchanged despite the reduction in the amount of labeled data. This could imply that the dataset contains sufficient diversity and that the model has learned effective generalization strategies from the available data.

4. Discussion

The significance of this study lies in its exploration of both legacy and contemporary CNNs in the context of UAV image analysis for precision agriculture. The focus on cornfields in Mato Grosso do Sul, a region known for its agricultural productivity, brings practical relevance to the research. These tools and findings can improve the farming of similar environments, and, using multiple stages of growth, we intended to present some of the challenges of image segmenting multiple stages of plant growth. The challenges faced in the case of corn can be expected while training with other kinds of cultures. Development of this kind is suitable for developing robust, scalable solutions for precision agriculture where the ability to accurately segment plants at different growth stages with UAV field data can significantly impact yield predictions and resource-management strategies.

The study's results can also contribute to a deeper understanding of how different CNN architectures perform in semantic-segmentation tasks for complex images. The superior performance of SegFormer, which utilizes a transformer-based architecture, underscores the transformative impact of incorporating transformers in image segmentation tasks. Transformers' ability to model long-range dependencies and global context effectively allows for precise segmentation, even in complex and heterogeneous agricultural landscapes. The results indicate that SegFormer achieves the highest Intersection over Union (IoU) for corn at 78.81 and maintains competitive accuracy, making it a highly suitable choice for precision agriculture applications. DeepLabV3+, with its advanced Atrous Spatial Pyramid Pooling (ASPP) module, also demonstrates strong performance. Its architecture efficiently balances accuracy and computational efficiency, achieving a high IoU for both corn and background classes. This network's encoder–decoder structure allows it to capture rich semantic information while refining segmentation results, especially around object boundaries, which is crucial for accurately delineating crops or other complex kinds of objects from the background.

The integration of UAV imagery with neural networks offers a powerful tool for precision agriculture. High-resolution UAV images provide detailed spatial information that, when processed through robust segmentation models, can infer important insights about the quality and quantity of the target crop, such as yield estimation and the early detection of stress factors such as disease or pest infestations, all with low environmental impact. This capability is particularly relevant where diverse ecosystems and intensive agricultural activities demand precise data and timely decisions about any possible damage. Despite the promising results, several challenges remain. One significant challenge is the generalization of these models across different crops and varying environmental conditions. While this study focused on corn in a specific region, future research should aim to validate these findings across diverse crops and geographic areas, meaning that for future projects with similar processing architectures, it is good practice to use transfer-learning models. We destabilized the necessary data to reproduce the experiment and advance with future similar proposals on our repository, as presented in the Materials and Methods section. Additionally, the computational demands of transformer-based architectures, though manageable, may still pose limitations for widespread adoption, especially in resource-constrained settings.

SegFormer, with its transformer-based architecture, exhibits the highest segmentation performance, particularly in modeling long-range dependencies and global context. DeepLabV3+ also shows strong performance with its efficient balance of accuracy and computational demands. These findings suggest that integrating these advanced segmentation models with UAV imagery can significantly improve crop monitoring, yield estimation, and resource management in agriculture. The research also presents the importance of continued innovation in neural network architectures and their application in precision agriculture.

While the impact of wild pigs on crops is a key motivator for this research, the primary objective is not to develop a model specifically for assessing wildlife-induced damage. Instead, our focus is on segmenting corn from the background. Future research could explore a more detailed analysis of crop damage, incorporating images of various impacts beyond those caused by wildlife. Despite significant advancements in deep-learning models, UAV technology, and precision agriculture, a gap remains in integrating these tools to address challenges in crop monitoring

Improved segmentation models can greatly enhance agricultural practices by enabling precision resource allocation, early pest detection, optimized irrigation, and effective weed management, reducing costs and environmental impact. They also aid in crop monitoring, yield estimation, and soil health assessments, leading to more sustainable farming. While our model is not optimized to differentiate crop-growth stages using the BBCH scale—key for tracking development—future research could address this. Additionally, the model does not account for drought impacts, a crucial factor in crop health. Future work could integrate water-stress data to improve accuracy in drought-prone regions. This research approach contributes to this by offering adaptable methodologies that advance automated monitoring.

5. Conclusions

This study demonstrates the effectiveness of contemporary neural network architectures, particularly FCN, SegFormer, and DeepLabV3+, in the task of semantic segmenting of corn crops using UAV imagery. The results highlight the potential of these advanced models to enhance precision agriculture practices by providing accurate and detailed crop segmentation at various developmental stages.

By addressing current challenges and expanding the scope of research, future studies can further enhance the capabilities of crop segmentation technologies, contributing to more sustainable and productive agricultural practices. This work leverages advanced deep-learning techniques, specifically Vision Transformers (ViTs), for detailed vegetation mapping in a crucial agricultural landscape in Brazil. Using high-resolution UAV imagery, we demonstrate the feasibility of using semantic segmenting and possible pathways for more accurate crop monitoring as world food security becomes more challenging.

The methodology for manual labeling, dataset creation, and robust evaluation metrics sets a high standard for future research, ensuring reliability and validity. This is a key issue where this work can be improved. The process of data labeling for complex datasets, such as the one presented, is always laborious and ambiguous for the human operator, and we always miss most of the borders and shadows of the targeted objects. So, improved ways of making datasets and using newer deep-learning techniques should be directions for future work. Self-supervised learning and interactive annotation tools enhance label quality by making use of unlabeled data and offering real-time feedback. At the same time, automated approaches like active learning and weak supervision reduce manual labeling efforts by focusing on ambiguous samples and handling noisy labels.

Also, future studies should explore the relationship between the proximity of growing crops with environmentally sensible areas such as watercourses and how this can influence the effectiveness of conservation efforts. Additionally, research could identify critical phases in the crop-growth cycle that require specific physical interventions, such as the installation of fences or scarecrows, to protect the crop. Another promising area of investigation is the integration of upscaling with high-spatial-resolution satellite imagery to cover larger geographic areas, therefore enhancing agricultural monitoring and management on a broader scale.

Supplementary Materials: The following supporting information can be downloaded at: https://shorturl.at/R9YJQ. Supplementary materials have been included to provide additional data, resources, and context to support the findings presented in the main text. The suffix number in each folder name indicates the division of data into Train, Test, and Validation sets. For instance, `new_dataset_format_10` reflects the data split structure of 80_10_10. The folders labeled `original_patches_3` and `segmented_patches_3` contain the original dataset before it was divided for training purposes.

Author Contributions: Conceptualization J.A.C.M., A.Y.H.H. and L.A.P.; methodology, J.A.C.M. and A.O.P.; software, W.N.G. and D.A.S.; validation, A.M.d.A., A.P.M.R., L.A.P. and V.L.; formal analysis, A.M.d.A. and J.A.C.M.; investigation, J.A.C.M., J.M.J. and A.O.P.; resources, J.M.J.; data curation, J.A.C.M. and A.O.P.; writing—original draft preparation, J.A.C.M.; writing—review and editing, J.A.C.M. and J.M.J.; visualization, J.A.C.M. and W.N.G.; supervision, J.M.J., A.P.M.R., R.S.J. and H.P.; project administration, J.A.C.M. and J.M.J.; funding acquisition, J.A.C.M. and J.M.J. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge the support of UFMS (the Federal University of Mato Grosso do Sul) and CAPES (Finance code 001). And the concession of scholarships and research assistance provided by Fundect with the project-Fundect/CNPq Call No. 29/2022–Support Program for the Retention of Young Doctors in MS.

Data Availability Statement: 1. Data available in a publicly accessible repository https://shorturl.at/R9YJQ with the doi: 10.5281/*zenodo*.14019840. Under [license, e.g., CC BY 4.0]. 2. The files will also be shared in the Supplementary Materials.

Acknowledgments: The authors would like to acknowledge the Federal University of Mato Grosso do Sul (UFMS) Brazil https://www.ufms.br/. for their administrative and technical support during this research. We are also grateful to CAPES (Coordination for the Improvement of Higher Education Personnel) for the scholarships and funding that made this study possible, particularly through the CAPES PrInt program (Grant No. 88881.311850/2018-01). Our thanks extend to Fundect (Foundation to Support the Development of Education, Science, and Technology of the State of Mato Grosso do Sul) https://www.fundect.ms.gov.br/, for hosting the project number 29/2022 'Young Doctors in MS', that made this research possible. and also for their financial assistance under grants 83/013.192/2023, 71/001.902/2022, 83/013.192/2023, 71/700.004/2017, 71/033.021/2022, and 71/027.243/2022. Additionally, we also acknowledge the support from Embrapa (Brazilian Agricultural Research Corporation) Pantanal https://www.embrapa.br/pantanal, for their invaluable contributions and resources that greatly enhanced this research. And We also acknowledge the support provided by CNPq under grant numbers 308481/2022-4, 405997/2021-3, 403213/2023-1, 305296/2022-1, 303469/2022-6, 409432/2022-9, 305814/2023-0, and 317538/2021-7, which significantly contributed to the advancement of this study.

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

References

1. Hamuda, E.; Glavin, M.; Jones, E. A survey of image processing techniques for plant extraction and segmentation in the field. *Comput. Electron. Agric.* **2016**, *125*, 184–199. [CrossRef]
2. Bouguettaya, A.; Zarzour, H.; Kechida, A.; Taberkit, A.M. Deep learning techniques to classify agricultural crops through UAV imagery: A review. *Neural Comput. Appl.* **2022**, *34*, 9511–9536. [CrossRef] [PubMed]

3. Mahajan, S.; Das, A.; Sardana, H.K. Image acquisition techniques for assessment of legume quality. *Trends Food Sci. Technol.* **2015**, *42*, 116–133. [CrossRef]
4. Osco, L.P.; Nogueira, K.; Ramos, A.P.M.; Pinheiro, M.M.F.; Furuya, D.E.G.; Gonçalves, W.N.; de Castro Jorge, L.A.; Junior, J.M.; dos Santos, J.A. Semantic segmentation of citrus-orchard using deep neural networks and multispectral UAV-based imagery. *Precis. Agric.* **2021**, *22*, 1171–1188. [CrossRef]
5. Qin, R.; Liu, T. A Review of Landcover Classification with Very-High Resolution Remotely Sensed Optical Images—Analysis Unit, Model Scalability and Transferability. *Remote Sens.* **2022**, *14*, 646. [CrossRef]
6. Martins, J.A.C.; Menezes, G.; Goncalves, W.; Sant'Ana, D.A.; Osco, L.P.; Liesenberg, V.; Li, J.; Ma, L.; Oliveira, P.T.; Astolfi, G.; et al. Machine learning and SLIC for tree canopies segmentation in urban areas. *Ecol. Inform.* **2021**, *66*, 101465. [CrossRef]
7. Bai, X.; Cao, Z.G.; Wang, Y.; Yu, Z.; Zhang, X.; Li, C. Crop segmentation from images by morphology modeling in the CIE L* a* b* color space. *Comput. Electron. Agric.* **2013**, *99*, 21–34. [CrossRef]
8. Li, Y.; Huang, Z.; Cao, Z.; Lu, H.; Wang, H.; Zhang, S. Performance evaluation of crop segmentation algorithms. *IEEE Access* **2020**, *8*, 36210–36225. [CrossRef]
9. Patrício, D.I.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* **2018**, *153*, 69–81. [CrossRef]
10. Lee, E.K.; Zhang, W.J.; Zhang, X.; Adler, P.R.; Lin, S.; Feingold, B.J.; Khwaja, H.A.; Romeiko, X.X. Projecting life-cycle environmental impacts of corn production in the U.S. Midwest under future climate scenarios using a machine learning approach. *Sci. Total Environ.* **2020**, *714*, 136697. [CrossRef]
11. Erenstein, O.; Jaleta, M.; Sonder, K.; Mottaleb, K.; Prasanna, B. Global maize production, consumption and trade: Trends and R&D implications. *Food Secur.* **2022**, *14*, 1295–1319.
12. Fuglie, K.; Wang, S.L. Productivity growth in global agriculture shifting to developing countries. *Choices* **2012**, *27*, 1–7.
13. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090. [CrossRef]
14. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv: 2010.11929.
15. Islam, M.A.; Jia, S.; Bruce, N.D.B. How Much Position Information Do Convolutional Neural Networks Encode? *arXiv* **2020**, arXiv:2001.08248.
16. Barros, N.N.; Ruschel, R.C. Machine Learning for Whole-Building Life Cycle Assessment: A Systematic Literature Review. *Lect. Notes Civ. Eng.* **2021**, *98*, 109–122. [CrossRef]
17. Ramankutty, N.; Evan, A.T.; Monfreda, C.; Foley, J.A. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Glob. Biogeochem. Cycles* **2008**, *22*. [CrossRef]
18. Miller, R.S.; Sweeney, S.J.; Slootmaker, C.; Grear, D.A.; Di Salvo, P.A.; Kiser, D.; Shwiff, S.A. Cross-species transmission potential between wild pigs, livestock, poultry, wildlife, and humans: Implications for disease risk management in North America. *Sci. Rep.* **2017**, *7*, 7821. [CrossRef]
19. Bevins, S.N.; Pedersen, K.; Lutman, M.W.; Gidlewski, T.; Deliberto, T.J. Consequences associated with the recent range expansion of nonnative feral swine. *BioScience* **2014**, *64*, 291–299. [CrossRef]
20. Roda, F.; Roda, J.M. Signs of foraging by wild boar as an indication of disturbance to ground-nesting birds. *J. Vertebr. Biol.* **2024**, *73*, 23103-1. [CrossRef]
21. Friesenhahn, B.A.; Massey, L.D.; DeYoung, R.W.; Cherry, M.J.; Fischer, J.W.; Snow, N.P.; VerCauteren, K.C.; Perotto-Baldivieso, H.L. Using drones to detect and quantify wild pig damage and yield loss in corn fields throughout plant growth stages. *Wildl. Soc. Bull.* **2023**, *47*, e1437. [CrossRef]
22. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 801–818. [CrossRef]
23. Guo, Y.; Liu, Y.; Georgiou, T.; Lew, M.S. A review of semantic segmentation using deep neural networks. *Int. J. Multimed. Inf. Retr.* **2018**, *7*, 87–93. [CrossRef]
24. Zhang, C.; Kovacs, J. The application of small unmanned aerial systems for precision agriculture: A review. *Precis. Agric.* **2012**, *13*, 693–712. [CrossRef]
25. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A Review on UAV-Based Applications for Precision Agriculture. *Information* **2019**, *10*, 349. [CrossRef]
26. Ramos, A.P.M.; Gomes, F.D.G.; Pinheiro, M.M.F.; Furuya, D.E.G.; Gonçalvez, W.N.; Junior, J.M.; Michereff, M.F.F.; Blassioli-Moraes, M.C.; Borges, M.; Alaumann, R.A.; et al. Detecting the attack of the fall armyworm (Spodoptera frugiperda) in cotton plants with machine learning and spectral measurements. *Precis. Agric.* **2022**, *23*, 470–491. [CrossRef]
27. Boursianis, A.D.; Papadopoulou, M.S.; Diamantoulakis, P.; Liopa-Tsakalidi, A.; Barouchas, P.; Salahas, G.; Karagiannidis, G.; Wan, S.; Goudos, S.K. Internet of things (IoT) and agricultural unmanned aerial vehicles (UAVs) in smart farming: A comprehensive review. *Internet Things* **2020**, *18*, 100187. [CrossRef]
28. Barrios-Garcia, M.N.; Ballari, S.A. Impact of wild boar (*Sus scrofa*) in its introduced and native range: A review. *Biol. Invasions* **2012**, *14*, 2283–2300. [CrossRef]

29. Boyce, C.; Vercauteren, K.C.; Beasley, J.C. Timing and extent of crop damage by wild pigs (*Sus scrofa* Linnaeus) to corn and peanut fields. *Crop Prot.* **2020**, *133*, 105131. [CrossRef]
30. McKee, S.C.; Mayer, J.J.; Shwiff, S.A. Comprehensive economic impacts of wild pigs on producers of six crops in the South-Eastern US and California. *Agriculture* **2024**, *14*, 153. [CrossRef]
31. Agisoft, L. Metashape python reference. *Release* **2020**, *1*, 1–199.
32. Berg, S.; Kutra, D.; Kroeger, T.; Straehle, C.N.; Kausler, B.X.; Haubold, C.; Schiegg, M.; Ales, J.; Beier, T.; Rudy, M.; et al. Ilastik: Interactive machine learning for (bio) image analysis. *Nat. Methods* **2019**, *16*, 1226–1232. [CrossRef] [PubMed]
33. Meier, U. *Growth Stages of Mono- and Dicotyledonous Plants = Entwicklungsstadien Mono- und Dikotyler Pflanzen = Estadios de las Plantas Mono- y Dicotiledóneas = Stades Phénologiques des Mono- et Dicotylédones Cultivées*; Blackwell: Berlin, Germany, 1997.
34. Meier, U.; Bleiholder, H.; Buhr, L.; Feller, C.; Hack, H.; Heß, M.; Lancashire, P.D.; Schnock, U.; Stauß, R.; Van Den Boom, T.; et al. The BBCH system to coding the phenological growth stages of plants–history and publications. *J. Kult.* **2009**, *61*, 41–52.
35. Tottman, D. The decimal code for the growth stages of cereals, with illustrations. *Ann. Appl. Biol.* **1987**, *110*, 441–454. [CrossRef]
36. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef] [PubMed]
37. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.
38. Bottou, L. Online algorithms and stochastic approximations. In *Online Learning in Neural Networks*; Cambridge University Press: Cambridge, UK, 1998.
39. OpenMMLab. MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark. 2020. Available online: https://github.com/open-mmlab/mmsegmentation (accessed on 1 August 2024).