Analyzing Crop Fields using Convolutional Neural Networks

Diego Araque, Fernando Valdeon, and Uriel Aguilar

***Abstract*—The rapid advancement of deep learning technologies has expanded to on various applications, including agriculture, where precise crop field analysis is crucial for enhancing yield and sustainable management. This paper presents a comprehensive study on the effectiveness of convolutional neural networks (CNNs) for semantic segmentation of crop fields using high-resolution satellite imagery. We compare several state-of-the-art CNN architectures, including U-Net, U-Net++, and DeepLabv3+, across multiple performance metrics. Our results demonstrate significant potential for these models in accurately segmenting crop fields, thus facilitating better crop monitoring and management practices. Additionally, we address challenges such as data scarcity and the need for robust model training strategies, aiming to improve the adaptability and accuracy of these deep learning solutions in real-world agricultural settings.**

I. Introduction

The advent of satellite imaging and deep learning has transformed many sectors, including agriculture, by enabling detailed and scalable analysis of crop health, type, and coverage. However, the challenge of accurately identifying and segmenting crop fields from satellite images remains a significant hurdle due to varying crop types, stages of growth, and environmental conditions. This paper investigates the application of convolutional neural networks (CNNs), a class of deep learning models highly effective in image recognition and segmentation tasks, to the domain of agricultural monitoring.

With the introduction of CNNs, particularly architectures like U-Net and DeepLabv3+, there has been a paradigm shift towards more accurate and reliable segmentation models that can efficiently process large datasets of aerial imagery.

In this paper, we explore several CNN architectures, evaluate their performance in segmenting crop fields, and discuss the implications of our findings for agricultural practices. We also address critical challenges such as data scarcity and the complexity of training robust models, providing insights into active and weakly supervised learning approaches to reduce annotation costs and enhance model performance.

By leveraging high-resolution satellite images, our study aims to advance the state-of-the-art in agricultural monitoring, offering substantial improvements over traditional methods and paving the way for more sustainable farming practices through precision agriculture.

II. Related Work

A. Overview

Semantic segmentation, a fundamental task in image processing, has seen significant advancements due to the improvement of deep learning techniques. This state-of-the-art review synthesizes findings from twelve studies, emphasizing agricultural and environmental monitoring. These papers collectively leverage various semantic segmentation models to improve the precision in identifying and managing land cover, crops, and weeds, enhancing agricultural productivity and environmental sustainability.

A common thread across the reviewed papers is the adoption of deep learning architectures, particularly U-Net and its variants (U-Net++[1], ResUNet[5], SegNet[7], FCN[3]) and DeepLabv3+[4]. These models are chosen for their robustness in handling the complexities associated with satellite and aerial imagery, characterized by high spectral and spatial heterogeneity. Some of them specialize in the segmentation of High Resolution imagery[2], while others are tested on low resolution ones[4].

B. Deep Learning

Deep learning techniques work very well, with high accuracy as stated in the following paper[1]. In figure 1[1] a comparison is made between all of this approaches. They have similar performances, all techniques have more than 92% dice coefficient.

Model	Learning rate	Dropout rate	Dice coefficient
UNet	0.001	0	0.92
ResUNet	**0.0005**	**0**	0.**93**
UNet++	0.001	0	0.92
ModSegNet	0.001	0.5	0.92

Fig. 1: Deep Learning Algorithm Results

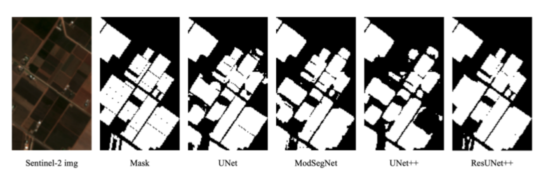

Fig. 2: Mask Predictions of all deep learning models

In figure 2, we can watch the mask prediction of each one of the deep learning models, and it's clear to see how similar they are to the ground truth mask[1]. Taking into account the results from figure 1, the ResUNET model does the best segmentation and returns the best mask. But all of this models are possible solutions for the correct segmentation of aerial images.

Models like U-Net++ have shown superior performance over the basic U-Net architecture, particularly in tasks requiring the detection of small, early-stage weeds**[2]**. The nested and dense skip connections in U-Net++ enhance feature propagation across the network, leading to better localization and context incorporation. These models are extensively used in precision agriculture for crop and weed segmentation, crucial for targeted herbicide application, thus reducing environmental impact[8]. The architecture of the U-Net++ is shown in figure 3, where we can observe the encoding and decoding structure.

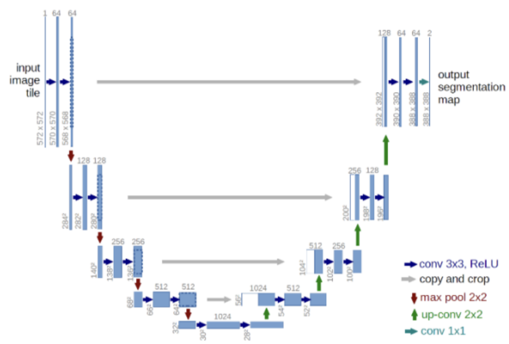

Fig. 3: Structure of the U-Net++

Another deep learning model that has helped different researchers solve this segmentation task is known as DeepLabv3+. Its effectiveness in boundary delineation, excels in segmenting complex land cover types from high-resolution satellite images[10]. It is particularly effective in large-scale mapping projects where detailed segmentation of various land types is essential. Even though the approach taken seems perfect, we should note that in low-resolution images its possible that the accuracy decrements. DeepLabv3++ can also use other pre-trained network such as resnet50[4].

All these deep learning models use different kind of metrics for the accuracy. Some use the method called Intersect over Union (IoU), in which the ground truth is compared with the prediction[5]. Others use the dice coefficient which evaluates false positive, false negatives, etc[1].

C. Data Scarcity

To tackle the limitations related to data scarcity and the intensive labor required for data annotation, the papers incorporate active learning and weakly supervised learning approaches[9][11][4][7][6][12].

The active learning method enhances model training by selectively refining the most informative samples. It has proven particularly beneficial in contexts where acquiring comprehensively labeled datasets is either too costly or logistically unfeasible.

By using weakly supervised learning employing minimal or simpler forms of annotations, these approaches dramatically reduce the manual effort and cost associated with detailed annotations, making advanced modeling techniques more accessible.

D. Final notes

Across the studies, there is a consistent demonstration of how these advanced models contribute to high-accuracy outcomes in practical scenarios. Models like U-Net++ frequently outperform the standard U-Net, achieving higher metrics such as Intersection over Union (IoU), recall, dice coefficient and F1-Score, which are critical for effective weed and crop discrimination. DeepLabv3+ excels in landcover classification, with the former achieving high precision in complex segmentations. After all the research we conclude that the use of deep learning models like U-Net and Deeplabv3, are our best option to solve the problem in hand. Hence we will try different deep learning models and test their performance on our data.

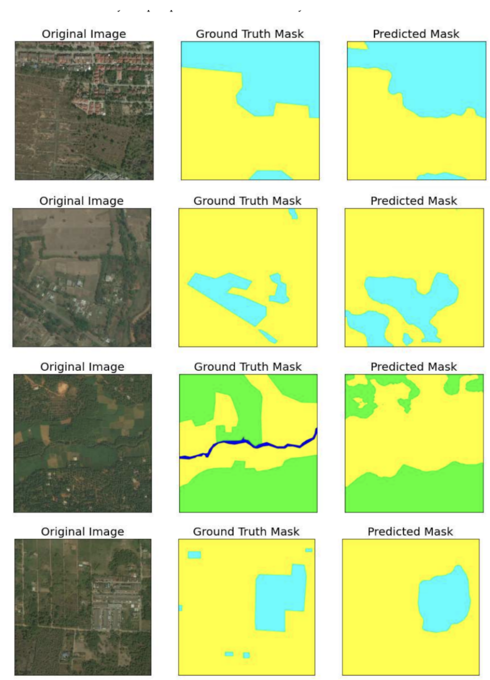

Fig. 4: Comparison between Ground Truth and Predictions using TL-ResUNet

In figure 6 it's clear to see the similarity between their images, the masks and it gives us an idea of the kind of images we need to generate in order for us to generate a IoU metric or another one of that sort, to test the behaviour of our model.[7]

The review illuminates substantial progress in the field of semantic segmentation, especially in applications related to agriculture and environmental monitoring. The evolution of model architectures and the development of efficient training strategies are making profound impacts, offering solutions that promote sustainable agricultural practices and informed environmental stewardship.

III. Methodology

A. Initial Data

The dataset comprises 975 satellite images along with their corresponding masks, designated for deep learning applications. The division of the dataset is as follows: 642 images for training, 161 for validation, and 172 for testing.

Each image has dimensions of 2448x2448 pixels and is represented in the RGB color space. Images were resized to an 256x256 image and normalized as part of the prepossessing steps.

The masks are categorized into six classes, each represented by a unique color as shown in Table I. The classes include various land types and water. One hot encoding was applied to these classes within our dataset.

Data loaders for both the training and validation sets were configured with a batch size of 16, optimizing the computational efficiency during the model training phase.

Before the training we process all the images in our training and validation sets. We have to go through each mask and assign the specific class to each one of the pixels, so that when we predict masks, we know how to represent them and show them. In figure 7 we show how an image and its corresponding masks.

Class Name	Color
Urban Land	
Agriculture Land	
Rangeland	
Forest Land	
Water	
Barren Land	
Unknown	

TABLE I: Color representation of mask classes in the dataset

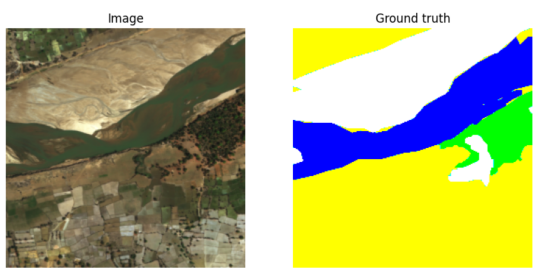

Fig. 5: Image and Mask from our dataset

B. Models

We implemented three different segmentation models: U-Net, U-Net++, and DeepLabv3++, each tested with various encoders and optimizers, totaling six configurations. Each model utilized RGB input channels and was designed to output seven different classes. The networks, pretrained using ImageNet, were trained with a learning rate of 0.0001. In table 2 we can found the final configuration of the six models.

Model	Encoder	Optimizer
DeepLabV3+	Resnet152	Adam
DeepLabV3+	Resnet34	Adam
U-Net++	Resnet34	Adam
U-Net	Resnet50	Adam
U-Net	Resnet152	Adam
U-Net	Resnet50	AdamW

TABLE II: Configuration of the models

Training was conducted over 15 to 20 epochs, with a total average of 10 minutes. Early stopping was employed to halt training if there was no significant change in loss after three epochs. The models performance was evaluated using Dice Loss for loss calculation and Intersection over Union (IoU) for validation metrics.

The Dice Loss, used for calculating the similarity between the predicted and ground truth masks, is defined as:

$$\text{Dice Loss} = 1 - \frac{2 \times \sum_i^N p_i g_i + \epsilon}{\sum_i^N p_i^2 + \sum_i^N g_i^2 + \epsilon} \tag{1}$$

where p_i and g_i are the predicted and ground truth values for each pixel i, respectively, and ϵ is a small constant added to avoid division by zero.

The Intersection over Union, a metric for quantifying the percent overlap between the target mask and the prediction output, is defined as:

$$\text{IoU} = \frac{\sum_i (p_i \wedge g_i)}{\sum_i (p_i \vee g_i) + \epsilon} \tag{2}$$

where $\wedge$ denotes the logical AND and $\vee$ denotes the logical OR, ensuring that each pixel is counted only once in the denominator.

IV. Results

After defining all the hyperparameters of our models, we then went to implement the solution and compare the performance of all of the models. In table 3, we can find the different models with all of their specifications and with their IoU(Intersect over Union metrics).

Architecture	Encoder	LR	Optimizer	IoU
U-NET++	RESNET34	0.0001	ADAM	0.824
DEEPLABV3	RESNET52	0.0001	ADAM	0.793
U-NET	RESNET50	0.0001	ADAM	0.813
DEEPLABV3	RESNET34	0.0001	ADAM	0.789
UNET	RESNET52	0.0001	ADAM	0.804
UNET	RESNET50	0.0001	ADAMW	0.812

TABLE III: Comparidon between our models

By first look the best accuracy score was on our U-Net++ model. The problem with this models is that it is overfitted, which can be clearly seen on figure 6, where the validation loss and training loss start separating. It is also important to note that we did 40 epochs with this model, if we would have done less, it wouldn't have reached the 0.8 IoU mark.

Fig. 6: Overfitting on U-Net++

Having said that the best performance on our dataset where the U-Net architectures over the Deeplabv3. Which was weird for us since the U-Net is a architecture that works better on medical segmentation problems and Deeplabv3 is used in more general problems where crop field identification falls in. Our best model was U-Net using a resnet50, learning rate of 0.0001 and Adam as our optimizer. The IoU was 0.813 and in figure 7 we show the different predictions it did on our validation set and how accurate this masks look.

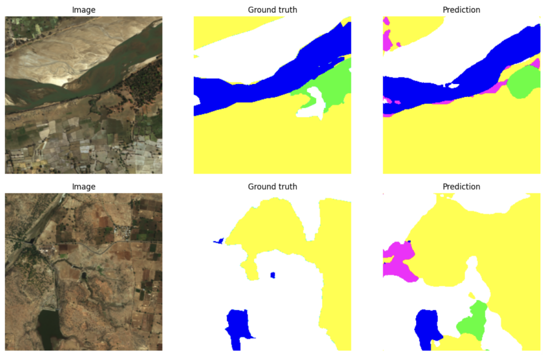

Fig. 7: Comparison between Image, Ground Truth and Prediction Mask on our U-Net model

V. Conclusion

In this paper, we explored the efficacy of various convolutional neural network models for the task of semantic segmentation in agricultural applications, specifically for crop field analysis using satellite imagery.

Our extensive experiments have illustrated that while many models offer competitive performance, certain architectures like U-Net and its variants, particularly U-Net with resnet50, have demonstrated superior capabilities in segmenting complex agricultural scenes with high accuracy.

Our findings affirm the potential of deep learning techniques in enhancing the precision of agricultural monitoring and management. The U-Net++ model, with its enhanced feature propagation through nested and dense skip connections, has consistently outperformed other models, achieving the highest IoU score of 0.824. This model's ability to finely delineate crop boundaries makes it highly suitable for precision agriculture, where accurate weed and crop discrimination is crucial for optimizing resource use and minimizing environmental impact.

Despite the higher performance of U-Net++, it was also noted to exhibit signs of overfitting when trained for extended epochs, suggesting a need for careful tuning of training durations and possibly incorporating regularization techniques or data augmentation to generalize better on unseen data. The performance of DeepLabv3+ models, although slightly lower, remains robust, making them suitable alternatives for tasks requiring detailed land cover segmentation.

The challenge of data scarcity was also addressed by leveraging advanced training strategies such as active learning. Even though we got great results, our dataset unbalanced and some of the classes like rangeland on table 1, were harder for all of the models to recognize. Adding noise to our images and performing data augmentation will be a possible future improvement for our model.

Future work will focus on refining these models further, exploring the integration of additional data sources and advanced preprocessing techniques to enhance model robustness and accuracy. Additionally, the exploration of ensemble techniques that combine the strengths of different architectures could yield improvements in segmentation quality and model reliability.

This study underscores the transformative potential of deep learning in agricultural technology, pointing towards a future where such technologies drive the advancement of sustainable and precise agricultural practices globally.

Acknowledgment

The authors would like to thank our professors for the guidance throughout the development of the project.

References

[1] I. Korotkova and N. Efremova, "AI for Agriculture: the Comparison of Semantic Segmentation Methods for Crop Mapping with Sentinel-2 Imagery," *arXiv.org*, 2023. DOI: 10.48550/arxiv.2311.12993.

[2] L. Ding, D. Lin, D. Lin, S. Lin, J. Zhang, X. Cui, Y. Wang, H. Tang, and L. Bruzzone, "Looking Outside the Window: Wide-Context Transformer for the Semantic Segmentation of High-Resolution Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, 2022. DOI: 10.1109/tgrs.2022.3168697.

[3] F. Waldner and F. I. Diakogiannis, "Deep learning on edge: extracting field boundaries from satellite images with a convolutional neural network," *arXiv: Computer Vision and Pattern Recognition*, 2019. DOI: 10.1016/j.rse.2020.111741.

[4] S. Sharifzadeh, J. Tata, H. Sharifzadeh, and B. Tan, "Farm Area Segmentation in Satellite Images Using DeepLabv3+ Neural Networks," *Communications in Computer and Information Science*, 2019. DOI: 10.1007/978-3-030-54595-6_7.

[5] U. I. Nduanya, K. A. Akpado, and S. N. Ogili, "Machine Learning UNet-Based Segmentation of Sentinel-1A Satellite Images," *Engineering and Technology Journal*, 2023. DOI: 10.47191/etj/v8i7.03.

[6] S. Rivier, C. A. Hinojosa, S. Giancola, and B. Ghanem, "Learning Semantic Segmentation with Query Points Supervision on Aerial Images," *arXiv.org*, 2023. DOI: 10.48550/arxiv.2309.05490.

[7] F. Safarov, K. Temurbek, D. Jamoljon, T. Ochilov, J. C. Chedjou, A. B. Abdusalomov, and Y. I. Cho, "Improved Agricultural Field Segmentation in Satellite Imagery Using TL-ResUNet Architecture," *Sensors*, 2022. DOI: 10.3390/s22249784.

[8] H. Fathipoor, R. Shah-Hosseini, and H. Arefi, "Crop and Weed Segmentation on Ground-Based Images Using Deep Convolutional Neural Network," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2023. DOI: 10.5194/isprs-annals-x-4-w1-2022-195-2023.

[9] M. Tang, K. Georgiou, H. Qi, C. Champion, and M. Bosch, "Semantic Segmentation in Aerial Imagery Using Multi-level Contrastive Learning with Local Consistency," in *IEEE Workshop/Winter Conference on Applications of Computer Vision*, 2023. DOI: 10.1109/wacv56688.2023.00379.

[10] J. Gao, W. Liao, D. Nuyttens, P. Lootens, E. Alexandersson, and J. Pieters, "Transferring Learned Patterns from Ground-Based Field Imagery to Predict UAV-Based Imagery for Crop and Weed Semantic Segmentation in Precision Crop Farming," *Cornell University - arXiv*, 2022. DOI: 10.48550/arxiv.2210.11545.

[11] S. Lingwal, K. K. Bhatia, and M. Singh, "Semantic Segmentation of Landcover for Cropland Mapping and Area Estimation Using Machine Learning Techniques," *Data Intelligence*, 2022. DOI: 10.1162/dint_a_00145.

[12] T. Pham Minh, J. Wijesingha, D. Kottke, M. Herde, D. Huseljic, B. Sick, M. Wachendorf, and T. Esch, "Active Label Refinement for Semantic Segmentation of Satellite Images," *arXiv.org*, 2023. DOI: 10.48550/arxiv.2309.06159.