Plot Segmentation in Agriculture Using Computer Vision Techniques: A Scientific Approach

Allan Casado[1], Cristiane Coutinho[1], Elias Biondo[1], Gábrio Silva[1], Giovana Thomé[1], Rafael Cabral[1], and Thomas Barton[1]

[1]Institute of Technology and Leadership

Abstract

The agricultural sector increasingly relies on precise land management strategies to optimize crop yield and maintain competitive production levels. In this context, accurately identifying productive plots within farms is critical, forming the basis for comprehensive yield analysis and strategic farming decisions. This research addresses the urgent need for precise plot segmentation in the Southern Region of Brazil, where unique topographical and production characteristics present significant challenges to existing computer vision techniques. This study introduces a novel approach to plot segmentation, implementing and comparing three machine learning models based on Convolutional Neural Networks (CNNs). Two of wich are a self-implemented baseline model mixing a UNet architecture with residual connections inspired from the ResNet network architecture (denominated ResUNet) trained with different datasets and the other model is a VGG-16 architecture based. Quantitative analysis reveals that the ResUNet model trained with satellite and AI generated imagery outperforms both the other two models. This enhanced accuracy is crucial in minimizing the loss of productive area in calculations, thereby ensuring the integrity of crop yield assessments. The impact of this research is multifaceted, offering immediate benefits to individual farmers in the Southern Region of Brazil by providing a reliable tool for plot segmentation. The successful application of these models paves the way for advancements in automated farm management systems, precision agriculture, and resource allocation strategies, ultimately contributing to the sustainability and profitability of the agricultural sector.

1 Introduction

Agriculture, a sector that blends tradition with innovation, is a cornerstone of human civilization. With the world's population projected to reach 9 billion by 2050, the Food and Agriculture Organization of the United Nations emphasizes the urgent need for a 70% increase in food production [1]. This global challenge necessitates a shift towards precision agriculture, which can enhance crop yields and promote sustainability through data-driven techniques. A key component of precision agriculture is the accurate segmentation of productive plots, which forms the basis of efficient land management, resource distribution, and yield optimization.

The Southern Region of Brazil presents a unique agricultural landscape characterized by diverse topography, climate, and crop management practices. These factors pose a significant challenge to traditional, manual plot surveying methods, which are often laborious and prone to human error. This context underscores the need for a scalable and precise automated solution. This research is driven by the goal of innovation in precision agriculture, focusing on developing a computer vision model that can accurately delineate productive plots. The approach used involves the creation of a baseline model, which integrates a UNet architecture first introduced in Convolutional Networks for Biomedical Image Segmentation [2] with residual connections present in the ResNet model architecture, and a pretrained model VGG16. This innovative approach to plot segmentation using machine learning models is a significant step forward in precision agriculture, promising to revolutionize how we manage and optimize crop production.

The comparative analysis of these models is not just a theoretical exercise but a practical tool for precision agriculture. It is predicated on their dice coefficient, loss and coverage, focusing on the fidelity of plot polygon identification. Moreover, the practicality of these models will be scrutinized through their performance against a dataset reflective of real-world plot variations, ensuring that our findings are not merely theoretical but are imbued with pragmatic significance.

The ramifications of this study are not just limited to the Southern Region of Brazil but are anticipated to ripple across multiple facets of agriculture globally. At the individual farm level, enhanced plot segmentation equates to more accurate yield forecasts and informed crop management decisions, thereby bolstering the economic sustainability of agricultural operations. Zooming out to a macroscopic lens, the tools developed herein have the potential to revolutionize the precision agriculture movement, offering adaptable solutions that can be tailored to various regions and conditions. By refining the precision of plot segmentation, we also champion the cause of sustainable agriculture, as meticulous plot boundaries are instrumental in optimizing vital resources such as water, fertilizers, and pesticides. This optimization is a stride towards minimizing waste and curtailing environmental degradation. In sum, the study aspires to contribute to the global endeavor of rendering agriculture more productive, efficient, and sustainable in the face of the multifaceted challenges posed by a rapidly changing world, inspiring hope for a brighter future in precision agriculture. The global impact of our research is significant, as it offers a novel and practical approach to plot segmentation that can be applied in various agricultural settings worldwide, contributing to the advancement of precision agriculture and sustainable farming practices.

2 Related Work

The advancement of precision agriculture is contingent on developing accurate and efficient methods for delineating agricultural plots. This section reviews recent contributions in agricultural plot segmentation using computer vision techniques, especially those employing Convolutional Neural Networks (CNNs) and transfer learning methods.

Qi et al. (2024) [3] explored the detailed and precise segmentation of agricultural plots from satellite imagery, utilizing ArcGIS for label creation and training dataset construction. The focus was on applying an optimized hybrid CNN with Transformer elements and attention blocks to enhance accuracy in plot detection, which is crucial for efficient agricultural monitoring. Their study underscores the need for more suitable plotsegment datasets due to unique terrain characteristics and the extensive labeling process required. Methodologically, the research included data collection, annotation, training dataset construction, preprocessing, and training various neural network architectures, including UNet, SegNet, DeeplabV3+, and optimized TransUNet. Key findings involved using Data Augmentation techniques and integrating the Convolutional Block Attention Module (CBAM) into the TransUNet to enhance segmentation capabilities. This work is relevant to our project as it presents a case of high accuracy in plots with high variability, similar to the diverse agricultural landscapes in Brazil.

Wang et al. (2022) [4] proposed a method for mapping rice paddy fields by combining agricultural field boundary extraction with high-resolution satellite imagery and pixel-wise crop classification using SAR images from the Sentinel-1 time series. They utilized a U-net-based fully convolutional model for image segmentation and a simple decision tree classification based on the phenological characteristics of rice. The combination of agricultural field maps with the rice pixel detection model showed promising improvements in the accuracy and resolution of rice paddy field mapping. The study provides an effective method for agricultural field boundary delineation and crop classification, using cutting-edge image processing and machine learning technology applicable across various agricultural regions to improve crop management and monitoring.

Lorena et al. [5] presented a technique called Change Vector Analysis to analyze land use/land cover variability and dynamics in Peixoto, Acre, using TM-Landsat data. Their methodology included Tasseled Cap Transformation and the calculation of Change Vector Magnitude to identify changes. This study contributes to our project by demonstrating how spectral transformations and temporal analysis can be applied to identify plots and locate land portions that have varied over time and could be adapted to enhance the detection and classification of different crop types and vegetative growth stages.

Persello et al. [6] introduced a strategy for detecting agricultural field boundaries using a fully convolutional network combined with combinatorial grouping. Extensive experimental analysis was conducted in real case studies in Nigeria and Mali, introducing a precision-recall framework for accuracy assessment in remote sensing. The proposed technique outperformed state-of-the-art algorithms, showcasing the potential of deep learning approaches in delineating agricultural fields in smallholder farms from very high-resolution satellite images.

Jong et al. [7] proposed an adversarial deep learning approach to improve field boundary delineation in ResUNets. Their adversarial settings significantly enhanced the quality of field boundary predictions and reduced the dependence on post-processing methods. The study highlighted the challenges of low-data settings for transfer learning, an essential consideration for regions with scarce training data, like some agricultural areas in Brazil. The research concluded that adversarial training is a promising way to improve the quality of field boundaries during prediction time.

These studies collectively inform our project by providing insights into the application of CNNs, attention mechanisms, data augmentation techniques, and adversarial training to improve plot segmentation. Adapting these methods to the unique conditions of Brazil's Southern Region is a critical step toward developing tailored models that can accurately delineate agricultural plots in complex terrains.

3 Methodology

In this study, it was employed Sentinel-2 satellite imagery to perform plot segmentation in the Southern Region of Brazil and also AI generated images for a bigger generalization capacity. Sentinel-2 provides high-resolution optical images suitable for agricultural monitoring due to their rich spectral information and frequent revisit time. The Sentinel-2 multi-spectral instrument (MSI) captures data in 13 spectral bands ranging from visible and near-infrared to shortwave infrared at different spatial resolutions (10m, 20m, and 60m).

For plot segmentation, we focused on images with a 10-meter resolution, which offers a detailed view of the agricultural landscapes. The quality of the imagery is ensured by the rigorous calibration of the satellite sensors and the systematic removal of atmospheric effects that could distort the reflectance values of the Earth's surface. Additionally, the dataset was filtered to include only images with cloud coverage of less than 10%, thereby minimizing the impact of cloud shadows and ensuring the consistency of the input data for the segmentation models.

3.1 Data Preprocessing

The preprocessing of the images included several steps to ensure that the data fed into the machine-learning models was of high quality:

1. **Atmospheric Correction:** The raw satellite images are corrected using the Sen2Cor processor to convert top-of-atmosphere reflectance values into bottom-of-atmosphere, or surface reflectance values. This step is crucial for accurate land cover classification.

2. **Cloud Masking:** Clouds and their shadows are detected and masked using thresholding techniques and morphological operations on the brightness and cloud probability layers provided by Sentinel-2's cloud detection algorithm.

3. **Reprojection and Resampling:** Images are reprojected to a standard coordinate reference system (CRS) and resampled to ensure pixel alignment across the different spectral bands.

4. **Normalization:** Pixel values are normalized across the dataset to bring all input features to a similar scale, which is essential for the practical training of neural networks.

5. **Data Augmentation:** To increase the diversity of the training dataset and to prevent overfitting, we apply random transformations such as rotations, flips, and zooms to the imagery.

6. **Tiling:** Large satellite images are split into smaller, more manageable tiles that fit the input size requirements of the CNN models.

3.2 Data Generation

In order to expand the training dataset and train a model more capable of generalizing predictions, a dataset containing AI generated satelite images and masks was incorporated in on of the trainings of the ResUNet model described in the next section (demoninated Cross Training model). The main visual differences can be seen with the examples below:

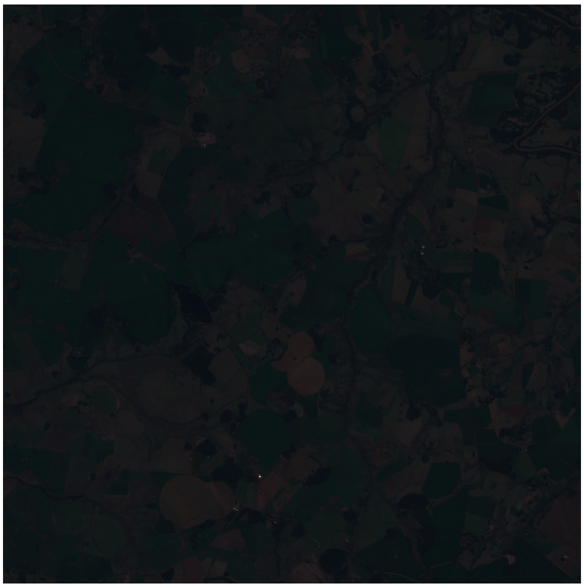

Figure 1: Sentinel image

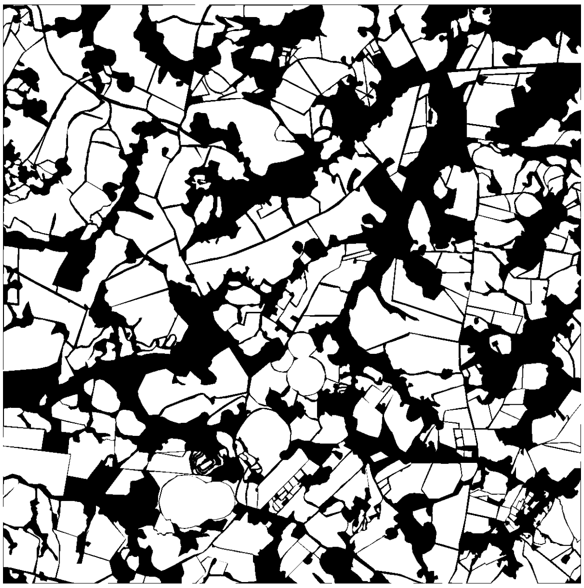

Figure 2: Sentinel image related mask

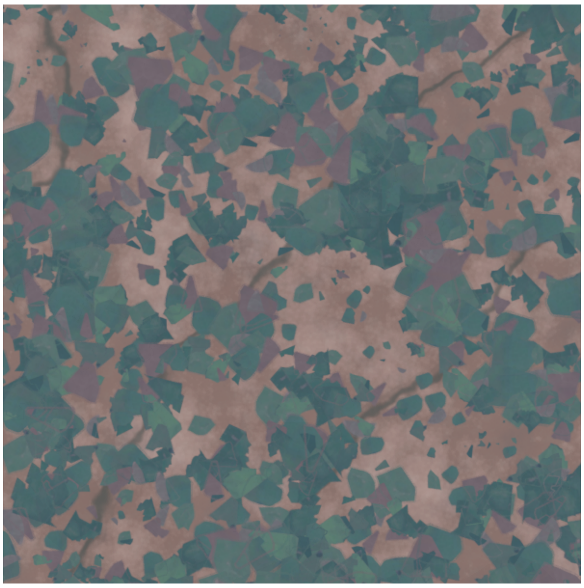

Figure 3: AI generated image

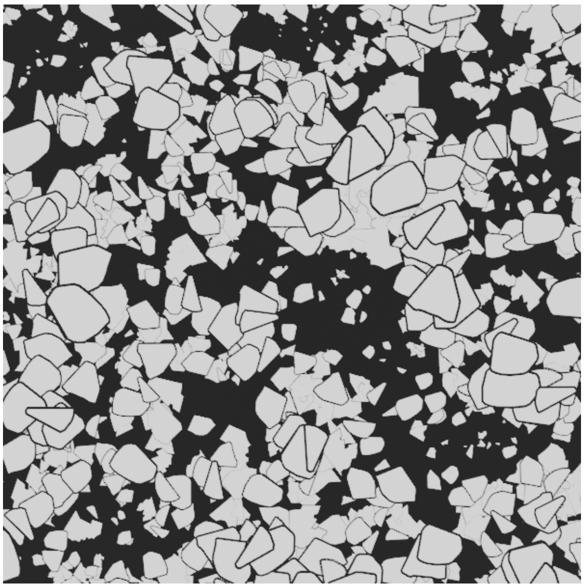

Figure 4: AI generated image related mask

3.3 Model Development

The baseline model architecture is a custom CNN that leverages the UNet structure, renowned for its effectiveness in biomedical image segmentation, which parallels the feature scale and complexity of agricultural plot segmentation. Alongside the UNet architecture, the idea of residual connections, characterized by skipping layers to foster efficient training in deep networks, was introduced to preserve data across the downsampling layers in the encoder part of the architecture, since the model's goal is to generate a mask the same size as the input images. This combination of the UNet and ResNet models derived the ResUNet architecture, focused on precise segmentation to generate accurate plot segmentation masks.

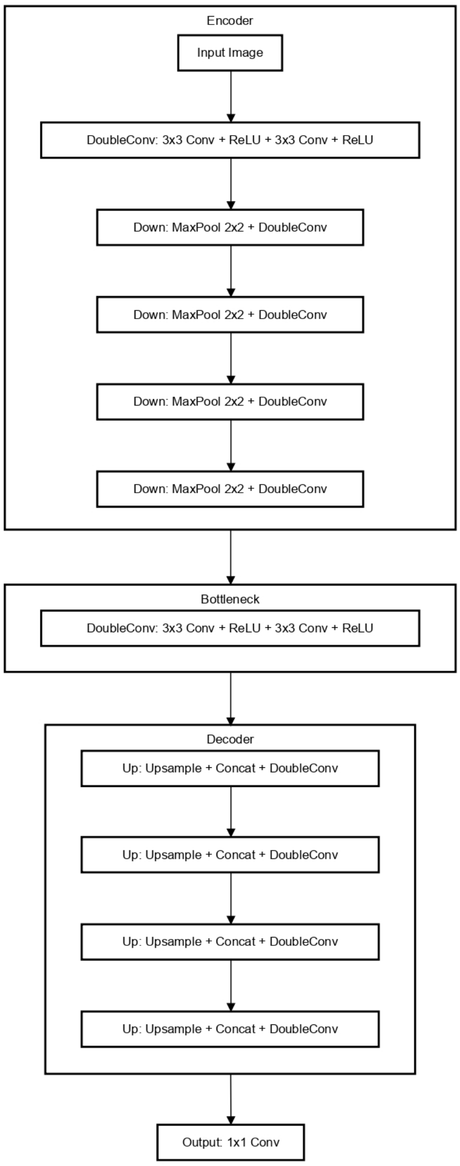

Figure 5: ResUNet utilized architecture

In the other hand, the pre-trained model uses the VGG16 architecture uses convolution layers with small sized filters, as said in [8]. The architecture is the following:

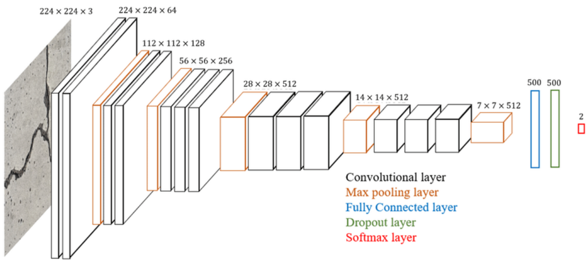

Figure 6: VGG16 architecture

3.4 Model Training and Validation

The models were trained with different datasets and split into training and validation subsets. The training process involved the following steps:

1. **Loss Function:** We used binary cross-entropy as the loss function for both models, suitable for binary classification tasks.

2. **Optimizer:** The Adam optimizer was chosen for its adaptive learning rate capabilities, which help it converge to the optimal solution faster.

3. **Metrics:** To evaluate the models, we focused on precision, recall, and the F1 score, emphasizing the fidelity of plot polygon identification.

4. **Validation:** The models were validated using a separate dataset to ensure that the performance metrics indicated the models' abilities to generalize to unseen data.

5. **Hyperparameter Tuning:** We conducted several experiments to fine-tune the hyperparameters of the models, such as the learning rate, batch size, and number of epochs.

With the difference between the training datasets being:

Model	Training dataset
ResUNet1	Sentinel's Satellite imagery
CrossTraining	Sentinel's Satellite imagery and AI generated images
VGG16	Sentinel's Satellite imagery

Table 1: Comparison of Models Based on Loss, Dice Coefficient, and Coverage in their final training epoch

3.5 Evaluation Indicators Using Dice Coefficient

The Dice Coefficient is a crucial evaluation metric used in image segmentation tasks to measure the overlap between predicted and ground truth segmentations. It quantifies the similarity between these two sets, making it instrumental in assessing the performance of segmentation models. The Dice Coefficient ranges from 0 to 1, with 1 indicating perfect overlap and 0 indicating no overlap [9].

Dice Coefficient is calculated using the following formula:

$$\text{Dice Coefficient} = \frac{2|A \cap B|}{|A| + |B|}$$

Where A is the predicted set and B is the ground truth set. In binary segmentation, it is often expressed as:

$$\text{Dice Coefficient} = \frac{2\sum_i p_i g_i}{\sum_i p_i + \sum_i g_i}$$

Where p_i represents the predicted binary value and g_i represents the ground truth binary value for the i-th pixel.

The Dice Loss, derived from the Dice Coefficient, is used as a loss function in model training to maximize the overlap between predicted and actual segmentations. The Dice Loss is defined as:

$$\text{Dice Loss} = 1 - \text{Dice Coefficient}$$

Alternatively, a small constant ϵ is added to avoid division by zero:

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

Coverage in segmentation tasks generally refers to how well the predicted segmentation covers the ground truth. This can be assessed through precision (the proportion of true positives among the predicted positives) and recall (the proportion of true positives that are correctly predicted). These metrics provide a more comprehensive understanding of the model's performance by highlighting both its ability to correctly identify positive instances and its tendency to avoid false positives.

In the context of the project "Plot Segmentation in Agriculture Using Computer Vision Techniques," we employed the Dice Coefficient and Dice Loss to evaluate and train our segmentation models. These metrics were chosen to ensure that the models accurately delineated the boundaries of agricultural plots, essential for precise segmentation.

By using the Dice Coefficient as an evaluation indicator and Dice Loss for model training, our approach focused on optimizing the overlap between predicted and ground truth segmentations. This method proved effective in enhancing the accuracy and reliability of our models, thereby contributing significantly to the efficiency of agricultural plot segmentation.

The adoption of the Dice Coefficient and Dice Loss in our project facilitated the development of robust segmentation models capable of generalizing well to new data, ultimately aiding in the optimization of agricultural management and sustainability.

3.6 Computational Environment

All experiments were conducted using Python 3.10.12, with the following libraries and their respective versions:

- PyTorch 2.0
- Keras 2.4.3
- NumPy 1.19.5

- OpenCV 4.5.1
- Scikit-learn 0.24.1

To expedite the training process, the models were trained on a machine equipped with 12 Intel(R) Xeon(R) CPUs (2.20GHz), 83.5 GB of RAM, and a NVIDIA A100-SXM4-40GB GPU.

3.7 Performance and Evaluation

The performance of the models was evaluated using the metrics mentioned above. In addition to these quantitative measures, we conducted a qualitative analysis by visually inspecting the segmented plots. This allowed us to assess the models' segmentation accuracy in real-world conditions.

4 Results

4.1 Model Performance

The performance of the implemented models was evaluated based on their respective dice coefficient and dice loss over training epochs and calculating the coverage of each generated mask compared to the original.

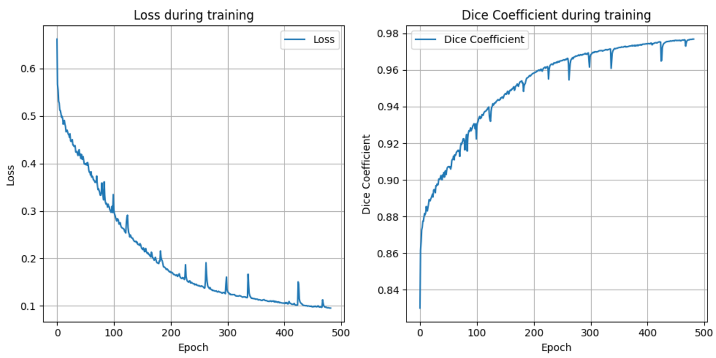

Figure 7: Epoch versus loss and dice coefficient for the ResUNet model trained with Sentinel-2 images.

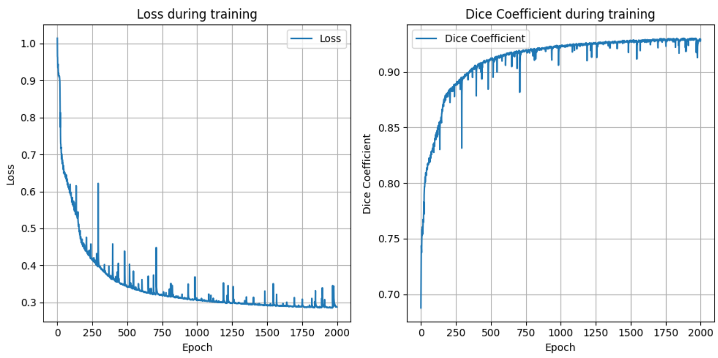

Figure 8: Epoch versus loss and dice coefficient for the ResUNet Cross Training model trained with synthetic generated images and Sentinel-2 images.

Figure 9: Epoch versus dice loss for the VGG16 model trained with Sentinel-2 images.

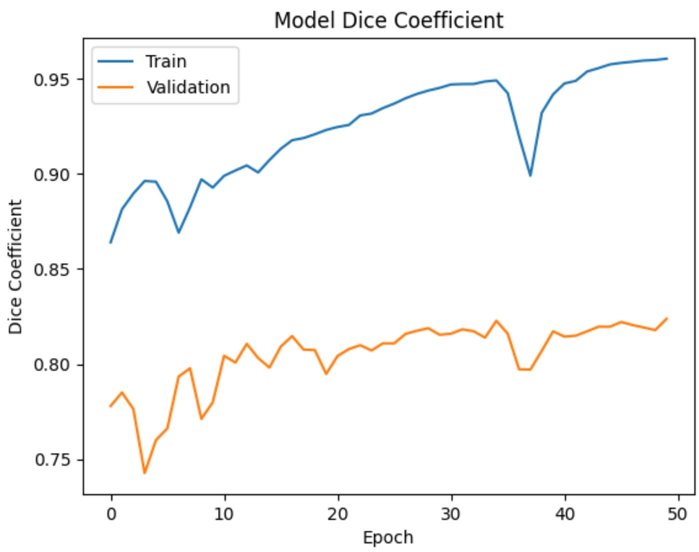

Figure 10: Epoch versus dice coefficient for the VGG16 model trained with Sentinel-2 images.

Model	Coverage (%)
ResUNet	92.05
ResUNet CrossTraining	20.79
VGG16	88.64

Table 2: Comparison of Models Based Coverage in their final training epoch

4.2 Qualitative Analysis

Qualitative comparisons between the predicted segmentation maps and the ground truth for both models are presented in Figures. Both images were chosen randomly among the

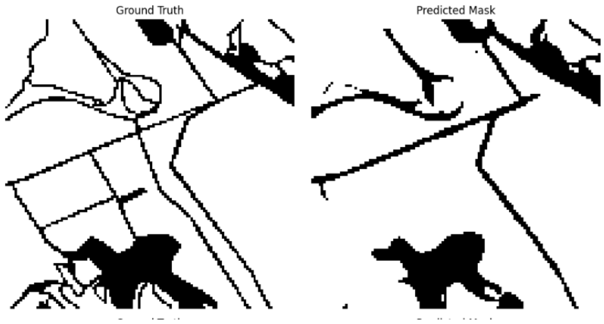

Figure 12: Qualitative comparison of predicted mask with the groud truth for ResUNet CrossTraining model

dataset.

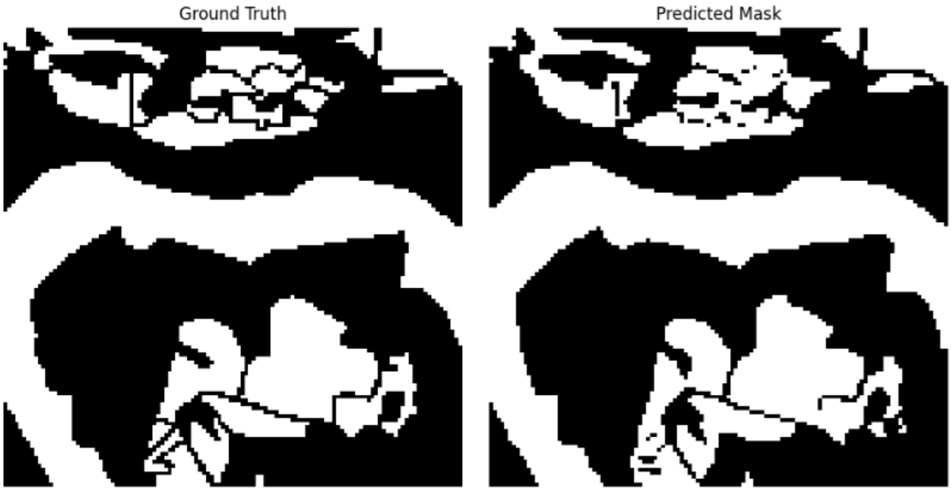

Figure 11: Qualitative comparison of predicted mask with the groud truth for ResUNet model

5 Training time

The training time of the models is an essential factor to consider, especially when scaling the solution to larger datasets or deploying it in real-time applications. The table below summarizes the training times for the different models implemented in the study:

Model	Epochs	Training Time
ResUNet	200	20 minutes
ResUNet CrossTraining	150	18 minutes
VGG16	50	68.47 seconds

Table 3: Comparison of Models Based on Training Time

The ResUNet model trained for 200 epochs took approximately 20 minutes, highlighting the computational intensity of training deep learning models with complex architectures. The ResUNet CrossTraining model, trained for 150 epochs, required 18 minutes, slightly less time due to fewer epochs but still substantial given the inclusion of synthetic data.

In contrast, the VGG16 model trained for 50 epochs required only 68.47 seconds, demonstrating the efficiency of using pre-trained models and transfer learning techniques.

Overall, while the custom ResUNet models provide high accuracy and precise segmentation, the VGG16 model offers a faster training alternative.

6 Discussion

This study highlights the advancements in precision agriculture through accurate plot segmentation using computer vision techniques. The comparative analysis of the ResUNet and VGG16 models revealed significant insights.

The ResUNet model trained with Sentinel-2 images demonstrated superior performance with a Dice Coefficient of 0.9703 and a Coverage of 92.05%, compared to the ResUNet CrossTraining model, which had a Dice Coefficient of 0.9291 and a Coverage of 20.79%. This indicates the importance of high-quality, real-world data for training segmentation models.

The VGG16 model, despite having a faster training time of 68.47 seconds for 50 epochs, showed a Dice Coefficient of 0.9645 and a Coverage of 88.64%, making it a viable option when computational efficiency is prioritized.

The integration of residual connections in the ResUNet architecture improved data preservation across downsampling layers, enhancing segmentation precision. The use of synthetic AI-generated data, although less effective than real satellite imagery, suggests a potential avenue for generalizing results in any given region.

The qualitative analysis confirmed the accuracy of the models, with the ResUNet model's predictions closely aligning with ground truth masks. This accuracy is crucial for practical applications in agriculture, ensuring efficient resource management and yield optimization.

Overall, the study's findings underscore the scalability of these models across diverse agricultural landscapes, contributing to the economic sustainability of agricultural operations.

7 Conclusion

The study demonstrated the effectiveness of computer vision techniques in agricultural plot segmentation. The ResUNet model trained with real Sentinel-2 images outperformed other models, achieving a higher Dice Coefficient and Coverage, thus ensuring precise plot delineation. The VGG16 model provided a faster training alternative, suitable for scenarios where computational efficiency is essential.

The findings highlight the potential of these models to enhance resource allocation, crop monitoring, and yield estimation in precision agriculture. Future research should explore the scalability across different regions and crop types, incorporating additional data sources for improved accuracy.

Overall, this study significantly contributes to the field of precision agriculture, promoting sustainability and efficiency in meeting the food demands of a growing global population.

References

[1] Subunit exchange in spinach short-form rubisco activase. CORE, 2017. Accessed: 2024-05-10.

[2] Thomas Brox Olaf Ronneberger, Philipp Fischer. Convolutional networks for biomedical image segmentation. *arXiv - Cornell University*, 2015.

[3] L. Qi, D. Zuo, Y. Wang, Y. Tao, R. Tang, J. Shi, J. Gong, and B. Li. Convolutional neural network-based method for agriculture plot segmentation in remote sensing images. *Remote Sensing*, 16:346, 2024.

[4] M. Wang, J. Wang, Y. Cui, J. Liu, and L. Chen. Agricultural field boundary delineation with satellite image segmentation for high-resolution crop mapping: A case study of rice paddy. *Agronomy*, 12:2342, 2022.

[5] R. B. Lorena, J. R. Santos, Y. E. Shimabukuro, I. F. Brown, and H. J. H. Kux. A change vector analysis technique to monitor land use/land cover in sw brazilian amazon: Acre state. ResearchGate.

[6] C. Persello, V. A. Tolpekin, J. R. Bergado, and R. A. de By. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sensing of Environment*, 229:1–15, 2019.

[7] M. Jong, K. Guan, S. Wang, Y. Huang, and B. Peng. Improving field boundary delineation in resunets via adversarial deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 182:1–13, 2022.

[8] Andrew Zisserman Karen Simonyan. Very deep convolutional networks for large-scale image recognition. *arXiv - Cornell University*, 2014.

[9] Mostafa Ibrahim. Understanding the difference in performance between binary cross-entropy and categorical cross-entropy. *Journal of Machine Learning Research*, 2024. Explicação detalhada das funções de perda de entropia cruzada binária e categórica e suas aplicações em aprendizado de máquina.