Divide and conquer: Using plots to understand agricultural activities

Ian Nunes[1], Hugo Neves de Oliveira[2], Edemir Ferreira de Andrade Junior[1], and Octavio Costa de Oliveira[1]

[1] Brazilian Institute of Geography and Statistics (IBGE), Rio de Janeiro, Brazil
{ian.nunes,edemir.ferreira,octavio.oliveira}@ibge.gov.br
[2] Federal University of Viçosa (UFV), Minas Gerais, Brazil
hugo.n.oliveira@ufv.br

Abstract

The emergence of Deep Learning techniques combined with Big Data allows for the extraction of high-level semantic features directly from massive volumes of data. While this paradigm has revolutionized Computer Vision using natural RGB images, specialized domains such as Remote Sensing often lack massive labeled datasets. Consequently, training neural networks in these fields requires methods designed to go beyond standard supervised learning. In this context, the Brazilian Agricultural Census presents both a significant challenge and a unique opportunity, covering 8.5 million km^2 and over 5 million establishments. This work details a strategy using Remote Sensing and Computer Vision for automated agricultural plot delineation. By utilizing Census outcomes to guide crop mapping and yield estimation, while simultaneously using automatic field boundary delineation as input for Census fieldwork, this framework enables the collection of high-quality ground truth data. The proposed methodology is capable of transforming and improving the production of official statistics across Brazil's diverse biomes. To demonstrate the robustness of the approach, we provide a detailed analysis of results from two representative municipalities.

1 Introduction

The agricultural sector is a cornerstone of the Brazilian economy, representing nearly 25% to 30% of the national GDP and playing a vital role in global food security [5, 2, 12]. With the capacity to provide food, fiber, and bioenergy for approximately 800 million people, Brazil has established itself as a global leader in the sector. However, a significant gap persists: the lack of a robust, official monitoring system for agricultural activities. Currently, the predominance of private initiatives limits widespread access to evidence-based technologies and fails to meet the rigorous requirements for producing official national statistics.

This work proposes leveraging the 12th Agricultural Census as a transformative catalyst for the production of official agricultural statistics in Brazil. We outline

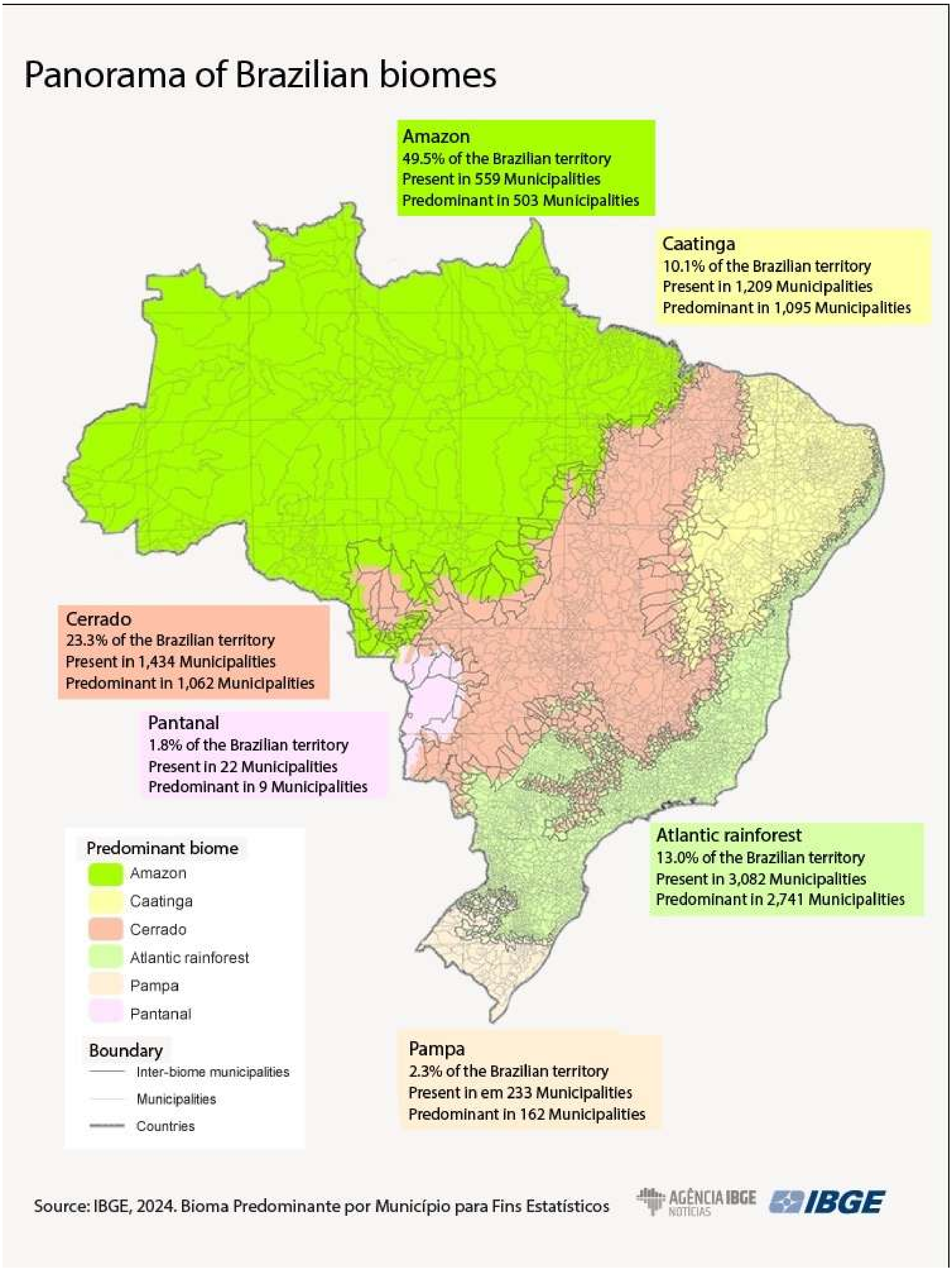


Figure 1: The diversity of Brazilian biomes and the scale of territorial monitoring [13]

a comprehensive framework for data generation, methodological approaches, and the orchestration necessary to support a modern, timely information system. The primary objective of this study is to present a strategic roadmap for modernizing Brazilian agricultural statistical production.

The task of monitoring Brazil's agricultural landscape is inherently multifaceted, shaped by immense territorial scale and environmental diversity. As shown in Figure 1, the country comprises six distinct biomes: Amazon, Cerrado, Caatinga, Atlantic Forest, Pantanal, and Pampa. Each biome exhibits unique vegetation patterns, seasonality, and varying degrees of anthropogenic pressure. Consequently, monitoring efforts require algorithms specifically calibrated to the ecological and social nuances of each region.

Beyond geographic diversity, complex agricultural dynamics further complicate analysis. The prevalence of double and triple cropping—such as the sequential planting of soy, corn, and rice on the same plot within a single year—demands high-resolution temporal models capable of distinguishing between multiple crop cycles. Furthermore, persistent cloud cover in critical regions like the Amazon and the Cerrado during rainy seasons creates significant blind spots in optical satellite data, presenting a major technical hurdle for continuous land-use monitoring.

In addition to environmental factors, land fragmentation and administrative gaps pose significant challenges. Brazil's rural landscape is a mosaic of millions of properties ranging from small-scale family farms to vast industrial estates. Although the National Rural Environmental Registry (CAR) is a vital tool, the absence of a fully integrated and verified administrative record complicates data cross-referencing and accountability.

Finally, technical capacity and resource accessibility remain critical barriers. There is a persistent skill gap within public institutions, characterized by a shortage of personnel trained in complex Geographic Information Systems (GIS) and Artificial Intelligence (AI). This is exacerbated by the prohibitive costs of high-resolution imagery and specialized hardware/software, which often limit the institutional reach of advanced monitoring technologies.

Synthesis Ultimately, the challenge of monitoring Brazilian agriculture transcends mere geographic size; it is a convergence of environmental heterogeneity, intensive production systems, and structural bottlenecks. Any effective solution must be scalable, adaptable to regional specificities, and resilient enough to operate amidst data gaps and institutional constraints.

2 Proposed Methodological Framework: An Integrated Cycle of Earth Observation and Field Enumeration

To address the challenges of scale and dynamism inherent in Brazilian agriculture, this work proposes a cyclical and integrated framework that fuses **Artificial Intelligence (AI)**, **Remote Sensing (RS)**, and **in-situ data collection**. Unlike traditional linear pipelines, where fieldwork and digital analysis are often treated as separate silos, the proposed architecture establishes a continuous feedback loop. As can be seen in Figure 1, the workflow is anchored by the automated delineation of agricultural plots, which serves as the common spatial unit for both the Agricultural Census and continuous crop monitoring.

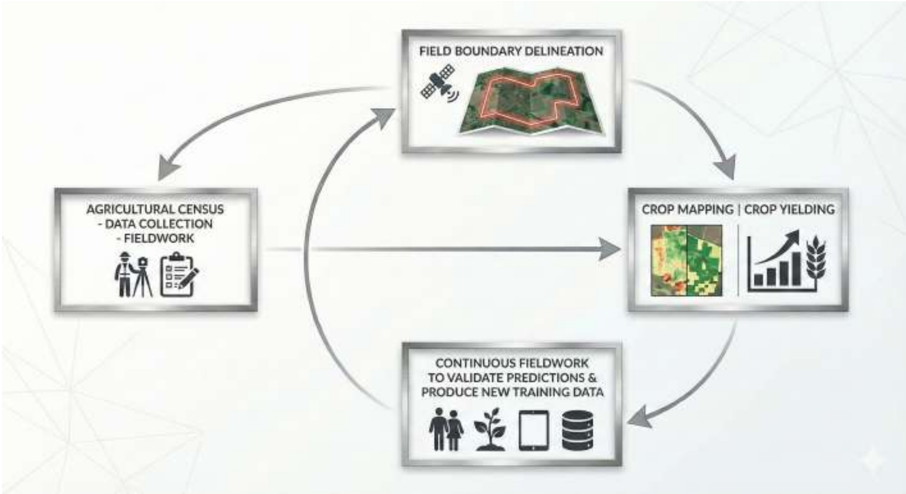


Figure 2: Schema for the information production.

2.1 Workflow Description

The process begins with **AI Field Boundary Delineation (FBD)**. Utilizing medium-resolution satellite imagery and specialized **field boundary delineation models**, the system automatically infers the geometry of agricultural plots. This vector layer serves as the spatial backbone of the entire system, feeding into two distinct but interconnected operational streams:

1. **Census & Field Calibration:** The inferred boundaries provide enumerators with a pre-mapped "digital twin" of the territory. During the **Agricultural Census**, field agents validate these boundaries and attribute specific production data to them, explicitly recording **what is produced (crop type) and the volume produced (yield)**. This effectively transforms the Census from a mere statistical survey into a massive campaign of ground-truth generation.

2. **Crop Mapping & Yield Estimation:** Simultaneously, these boundaries define the areas of interest for temporal spectral analysis. Within each delineated plot, algorithms analyze time-series data—including the **Normalized Difference Vegetation Index (NDVI)** [19], the **Enhanced Vegetation Index (EVI)** [11], and **Synthetic Aperture Radar (SAR)** backscatter [20]—to classify crop types and estimate yields, free from the noise of mixed pixels at the edges.

The cycle closes with **Continuous Fieldwork**. Discrepancies between model predictions (e.g., low confidence in crop classification) trigger targeted field inspections. The data collected in these missions verifies the inference and, crucially, produces new labeled samples to retrain and refine the AI models, creating a virtuous cycle of active learning.

2.2 Advantages and Strategic Benefits

The adoption of this circular framework offers significant advantages over traditional methods:

- **Spatial Consistency:** By using FBD as the input for the Census, we ensure that statistical data is intrinsically linked to physical geography. This eliminates the disconnect between administrative records and the actual land use.

- **Resource Optimization:** The system directs human effort where it is most needed. Instead of sweeping entire regions blindly, field teams can be deployed specifically to validate areas with high model uncertainty or rapid land-use change.
- **Temporal Granularity:** While the Census provides a decennial snapshot, the "Crop Mapping & Yield Estimation" branch allows for seasonal or even monthly updates of agricultural production, providing policymakers with near real-time data.
- **Evolving Accuracy:** The feedback loop ensures the system is anti-fragile; it improves with use. As the model encounters new biomes or crop varieties, the continuous injection of validated field data mitigates model drift.

2.3 Challenges and Critical Bottlenecks

Despite its potential, the implementation of this framework imposes non-trivial challenges:

- **Domain Shift and Generalization:** A boundary delineation model trained on the geometric fields of the Center-West (Cerrado) may struggle to generalize to the irregular, fragmented plots of the Northeast (Caatinga) or the South. Ensuring model robustness across diverse biomes requires sophisticated domain adaptation techniques.
- **Computational High-Availability:** Processing petabytes of imagery to update boundaries and classifications on a continental scale demands a robust, high-performance computing infrastructure (HPC) and optimized orchestration pipelines.
- **Synchronization Latency:** The lag between identifying an anomaly via satellite and deploying a field team must be minimized. If the crop cycle ends before validation occurs, the ground truth is lost, breaking the feedback loop.

3 Field Boundary Delineation (FBD)

The accurate delineation of agricultural field boundaries is the cornerstone of our proposed monitoring framework. It defines the fundamental spatial unit for subsequent census enumeration and yield estimation. To address the challenge of segmenting diverse agricultural landscapes across Brazil, we employed a state-of-the-art Deep Learning approach, combining high-resolution architectures with self-supervised pre-training strategies.

3.1 Data Curation and Pre-processing

High-quality input data is critical for training robust segmentation models. Our dataset comprises three distinct components:

3.1.1 Earth Observation Imagery

Input data consists of medium-resolution optical imagery from the Sentinel-2 constellation. To mitigate the persistent cloud cover typical of tropical regions and ensure temporal consistency, we constructed monthly mosaics. Specifically, we utilized the median pixel composition technique over the time series individually for July, August, and September 2023. This temporal window captures critical phenological stages for winter crops and preparation for summer planting. The median composite approach effectively filters out transient noise (clouds and shadows) while preserving the spectral integrity of the agricultural features.

3.1.2 Ground Truth and Annotations

The supervised component of our model relies on a proprietary dataset of unprecedented scale produced by the Brazilian Institute of Geography and Statistics (IBGE). This dataset contains over 600,000 manually annotated polygons, validated by technicians with local domain expertise. To ensure a rigorous evaluation of the model's generalization capabilities, we implemented a spatial stratification based on experts' directions. Municipalities used for training were separated from those used for testing, preventing autocorrelation leakage and ensuring that test metrics reflect true performance in unseen territories.

3.1.3 Benchmarking Data

For comparative analysis, we utilized the VARDA Foundation Global FieldID dataset[1] (July 2023 snapshot). This external dataset serves as a baseline to benchmark our model against global standards for field boundary delineation.

3.2 Architecture

Standard Convolutional Neural Networks (CNNs)[15] for segmentation, such as U-Net[18] or ResNet-50 [10], typically employ an encoder-decoder structure that downsamples the input to extract semantic context, recovering spatial resolution only in the final stages. This process often results in the loss of fine spatial details, leading to "blurred" boundaries—a critical failure mode for cadastral applications. To overcome this limitation, we adopted the High-Resolution Network (HRNet) [21]. The core advantage of HRNet is its ability to maintain high-resolution representations throughout the entire forward pass. It achieves this by connecting high-to-low resolution convolution streams in parallel and repeatedly exchanging information across resolutions (multi-scale fusion). This architecture allows the model to capture robust semantic features (from low-resolution streams) without sacrificing the spatial precision required to delineate narrow boundaries between adjacent plots (from high-resolution streams), as presented in Figure 3.

Complementing this architectural choice, we developed a proprietary self-supervised pre-training method designed to optimize the network's initialization. This novel approach enables the model to learn robust feature representations from unlabeled data before the fine-tuning stage. The specific impact of this method on segmentation performance, particularly in challenging biomes, constitutes a core contribution to the process and will be presented in the results section.

[1] `https://fieldid.varda.ag/hub/downloads`

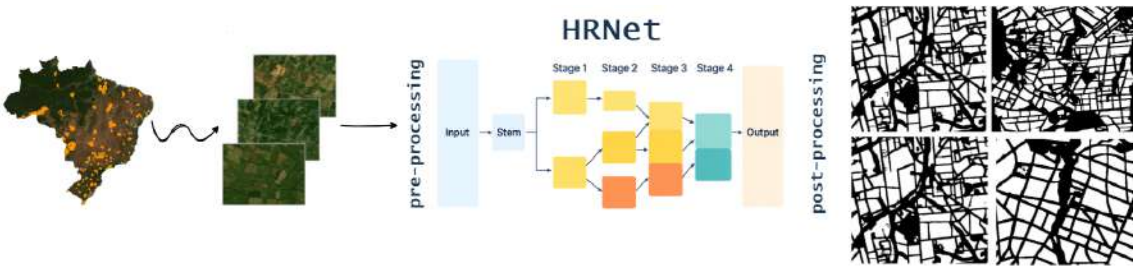


Figure 3: The proposed execution pipeline: from temporal pre-processing of Sentinel-2 mosaics to high-fidelity segmentation using HRNet.

3.3 Self-Supervised Pre-training (SSL)

To mitigate the domain shift across biomes and improve model generalization, we implemented a Self-Supervised Learning approach. The literature on Representation Learning via SSL is extensive and broadly categorized into three primary streams:

1. **Transformation Prediction-based SSL:** Approaches that learn representations by solving pretext tasks related to geometric or radiometric transformations [16, 17, 7];
2. **Similarity Learning-based SSL:** Methods focused on contrastive learning, aiming to maximize agreement between differently augmented views of the same data point while distancing others [3, 1, 4];
3. **Masked Image Modeling (MIM):** A generative paradigm where the model learns to reconstruct masked portions of the input image, encouraging the capture of dense contextual features [8, 6, 22].

Leveraging this theoretical foundation, our pipeline utilizes a diverse dataset comprising agricultural landscapes from both Brazil and Europe to pre-train the network backbone, ensuring robust feature extraction before fine-tuning on the specific IBGE annotations.

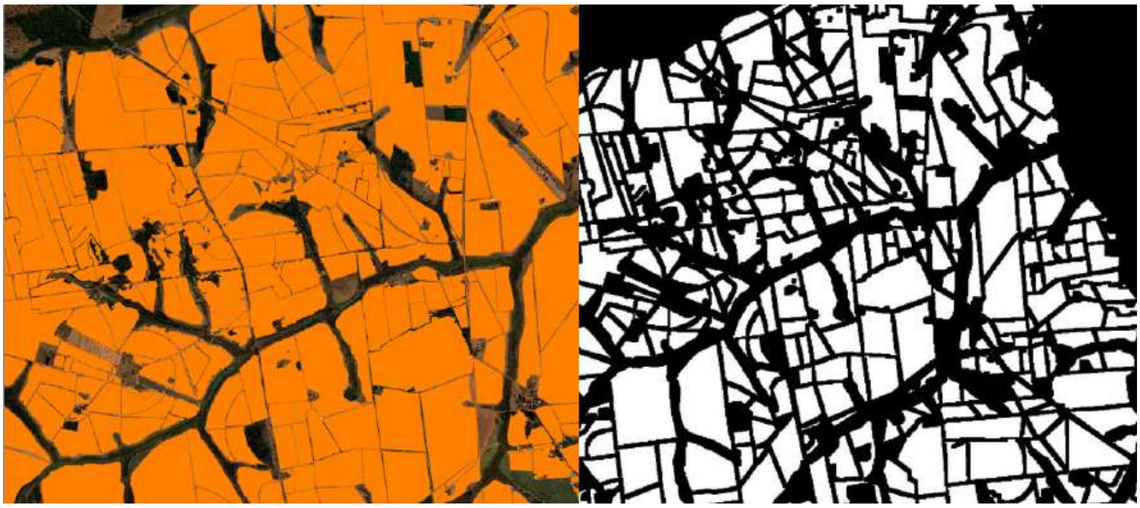
4 Operational Tests and Results

The evaluation of the proposed framework was conducted to assess not only the pixel-level accuracy but, more importantly, the instance-level reliability required for a National Census. We compared our proposed method against two baselines: the **VARDA** global foundation model and a standard **HRNet** trained via fully supervised learning without SSL pre-training.

4.1 Qualitative Analysis

The visual inspection of the segmentation results, presented in Figure 4, reveals the impact of the domain-specific design. In regions characterized by high geometric irregularity and small-scale farming, such as the Northeast of Brazil, the global baseline (VARDA) frequently fails to detect boundaries, merging adjacent smallholder plots into single amorphous regions (under-segmentation).

In contrast, the proprietary model demonstrates a superior ability to delineate complex boundaries. The Self-Supervised Learning (SSL) pre-training effectively allows the model to distinguish between subtle textural changes—such as the transition from a fallow field to native vegetation—even when spectral differences are minimal. This results in sharper boundary definitions and a significant reduction in blob-like artifacts common in standard semantic segmentation.



(a) An example of the mask on the left, and the predictions on the right.

(b) A second example of the mask on the left, and the predictions on the right.

Figure 4: Visual comparison of segmentation results. Note how the proposed method preserves the boundary integrity of small, irregular plots where the baseline (left) tends to merge them. Predictions are semantically close to the ground truth.

4.2 Evaluation Metrics

To rigorously assess the performance of the proposed framework, we employed a comprehensive set of metrics divided into two distinct categories: **Pixel-Level Metrics** and **Instance-Level Metrics**. While the former serves as a general proxy for segmentation quality, the latter is crucial for this study as it aligns more closely with human perception and the specific requirements of the Agricultural Census (e.g., distinguishing adjacent plots).

4.2.1 Pixel-Level Metrics

These metrics evaluate the classification accuracy of individual pixels, ignoring the concept of distinct objects. We report:

- **Jaccard Index (IoU):** Reported in both its binary and weighted variations, measuring the overlap between the predicted segmentation mask and the ground truth.

4.2.2 Instance-Level Metrics

Given that the primary goal is to delineate individual agricultural plots, instance-level metrics are prioritized in our analysis (Section 4.3). Unlike pixel metrics, these indices penalize the merging of neighboring fields (under-segmentation) or the fracturing of a single field (over-segmentation).

Standard Instance Metrics: We utilize the standard COCO metrics [9]:

- **Average Precision (AP):** Measures the precision of detection at various IoU thresholds.
- **Average Recall (AR):** Measures the proportion of ground truth objects correctly detected.

Note on Metric Prioritization: In the context of field boundary delineation, missing a plot is often more detrimental than slightly over-estimating boundaries. Therefore, **Average Recall (AR)** is considered the most critical metric for our comparisons. Higher values of AP and AR indicate better performance.

Segmentation Consistency Metrics: To capture specific geometric errors, we adopted the metrics [14]:

- **Global Over-Segmentation (GOS):** Quantifies the degree to which single plots are incorrectly split into multiple fragments.
- **Global Under-Segmentation (GUS):** Quantifies the merging of distinct adjacent plots into a single object.
- **Global Total-Segmentation (GTS):** A combined error metric.

For GOS, GUS, and GTS, *lower* values indicate better performance. These metrics provide a granular understanding of the model's topological accuracy, offering insights that approximate visual validation by human experts.

4.3 Quantitative Performance

Table 1 details the quantitative performance across five municipalities representing distinct agricultural realities. A key aspect of our evaluation is the breakdown of performance by instance size, where **AR50 (S)** measures recall for small plots (area pixels) and **AR50 (M)** for medium plots (area between and pixels).

Impact on Family Farming and Smallholdings: A crucial finding for the Brazilian context is the performance on small agricultural plots. Considering that over 70% of Brazil's agricultural holdings are family-based—often characterized by smaller, fragmented areas—the model's ability to recover these specific instances is vital. Our method demonstrates a consistent and significant advantage over the VARDA baseline in the **AR50 (S)** metric across all municipalities. For instance, in **Ibiporã**,

Source	Jacc.(w)	GOS ↓	GUS ↓	AP50 ↑	AR50 ↑	AR50 (S) ↑	AR50 (M) ↑
Chorozinho, CE							
VARDA	0.777	0.064	0.117	0.004	0.027	0.022	0.053
HRNet	0.935	0.067	0.035	0.135	0.243	0.213	0.364
Ours	**0.937**	**0.039**	**0.044**	**0.216**	**0.341**	**0.324**	**0.393**
Jaboticabal, SP							
VARDA	0.838	0.106	0.107	0.397	0.580	0.526	0.610
HRNet	0.882	0.121	0.073	0.618	0.666	0.588	0.795
Ours	**0.927**	**0.061**	**0.051**	**0.759**	**0.828**	**0.783**	**0.848**
Irapuã, SP							
VARDA	0.900	0.058	0.060	0.456	0.645	0.611	0.666
HRNet	0.955	**0.031**	0.026	0.723	0.765	0.719	**0.842**
Ours	**0.956**	0.032	**0.024**	**0.726**	**0.812**	**0.797**	0.804
Pitangueiras, SP							
VARDA	0.884	0.070	0.072	0.498	0.670	0.612	0.705
HRNet	0.941	0.038	0.036	0.742	0.777	0.703	0.873
Ours	**0.952**	**0.028**	**0.030**	**0.788**	**0.833**	**0.787**	**0.878**
Ibiporã, PR							
VARDA	0.877	0.072	0.075	0.207	0.399	0.338	0.481
HRNet	0.934	0.069	0.042	0.440	0.548	0.474	0.618
Ours	**0.946**	**0.056**	**0.034**	**0.523**	**0.682**	**0.621**	**0.786**

Table 1: Detailed performance comparison. Arrows indicate whether lower (↓) or higher (↑) values are better. (S) and (M) denote Small and Medium instances.

the AR50 (S) jumps from 0.338 (VARDA) to 0.621 (Ours), and in the challenging landscape of **Chorozinho**, it rises from a negligible 0.022 to 0.324. This highlights that our approach does not merely optimize for large, easy-to-detect monocultures but provides the necessary granularity to include family farming in the automated census pipeline.

Robustness to Domain Shift: The generalizability of the model is most evident in **Chorozinho (CE)**. Here, the domain shift causes the VARDA baseline to collapse (AP50 of 0.004). While the supervised HRNet recovers some capability, our method achieves a further substantial leap, validating that Self-Supervised Learning is essential for generalizing to the diverse biomes of Brazil where labeled data is scarce.

Boundary Precision and Segmentation Quality: Beyond detection, the topological quality of boundaries is superior. In technified regions like **Jaboticabal** and **Pitangueiras**, our method consistently yields the lowest Global Under-Segmentation (GUS) scores (e.g., 0.030 in Pitangueiras vs. 0.072 for VARDA). This low GUS indicates that the model successfully delineates adjacent fields rather than merging them, a critical requirement for accurate area estimation and land tenure mapping.

5 Conclusion and Strategic Outlook

The modernization of Brazil's National Statistical System requires more than just digitalization; it demands a structural integration of advanced data science with the capillarity of field enumeration. This work not only validated a robust Deep Learning model for Field Boundary Delineation (FBD) but also established a comprehensive **Integrated Cycle of Earth Observation and Field Enumeration**.

Our results unequivocally demonstrate that the proposed framework, powered by High-Resolution Networks (HRNet) and Self-Supervised Learning (SSL), overcomes the limitations of global baselines, particularly in fragmented landscapes like the Caatinga. However, the core contribution lies in the methodological orchestration: by using these automated plots as the "spatial backbone" for the 12th Agricultural Census, we ensure **spatial consistency** between administrative records and physical land use.

This approach transforms the Census from a static survey into a dynamic engine for ground-truth generation, where the Census enables resource optimization. The resulting feedback loop validates the AI boundaries, and the AI guides the fieldwork optimization and continuous model refinement. By adopting this "Divide and Conquer" strategy, we pave the way for the next phase of the roadmap: orchestrating the **Crop Mapping and Yield Estimation** pipeline. Consequently, future work will address three strategic pillars:

1. A **complete and detailed analysis** covering all Brazilian biomes and the consolidation of the large-scale training set currently under development;

2. An **extensive benchmark** comparing the proposed framework against other state-of-the-art segmentation methods using the proprietary dataset;

3. The development of downstream applications utilizing these plots for **crop mapping, yield modeling, and precise agricultural land use area estimation**.

References

[1] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *NeurIPS*, 33:9912–9924, 2020.

[2] CEPEA. Pib do agronegócio brasileiro, 2025.

[3] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607. PMLR, 2020.

[4] X. Chen and K. He. Exploring simple siamese representation learning. In *CVPR*, pages 15750–15758, 2021.

[5] E. Contini and A. Aragão. O agro brasileiro alimenta 800 milhões de pessoas. *Brasília: Embrapa*, 2021.

[6] P. Gao, T. Ma, H. Li, Z. Lin, J. Dai, and Y. Qiao. Convmae: Masked convolution meets masked autoencoders. *arXiv preprint arXiv:2205.03892*, 2022.

[7] S. Gidaris, P. Singh, and N. Komodakis. Unsupervised representation learning by predicting image rotations. In *ICLR*, 2018.

[8] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick. Masked autoencoders are scalable vision learners. In *CVPR*, pages 16000–16009, 2022.

[9] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *ICCV*, pages 2961–2969, 2017.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[11] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira. Overview of the radiometric and biophysical performance of the modis ecosystem products. *Remote sensing of environment*, 83(1-2):195–213, 2002.

[12] IBGE. Sistema de contas nacionais do brasil, 2025.

[13] Instituto Brasileiro de Geografia e Estatística - IBGE. *Bioma predominante por município para fins estatísticos*. IBGE, Rio de Janeiro, 2024. Accessed on: 25 Jan. 2026.

[14] M. Li, J. Long, A. Stein, and X. Wang. Using a semantic edge-aware multi-task neural network to delineate agricultural parcels from remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 200:24–40, 2023.

[15] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, pages 3431–3440, 2015.

[16] M. Noroozi and P. Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In *ECCV*, pages 69–84. Springer, 2016.

[17] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In *CVPR*, pages 2536–2544, 2016.

[18] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015.

[19] J. W. Rouse, R. Haas, J. Schell, and D. Deering. Monitoring vegetation systems in the great plains with erts. In *Third Earth Resources Technology Satellite-1 Symposium*, volume 1, pages 309–317. NASA, 1974.

[20] F. T. Ulaby, R. K. Moore, and A. K. Fung. *Microwave remote sensing: Active and passive. Vol. 1: Microwave remote sensing fundamentals and radiometry.* Artech House, Norwood, MA, 1981.

[21] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao. Deep high-resolution representation learning for visual recognition. *TPAMI*, 2019.

[22] Z. Xie, Z. Zhang, Y. Cao, Y. Lin, J. Bao, Z. Yao, Q. Dai, and H. Hu. Simmim: A simple framework for masked image modeling. In *CVPR*, pages 9653–9663, 2022.