Revista de Informática Teórica e Aplicada - RITA - ISSN 2175-2745
Vol. 33, Num. 2 (2026) 202-209

RESEARCH ARTICLE

# Semantic Segmentation of Coffee Regions From Satellite Images Using Deep Learning

## Segmentação Semântica de Áreas de Café a Partir de Imagens de Satélite Utilizando Aprendizado Profundo

João Marcelo Ribeiro[1], Vitor Augusto Negrão de Assis[1], Lais Bueno Palos[1], Jorge Otávio de Souza Nogueira [1], Jhonathas Emiliano Candido[1], André Ricardo Backes[2]*

**Abstract:** Coffee is one of the main agricultural products exported by Brazil, with Minas Gerais standing out as the largest producer and responsible for a significant portion of the country's production. Due to its importance, this work investigates various Convolutional Neural Networks (CNNs) for mapping coffee plantation areas in the municipality of Muzambinho, MG. Using satellite images, we evaluated four semantic segmentation network architectures: FCN, MA-NET, SegNet, and DeepLabV3. Experiments using a dataset with 100 satellite images demonstrated that the FCN and MA-NET networks achieve better performance in identifying crops than SegNet and DeepLabV3. The latter, however, presented generalization challenges in the test set, indicating sensitivity to variations in lighting and shading. The study highlights the potential of artificial intelligence combined with remote sensing to improve agricultural management, and the need for more robust and diversified datasets to optimize the generalization and robustness of the models in real conditions.
**Keywords:** coffee cultivation — remote sensing — convolutional neural networks (CNNs) — semantic segmentation

**Resumo:** O café é um dos principais produtos agrícolas exportados pelo Brasil, destacando-se Minas Gerais como o maior produtor e responsável por uma parcela significativa da produção nacional. Devido à sua importância, este trabalho investiga diversas Redes Neurais Convolucionais (CNNs) para o mapeamento de áreas de cafeicultura no município de Muzambinho, MG. Utilizando imagens de satélite, foram avaliadas quatro arquiteturas de redes de segmentação semântica: FCN, MA-NET, SegNet e DeepLabV3. Experimentos com um conjunto de dados contendo 100 imagens de satélite demonstraram que as redes FCN e MA-NET apresentam melhor desempenho na identificação de lavouras em comparação com a SegNet e a DeepLabV3. Esta última, entretanto, apresentou desafios de generalização no conjunto de teste, indicando sensibilidade a variações de iluminação e sombreamento. O estudo destaca o potencial da inteligência artificial combinada com o sensoriamento remoto para aprimorar o gerenciamento agrícola, bem como a necessidade de conjuntos de dados mais robustos e diversificados para otimizar a generalização e a robustez dos modelos em condições reais.
**Palavras-Chave:** cafeicultura — sensoriamento remoto — redes neurais convolucionais (CNNs) — segmantação semântica

[1] *Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas (IFSULDEMINAS), Brazil*
[2] *Departamento de Computação, Universidade Federal de São Carlos (UFSCar), Brazil*
***Corresponding author**: arbackes@yahoo.com.br*
**DOI:** http://dx.doi.org/10.22456/2175-2745.150916 • **Received:** 14/10/2025 • **Accepted:** 27/10/2025


## 1. Introduction

Coffee is an essential global commodity, with Brazil standing out as the world's largest producer [1, 2]. The state of Minas Gerais leads production, with its southern region contributing significantly and having historical relevance to coffee farming [3, 4, 5]. The accurate identification of coffee growing areas is essential for agricultural planning and sustainable management, especially in municipalities such as Muzambinho, MG, which dedicates a large part of its territory to coffee [6].

However, traditional mapping approaches, which rely predominantly on human intervention, often have low scalability, limitations in terms of accuracy, and subjectivity in visual interpretation. These methodologies are largely based on manual interpretation of images and require specialized knowledge to identify and classify coffee crops, which makes the process time-consuming and prone to errors. Improvements in these classifications have occurred with the use of supervised and unsupervised classification algorithms, such as k-means, for example, which was widely used to assist in this task. Recently, advances in remote sensing and artificial intelligence have provided new opportunities for mapping and managing coffee plantations, using satellite and UAV data with advanced techniques such as Deep Learning and Random Forest [7, 8, 9].

Traditional machine learning has been widely used in remote sensing applications, with techniques such as Random Forest and Support Vector Machines achieving good results in classification tasks [10, 11]. Deep learning, on the other hand, represents a more comprehensive approach, based on deep neural networks capable of learning hierarchical representations directly from data [12]. Unlike traditional methods, which rely heavily on manual descriptors, deep learning architectures such as Convolutional Neural Networks (CNNs) automatically extract relevant patterns, showing promise for agricultural segmentation problems.

Traditional mapping approaches are limited in accuracy and scalability. In contrast, Artificial Intelligence (AI) has revolutionized many areas, including agriculture [13]. Convolutional Neural Networks (CNNs), in particular, are effective in image analysis and segmentation. The integration of remote sensing with AI allows for accurate and automated mapping, overcoming challenges such as spectral confusion [14, 15].

In this work, we investigate the application of different CNN architectures – FCN, MA-NET, SegNet, and DeepLabV3 – for mapping coffee plantations in Muzambinho, MG, using Google Earth images processed in QGIS software. Our main goal is to train and compare the performance of these CNNs in identifying and segmenting coffee areas, aiming to provide subsidies for more efficient coffee farming in the region.

The remainder of this paper is organized as follows: we present the current status of the research area in Section 2. Section 3 describes the material and methods used in our work. In Section 4 we describe the experiments we conducted. Section 5 presents and discusses the results, while Section 6 concludes the paper.

## 2. Related Work

Coffee plantation segmentation using remote sensing and deep learning has established itself as a relevant research area. Initial approaches relied on traditional machine learning methods, such as Support Vector Machines (SVM) and Random Forest, typically combined with manual descriptors extracted from multispectral or hyperspectral images [16]. Although these methods achieved satisfactory results in some scenarios, they lacked generalization due to variability in planting patterns, lighting conditions, and seasonal effects.

Recently, Convolutional Neural Networks (CNNs) have been widely applied to classification and segmentation tasks in the agricultural context, including the mapping of coffee plantations [17]. CNNs demonstrate superior performance by automatically extracting relevant hierarchical patterns directly from data [12]. Architectures such as U-Net and its variants have achieved impressive results in pixel-level classification, benefiting from the ability to capture both local texture information and the global context of the scene [18].

CNNs have proven versatile in agricultural applications, such as the use of *Fully Convolutional Networks* (FCN) for mapping rice crops [19] and U-Net for segmenting corn and soybeans [20], highlighting their effectiveness in distinguishing crops with complex spectral characteristics. In the context of coffee, CNNs have been successfully applied to disease detection in coffee leaves [21] and fruit ripeness classification [22]. Despite these advances, CNN-based models often struggle in landscapes with high topographic variability and heterogeneity. The limited ability to capture long-range dependencies, such as the relationship between a crop and its distant surroundings, can lead to segmentation errors in complex terrain.

This limitation in capturing long-range contextual dependencies contributed to the emergence of Transformers. By adapting the self-attention mechanism [23], the Transformer overcomes the challenge posed by CNNs, allowing the model to assign weights to each pixel based on its relationship to all others, thus capturing the global context of the image. This architecture has been successfully adapted for remote sensing [24] and is particularly promising in agriculture for dealing with the spatial variability of crops [25]. Silva et al. [26] applied a Transformer-based segmentation approach to coffee crops in Brazil, demonstrating that the fusion of long-range contextual information with multispectral features resulted in more robust and accurate segmentation, similar to that observed in segmentation studies of crops such as cotton [27].

This work contributes to this field by conducting a comparative evaluation of four CNN-based architectures, FCN, MA-Net, Segnet, and DeepLabV3, applied to the segmentation of coffee fields in Brazilian regions. Unlike studies that focus on a single architecture or limited experiments, our analysis offers a systematic comparison that considers both quantitative and qualitative results. Additionally, we discuss the limitations imposed by the small dataset size and propose future research directions that can strengthen the robustness of CNN-based models in this domain.

## 3. Material and Methods

### 3.1 Semantic Segmentation

Semantic segmentation is an important task in computer vision. It aims to assign a class label to every pixel in an image. Unlike image classification, which predicts a single label for the entire image, or object detection, which localizes objects with bounding boxes, semantic segmentation provides a much finer level of understanding by distinguishing different regions of interest within the image. This process enables detailed analysis of the structure of the image and it is essential in applications such as medical imaging [28], autonomous driving [29], and satellite imagery interpretation [30].

Over the years, many methods have been proposed to accomplish this task. Modern approaches rely on deep learning, particularly convolutional neural networks (CNNs) and their extensions, which allow the extraction of rich hierarchical features and the generation of dense prediction maps with high accuracy.

### 3.2 Convolutional Neural Networks (CNNs)

Artificial Intelligence (AI) aims to build intelligent computer systems capable of simulating aspects of human reasoning [31, 32], enabling the efficient processing of large volumes of data to identify complex patterns. Among the techniques used, Artificial Neural Networks (ANNs) stand out as models inspired by the human brain, learning and storing knowledge through their connections [33, 34].

Among many ANN approaches, Convolutional Neural Networks (CNNs) are widely used in image processing for tasks such as classification and segmentation, recognizing visual patterns, and automatically extracting features [35]. Their basic structure includes convolutional layers that apply filters to extract local features, pooling layers that reduce the dimension of the data, and dense layers that generate the final output of the model. The combination of these layers allows CNNs to learn hierarchical representations of the data [36].

Segmentation is usually the first step in computer vision applications. It divides the content of an image into various regions with similar characteristics but distinct from each other. For this task, several CNN architectures have been developed:

- **Fully Convolutional Networks (FCN)**: Unlike traditional CNNs, FCNs replace fully connected layers with convolutional layers, enabling the network to produce spatial output maps. This structure takes input images of arbitrary size and outputs dense, pixel-wise predictions for image segmentation tasks [37].

- **MA-Net**: enhances segmentation performance by integrating multi-scale feature extraction with attention mechanisms. It captures fine-grained spatial details and global contextual information simultaneously. The key components include position and channel attention modules, which help the model focus on relevant spatial locations and feature channels, making it effective for segmenting complex structures [38, 39].

- **SegNet**: is an encoder-decoder architecture designed for semantic segmentation. The encoder is based on a VGG-like architecture and captures high-level features, while the decoder upsamples these features to the original resolution using pooling indices from the encoder. This approach preserves spatial details efficiently, making it suitable for real-time applications [40].

- **DeepLabV3**: DeepLabV3 is a powerful semantic segmentation model that uses atrous (dilated) convolutions to capture multi-scale context without reducing spatial resolution. It incorporates Atrous Spatial Pyramid Pooling (ASPP) to extract features at multiple receptive fields. DeepLabV3 provides high accuracy in challenging segmentation tasks, particularly in cases with objects at various scales [41].

### 3.3 Dataset Collection and Preparation

The study area was delimited in QGIS as the municipality of Muzambinho, MG (21.37°S, 46.52°W), based on IBGE data [42]. We built a dataset containing 100 satellite images obtained from *Google Earth* (via plugin in QGIS), all in PNG format, RGB images with $512 \times 512$ pixels size and scale 1:800. For each image, we manually created a binary segmentation mask representing the coffee areas. For the experiments, we divided these images into three sets: 60 images for training, 30 for validation, and 10 for testing or prediction.

The images were acquired from June to August 2024, ensuring maximum visual contrast between the mature crops and the surrounding vegetation. Images taken during the same period of the year and under similar phenological conditions were considered, encompassing variations in lighting and soil cover. The small number of samples represents a limitation, as deep learning models typically require larger datasets. This factor increases the risk of overfitting, even though the results indicated consistency between the training, validation, and test sets. Therefore, we recognize that generalization may be limited in more diverse scenarios.

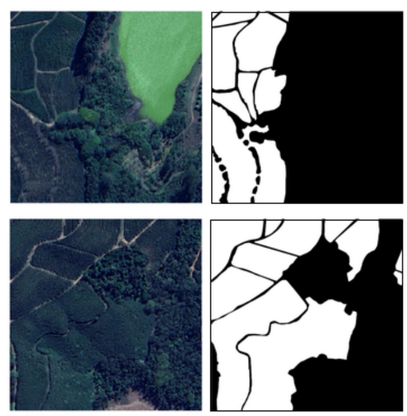

**Figure 1.** Example of coffee are images and their respective masks.

## 4. Experiments

In our experiments, we aimed to train and evaluate different architectures of Convolutional Neural Networks (FCN, MA-NET, SegNet, and DeepLabV3) in the task of segmenting coffee areas. The data set (100 images) was divided into 60 images for training, 30 for validation, and 10 for testing. Due to computational cost, the batch size was set to 4, meaning that the model processes 4 pairs of images per iteration. We performed training for 500 epochs, Adam optimizer, and a learning rate of 0.001.

The training was performed entirely with the Muzambinho image dataset, created specifically for this purpose, and the networks were trained from scratch. It is crucial to note that we did not apply data augmentation to the images, which were used in their original form extracted from QGIS.

To evaluate the performance of each CNN model we used the Intersection over Union (IoU) metric, one of the most widely used metrics for measuring segmentation accuracy [43, 44]. It measures the overlap between the predicted ($A$) and the expected result ($B$), by the ratio between the intersection area and the union area:

$$\text{IoU} = \frac{A \cap B}{A \cup B} \tag{1}$$

A higher IoU value indicates greater accuracy [45]. In addition to the IoU metric, we also evaluated the models using Precision, Recall, and the Dice coefficient. These complementary metrics allow for a more transparent and balanced analysis of performance, considering both pixel-level classification successes and errors.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

where $TP$ is the number of true positives, $FP$ is the number of false positives and $FN$ is the number of false negatives. The Dice coefficient, $D$, is defined as

$$D = 2\frac{|A \cap B|}{|A|+|B|}, \tag{4}$$

where $D$, $0 \leq D \leq 1$, is the level of similarity between images.

We conducted the experiments on a PC with a Windows SO. The main tools used included Visual Studio Code (VS Code 1.96.4) for development in Python (3.12.5), with TensorFlow and Keras (3.7.0). To obtain and manipulate the images, we used QGIS Desktop (3.34.11) and *Google Earth*. We used GIMP (2.10.38) to manually create the segmentation masks.

# 5. Results and Discussion

**Table 1.** Average results obtained for each architecture.

| Model | IoU | Precision | Recall | Dice |
|---|---|---|---|---|
| FCN | 0.728 | 0.807 | 0.885 | 0.836 |
| MA-Net | 0.813 | 0.856 | 0.945 | 0.895 |
| SegNet | 0.696 | 0.763 | 0.890 | 0.810 |
| DeepLabV3 | 0.188 | 0.287 | 0.502 | 0.310 |

Comparison of segmentation performance across the four CNN models reveals clear differences in their ability to balance precision, recall, and overall segmentation quality, as shown in Table 1. Figure 3 shows the IoU curve obtained during the training of the FCN. MA-Net achieves the highest IoU score (0.813) and Dice coefficient (0.895), demonstrating its effectiveness in segmenting coffee regions. These results indicate that MA-Net is highly accurate and consistent for coffee plantation segmentation, capturing most relevant regions while maintaining low false-positive rates. It also shows the highest recall (0.945) and precision (0.856), which indicates excellent sensitivity and object coverage. Its strong performance can be attributed to its attention-guided mechanism, which effectively enhances spatial detail and feature representation. FCN achieves the second highest IoU score (0.728) and Dice coefficient (0.836), demonstrating its effectiveness in segmenting coffee regions, even though it uses less parameters than compared approaches (Table 2). Its balanced performance across precision (0.807) and recall (0.885) suggests that it maintains a good equilibrium between avoiding false positives and capturing most relevant regions, while using $\approx 6\times$ less parameters than MA-Net. SegNet shows moderate performance, similar to FCN, obtaining an IoU of 0.696 and a Dice score of 0.810. Its recall (0.890) shows that it successfully captures relevant areas, but at the cost of lower precision (0.763), implying a higher rate of false positives compared to FCN and MA-Net. Although, SegNet presents an architecture much simpler and makes no use of attention. Finally, DeepLabV3 stands out for presenting the lowest IoU (0.188), demonstrating a critical limitation regarding segmentation and poor generalization capabilities, even though it uses more parameters. one possible explanation may lies to inadequate adaptation of its atrous convolution strategy or poor parameter optimization for this dataset.

**Table 2.** Number of parameters of each CNN.

| Model | Number of parameters |
|---|---|
| FCN | 297.089 |
| MA-Net | 1.968.513 |
| SegNet | 297.089 |
| DeepLabV3 | 39.633.729 |

The qualitative evaluation of the segmentation obtained by each CNN shows these differences in results. Figure 2 shows a qualitative analysis of FCN segmentation results. Visually, the segmentation shows high coherence and accuracy with the ground truth, indicating that the network effectively groups coffee regions. Although there may be small imperfections at the edges or minimal uncaptured areas, the FCN exhibits a remarkable ability to learn and replicate the spatial characteristics of objects with fidelity. The main challenges involved false positives in areas with similar visual appearance but without active crops, and false negatives in regions with intense shadows.

The results obtained by the MA-Net network (Figure 2) show improvement over the FCN, with fewer false positive regions, which are now concentrated to the boundaries of the coffee plantation area and small areas of similar appearance.

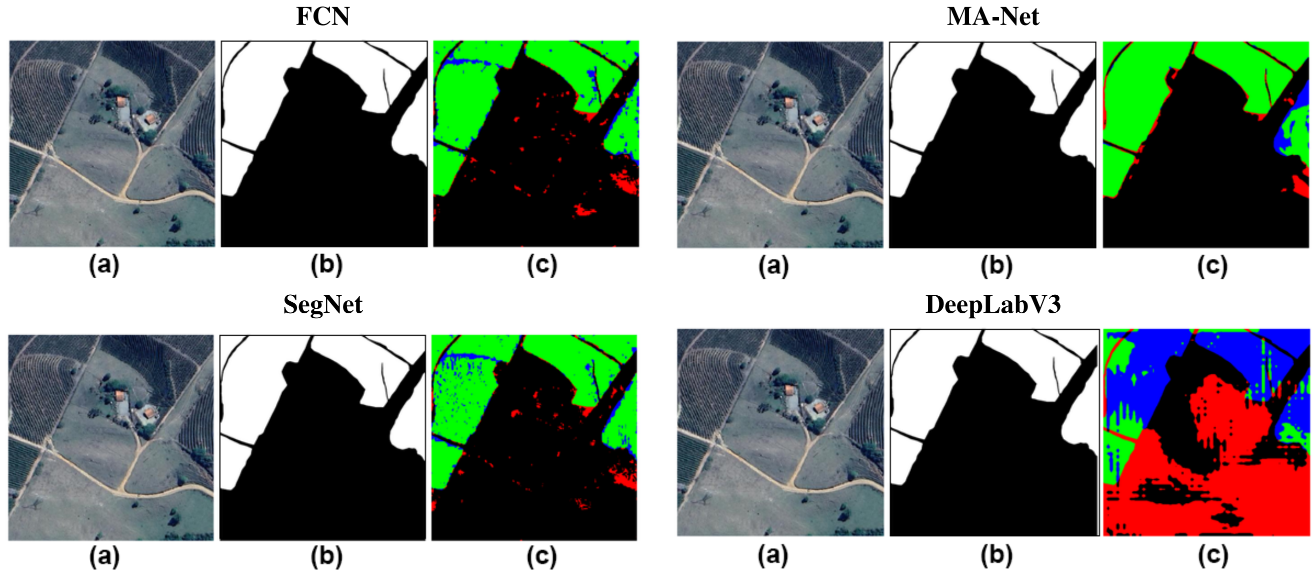


**Figure 2.** Segmentation example obtained by each model: (a) Original Image; (b) Expected result (mask); (C) Segmentation result. Colors represent true positive (green), false negative (blue) and false positive (red).

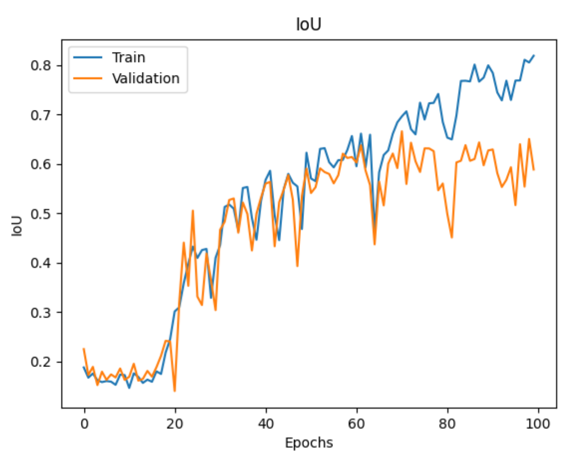


**Figure 3.** IoU curve obtained during the training of the FCN.

Such results indicate that attention mechanisms improve the discrimination between different areas, thus improving segmentation and the IoU achieved by the network. However, shadow and lighting variations seem to be the main limitations identified for this network.

SegNet results (Figure 2) show that this network was able to delineate the areas of interest coherently and less noisily. The segmented shapes more closely resemble the expected results, especially in the larger regions. However, this CNN presents many false negatives inside a coffee area, resulting in fragmented regions. Additionally, some edges are not perfectly smooth or precise and areas without active cultivation are marked as coffee (false positive), an appearance similar to those of FCN.

Finally, Figure 2 shows a segmentation example obtained by DeepLabV3. Visually, it is evident that the segmentation produced by this network is highly imprecise. Large areas without active cultivation are marked as coffee, while coffee areas are missed by the network, indicating severe difficulty in grouping pixels correctly to form cohesive and meaningful regions. Correctly detected coffee areas are smaller compared to the results from other networks and highly fragmented.

Figure 4 shows a comparison using different samples of the dataset. This visual comparison among CNN models shows that FCN and MA-NET are the most effective to segment coffee crops, followed by SegNet. The unsatisfactory performance of DeepLabV3 highlights the importance of diversity and quality of the training dataset to avoid overfitting and ensure robustness in generalization. Common challenges identified include distinguishing between background and coffee regions with similar appearances and the significant influence of shadows and lighting on segmentation accuracy.

## 6. Conclusion

In this work, we evaluated different CNNs to identify coffee areas from satellite images. The combination of CNNs and remote sensing techniques has significant potential for identification and quantification of cultivation areas, contributing to sustainability and more efficient agricultural management. FCN and MA-NET models presented the best results, followed by SegNet, validating their effectiveness in the proposed task. However, challenges such as shadows, lighting variations, and visual similarity with uncultivated areas negatively impacted the performance of the models. DeepLabV3, in particular, presented difficulties in generalization and was sensitive to these variations, indicating that it may not be the most appropriate architecture for this scenario.

This study has some limitations that should be considered. The first concerns the small size of the dataset, which may

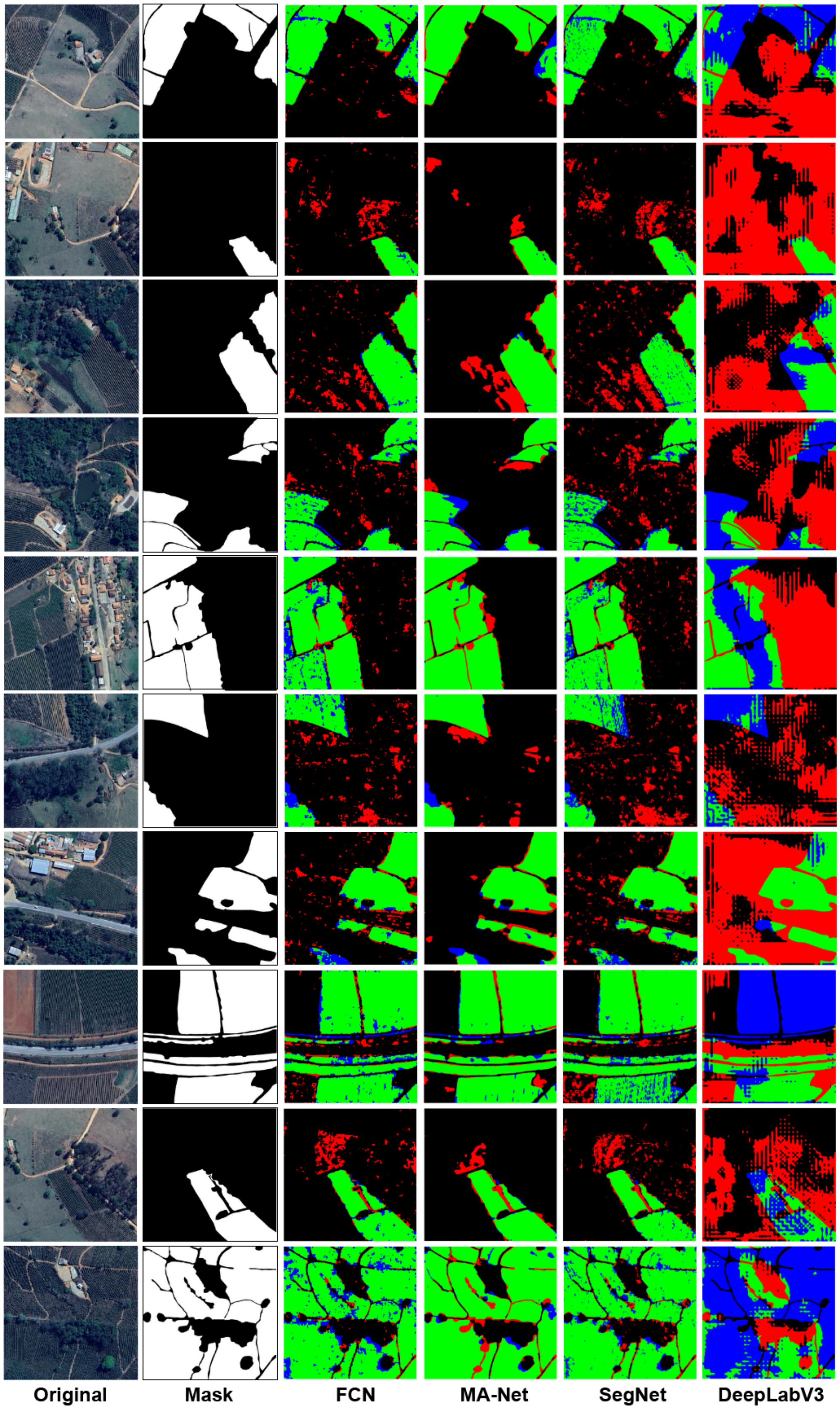


**Figure 4.** Comparison of the segmentation results obtained for various samples. Colors represent true positive (green), false negative (blue) and false positive (red).

restricts the generalization capacity and lead to overfitting. Another limitation is the lack of data augmentation techniques, which could increase sample variability and improve model robustness. Future analyses could explore additional indicators related to computational cost and training time. Future work aims to expand the dataset with images acquired at different times and capture conditions, apply data augmentation strategies, and investigate the use of more recent CNN models. Comparisons with hybrid or Transformer-based approaches are also planned, as well as assessing the method's scalability in agricultural monitoring scenarios.

## Acknowledgements

André R. Backes gratefully acknowledges the financial support of CNPq (Grant #302790/2024-1). This study was financed in part by CAPES - Finance Code 001.

## Author Contributions

João Marcelo Ribeiro: conception, design, experiments and writing; Vitor Augusto Negrão de Assis: experiments, statistical analysis and figures; Lais Bueno Palos: experiments, statistical analysis and figures; Jorge Otávio de Souza Nogueira: experiments, statistical analysis and figures; Jhonathas Emiliano Candido: experiments, statistical analysis and figures; André Ricardo Backes: supervision, writing and funding.

## References

[1] STATISTA. *Coffee Production Worldwide in 2020, by Leading Country*. 2024. Acesso em: 15 abr. 2024. Disponível em: ⟨https://www.statista.com/statistics/277137/world-coffee-production-by-leading-countries/⟩.

[2] Organização Internacional do Café (OIC). *Relatório sobre o mercado do café (janeiro/2024)*. 2024.

[3] ALVES, D. P.; LINDNER, E. O mercado do café no brasil e a importância de minas gerais. *Revista Brasileira de Economia Agrícola*, v. 15, n. 3, p. 456–478, 2020.

[4] COOXUPé. *Produção de Café Aumenta 13 Vezes em Quase 50 Anos no Sul de Minas*. 2016. Acesso em: 15 abr. 2024. Disponível em: ⟨https://www.cooxupe.com.br/noticias/producao-de-cafe-aumenta-13-vezes-em-quase-50-anos-no-sul-de-minas/⟩.

[5] PAULA, R. Z. A. d. Região e regionalização: Um estudo da formação regional da zona da mata de minas gerais. *Revista de História Econômica & Economia Regional Aplicada*, v. 1, n. 1, p. 66–80, 2006.

[6] GONçALVES, H. T. et al. Caracterização da cultura cafeeira em relação à altimetria e declividade no município de muzambinho-mg. In: IFSULDEMINAS. *Jornada Científica e Tecnológica e Simpósio da Pós-Graduação do IFSULDEMINAS*. Muzambinho, MG, Brazil, 2020.

[7] MASKELL, G. M. et al. Integration of sentinel optical and radar data for mapping smallholder coffee production systems in vietnam. *Remote Sensing of Environment*, Elsevier, v. 266, p. 112709, 2021.

[8] ARRIOLA-VALVERDE, S. et al. A comparative study of deep learning frameworks applied to coffee plant detection from close-range uas-rgb imagery in costa rica. *Remote Sensing*, MDPI, v. 16, n. 24, p. 4617, 2024.

[9] WANG, X. et al. A coffee plant counting method based on dual-channel nms and yolov9 leveraging uav multispectral imaging. *Remote Sensing*, MDPI, v. 16, n. 20, p. 3810, 2024.

[10] ZHANG, C.; KOVACS, J. M. The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture*, Springer, v. 13, p. 693–712, 2012.

[11] KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, Elsevier, v. 147, p. 70–90, 2018.

[12] MA, L. et al. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 152, p. 166–177, 2019.

[13] McKinsey & Company. *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value*. 2024. Acesso em: 26 out. 2024. Disponível em: ⟨https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/⟩.

[14] CHAGAS, C. S. et al. Atributos topográficos e dados do landsat 7 no mapeamento digital de solos com uso de redes neurais. *Pesquisa Agropecuária Brasileira*, Brasília, v. 45, n. 5, p. 497–507, 2010.

[15] VIEIRA, C. A. O.; LACERDA, J. D.; BOTELHO, M. L. *Sensoriamento Remoto e Sistemas de Informação Geográfica: Fundamentos e Aplicações*. 2. ed. São Paulo, Brazil: Érica, 2009.

[16] KAWAKUBO, F. S.; MACHADO, N. G. Mapping coffee crops in southeastern brazil using spectral mixture analysis and data mining classification. *International Journal of Remote Sensing*, Taylor & Francis, v. 37, n. 18, p. 4390–4405, 2016.

[17] PEREIRA, F. V.; SILVA, J. P.; OLIVEIRA, M. R. Estimating coffee crop parameters through multispectral imaging and machine learning algorithms. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-3-2024, p. 317–324, 2024.

[18] ALAMANDA, F.; RACHMAT, A.; WIBOWO, B. Visual segmentation and classification of coffee beans after roasting. *Journal of Applied Informatics and Computing*, v. 8, n. 2, p. 111–120, 2024.

[19] LU, S.; WANG, H.; LIU, X. Rice plantation mapping using fully convolutional networks on sentinel-1 sar and sentinel-2 optical data. *Remote Sensing*, MDPI, v. 12, n. 18, p. 2975, 2020.

[20] SUN, H.; ZHANG, Y.; LIN, W. Deep learning for crop type classification and segmentation in agriculture using remote sensing imagery. *Sensors*, MDPI, v. 20, n. 19, p. 5559, 2020.

[21] SANTOS, A.; SOUZA, B. A. de; GOMES, H. L. Coffee leaf disease detection using a deep convolutional neural network. In: IEEE. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. [S.l.], 2020. p. 188–192.

[22] BAZAME, R.; SOUZA, B. A. de; SANTOS, A. Deep learning for coffee cherry maturation classification using computer vision. *Computers and Electronics in Agriculture*, Elsevier, v. 188, p. 106338, 2021.

[23] DOSOVITSKIY, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. In: *International Conference on Learning Representations (ICLR)*. [S.l.: s.n.], 2021.

[24] HE, X. et al. Multi-scale vision transformer for semantic segmentation in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, IEEE, v. 60, p. 1–13, 2022.

[25] AFONSO, M. et al. Vision transformers for crop classification in satellite imagery: a comprehensive study. *Remote Sensing*, MDPI, v. 15, n. 6, p. 1672, 2023.

[26] SILVA, L.; SANTOS, J. a.; PEREIRA, M. Transformer-based panoptic segmentation for coffee crop mapping in brazilian regions. *IEEE Geoscience and Remote Sensing Letters*, IEEE, v. 20, p. 1–5, 2023.

[27] LIU, X.; WANG, H.; LU, S. A transformer-based model for cotton plantation segmentation in high-resolution remote sensing imagery. *Remote Sensing*, MDPI, v. 13, n. 17, p. 3430, 2021.

[28] GARIA, L.; MUTHUSAMY, H. Dual-tree complex wavelet pooling and attention-based modified u-net architecture for automated breast thermogram segmentation and classification. *Journal of Imaging Informatics in Medicine*, Springer, p. 1–15, 2024.

[29] YAO, S. et al. Radar-camera fusion for object detection and semantic segmentation in autonomous driving: A comprehensive review. *IEEE Transactions on Intelligent Vehicles*, IEEE, v. 9, n. 1, p. 2094–2128, 2023.

[30] ARAKAKI, L. G. et al. Evaluation of u-net backbones for cloud segmentation in satellite images. In: *VISIGRAPP (4: VISAPP)*. [S.l.: s.n.], 2023. p. 452–458.

[31] MCCARTHY, J. *What is Artificial Intelligence?* 2007. Computer Science Department, Stanford University.

[32] GOLDSCHMIDT, R. R. *Uma Introdução à Inteligência Computacional: Fundamentos, Ferramentas e Aplicações*. 1. ed. Rio de Janeiro, Brazil: IST-Rio, 2010.

[33] FLECK, L. et al. Redes neurais artificiais: Princípios básicos. *Revista Eletrônica Científica Inovação e Tecnologia*, Medianeira, v. 1, n. 13, p. 47–57, 2016.

[34] HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2. ed. Upper Saddle River, NJ, USA: Prentice Hall, 2001.

[35] GU, J. et al. *Recent advances in convolutional neural networks*. 2015. Disponível em: ⟨http://arxiv.org/abs/1512.07108⟩. Acesso em: 03 jun. 2025.

[36] LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998.

[37] LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: [s.n.], 2015. p. 3431–3440.

[38] FAN, T. et al. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, v. 8, p. 179656–179665, 2020.

[39] CHANG, J. et al. Multi-scale attention network for building extraction from high-resolution remote sensing images. *Sensors*, v. 24, n. 3, p. 1010, 2024.

[40] BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. Segnet: A deep convolutional encoder–decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 39, n. 12, p. 2481–2495, 2017.

[41] CHEN, L.-C. et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 40, n. 4, p. 834–848, 2017.

[42] Instituto Brasileiro de Geografia e Estatística (IBGE). *Portal de Mapas*. 2014. Acesso em: 08 out. 2024. Disponível em: ⟨https://portaldemapas.ibge.gov.br/portal.php#mapa207423⟩.

[43] ZHENG, Z. et al. Distance-iou loss: Faster and better learning for bounding box regression. In: *AAAI Conference on Artificial Intelligence (AAAI-20)*. [S.l.: s.n.], 2020. v. 34, p. 12993–13000.

[44] REZATOFIGHI, H. et al. Generalized intersection over union: A metric and a loss for bounding box regression. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: [s.n.], 2019. p. 658–666.

[45] JORDAN, J. *Evaluating image segmentation models*. 2018. Acesso em: 03 jun. 2025. Disponível em: ⟨https://www.jeremyjordan.me/evaluating-image-segmentation-models⟩.