Exploring the Segment Anything Model for Mapping Urban Tree Cover in Orbital Imagery

Gleison Marrafon [1], Vagner Souza Machado [2], Lucas Prado Osco [3], José Marcato Junior [4], Wesley Nunes Gonçalves [4], Ana Paula Marques Ramos [5]

[1] Undergraduate student in Surveying and Cartographic Engineering at the Faculty of Science and Technology, São Paulo State University (UNESP), Campus of Presidente Prudente, Rua Roberto Simonsen, 305, Presidente Prudente, São Paulo, Brazil, ZIP code 19060-900 – g.marrafon@unesp.br
[2] Ph.D. student in the Graduate Program in Environmental and Regional Development, Western São Paulo State University (UNOESTE), Campus II, Raposo Tavares Highway, km 572, Limoeiro District, Presidente Prudente, São Paulo, Brazil, ZIP code 19067-175 – vagcarto@gmail.com
[3] Western São Paulo State University (UNOESTE), Campus II, Raposo Tavares Highway, km 572, Limoeiro District, Presidente Prudente, São Paulo, Brazil, ZIP code 19067-175 – lucasosco@unoeste.br
[4] Federal University of Mato Grosso do Sul, Cidade Universitária, Av. Costa e Silva, Campo Grande, MS, Brazil, ZIP code 79070-900 – (jose.marcato, wesley.goncalves)@ufms.br
[5]São Paulo State University (UNESP), Campus of Presidente Prudente, Rua Roberto Simonsen, 305, Presidente Prudente, São Paulo, Brazil, ZIP code 19060-900 – marques.ramos@unesp.br

Keywords: Multispectral images, semantic segmentation, Transformer Vision Models, urban tree canopy, urban planning.

Abstract

Urban tree vegetation plays a key role in sustainable urban planning and ecosystem service provision. This study evaluates the performance of the Segment Anything Model (SAM), developed by Meta AI, in the segmentation of urban tree vegetation from orbital PlanetScope imagery. These images were selected due to their high spatial and temporal resolution, which makes them particularly suitable for urban applications. SAM was applied in zero-shot mode, guided by geometric prompts over representative tree-covered areas. The analysis was conducted across three Brazilian cities—Corumbá (MS), Rio Verde (GO), and Valparaíso de Goiás (GO)—using different spectral band compositions. SAM's performance was evaluated through a combined quantitative and qualitative approach, using reference masks derived from manually annotated tree canopy polygons. Although SAM had not been previously trained on satellite imagery, it achieved an F1-scores close to 70% and recall values around 75%, independently of the spectral band composition provided as input. These results demonstrate the model's generalization ability—even under spectrally constrained scenarios involving only three bands. Qualitative analysis confirmed spatial consistency in tree crown delineation, particularly in homogeneous areas, while over-segmentation was observed in spectrally heterogeneous environments. While the results are promising for exploratory and semi-automated vegetation mapping, they also underscore need for fine-tuning SAM on satellite data to enhance spatial precision and thematic discrimination. Overall, SAM's modular and prompt-based architecture offers a robust foundation for scalable, supervised remote sensing workflows focused on urban vegetation monitoring.

1. Introduction

Urban tree vegetation plays a strategic role in maintaining the ecological balance of cities by providing ecosystem services relevant to human well-being (Nowak et al., 2014; Rahman et al. 2024). Mapping these features is essential for sustainable urban planning, enabling the quantification and monitoring of such structures over time. In this context, remote sensing data offer efficient solutions for large-scale vegetation mapping, with the potential to differentiate structural and spectral patterns associated with vegetation.

The advancement of artificial intelligence approaches, particularly deep neural networks, has expanded the possibilities for automated information extraction from remote sensing data. In image segmentation tasks, deep learning has been employed to identify and delineate objects at the pixel level, serving as a promising alternative for vegetation mapping using remote sensing data (Osco et al., 2021). However, deep learning methods often require large volumes of labelled data for supervised training, which limits their operational scalability (Trask, 2019; Osco et al., 2021).

As an alternative, the Segment Anything Model (SAM), released in 2023 by Meta AI, proposes a segmentation approach capable of operating in zero-shot mode—that is, without the need for task-specific training for a given class (Kirillov et al., 2023). SAM uses geometric and textual prompts to segment objects in images, returning binary masks as output, and was adapted for use with remote sensing imagery (Wang et al., 2023; Osco et al., 2023).

This study aims to evaluate the performance of SAM, operating in zero-shot mode with geometric prompts, for the segmentation of urban tree vegetation from multispectral imagery with varying spectral compositions and ecological characteristics. PlanetScope imagery was selected due to its high spatial and temporal resolution, which is essential for capturing the dynamic and heterogeneous nature of urban environments. This analysis allows to investigate the influence of spectral variability on the model's efficiency in distinct urban contexts.

2. Methodology

The study was conducted in three Brazilian municipalities with distinct ecological characteristics: Rio Verde (GO), Valparaíso de Goiás (GO), and Corumbá (MS). The first two are in the Cerrado biome, while the latter is part of the Pantanal biome. Each area covers approximately 25 km², encompassing urban contexts with different patterns of tree vegetation. High-resolution orthorectified PlanetScope multispectral images (~3 m) (Planet Lab, 2023) were used, acquired between May 2024 and February 2025 by the PlanetScope constellation. For each city, specific spectral band composites were generated (Figure 1)

to evaluate the influence of spectral information on the model's performance in different urban landscape contexts:

i) Rio Verde (GO): Red (R), Green (G), and Blue (B) - standard true-color composite representing human visual perception.
ii) Valparaíso de Goiás (GO): Near-Infrared (NIR), Red, and Green - standard false-color infrared composite commonly used in vegetation analysis.
iii) Corumbá (MS): Red-Edge, Near-Infrared and Red - optimized for vegetation sensitivity through enhanced infrared response.

The images were subjected to standardized radiometric calibration, with band rescaling to 8 bits and the application of linear contrast stretching between the 2nd and 98th percentiles. This approach allowed the exclusion of extreme values and normalization of the digital value range, enhancing spectral difference critical for vegetation discrimination, while also standardizing data format for neural network processing. Metadata were modified to ensure that zero values were interpreted as valid data rather than as NoData, ensuring compatibility with the segmentation workflow.

For reference (Ground Truth), urban vegetation present in the images was manually annotated using polygons in the GIS software ArcGIS Pro 3.3.2, following strict visual criteria and a minimum mapping threshold of 25 pixels per instance, to ensure consistency and reliability in the segmentation process (Figure 2). It is worth noting that only tree vegetation located within or near the municipal boundaries was annotated, as the study focuses on the analysis of urban vegetation. The geometries were stored in GeoJSON format and converted into binary raster masks, matched in resolution and spatial extent to the original images.

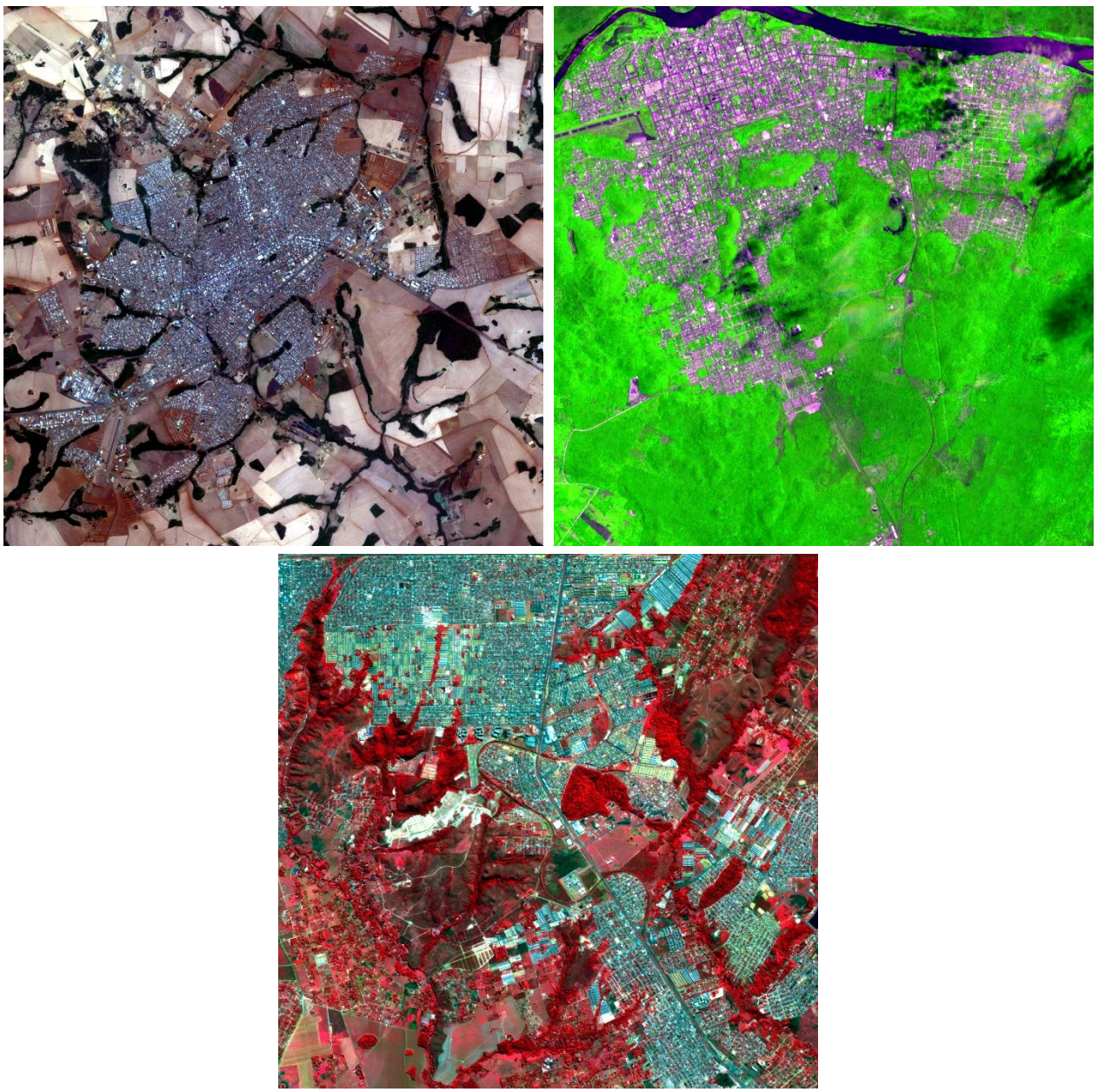

Figure 1. PlanetScope multispectral compositions for the study areas: *(top left)* Rio Verde (GO) – true-color composite (642:Red, Green, Blue), representing human visual perception; *(top right)* Corumbá (MS) – enhanced vegetation-sensitive composite (786: Red-Edge, Near-Infrared, Red), optimized for highlighting subtle spectral responses in plant canopies; *(bottom center)* Valparaíso de Goiás (GO) – false-color infrared composite (864:Near-Infrared, Red, Green), commonly used for vegetation analysis.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-241-2026

Figure 2. Examples of manual annotations of *tree vegetation* in three different spectral compositions (left: RGB; middle: Red-Edge-NIR-R; right: NIR-R-G).

Segmentation was performed using the SAM, specifically the ViT-H variant adapted by Osco et al. (2023), available in the GitHub repository (https://github.com/opengeos/segment-geospatial). The input strategy was relied on bounding boxes derived from reference polygons, where the minimum enclosing rectangle for each instance was extracted and used as a prompt for the model. The model exports the georeferenced mask that yields the highest similarity score, allowing for direct geometric comparison with the reference data. To ensure comparability, all reference polygons were rasterized to produce binary masks with the same spatial properties as the predicted masks.

Segmentation assessment was conducted through pixel-by-pixel comparison between predicted and reference masks, generating the elements of the confusion matrix: true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). Based on these values, the following metrics were computed: Accuracy (1), Precision (2), Recall (3), F1-Score (4), Intersection over Union (IoU) (5) and False Detection Rate (also called False Discovery Rate – FDR. It means 1-Precision) (6).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ Score = \frac{2 * TP}{2 * TP + FP + FN} \quad (4)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (5)$$

$$False\ Detection\ Rate = \frac{FP}{FP + TP} \quad (6)$$

The metrics were computed separately for each study area, with the results were presented in both tabular form (Table 1) and as graphical representations (Figure 3).

3. Results and discussion

SAM was employed in this study in zero-shot mode. Therefore, the results presented here reflect solely the model's capability when applied to an unfamiliar domain. Accordingly, the predicted masks were generated and compared on a pixel-by-pixel basis with the reference masks, enabling the quantitative analysis of the results through image segmentation metrics.

3.1 Quantitative analysis

In Corumbá (Table 1), the SAM model achieved a balanced performance, with an F1-score around 70% and IoU of 52.77%, indicating good segmentation quality. Recall was relatively high (close to 75%), suggesting a lower occurrence of false negatives, reflecting the model's ability to correctly identify most vegetation pixels. The false detection rate (36.12%) was the lowest among the three spectral band compositions, reinforcing the reliability of the predictions.

Metrics (%)	Rio Verde (RGB)	Corumbá (Red-Edge, NIR-R)	Valparaíso de Goiás (NIR-R-G)
Accuracy	81.27	75.21	84.12
Precision	60.19	63.88	57.24
Recall	81.27	75.21	84.12
F1-score	69.16	69.08	68.13
IoU	52.86	52.77	51.66
False Detection Rate	39.81	36.12	42.76

Table 1. Performance metrics for tree vegetation segmentation using the SAM model in the three study areas.

In Rio Verde (GO), the SAM demonstrated high accuracy in mapping tree vegetation, as reflected by a high recall (81.27%) (Table 1). However, the model exhibited a tendency toward over-segmentation, evidenced by a moderate precision of 60.19%. These results indicate that, although the model was effective in detecting vegetation, it also produced a relatively high number of false positives (34.96%, as shown in Figure 4). This limitation is reflected in the F1-score of 69.19%. The false detection rate (39.81%) supports this interpretation, as it is higher than the rate observed in Corumbá.

In Rio Verde, SAM received only RGB spectral bands, which are generally less effective at highlighting vegetation compared to infrared-based compositions, such as those used for Corumbá (MS). Nevertheless, the city's well-defined urban layout and relatively high vegetation contrast may have contributed to the model's high recall (81.27%), despite the absence of spectrally enhanced inputs. In Valparaíso de Goiás, the model received NIR-R-G spectral inputs and achieved the highest accuracy (84.12%) and recall (84.12%) among the three experiments (Table 1), indicating strong sensitivity to the tree vegetation class. However, this was accompanied by the highest incidence of false positives (39.58%, as shown Figure 4), resulting in the lowest precision (57.24%). Consequently, the F1-score (68.13%) and IoU (51.66%) remained comparable to those observed in the other study areas. The false detection rate (42.76%) was the highest overall, further reinforcing the model's tendency to overestimate vegetation in this region.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-241-2026 | © Author(s) 2026. CC BY 4.0 License.

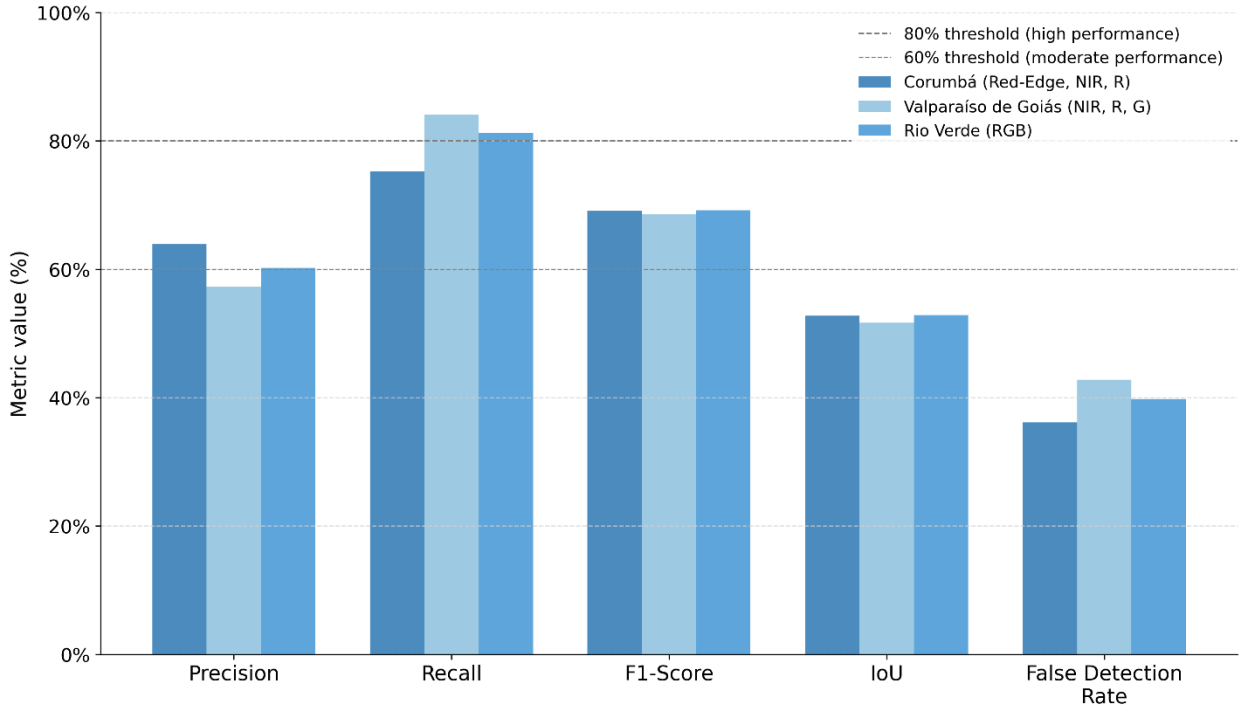

Figure 3. Performance of the Segment Anything Model in the segmentation of the tree vegetation class across the three study areas: Rio Verde (RGB), Valparaíso de Goiás (Near-Infrared, Red, Green), and Corumbá (Red-Edge, Near-Infrared, Red). The metrics were computed based on pixel-by-pixel comparison between predicted masks and reference masks. The dotted lines represent reference thresholds of 60% and 80%, used to classify performance as moderate or high, respectively.

Moreover, it is important to consider the spatial resolution of PlanetScope imagery (3 m), which, while suitable for identifying clusters of urban tree cover, presents limitations for isolating individual trees or small vegetation fragments. This constraint may have contributed to inaccuracies in mixed-use or structurally heterogeneous areas, where distinct targets are represented by only a few pixels. As a result, part of the false positives and false negatives observed can be attributed not only to the spectral behavior of the targets but also to the challenges of spatial discrimination in complex urban environments.

As a complementary analysis to the metrics presented in Table 1, several standard performance metrics were further examined using two reference thresholds—60% and 80%—to classify results as moderate or high. Mapping urban tree vegetation in orbital imagery remains a complex task due to the high spectral and spatial heterogeneity commonly found in urban environments.

Figure 3 illustrates the performance of the SAM in segmenting tree vegetation across the three study areas, evaluated using these standard metrics. Despite operating in a zero-shot setting—without any prior exposure to satellite imagery or task-specific fine-tuning, SAM achieved recall above the 80% threshold in almost all cases, indicating strong sensitivity to vegetation targets. However, precision values remained below the 80% threshold and, in some cases, only slightly above the moderate-performance threshold of 60%, suggesting a tendency toward over-segmentation. F1-scores and IoU metrics consistently fell between the 60% and 80% thresholds, reflecting a balanced but non-optimal trade-off between omission and commission errors. Notably, false detection rates exceeded 40% in Valparaíso de Goiás, reinforcing this tendency. These results highlight SAM's potential for generalization in remote sensing applications, even under constrained spectral input and without retraining, though further optimization is needed to improve precision. A qualitative analysis is presented in the following section to provide a visual and interpretative assessment of the segmentation results across the study areas.

3.2 Qualitative analysis

A visual analysis of the semantic segmentation results produced by SAM is presented in Figure 5, offering insights into the model's behaviour across different urban environments. For the municipality of Rio Verde (GO), the orbital image used was a natural RGB composite, which presents spectral limitations for accurately distinguishing vegetation cover. The comparison between the reference mask and mask generated by SAM reveals satisfactory performance in areas with dense and visually prominent vegetation, suggesting that the model was able to identify consistent patterns within the visible spectrum. However, the absence of bands beyond the visible range—particularly near-infrared—compromised the model's ability to discriminative vegetation in mixed-use areas, where tree cover is fragmented, shaded, or interspersed with anthropogenic surfaces. In these regions, segmentation inconsistencies such as imprecise boundaries and overfilling were observed, especially along the edges of vegetated objects.

The qualitative analysis for the municipality of Valparaíso de Goiás (GO), based on a composite of NIR-Red-Green bands, indicates that the semantic segmentation model performed consistently in identifying denser vegetated areas. The inclusion of the near-infrared band enhanced the discrimination of active vegetation, enabling the model to more confidently detect photosynthetically active vegetation. In regions with continuous vegetation cover, there was a strong agreement between the predicted and reference masks, with minimal noise (Figure 5).

https://doi.org/10.5194/isprs-annals-X-3-W4-2025-241-2026

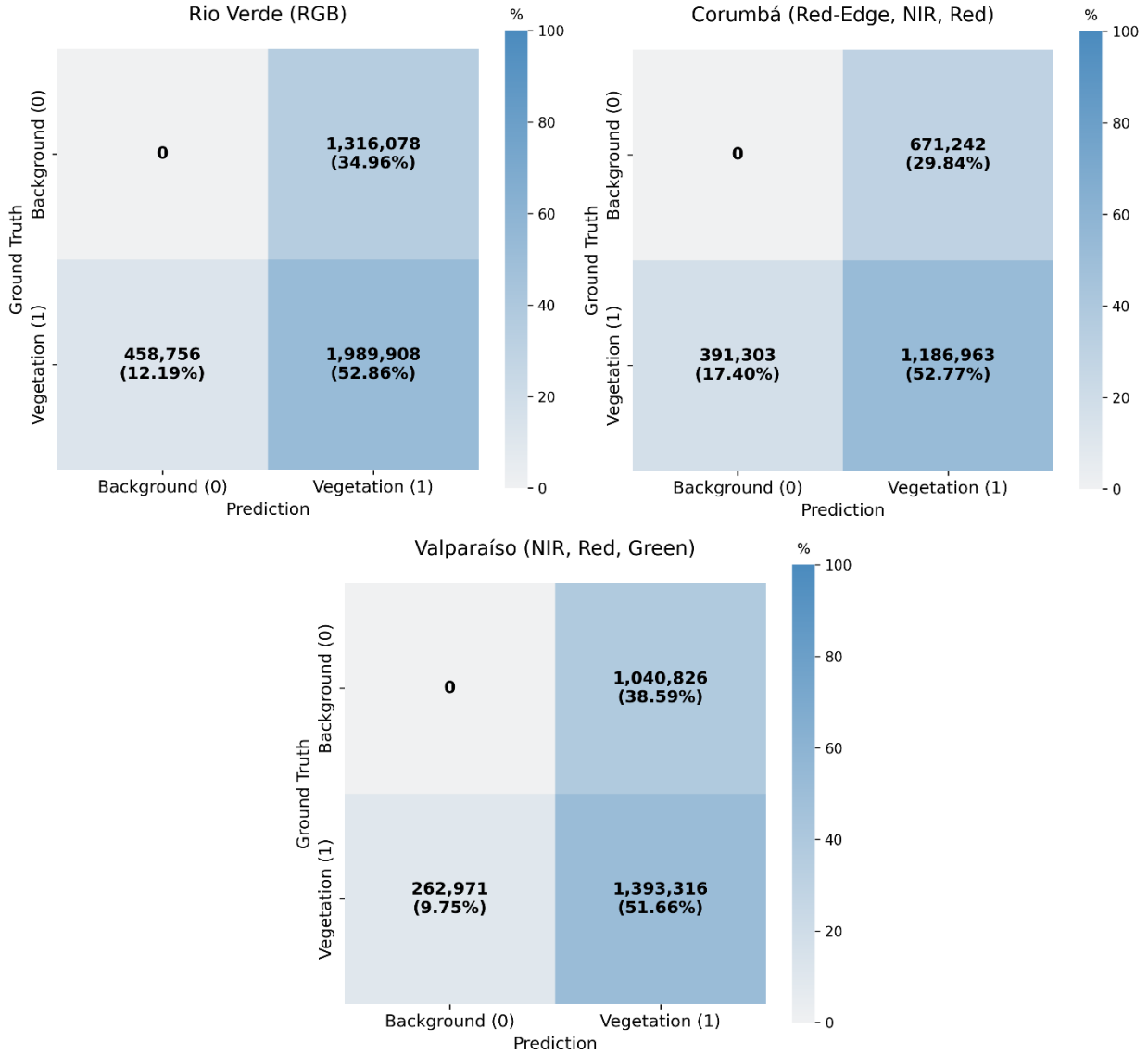

Figure 4. Confusion matrices for tree vegetation segmentation in Rio Verde, Corumbá, and Valparaíso de Goiás. TP, FP, and FN are shown as absolute values and percentages. Only class 1 (vegetation) is considered; background is ignored.

Despite the spectral advantages of the NIR-Red-Green composite, the model exhibited difficulties in transition zones between vegetation and artificial surfaces. False positives were frequently observed in paved regions or on structures composed of materials with high NIR reflectance, such as light-coloured roofs and concrete, leading to misclassification as vegetation. Conversely, sparse or partially shaded vegetation was often under-detected, resulting in false negatives. These findings suggest that, although the model benefited from enhanced spectral input, challenges persist inaccurately segmenting vegetation in densely built areas with complex spectral and structural patterns. Therefore, complementary strategies may be required to improve segmentation robustness in urban contexts such as Valparaíso de Goiás (GO).

For the images composition (Red-Edge, Near-Infrared, Red) of Corumbá, the comparison between the reference and predicted masks reveals moderate overall agreement, with good performance in detecting areas of dense and continuous vegetation. The model successfully captured spectral patterns associated with photosynthetically active vegetation, particularly those reflected in the Near-Infrared and Red-edge bands, suggesting that these bands contributed positively to class discrimination. However, segmentation inaccuracies were observed in transitional zones—such as edges the interfaces between vegetation and water bodies, urban surfaces, or exposed soils—including smoothed objects boundaries and, in some cases, slight spatial misalignments relative to the reference mask.

3.3 Comparative Analysis

Despite the challenging conditions posed by urban complexity and spectral ambiguity, the Segment Anything Model (SAM) demonstrated the ability to delineate vegetated structures in a manner broadly consistent with the reference masks. As summarized in Figure 3, most performance metrics across the three study areas exceeded the 60% threshold—considered indicative of moderate performance. Notably, Recall values surpassed 80% for the images from Valparaíso de Goiás and Rio Verde, suggesting high sensitivity to the vegetation class in those regions.

Among the study areas, Rio Verde exhibited the best balance between detection and selectivity, despite relying solely on RGB

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-241-2026 | © Author(s) 2026. CC BY 4.0 License.

bands. This highlights SAM's relative robustness under spectrally constrained conditions. Valparaíso, although presenting extensive vegetative cover and benefiting from NIR-based inputs, displayed lower selectivity, leading to a greater incidence of false positives and reduced precision. Corumbá showed a more balanced behaviour, yet its performance was partially affected by adverse atmospheric conditions—such as cloud shadows—that impaired the vegetation's spectral response. The systematic presence of false positives, observed in all study areas (Figure 4), negatively impacted precision-based metrics such as F1-Score and IoU. This trend reflects a key limitation of the zero-shot approach adopted in this study—wherein SAM was applied without any task-specific fine-tuning or retraining using remote sensing imagery. False positives were frequently associated with surfaces that strongly reflect in the NIR or Red-edge bands, including bare soils and artificial materials, as well as RGB targets with greenish or brownish tones (e.g., certain roofs or pavements), which the model occasionally misclassified as vegetation (Figure 5).

Conversely, false negatives were predominantly located in regions with sparse, shaded, or early-stage vegetation, which tend to exhibit weak spectral responses—particularly in Red-edge and NIR wavelengths. This hindered the model's ability to consistently identify such areas, especially in transition zones or mixed land cover. In summary, while the model performed well in homogeneous, high-contrast environments, its accuracy declined in structurally and spectrally complex contexts.

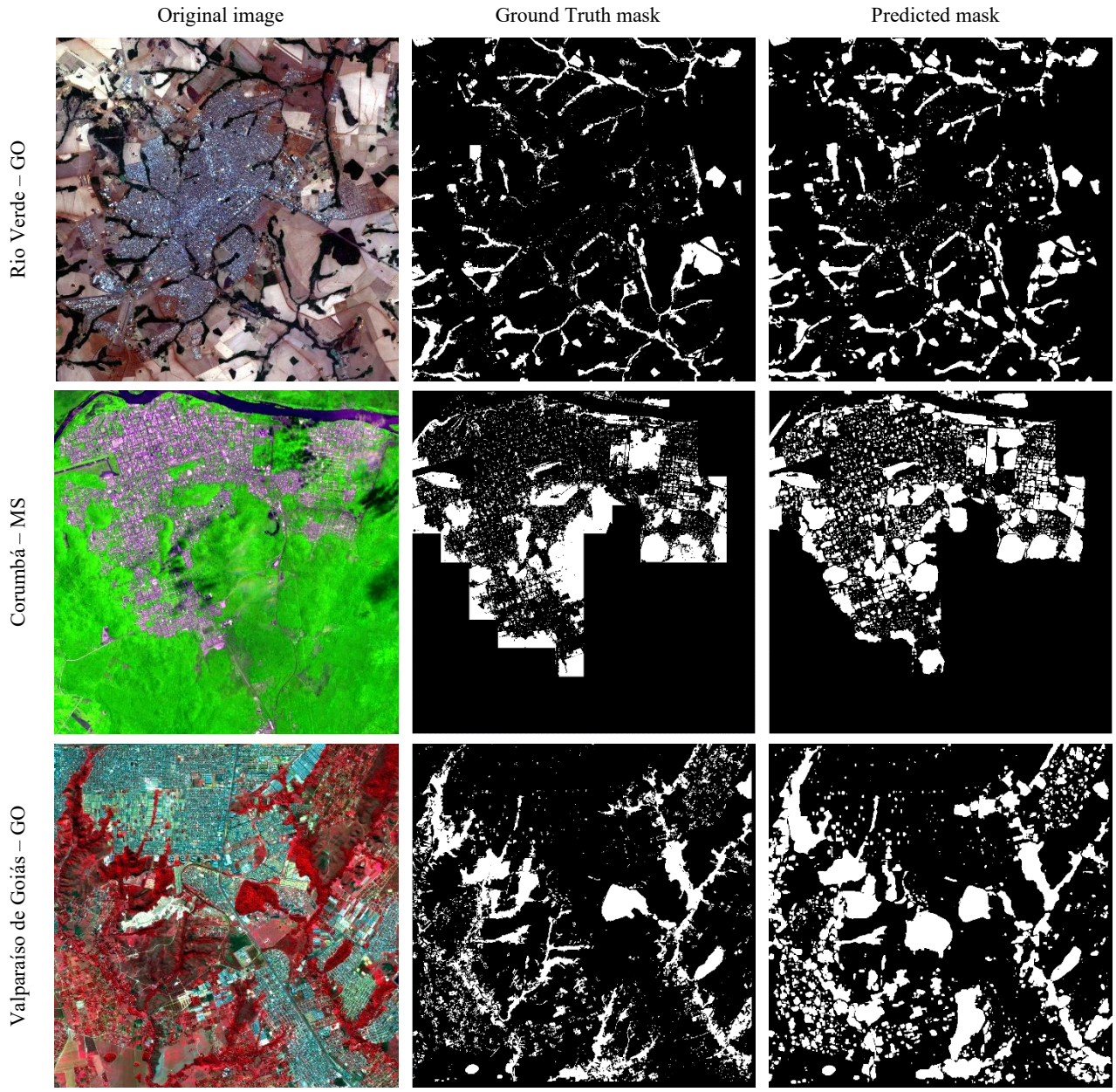

Figure 5. Visual comparison of the segmentation results for the *tree vegetation* class in the three study areas. The first column shows the original orbital images used as input to the model, with distinct spectral compositions: (top) Rio Verde– RGB; (middle) Valparaíso de Goiás– Near-Infrared, Red, and Green; (bottom) Corumbá– Red-Edge, Near-Infrared, and Red. The second column displays the manually created ground truth masks based on visual interpretation. The third column presents the predicted masks generated by the SAM model, applied in zero-shot mode. White areas in the masks indicate pixels classified as vegetation.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-241-2026 | © Author(s) 2026. CC BY 4.0 License.

Given that SAM was deployed in a zero-shot setting, the overall results—characterized by F1-scores exceeding 68% across all experiments (Table 1)—are promising and indicate considerable generalization capability. However, to improve segmentation reliability, especially in heterogeneous or spectrally ambiguous areas, complementary strategies should be considered. These may include fine-tuning the model on domain-specific data, incorporating additional spectral bands or vegetation indices, and leveraging prompt engineering techniques to reduce over-segmentation.

This study highlights SAM's potential in zero-shot mode for urban tree vegetation mapping, but several avenues remain for future work. Exploring prompting strategies beyond bounding boxes could better represent operational scenarios. Incorporating instance-based metrics such as Average Precision (AP) would add object-level insights, while comparisons with supervised models (e.g., U-Net, DeepLab) could provide useful baselines despite their different training requirements. In addition, lightweight fine-tuning or prompt tuning tailored to remote sensing data may improve spatial precision in complex urban environments.

4. Conclusion

This study assessed the applicability of the Segment Anything Model (SAM), applied in a zero-shot mode, for the semantic segmentation of urban tree vegetation using orbital imagery. The evaluation was conducted three Brazilian municipalities - Corumbá (MS), Rio Verde (GO), and Valparaíso de Goiás (GO) – each characterized by distinct spectral configurations, urban morphologies, and ecological conditions.

Despite the absence of any supervised training or domain-specific adaptation to remote sensing data, SAM demonstrated consistent performance in identifying vegetated areas. The quantitative results revealed that the model's effectiveness varied according to spectral composition and vegetation density. Scenes enriched with infrared bands and containing dense vegetation cover yield more balanced segmentation outcomes. However, systematic false positives – particularly in spectrally ambiguous or mixed-use-areas - negatively affected precision-based metrics like F1-score. In addition, segmentation accuracy declined in regions with could shadows or abrupt transitions between tree and urban infrastructure.

The qualitative analysis further supported these observations. SAM showed sensitivity to structural and spectral vegetation patterns, performing well in heterogeneous regions with high contrast. Nonetheless, challenges were evident in heterogeneous urban contexts, specially where vegetation was sparce, degraded, or interspersed with built-up-features. Common errors included over-segmentation and omission of subtle tree features, indicating limitations in the model's capacity to generalize across complex spatial configurations.

Taken together, the findings suggest that SAM, even when used without fine-tuning or prior exposure to orbital imagery, possesses potential as a generic segmentation tool in remote sensing applications. Its ability to extract meaningful vegetation patterns across diverse urban landscapes highlights its versality and scalability. However, to enable its adoption in operational urban vegetation mapping, refinement strategies are recommended- particularly supervised fine-tuning using domain-specific datasets and the integration of tailored prompt configurations. These enhancements would likely improve the model's class discrimination, spatial precision, and overall reliability in applied geospatial workflows.

Acknowledgements

We acknowledge the support of the Coordination for the Improvement of Higher Education Personnel (CAPES) for awarding a master's scholarship. We also thank the National Council for Scientific and Technological Development (CNPq) for financial support under grants 305814/2023-0, 403213/2023-1, 308481/2022-4, 305296/2022-1, and 135614/2024-4. We also acknowledge the Faculty of Science and Technology of São Paulo State University (UNESP) for financial support under Call No. 03/2024.

References

Ji, Wei et al. Segment anything is not always perfect: An investigation of SAM on different real-world applications. arXiv preprint arXiv:2304.05750, 2023.

Kirillov, Alexander et al. Segment anything. arXiv https://arxiv.org/pdf/2304.02643.pdf, 2023.

Nowak, D. J., et al., 2014. Tree and forest effects on air quality and human health in the United States. *Environmental Pollution*, 193, 119–129.

Osco, Lucas Prado et al. A review on deep learning in UAV remote sensing. International Journal of Applied Earth Observation and Geoinformation, v. 102, p. 102456, 2021.

Osco, Lucas; Wu, Q.; Lemos, E. L.; Goncalves, W. N.; Ana Marques Ramos; Li, Jonathan; Marcato, José. The Segment Anything Model (SAM) for Remote Sensing Applications: From Zero to One Shot. International Journal of Applied Earth Observation and Geoinformation., v.124, 2023.

PLANET LABS PBC. *PlanetScope Product Specifications*. 2023. https://pal.planet.com/assets/share/asset/xgfpsxmnzy.

Rahman, M.A., Arndt, S., Bravo, F., Cheung, P.K., van Doorn, N., Franceschi, E., del Río, M., Livesley, S.J., Moser-Reischl, A., Pattnaik, N., Rötzer, T., Paeth, H., Pauleit, S., Preisler, Y., Pretzsch, H., Tan, P.Y., Cohen, S., Szota, C. & Torquato, P.R., 2024. More than a canopy-cover metric: Influence of canopy quality, water-use strategies and site climate on urban-forest cooling potential. Landscape and Urban Planning, 248, 105089. https://doi.org/10.1016/j.landurbplan.2024.105089

Trask, A. Grokking Deep Learning. Manning Publications. 335p. 2019.

Wang, Di et al. Scaling-up Remote Sensing Segmentation Dataset with Segment Anything Model. arXiv preprint arXiv:2305.02034, 2023.