Mapping of urban tree canopy in high-resolution aerial imagery using deep neural networks

Brian Leite Machado[1], Rafael Ochi Kikuti[1], Lucas Prado Osco[2], José Marcato Junior[3], Wesley Nunes Gonçalves[3], Ana Paula Marques Ramos[1]

[1]São Paulo State University (UNESP), [2]Universidade do Oeste Paulista (Unoeste), [3]Federal University of Mato Grosso do Sul (UFMS)

Keywords: Urban forestry, Semantic segmentation, DeepLabV3, Aerial high-resolution imagery, Deep learning.

Abstract

While deep learning has proven effective for urban tree mapping, there is a critical lack of validated benchmarks and comparative methodological studies for the diverse urban landscapes of Brazil. To address this gap, this work presents a deep-learning workflow that produces such maps from 25 cm RGB orthophotos. Images covering ten São Paulo cities were compiled; seven were used for training/validation and three for independent testing. The DeepLabV3 architecture with a ResNet-152 backbone was assessed under three loss configurations: (i) Balanced Cross-Entropy (BCE) baseline, (ii) BCE plus PointRend boundary refinement, and (iii) BCE combined with a 0.5-weighted Dice term. The BCE baseline delivered the top mean IoU (0.83) and F1-Score (0.91). PointRend increased recall but introduced systematic false positives in heterogeneous roofs and shaded riparian zones. The BCE+Dice variant recovered recall without raising commission error, achieving the highest balanced accuracy (0.96). The workflow delineates canopy with fine spatial detail and processes 2.8×10^6 m² in under 30 minutes on a single RTX 4000 Ada workstation, demonstrating a practical, scalable solution for statewide tree-inventory production.

1. Introduction

Urban trees play a strategic role in city sustainability by delivering ecosystem services that mitigate the urban-heat-island effect, filter pollutants, reduce runoff, and enhance psychological well-being. Studies have shown a direct link between canopy cover and urban quality of life (Rahman et al., 2024; Jones, 2021). However, canopy distribution is often uneven, raising socio-environmental justice concerns that require precise mapping to inform policy (Locke et al., 2023).

Mapping urban vegetation at municipal scale is challenging due to spectral complexity, building shadows, and similar targets like lawns and green roofs. Advances in high-resolution aerial imagery—with ground sampling distances (GSD) of just a few centimetres—now enable individual tree detection. Yet this also increases data volume and processing complexity. Traditional classifiers (e.g., SVM, RF) often struggle to generalize across varied urban contexts, resulting in spatial inconsistency (Morgan et al., 2024). In 2022, the Instituto Geográfico e Cartográfico de São Paulo (IGC-SP) released full RGB coverage at 25 cm GSD across the state. While spatially detailed, this dataset requires robust computational tools.

Deep neural networks have transformed semantic segmentation by uniting automatic feature extraction with multi-scale contextual learning. The DeepLab family pioneered the use of dilated convolutions and the Atrous Spatial Pyramid Pooling (ASPP) module, achieving state-of-the-art results across benchmarks (Chen et al., 2017). Recent variants, including DeepLabV3+ and custom adaptations, remain top performers in high-resolution remote sensing applications (Wang et al., 2024). Among available backbones, ResNet-152 offers sufficient depth to resolve fine canopy structures while preserving global context and has been recommended for its strong inter-site generalization in tree crown mapping (Weinstein et al., 2020).

While these advances are significant, three specific research gaps persist, particularly within the Latin American context. First, a geographical gap exists, as most high-resolution validation studies are concentrated in North American and European cities, neglecting the unique spectral and morphological characteristics of Brazilian urban areas. Second, a methodological gap exists regarding the practical trade-offs of different model configurations; few studies have directly compared the impact of boundary-refinement modules like PointRend against shape-aware loss functions (e.g., Dice loss) for the specific task of tree crown delineation. Finally, an implementation gap often separates academic research from operational deployment, highlighting the need for workflows validated within GIS platforms widely used by municipal agencies.

Despite such advances, state-wide applications in Brazil remain scarce. Recent efforts have focused on national-scale automated mapping but lack detailed municipal validation and controlled comparisons across biomes and contrasting urban contexts (Guo et al., 2023; Dalagnol et al., 2023). To address these specific gaps, this study presents three core contributions: (1) it provides a rigorous performance benchmark for tree canopy segmentation across ten representative cities in São Paulo, a major yet under-studied region; (2) it delivers a direct comparative analysis of three distinct loss configurations, offering practical insights into model optimization; and (3) it details an operational workflow developed entirely within ArcGIS Pro, demonstrating a scalable and deployable solution for statewide mapping. Ground truth masks were manually generated, enabling cross-city performance assessment and supporting statewide operational deployment. The workflow was developed entirely within ArcGIS Pro using a DeepLabV3 model with a ResNet-152 backbone and trained on a pixel-level dataset compiled from ten representative cities across the Mata Atlântica and Cerrado biomes. The resulting model was evaluated through systematic cross-city testing, demonstrating its potential for high-resolution, scalable mapping of urban tree canopy.

2. Materials and methods

2.1 Study Area and Dataset

Figure 1 locates the ten municipalities selected across São Paulo State: seven constitute the training/validation set, whereas the remaining three form an independent test set for the semantic-

segmentation experiments. The imagery dataset comprises true-colour (RGB) orthophotos with 25 cm ground-sampling distance (GSD), acquired between September and November 2022 under a statewide survey commissioned by the São Paulo State Geographic and Cartographic Institute (IGC-SP). All scenes were orthorectified, radiometrically normalized, and delivered in the SIRGAS 2000 reference frame (UTM zones 22S and 23S). Table 1 lists the main attributes of each municipality—dataset split (train or test), biome, Köppen climate class, median elevation, and total canopy area.

To maximise spectral diversity in the test set, we computed an RGB histogram for every annotated crown in each city and measured pairwise distances between the city-level histograms. The three cities with the most divergent spectral signatures—Santos, Jaguariúna, and Itapetininga—were reserved for independent testing, while the remaining seven supported training and validation. This choice captures (i) both dominant state biomes (Atlantic Forest and Cerrado), (ii) the two prevailing climate classes (Cfa: humid subtropical; Cwa: dry-winter subtropical), and (iii) an elevation range from 2 m (coastal Santos) to 660 m (Itapetininga). Consequently, the segmentation models are evaluated under a broad spectrum of spectral and structural conditions, reinforcing the assessment of their generalisation capability (see Table 1).

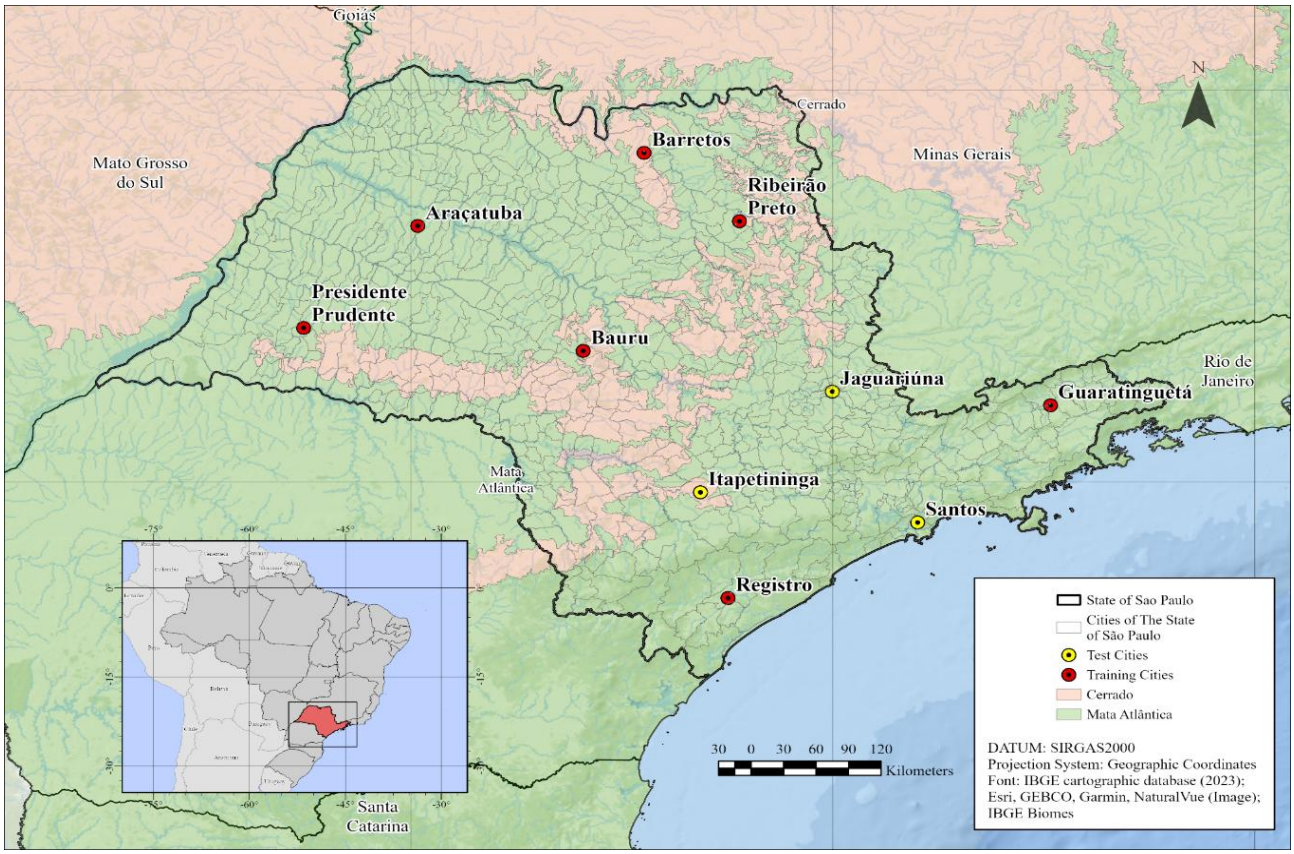

Figure 1. Spatial distribution of case-study cities in São Paulo State, Brazil.

City	Group	Biome	Köppen	Average Elevation (m)	Labeled Area (ha)
Araçatuba	Train	Mata Atlântica	Aw	390	22.23
Barretos	Train	Cerrado	Aw	530	18.91
Bauru	Train	Cerrado	Cwa	560	40.40
Guaratinguetá	Train	Mata Atlântica	Cwa	540	14.13
Prudente	Train	Mata Atlântica	Cfa	430	38.65
Registro	Train	Mata Atlântica	Cwa	20	15.03
Ribeirão Preto	Train	Mata Atlântica	Cwa	550	25.08
Itapetininga	Test	Cerrado	Cwa	660	26.11
Jaguariúna	Test	Mata Atlântica	Cfa	600	26.45
Santos	Test	Mata Atlântica	Cfa	2	22.26

Table 1. Metadata for case-study cities and their corresponding tree-cover annotations.

2.2 Ground-Truth Annotation and Image Tiling

A total of 460 patches (256 × 256 pixels; 25 cm GSD) were extracted from RGB orthophotos covering ten cities. To capture the full diversity of urban tree canopies, ground-truth masks were delineated across ecologically distinct regions—from the humid Atlantic Forest lowlands to the drier Cerrado highlands—ensuring variability in biome, climate, and elevation.

In each city, tree-canopy fragments within a 512.19 m × 768.29 m rectangle (~39.4 ha) were manually annotated. Using the high-resolution orthophotos as a base, individual crowns or contiguous crown clusters were digitized with a one-pixel snap tolerance, guaranteeing precise alignment between polygon vertices and the raster grid. This process yielded over 10 million tree-labeled pixels, equivalent to approximately 250 hectares of canopy area. Figure 2 presents an example orthophoto alongside its corresponding mask, highlighting both the spatial detail and annotation precision.

A detailed protocol (adapted from Weinstein et al., 2020) defined the class "tree" as any woody perennial with a projected crown diameter ≥ 1 m, visibly distinct from shrubs, grasses, or herbaceous cover. The guidelines specified: (i) include crowns partially obscured by shadow or buildings if ≥ 50% of the crown margin is visible; (ii) exclude streetlights or green roofs; and (iii) merge adjacent treetops from the same set into a single polygon if separation is impossible at 25 cm resolution.

Figure 2. 25 cm-GSD orthophoto and its manually annotated tree-crown mask.

Manually delineated tree fragments were rasterized into single-band, 8-bit masks, where canopy pixels were assigned a value of 1 and background pixels 0, ensuring perfect alignment between images and labels. The resulting dataset is highly imbalanced: only 3.3% of pixels correspond to trees, while 96.7% represent background. To evaluate which training strategy most effectively mitigates this imbalance, three configurations were tested: (i) balanced cross-entropy with PointRend (Kirillov et al., 2020) refinement (Run A), (ii) a baseline using only balanced cross-entropy (Run B), and (iii) a composite loss combining balanced cross-entropy with 50% soft Dice (Run C).

2.3 Data Augmentation

The default ArcGIS Pro pipeline applied random augmentations to every 256 × 256 pixels image–mask pair at each epoch: crop (retain ≈ 80–100 % of the tile), dihedral flips/rotations, ±20 % brightness or contrast shifts, and uniform zoom (0.75–1.25×). These on-the-fly, label-synchronized transforms expose the model to varied geometry and radiometry, improving generalization to the three unseen test cities.

2.4 Network Architecture

The segmentation model adopted is DeepLabV3, shipped with the ArcGIS Pro deep learning toolbox. DeepLabV3 preserves fine spatial detail by replacing the final strided convolutions of the backbone with atrous (dilated) convolutions and then aggregating multi-scale context through an ASPP head that uses dilation rates of 6, 12, and 18 together with a global image-level pooling branch (Esri, 2024a).

A 152-layer ResNet pre-trained on ImageNet serves as the backbone. Feature maps emerging from the backbone enter the improved ASPP module; its concatenated output is compressed with a 1 × 1 convolution, batch-normalized, activated with ReLU, and regularized with 0.5 dropout. A final 1 × 1 convolution produces the logits map, which is bilinearly up-sampled to the original 256 × 256 pixels tile size, and a sigmoid activation yields per-pixel probabilities for the binary classes, tree, and background. The ArcGIS implementation omits the fully connected Conditional Random Field found in earlier DeepLab versions. Additionally**,** it incorporates batch normalisation inside ASPP for greater training stability (Esri, 2024a). The optional PointRend refinement was tested and disabled because it increased inference time while providing negligible IoU improvement in preliminary trials (Esri, 2024a).

2.5 Experimental Procedure

All experiments were conducted in ArcGIS Pro 3.5 using Esri's Deep Learning Libraries Installer, which extends the default *arcgispro-py3* environment with 254 pre-packaged libraries. Computations were performed on a workstation equipped with an Intel Core i7-13700K processor, 64 GB of RAM, and an NVIDIA RTX A4000 GPU (16 GB VRAM). ResNet weights were initialized from ImageNet and fine-tuned using discriminative learning rates: 1×10^{-3} for the ASPP head and classifier, and 1×10^{-4} for earlier backbone layers. The resulting model comprises approximately 60 million trainable parameters.

We evaluated three semantic segmentation pipelines that differed solely in their loss function configurations. Accordingly, three DeepLabV3 models were fine-tuned under identical settings: ResNet-152 backbone, AdamW optimizer with one-cycle learning rate policy, batch size of 8, on-the-fly data augmentation (Section 2.3), and early stopping triggered by ≥10 epochs without validation loss improvement. Seventy percent of the cities were allocated for model development, with 10% of that subset set aside for validation. The remaining 30% were reserved for independent testing. **Run A** utilized the PointRend module with the default balanced cross-entropy (BCE) loss. **Run B**, the baseline, used only the BCE loss, with both the PointRend module and the Dice loss term disabled. **Run C** created a composite loss by setting the Dice loss fraction to 0.5 (50% BCE, 50% soft-Dice), while keeping the PointRend module disabled.

2.5.1 Run A — PointRend enabled, Balanced Cross-Entropy loss

PointRend was activated to test whether a dedicated boundary-refinement step could recover fine crown edges that the coarse 1/16-resolution feature map of DeepLabV3 might miss. In the ArcGIS implementation, PointRend is applied after the usual balanced cross-entropy (BCE) optimisation: the K most uncertain pixels are resampled and re-rendered at full resolution, iteratively sharpening object borders without altering the loss function itself (ArcGIS Developers, 2024a). With the PointRend module enabled, the model was trained for 20 epochs with a frozen

backbone. Validation loss dropped sharply over the first five epochs and then plateaued, reaching its minimum (0.080 ± 0.005) at epoch 17, with pixel accuracy of 0.967 and Dice score of 0.753. PointRend increased GPU memory usage from 9.4 GB to ~12.7 GB and extended iteration time by approximately 18%, but training remained stable: loss curves for training and validation ran parallel, suggesting minimal overfitting. Total convergence time for the 20 epochs was 21 minutes.

2.5.2 Run B — Baseline configuration, Balanced Cross-Entropy loss without PointRend

The run B — used as reference for all subsequent comparisons — relied strictly on the default DeepLabV3 settings in ArcGIS Pro. Both the PointRend module and the Dice loss term were not used, so optimization proceeded with the balanced BCE loss only. The memory footprint stabilized at 9.4 GB, allowing the planned batch size of eight tiles on the RTX 4000 Ada without gradient accumulation. Convergence occurred at epoch 17, the same point as Run A, but with marginally lower wall-clock time due to the absence of PointRend refinement.

2.5.3 Run C — Dice–BCE composite loss, PointRend disabled

The third experiment kept the PointRend module disabled but set the Dice loss fraction to 0.5, instructing the model to back-propagate a loss of equal parts balanced BCE and soft Dice. The Dice component encourages overlap between predicted and reference masks, thus penalizing fragmented or overly smooth crowns. Internal validation showed slower initial learning, common when Dice is introduced from epoch 1, yet continued to improve after the BCE-only runs had plateaued, eventually triggering early stopping at epoch 52. GPU usage (9.5 GB) and iteration time (-3 % versus baseline) remained virtually unchanged because Dice adds computations only during the loss step, not in the forward path. This run, therefore, isolates the effect of the composite loss, free of PointRend's boundary refinement, setting the stage for the comparative results reported in the next section.

2.6 Evaluation Metrics and Test Strategy

Performance was quantified with six standard metrics — Intersection over Union (IoU), Precision, Recall, F1-Score, Overall Accuracy, Balanced Accuracy, Symmetric Difference calculated separately for each of the three test-set cities and every model variant (Runs A, B, and C). The city-level scores were then averaged, yielding a single value for each metric per run. Pixel-wise counts of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) were extracted from a confusion matrix generated for every model–city pair. Balanced Accuracy was defined as the average of recall for each class (tree and background), which is better suited for imbalanced datasets.

Together, these metrics capture pixel-level agreement (Overall Accuracy, Balanced Accuracy, Symmetric Difference) and canopy-specific performance (IoU, Precision, Recall, F1-Score), providing a comprehensive assessment of each segmentation model. All computations were automated in the stand-alone Python script Metric_Comparison.py, which rasterises the ground-truth and prediction polygons to the orthophoto grid and then derives the complete pixel-level confusion matrix.

3. Results

3.1 Quantitative analysis

Figure 3 compares the six-evaluation metrics for the three configurations in each test city. Across the three configurations, the baseline model without PointRend (Run B) achieves the best mean performance, registering the highest Intersection-over-Union (0.832) and F1-Score (0.908). Its advantage over the PointRend variant is clearest in Jaguariúna, where additional false positives lower Run A's IoU despite its higher recall.

Run A indeed produces the greatest mean recall (0.935), showing that the refinement step recovers thin or fragmented crowns, but it does so at the expense of precision (0.846). Introducing a 0.5 Dice term (Run C) restores almost all the recall lost by removing PointRend (0.942) while moderating commission error, so balanced accuracy rises to 0.958—the best overall. This gain is most evident in Santos, where the Dice component reduces false positives along shaded coastal vegetation, lifting IoU from 0.799 (Run A) to 0.833. By contrast, Itapetininga remains an easy scene for every network ($\text{IoU} \geq 0.864$), indicating that its bright, isolated crowns are well represented in the training distribution.

3.2 Qualitative analysis

Figure 4 contrasts the symmetric difference (all prediction errors – FN in red and FP in green) maps of Run A (PointRend + BCE) and Run B (baseline BCE) for three zoomed-in excerpts of Jaguariúna and two others of Itapetininga. In Figure 4b, the ground-truth polygons delineate a dense riparian stand interrupted by a narrow water body that was intentionally excluded from the reference mask. PointRend merges the entire cluster into a single, spatially smooth region; the resulting union spills into the channel, producing a continuous red patch in the symmetric-difference layer (false positives and false negatives combined). The baseline network, by contrast, respects the spectral discontinuity introduced by the water and leaves a clear gap, eliminating the commission error responsible for the 5.8% IoU deficit observed for Run A in this city.

Figure 4c highlights a second failure mode of the refinement module. Here, PointRend extends the crown of an isolated tree outward to encompass the surrounding low-shrub understory, again visible as a contiguous red envelope in the error map. Without the refinement step, Run B preserves a sharper boundary between canopy and shrub layer, matching the vector mask and reducing both false positives along the crown edge and false negatives in the inter-crown gaps.

Together, these examples illustrate that, for heterogeneous peri-urban scenes, the spatial smoothing performed by PointRend can override legitimate high-frequency boundaries—bodies of water, abrupt height transitions, or shrub gaps—thereby degrading overall IoU despite its higher recall. The qualitative evidence in Figure 4, therefore, corroborates the quantitative Figure 3.

Figure 4a offers a counterexample: a second water channel whose spectral response is nearly indistinguishable from the surrounding canopy. Here, both networks—PointRend and baseline—misclassify the entire feature as tree, generating an identical red strip in the symmetric-difference map. This case underscores a residual limitation of the spectral-RGB input itself: when water is fully occluded by overhanging foliage or exhibits low-contrast reflections, neither architectural adjustment nor loss weighting suffices to separate it from dense crowns advantage of the baseline configuration reported Figure 4d and 4e highlights two error cases

in Itapetininga, the best-performing city in the quantitative evaluation.

In Figure 4d a single broad-leaf tree, spectrally different than the surrounding dense canopy, lies at the margin of a forest patch. PointRend (Run A) preserves this crown, but the baseline model (Run B) erodes its outline and omits most of its pixels, indicating that the refinement step can occasionally recover spectrally deviant trees at stand edges. Figure 4e, both networks fail: a strip of low, shrubby vegetation on a gentle embankment is absorbed into the adjacent closed canopy, producing a continuous false-positive ribbon. This example shows that, even in the easiest test scene, the model struggles to separate understory or heterogeneous height layers when spectral contrast with the upper canopy is weak, a limitation common to all three configurations.

Similar artifacts—spectrally different edge crowns omitted, low-shrub strips merged with adjacent canopy, and small water bodies embedded within dense tree stands misclassified as vegetation—were likewise detected in Santos (not shown), confirming that the qualitative error modes documented for Jaguariúna and Itapetininga recur across all test cities.

4. Discussion

Overall, the baseline DeepLabV3 trained with Balanced Cross-Entropy alone delivered the highest mean IoU and the smallest commission errors, making it the most reliable configuration for large-scale canopy mapping. While adding PointRend boosted recall and mixing Dice loss improved class balance, both alternatives sometimes over-smoothed fine edges, so the simpler baseline ultimately offered the best trade-off between accuracy and robustness. The experiments show that the DeepLabV3–ResNet–152 was capable of segmenting trees across São Paulo State with a mean IoU of 0.83, a figure that exceeds most sub-meter studies based on satellite data and remains consistent across altitude, biome, and Köppen-class gradients (Guo et al., 2023; Morgan et al., 2024; Dalagnol et al., 2023). Boundary refinement with PointRend merged spectrally dissimilar elements—water pockets inside riparian forest, shrub belts at canopy edges—thereby depressing IoU in the most heterogeneous scene (Jaguariúna) by six points. Replacing PointRend with a 0.5 Dice term restored nearly all of the lost recall while keeping commission error low, giving the highest balanced accuracy (0.96).

These findings confirm that, for dense urban canopies, shape-aware loss functions are more beneficial than late-stage contour sharpening, and they do so without the 35 % memory and 18 % runtime penalty imposed by PointRend. This conclusion provides a meaningful methodological insight for practitioners, suggesting that for complex natural targets, optimizing the loss function is a more efficient strategy for improving performance than relying on post-hoc refinement modules.
Small water bodies hidden by overhanging crowns, bright outlier trees at stand borders, and low shrubs abutting tall canopies are all mislabeled in at least one city. Mitigation will require either near-infrared or LiDAR (Light Detection and Ranging) height cues as well as targeted sampling of such edge cases to reduce the present 3.3: 96.7 tree–background imbalance. Even so, the baseline network processes a 280 ha in under 30 minutes on a single RTX 4000 Ada GPU and integrates seamlessly into the ArcGIS Pro workflow already used by state agencies, making it an immediately deployable tool for auditing greening targets, locating planting deficits, and refining urban heat-exposure models. This directly addresses the implementation gap between academic models and operational municipal workflows, providing a validated tool ready for practical use.

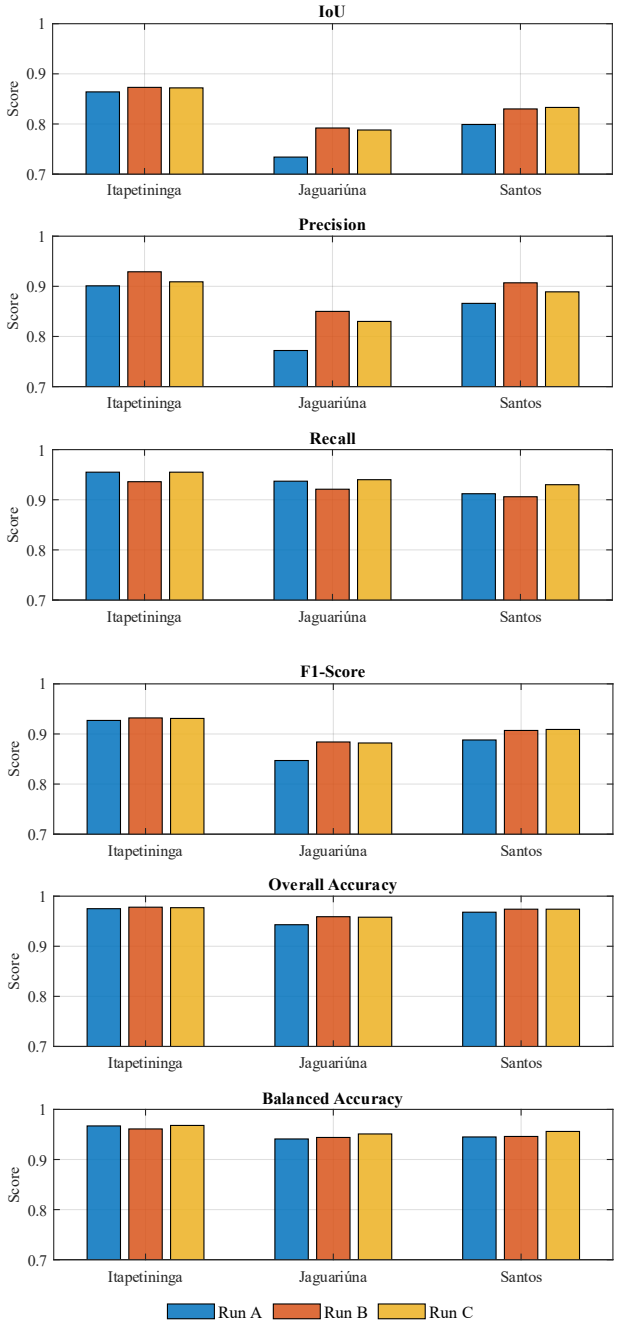

Figure 3. Comparison Metrics of Runs A, B, and C for the three test cities.

5. Conclusion

This This study presented a deep learning workflow that successfully segments urban tree cover across São Paulo, Brazil.

More importantly, it contributes a regional performance benchmark and a direct comparative analysis of three model configurations, offering clear methodological guidance. The baseline Balanced Cross-Entropy proved most robust, while the addition of a Dice loss term provided the best-balanced accuracy, demonstrating a practical path to optimizing class balance without introducing boundary errors.

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-W4-2025-219-2026 | © Author(s) 2026. CC BY 4.0 License.

Future work will advance this research in several directions. We will explore the integration of near-infrared data or LiDAR-derived canopy height information to address remaining classification ambiguities.

To further mitigate the significant class imbalance in the training data, we will also investigate sampling strategies, such as exclusively training on image patches that contain a minimum percentage of tree canopy pixels.
Furthermore, we plan to test the model's generalizability across more diverse geographic biomes in Brazil to produce a national-scale open dataset of tree canopy annotations, aiming to foster broader research and enhance the value of this work for the scientific community.

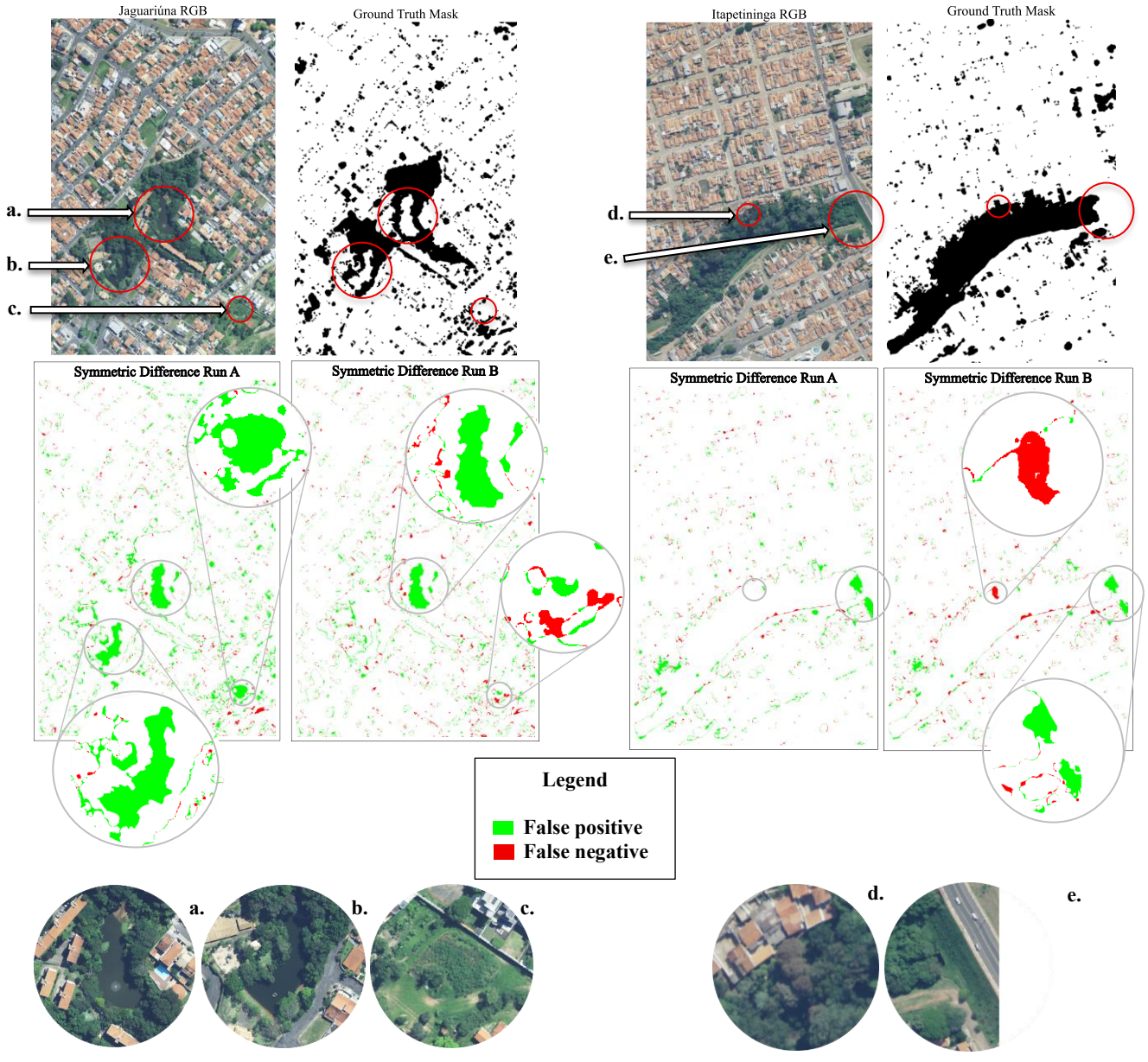

Figure 4. Qualitative comparison of prediction errors for the PointRend model (Run A) and the baseline (Run B) in Jaguariúna and Itapetininga. Run C is omitted from this visual comparison as its error patterns were nearly identical to Run B, a finding supported by the quantitative metrics in Figure 3. Green indicates false positives (commission errors) and red indicates false negatives (omission errors).

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-W4-2025-219-2026 | © Author(s) 2026. CC BY 4.0 License.

6. References

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A.L., 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint, arXiv:1706.05587. https://doi.org/10.48550/arXiv.1706.05587.

Esri, 2024a. How DeepLabV3 works. ArcGIS Developers documentation. Available at: https://developers.arcgis.com/python/latest/guide/how-deeplabv3-works/ (accessed 6 Jun 2025).

Esri, 2024b. Deep Learning Frameworks for ArcGIS Pro 3.5 [software environment]. GitHub repository. Available at: https://github.com/Esri/deep-learning-frameworks (accessed 6 Jun 2025).

Guo, J., Xu, Q., Zeng, Y., Liu, Z. & Zhu, X.X., 2023. Nationwide urban tree canopy mapping and coverage assessment in Brazil from high-resolution remote-sensing images using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing, 198, 1–15. https://doi.org/10.1016/j.isprsjprs.2023.02.007

Jones, B.A., 2021. Planting urban trees to improve quality of life? The life-satisfaction impacts of urban afforestation. Forest Policy and Economics, 125, 102408. https://doi.org/10.1016/j.forpol.2021.102408

Locke, D.H., Grove, J.M. & Pickett, S.T.A., 2023. The role of urban tree canopies in environmental justice and conserving biodiversity. In: Lambert, M.R. & Schell, C.J. (eds) Urban Biodiversity and Equity. Oxford University Press, pp. –. https://doi.org/10.1093/oso/9780198877271.003.0006

Morgan, G.R., Zlotnick, D., North, L., Smith, C. & Stevenson, L., 2024. Deep learning for urban tree-canopy coverage analysis: A comparison and case study. Geomatics, 4, 412–432. https://doi.org/10.3390/geomatics4040022

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

Wang, Y., Yang, L., Liu, X. et al., 2024. An improved semantic segmentation algorithm for high-resolution remote-sensing images based on DeepLabV3+. Scientific Reports, 14, 9716. https://doi.org/10.1038/s41598-024-60375-1

Weinstein, B.G., Marconi, S., Bohlman, S.A., Zare, A. & White, E.P., 2020. Cross-site learning in deep-learning RGB tree-crown detection. Ecological Informatics, 56, 101061. https://doi.org/10.1016/j.ecoinf.2020.101061.