Evaluating Mask R-CNN for instance segmentation of ceramic roofs in a Brazilian urban area using UAV imagery

Marcelo Henrique de Oliveira[1], Danilo Marques de Magalhães[1], Antonio Vieira dos Santos Júnior[1], Diandra dos Santos Moura[1], Tomás Lyra Lutterbach[1], Vitória Diniz de Oliveira[1], Luisa da Cunha Vieira[2]

[1] Dept. of Geography and Environmental Planning, São Paulo State University, Rio Claro, Brazil – marcelo.h.oliveira@unesp.br, danilo.magalhaes@unesp.br, antonio.v.santos@unesp.br, diandra.moura@unesp.br, tomas.lutterbach@unesp.br, vitoria.d.oliveira@unesp.br
[2] Geotechnical Project Management, BVP Geotecnia e Hidrotecnia, Belo Horizonte, Brazil - luisa.vieira@bvp.eng.br

Keywords: Mask R-CNN, Deep Learning, Instance Segmentation, Ceramic Rooftops, UAV, ArcGIS Pro.

Abstract
The performance of the Mask R-CNN model for instance segmentation of ceramic rooftops was evaluated using a high-resolution orthomosaic generated from UAV-based photogrammetry. Model training and inference were performed in ArcGIS Pro 3.5.3 with a ResNet-50 backbone. The model demonstrated high detection reliability, achieving a Precision of 96.62%, a Recall of 78.81%, and an F1-score of 86.81% at an Intersection over Union (IoU) threshold of 0.5. Most omission errors were associated with light-colored, elongated rooftops, highlighting limitations in the representativeness of the training sample and morphological variability. Fragmentation of larger rooftops into multiple segments was also observed, which affected accuracy metrics. To address this, a topological post-processing step was implemented to merge overlapping polygons, thereby improving segmentation consistency. These results indicate that Mask R-CNN is effective for high-resolution rooftop mapping, especially in applications requiring high precision. The approach is operationally feasible and transferable to similar datasets, enabling scalable analyses. It serves as a complementary tool for urban mapping, supporting the monitoring of urban dynamics and the analysis of construction patterns related to building standards and socioeconomic conditions.

1. Introduction

Deep Learning (DL), a subfield of Artificial Intelligence (AI), utilizes deep neural networks to learn hierarchical representations directly from data. This approach contrasts with traditional Machine Learning methods that depend on manual feature engineering (Chollet, 2021). The rapid progress in DL has been facilitated by the availability of large datasets and advancements in computational resources, particularly Graphics Processing Units (GPUs), which support efficient large-scale data processing (Deng et al., 2009; Russakovsky et al., 2015; Ponti and Costa, 2017; Vali et al., 2020).

Convolutional Neural Networks (CNNs) are among the most widely used DL architectures in remote sensing. They are designed to extract spatial features through convolutional operations, pooling, and fully connected layers (Azarang and Kehtarnavaz, 2021; Chamma et al., 2021; Li et al., 2018). Convolutional layers utilize trainable kernels to process input feature maps, producing increasingly abstract representations as network depth increases. This hierarchical feature extraction allows CNNs to capture both low-level patterns, such as edges and textures, and high-level structures. As a result, CNNs are particularly effective in complex urban environments characterized by high spectral and geometric variability.

CNN-based methods have been successfully applied to a range of remote sensing tasks, such as land use and land cover classification, object detection, and image enhancement (He et al., 2014; Girshick, 2015; Girshick et al., 2013; Dong et al., 2016). These methods demonstrate strong performance across diverse data types and spatial scales.

Mask R-CNN, which extends the Faster R-CNN framework, is a CNN-based model developed for object detection and instance segmentation. Beyond predicting bounding boxes, it produces pixel-level masks for each detected object using a dedicated segmentation branch (Hemanth, 2020; He et al., 2017). The architecture integrates a convolutional backbone, a Region Proposal Network (RPN), and task-specific heads for classification, regression, and segmentation (Braga et al., 2020).

Mask R-CNN has demonstrated strong performance in remote sensing applications, including vegetation mapping, building extraction, and cloud detection. It frequently outperforms traditional approaches such as Support Vector Machines (SVM) and Random Forests (RF) (Song et al., 2020; Guirado et al., 2021; Wang et al., 2022). The model is adaptable to various spatial resolutions and imaging modalities (Carvalho et al., 2021). It has been successfully applied in UAV-based studies, urban analysis (Magalhães and Souza, 2025; Wu et al., 2020; Zhao et al., 2018), and rooftop inspection tasks (Staffa et al., 2020). Similarly, deep learning models such as U-Net have demonstrated superior performance compared with traditional methods in rooftop classification (Chamma et al., 2021).

Despite its high accuracy, Mask R-CNN faces several challenges. These include the need for extensive labeled datasets, significant computational demands, and difficulty distinguishing spectrally similar classes.

This differentiates it from approaches such as U-Net, which perform semantic segmentation by assigning class labels at the pixel level without explicitly separating individual object instances (Ronneberger et al., 2015). It also differs from object detection methods such as YOLO, which identify and localize individual objects using bounding boxes but do not provide pixel-level delineation of object boundaries (Redmon et al., 2016; Redmon and Farhadi, 2018).

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-XI-3-2026-69-2026 | © Author(s) 2026. CC BY 4.0 License.

In addition, it contrasts with Object-Based Image Analysis (OBIA), which segments imagery into objects based on spectral and spatial criteria; however, these segments do not necessarily correspond to real-world object instances and may be affected by over- or under-segmentation (Blaschke, 2010).

In Brazil, the diversity of anthropogenic land uses, as evidenced by the range of construction materials and their links to various socioeconomic strata, presents significant challenges for public administration in mapping territorial characteristics. These challenges are further exacerbated in regions experiencing informal or unregulated urban expansion, often driven by migration dynamics that increase housing demand and result in irregular settlements outside formal planning frameworks. This process contributes to the cadastral deficit, defined as the gap between the formal city, consisting of officially registered properties, and the real city, encompassing informal and unregistered construction (Santos et al., 2015).

To address these mapping challenges, remote sensing-based feature extraction combined with geospatial data processing has become increasingly valuable at the municipal scale. This integration improves the monitoring of spatial dynamics and supports the formulation of urban and environmental policies (Santos et al., 2015). Furthermore, incorporating Geographic Information Systems (GIS) and cadastral updating processes enhances territorial management by increasing data consistency and spatial coverage (Leite et al., 2018). Given these benefits, the application of such approaches is especially relevant in complex urban environments.

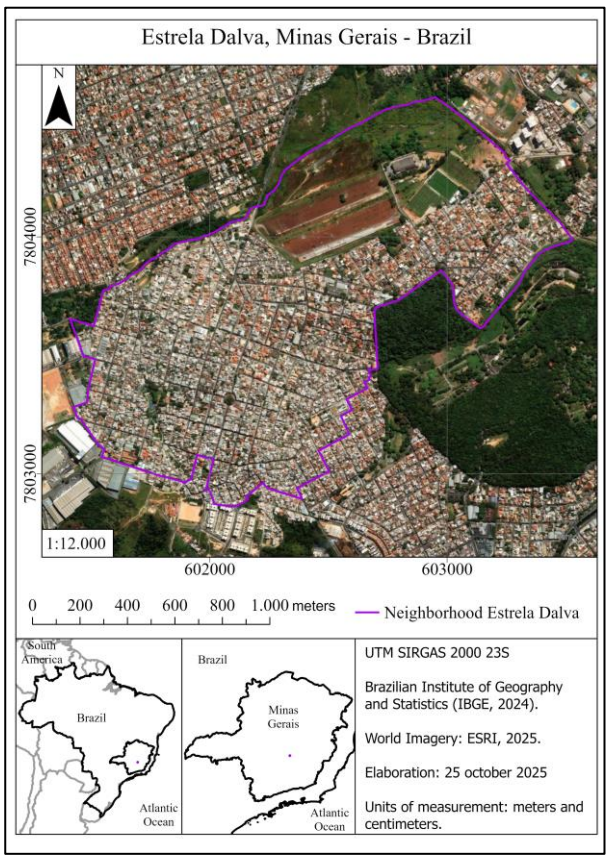

Figure 1. Location of study area.

Building on these methodological advances, this study evaluates the performance of the Mask R-CNN model for instance segmentation of ceramic rooftops using a UAV-derived orthomosaic. The results provide a foundation for assessing the applicability of this approach to very high-resolution imagery and demonstrate its potential to inform urban analysis and territorial planning.

The research was conducted in the Estrela Dalva neighborhood of Contagem, Minas Gerais, Brazil (Figure 1), selected for its heterogeneous urban environment characterized by contrasting building typologies and socioeconomic conditions. In this area, regular urban structures coexist with informal settlements, creating a complex landscape with substantial variability in shape, materials, and spectral response—factors that complicate feature detection and classification.

2. Materials and Methods

The orthomosaic used in this study was generated by Magalhães (2021) from an aerial photogrammetric survey conducted with a DJI Phantom 4 Pro UAV, which features a 1-inch CMOS sensor, 20 MP resolution, and an 84-degree field-of-view lens. Implementation of the Terrain Aware functionality maintained a constant flight altitude relative to the terrain, thereby ensuring uniform scale and high geometric fidelity in the final product.

Following orthomosaic acquisition, ArcGIS Pro was used for manual labeling, model training, and deep-learning inference. A vector dataset was constructed to delineate ceramic-tiled rooftops within the study area and served as the labeled dataset for supervised learning. Very small structures, including entrance canopies and minor roof extensions, were excluded to maintain consistency in object representation.

A total of 3,476 ceramic rooftop samples were labeled throughout the complete orthomosaic, extending beyond the study area boundaries. Model training utilized this expanded dataset to enhance sample diversity and improve generalization. The orthomosaic was then cropped to the Estrela Dalva neighborhood, where 1,921 ceramic rooftops were identified and used as ground truth for model evaluation.

The ceramic tile class was selected for its high visual discriminability, as demonstrated by distinct spectral and tonal contrasts compared to other roofing materials, such as fiber cement and metal. In Brazil, this roofing type also indicates higher construction standards, supporting its use as a proxy in exploratory socioeconomic analyses (Figure 2).

Model training was conducted using the Train Deep Learning Model tool in ArcGIS Pro, employing the Mask R-CNN architecture with a ResNet-50 backbone. Training was set for 20 epochs and concluded at the 18th epoch when validation loss stabilized. A learning rate of 0.1, a batch size of 4, and a chip size of 256 pixels were used. The dataset was divided into training and validation subsets, with 10% allocated for validation. Data augmentation followed the default ArcGIS Pro configuration, incorporating random cropping, dihedral affine transformations, brightness and contrast adjustments, and zoom operations to introduce geometric and radiometric variability and enhance model robustness (ESRI, 2025).

Object detection was performed using the Detect Objects Using Deep Learning tool in ArcGIS Pro, applying the trained model to identify ceramic rooftops. Inference parameters included a

https://doi.org/10.5194/isprs-annals-XI-3-2026-69-2026 | © Author(s) 2026. CC BY 4.0 License.

padding of 56 pixels, a confidence threshold of 0.9, and a tile size of 224 pixels. Model performance was assessed using Precision, Recall, and F1-score, with an Intersection over Union (IoU) threshold of 0.5.

Figure 2. Example of ceramic rooftop labeling, illustrating the variation in building patterns observed throughout the study area.

During processing, limitations emerged in segmenting larger rooftops, as the model produced multiple adjacent, partially overlapping polygons for the same object. While these segments were spatially aligned with the rooftops, their smaller individual areas adversely impacted accuracy metrics under the adopted IoU threshold.

To resolve this issue, a post-processing step involving topological correction was implemented to merge adjacent and overlapping polygons into a single feature. This approach improved segmentation consistency and reduced the total number of polygons from 2,252 to 1,567, thereby eliminating 685 redundant segments.

All experiments were conducted on a workstation featuring an Intel Core i7-12700F processor, an NVIDIA GeForce RTX 3060 GPU, and 32 GB of RAM. Model training took approximately two hours, and inference took about seven minutes. The trained model is applicable to other datasets with similar spectral and spatial characteristics, supporting its transferability to comparable urban environments.

3. Results

The model achieved 1,514 True Positives (TP), 53 False Positives (FP), and 407 False Negatives (FN), demonstrating strong performance in the instance segmentation of ceramic rooftops. The low incidence of false positives indicates effective discrimination between ceramic rooftops and visually similar non-target features, such as reddish pavement and exposed soil. Most errors were omission errors, corresponding to rooftops present in the image but not detected by the model. These missed detections were primarily associated with rooftops exhibiting lighter tones and with narrow or elongated structures, which were underrepresented in the training dataset (Figure 3). This outcome indicates that the model was more responsive to the dominant spectral and morphological characteristics of the labeled samples and less robust for less frequent rooftop configurations.

Figure 3. Rooftops predicted by the model. Lilac circles highlight light-colored or elongated rooftops that the model did not identify.

The accuracy metrics corroborate this pattern (Table 1). Precision reached 96.62%, indicating that most detected objects were true ceramic rooftops and confirming the reliability of the mapped outputs. Recall was 78.81%, demonstrating that approximately four out of five ceramic rooftops in the study area were successfully identified. The difference between Precision and Recall indicates that the main limitation of the model is not confusing other objects but missing some rooftops. The resulting F1-score of 86.81% reflects balanced overall performance, though still limited by omission errors. These findings indicate that further improvements in Recall may be achieved by developing a more representative training set, particularly by increasing the number of samples of light-colored, elongated, and small rooftop structures, and by refining inference parameters such as confidence threshold, tile size, and padding to enhance detection sensitivity and reduce missed objects.

Metric	Precision	Recall	F1-sc.	TP	FP	FN
Result	96.62%	78.81%	86.81%	1514	53	407

Table 1. Accuracy assessment metrics.

Topological correction was necessary to achieve a more accurate representation of model performance. Prior to post-processing, the model often generated multiple adjacent or partially overlapping segments across large rooftops, resulting in fragmented objects that should be represented as single entities (Figure 4). While these fragments were spatially consistent with the target rooftops, many did not meet the adopted Intersection over Union (IoU) threshold of 0.5, thereby artificially inflating the number of false positives in the accuracy assessment. Merging adjacent and overlapping polygons reduced geometric fragmentation and yielded a more coherent object-based result, aligning more closely with the intended instance segmentation task.

From an applied perspective, the high Precision achieved in this study demonstrates that the mapped rooftops can support spatial analyses requiring reliable identification of ceramic roofing patterns, particularly in heterogeneous urban environments.

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-XI-3-2026-69-2026 | © Author(s) 2026. CC BY 4.0 License.

Accordingly, this method is intended as a complementary mapping tool rather than a replacement for technical cadastral procedures. It can assist in territorial analysis, urban characterization, and the identification of spatial patterns associated with building standards. However, the moderate Recall indicates that the outputs should be interpreted with caution in applications requiring exhaustive inventories, as a portion of the rooftop stock remained undetected.

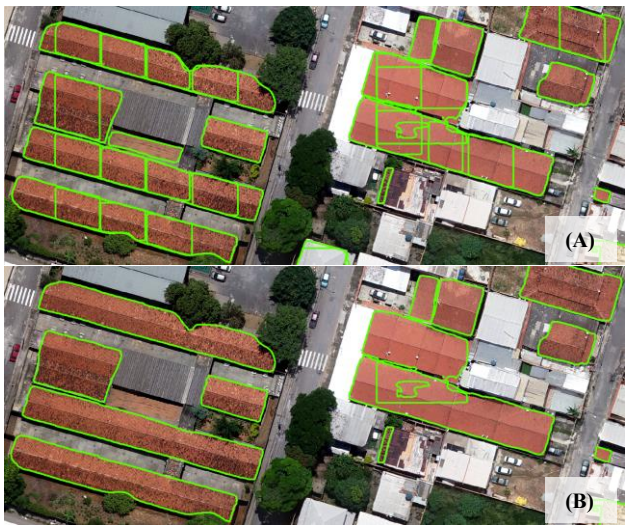

Figure 4. Topological adjustment of generated segments. (A) Segmentation produced by the model. (B) Output after applying topological correction.

This distinction is especially significant for cadastral and territorial management applications. Since roofing materials may be linked to building standards and, indirectly, to assessed property value and municipal taxation bases (Souza, 2017; Santos et al., 2015), a high-precision product is valuable for supporting screening, prioritization, and spatial updating routines. Nevertheless, the observed omission rate underscores the necessity for manual validation and integration with existing GIS and cadastral databases prior to operational deployment. In this context, deep learning-based mapping should be regarded as a tool to facilitate continuous cadastral updating and spatial diagnosis, reducing information gaps while maintaining technical oversight of the final dataset.

4. Conclusions

The performance of the Mask R-CNN model was evaluated for instance segmentation of ceramic rooftops in a UAV-derived orthomosaic. The model achieved high Precision (96.62%) and moderate Recall (78.81%), resulting in an F1-score of 86.81%. These results indicate reliable detection, although omission errors persist.

Most errors occurred on light-colored, elongated rooftops, highlighting limitations in the training dataset's representativeness and the model's sensitivity to spectral and morphological variability. Similar challenges have been documented in related studies, including Magalhães and Souza (2025) for tree detection and Chamma et al. (2021) for rooftop classification, both of which emphasize persistent difficulties in detecting small or morphologically complex objects. These findings underscore the need to increase sample diversity and refine training strategies to enhance model generalization.

The ResNet-50 backbone provided a suitable balance between performance and computational efficiency. Deeper architectures may yield incremental improvements in segmentation accuracy, particularly for complex object geometries. However, these potential gains must be balanced against increased computational costs, especially in operational settings.

Fragmentation observed in larger rooftops demonstrates that, despite accurate localization, instance segmentation models may require post-processing to ensure geometric consistency. Applying topological correction was essential for obtaining a more reliable representation of model performance and for reducing artificial inflation of error metrics under the IoU ≥ 0.5 criterion.

From an operational perspective, implementing Mask R-CNN within ArcGIS Pro demonstrates the accessibility of deep learning workflows in commercial geospatial platforms. This integration reduces technical barriers to adoption and facilitates the broader use of deep learning techniques in remote sensing applications. Additionally, UAV-derived orthomosaics proved highly effective for detailed urban mapping, suggesting that standardized acquisition protocols, such as orthomosaics with approximately 5 cm spatial resolution, can support consistent model performance. Once trained, the model can be applied to new datasets through transfer learning, enabling scalable spatial and temporal analyses.

These characteristics make the approach particularly suitable for territorial management applications. High-precision outputs support monitoring urban expansion and identifying construction patterns associated with building standards, which are indirectly related to thermal comfort and socioeconomic conditions. In the Brazilian context, where rapid and often informal urban transformation is prevalent, such tools offer a cost-effective means to support spatial diagnosis and updating processes, especially when integrated with existing GIS and cadastral systems.

However, the moderate Recall suggests that this method should serve as a complementary tool rather than a replacement for technical cadastral procedures. Integrating automated outputs with validation routines remains essential to ensure completeness and reliability.

Future research should prioritize improving sample representativeness, particularly for underrepresented rooftop types, and explore advanced post-processing and model refinement strategies to enhance segmentation accuracy in complex urban environments.

Acknowledgements

The authors acknowledge the support of the São Paulo Research Foundation (FAPESP), Brazil (Grant 2025/24168-5).

References

Azarang, A., Kehtarnavaz, N. 2021: *Image Fusion in Remote Sensing: Conventional and Deep Learning Approaches.* Synthesis Lectures on Image, Video, and Multimedia Processing, Vol. 10. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-031-02256-2.

Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004.

Braga, J.R.G., Peripato, V., Dalagnol, R., Ferreira, M.P., Tarabalka, Y., Aragão, L.E.O.C., Campos Velho, H.F., Shiguemori, E.H., Wagner, F.H. 2020. Tree Crown Delineation Algorithm Based on a Convolutional Neural Network. *Remote Sensing*, 12(8), 1288. https://doi.org/10.3390/rs12081288.

Carvalho, O.L.F., Carvalho Júnior, O.A., Albuquerque, A.O., Bem, P.P., Silva, C.R., Ferreira, P.H.G., Moura, R.D.S., Gomes, R.A.T., Guimarães, R.F., Borges, D.L., 2021. Instance Segmentation for Large, Multi-Channel Remote Sensing Imagery Using Mask-RCNN and a Mosaicking Approach. *Remote Sensing*, 13(1), 39. https://doi.org/10.3390/rs13010039.

Chamma, W.D.S., Batistella, D., Crisgiovanni, E.L., Victorino, H.S., Lima, V.A. Aprendizado de máquina aplicado em imagens de satélite para classificação de telhados. *Brazilian Journal of Development,* 7(7), 72558-72576, 2021. https://doi.org/10.34117/bjdv7n7-437.

Chollet, F. 2021. *Deep Learning with Python* (2nd ed.). Manning Publications.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

Dong, C., Chen, C. L., He, K., Tang, X. 2016. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 295–307. https://doi.org/10.1109/TPAMI.2015.2439281.

Esri, 2025. Train Deep Learning Model (Image Analyst). *ArcGIS Pro Tool Reference*. Available at: https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/train-deep-learning-model.htm

Girshick, R. 2015. Fast r-cnn. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1440-1448. https://doi.org/10.48550/arXiv.1504.08083.

Girshick, R., Donahue, J., Darrell, T., Malik, J. 2013. Rich feature hierarchies for accurate object detection and semantic segmentation. *Paper presented at Proceedings of the 20134 I.E. Conference on Computer Vision and Pattern Recognition, Portland, OR, USA*, 580–587. https://doi.org/10.48550/arXiv.1311.2524.

Guirado, E., Blanco-Sacristán, J., Rodríguez-Caballero, E., Tabik, S., Alcaraz-Segura, D., Martínez-Valderrama, J., Cabello, J., 2021. Mask R-CNN and OBIA Fusion Improves the Segmentation of Scattered Vegetation in Very High-Resolution Optical Sensors. *Sensors*, 21(1), 320. https://doi.org/10.3390/s21010320.

He, K., Gkioxari, G., Dollár, P., Girshick, R. 2017. Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2980-2988. https://doi.org/10.1109/ICCV.2017.322.

He, K., Zhang, X., Ren, S., Sun, J. 2014. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824.

Hemanth, D.J. (Ed.). 2020. *Artificial intelligence techniques for satellite image analysis* (1st ed.). Springer Cham. https://doi.org/10.1007/978-3-030-24178-0.

Leite, M.E., Rodrigues, H.L.A., Borges, M.G., 2018. Atualização do cadastro imobiliário por sensoriamento remoto e os impactos fiscais. *InterEspaço: Revista de Geografia e Interdisciplinaridade*, 4(13), 07-25. https://doi.org/10.18764/2446-6549.v4n13p07-25.

Li, Y., Zhang, H., Xue, X., Jiang, Y., Shen, Q., 2018. Deep learning for remote sensing image classification: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(6), e1264. https://doi.org/10.1002/widm.1264.

Magalhães, D.M., 2021. *Use of drones to support territorial planning: from data collection to geovisualization*. PhD thesis, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil. https://hdl.handle.net/1843/36455.

Magalhães, D.M., Souza, J.P., 2025. Detection of individual trees in drone imagery using deep learning: a case study in Belo Horizonte (MG). In: *XXI Brazilian Symp. Remote Sens*., Salvador, Brazil, 2025. https://proceedings.science/sbsr-2025/trabalhos/deteccao-de-individuos-arboreos-em-imagens-de-drone-usando-deep-learning-estudo?lang=pt-br (27 March 2026).

Ponti, M.A., Costa, G.B.P., 2017. Como funciona o Deep Learning. In: *Tópicos em Gerenciamento de Dados e Informações*. Sociedade Brasileira de Computação (SBC), Porto Alegre, 31 pp.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (CVPR), 779–788. https://doi.org/10.1109/CVPR.2016.91.

Redmon, J., Farhadi, A., 2018. YOLOv3: An Incremental Improvement. *arXiv preprint*. https://doi.org/10.48550/arXiv.1804.02767.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (MICCAI), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.*, 115(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Santos, T., Pelegrina, M., Julião, R.P., 2015. Atualização cadastral dirigida, utilizando imagens de satélite de alta resolução espacial. *Revista Brasileira de Cartografia*, 67(2), 435-444. https://doi.org/10.14393/rbcv67n2-44671.

Song, S., Liu, J., Liu, Y., Feng, G., Han, H., Yao, Y., Du, M., 2020. Intelligent Object Recognition of Urban Water Bodies via Mask R-CNN. *Sensors*, 20(2), 397. https://doi.org/10.3390/s20020397.

Souza, A.C.P., 2017. *Avaliação espacial imobiliária e suas implicações sobre a arrecadação do IPTU: um estudo de caso no município de Contagem/MG*. Master’s Thesis, Universidade Federal de Viçosa, Florestal, Brazil. http://www.locus.ufv.br/handle/123456789/20698.

Staffa, L.B.J., Sá, L.S.V., Lima, M.I.S.C., Costa, D.B., 2020. USO DE TÉCNICAS DE PROCESSAMENTO DE IMAGEM PARA INSPEÇÃO DE ESTRUTURAS DE TELHADOS DE EDIFICAÇÕES PARA FINS DE ASSISTÊNCIA TÉCNICA. ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO, 18(1), 1–8. https://doi.org/10.46421/entac.v18i.1178.

Vali, A., Comai, S., Matteucci, M., 2020. Deep learning for land use and land cover classification based on hyperspectral and multispectral Earth observation data: A review. *Remote Sens.*, 12(15), 2495. https://doi.org/10.3390/rs12152495.

Wang, Y., Li, S., Teng, F., Lin, Y., Wang, M., Cai, H., 2022. Improved Mask R-CNN for Rural Building Roof Type Recognition from UAV High-Resolution Images: A Case Study in Hunan Province, China. *Remote Sensing*, 14(2), 265. https://doi.org/10.3390/rs14020265.

Wu, W., Gao, X., Fan, J., Xia, L., Luo, J., Zhou, Y. 2020. Improved mask R-CNN-based cloud masking method for remote sensing images. *International Journal of Remote Sensing*, 41(23), 8910-8933. https://doi.org/10.1080/01431161.2020.1792576.

Zhao, K., Kang, J., Jung, J., Sohn, G. 2018. Building Extraction from Satellite Images Using Mask R-CNN with Building Boundary Regularization. *2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* 247-251. https://doi.org/10.1109/CVPRW.2018.00045.