UNIVERSIDADE ESTADUAL PAULISTA - UNESP
Instituto de Biociências, Letras e Ciências Exatas- campus de São José do Rio Preto

GIOVANA AUGUSTA BENVENUTO

Deep Learning Applied to Deforestation Detection in Brazilian Biomes :
An Approach Based on Convolutional Neural Networks, Vision Transformers and Attention Mechanisms

São José do Rio Preto
2026

Giovana Augusta Benvenuto

Deep Learning Applied to Deforestation Detection in Brazilian Biomes :
An Approach Based on Convolutional Neural Networks, Vision Transformers and Attention Mechanisms

Tese, apresentada à Universidade Estadual Paulista (UNESP), Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, para obtenção do título de Doutora em Ciência da Computação.
Área de Concentração: Computação Aplicada

Orientador: Prof° Dr. Wallace Correa de Oliveira Casaca
Coorientador: Prof° Dr. Rogério Galante Negri

São José do Rio Preto
2026

B478d

Benvenuto, Giovana Augusta

Deep learning applied to deforestation detection in brazilian biomes : an approach based on convolutional neural networks, vision transformers and attention mechanisms / Giovana Augusta Benvenuto. -- São José do Rio Preto, 2026

141 p. : il., tabs., mapas

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Wallace Correa de Oliveira Casaca

Coorientador: Rogério Galante Negri

1. Visão Computacional. 2. Inteligência Artificial. 3. Aprendizagem Profunda. 4. Sensoriamento Remoto. 5. Desmatamento. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

IMPACTO POTENCIAL DESTA PESQUISA

A pesquisa desenvolvida neste trabalho apresenta impacto científico e tecnológico, ao propor novos modelos de aprendizado profundo para a segmentação de áreas de desmatamento. Os modelos desenvolvidos foram aplicados a biomas nacionais e demonstraram vantagens para aplicações em tempo real, alcançando resultados competitivos com menor demanda de recursos computacionais. Dessa forma, as abordagens propostas podem contribuir para o monitoramento de áreas de interesse de maneira viável e com menor custo computacional, apoiando iniciativas de acompanhamento e gestão ambiental.

POTENTIAL IMPACT OF THIS RESEARCH

The research developed in this work presents both scientific and technological impact by proposing new deep learning models for the segmentation of deforested areas. The developed models were applied to national biomes and demonstrated advantages for real-time applications, achieving competitive results while requiring lower computational resources. Thus, the proposed approaches can contribute to the monitoring of areas of interest in a feasible and cost-effective manner, supporting environmental monitoring and management initiatives.

GIOVANA AUGUSTA BENVENUTO

DEEP LEARNING APPLIED TO DEFORESTATION DETECTION IN BRAZILIAN BIOMES :
An Approach Based on Convolutional Neural Networks, Vision Transformers and Attention Mechanisms

Tese apresentado(a) à Universidade Estadual Paulista (UNESP), Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, para obtenção do título de Doutora em Ciência da Computação.

Área de Concentração: Computação Aplicada

Data de defesa: 08/04/2026

BANCA EXAMINADORA

____

Profº Dr. Wallace Correa de Oliveira Casaca
UNESP – Instituto de Biociências, Letras e Ciências Exatas – Campus de São José do Rio Preto

Profº Dr. Thales Sehn Korting
Instituto Nacional de Pesquisas Espaciais

Profº Dr. Michael Macedo Diniz
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo

Profº Dr. Danilo Medeiros Eler
Faculdade de Ciências e Tecnologia

Profª Dr. Rogéria Cristiane Gratão de Souza
Instituto de Biociências, Letras e Ciências Exatas

Dedico este trabalho à minha família

ACKNOWLEDGEMENTS

Agradeço, primeiramente a Deus por tudo.

Aos meus pais, Angela e Luiz, pela compreensão com meus horários trocados, pelo apoio e preocupação com meus avanços, pela criação que me permitiu sonhar em ser mais, e principalmente pelo amor incondicional que nunca me deixou duvidar de mim. Ao Loki, que é parte dessa família, nos trazendo sempre alegria.

Aos professores Wallace, Rogério e Marilaine pela orientação constante, compreensão e humanidade ao longo dessa formação. Não apenas pela inspiração profissional que proporcionam, mas principalmente pelo exemplo de pessoas conscientes, solidárias e acolhedoras que são.

Agradeço também a todas as integrantes do projeto de extensão GECET (Garotas nas Engenharias, Ciências Exatas e Tecnologias) pelo incentivo e constante luta para divulgar e defender a atuação feminina nas áreas de exatas.

Aos meus amigos que sempre estiveram ao meu lado, me escutando, me acalmando, rindo comigo, confiando em mim e na minha capacidade, e me mostrando isso quando eu não conseguia enxergar.

À UNESP pela formação acadêmica e por proporcionar um ambiente de aprendizado, inclusão e troca de experiências.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Por fim, agradeço imensamente a todos que de alguma forma contribuíram para tornar esse sonho possível.

“...mas aquele que perseverar até o fim, será salvo.“
São Mateus 10, 22

RESUMO

O Brasil abriga dois dos biomas mais biodiversos do planeta: a Amazônia e a Mata Atlântica. Apesar de sua importância ecológica e climática, essas regiões têm sido severamente impactadas nas últimas décadas pelo avanço do desmatamento, majoritariamente associado à ação humana. Nesse contexto, o uso de tecnologias de Sensoriamento Remoto (SR) aliado a métodos avançados de Inteligência Artificial (IA) tem se mostrado uma estratégia promissora para ampliar a capacidade de monitoramento ambiental em larga escala. Esta tese investiga o uso de técnicas de Aprendizado Profundo para a segmentação automática de áreas desmatadas nesses biomas. Para isso, propõe-se uma arquitetura híbrida que combina redes neurais convolucionais com módulos baseados em *transformers* e mecanismos de atenção, explorando simultaneamente características espaciais e temporais extraídas de imagens de satélite multiespectrais. A metodologia considera dados bi-temporais, permitindo capturar mudanças na cobertura vegetal e aprimorar a identificação de padrões associados ao desmatamento. Os modelos propostos foram avaliados por meio de uma extensa bateria experimental, sendo comparados a treze modelos consolidados da literatura em diferentes cenários experimentais. Os resultados demonstram que as abordagens desenvolvidas alcançam desempenho competitivo, apresentando o melhor equilíbrio entre acurácia de segmentação e eficiência computacional. Além das contribuições científicas, esta pesquisa também visa fomentar o desenvolvimento de soluções tecnológicas aplicáveis ao monitoramento ambiental, contribuindo para o avanço do uso de IA em aplicações de SR no Brasil.

Palavras-Chave: visão computacional; inteligência artificial; aprendizagem profunda; sensoriamento remoto; desmatamento.

ABSTRACT

Brazil hosts two of the most biodiverse biomes on the planet: the Amazon and the Atlantic Forest. Despite their ecological and climatic importance, these regions have been severely impacted in recent decades by the advance of deforestation, largely associated with human activities. In this context, the use of Remote Sensing (RS) technologies combined with advanced Artificial Intelligence (AI) methods has emerged as a promising strategy to expand large-scale environmental monitoring capabilities. This doctoral thesis investigates the use of Deep Learning techniques for the automatic detection and segmentation of deforested areas in these biomes. To this end, a hybrid architecture is proposed that combines Convolutional Neural Networks with transformer-based modules and attention mechanisms, simultaneously exploring spatial and temporal features extracted from multispectral satellite imagery. The methodology considers bi-temporal data, enabling the capture of changes in vegetation cover and improving the identification of patterns associated with deforestation. The proposed model was evaluated through an extensive experimental framework and compared with thirteen well-established models from the literature across different experimental scenarios. The results demonstrate that the proposed approach achieves competitive performance, presenting the best trade-off between segmentation accuracy and computational efficiency. In addition to its scientific contributions, this research also aims to foster the development of technological solutions applicable to environmental monitoring, contributing to the advancement of AI-based RS applications in Brazil.

Keywords: computational vision; artificial intelligence; deep learning; remote sensing; deforestation.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AI Artificial Intelligence.

C2f Cross-Stage Partial Network Fusion.

C3k2 Cross Stage Partial with kernel size 2.

CE Cross-Entropy.

CNN Convolutional Neural Network.

DL Deep Learning.

DNN Deep Neural Networks.

EF Early Fusion.

FCN Fully Convolutional Networks.

FFNs Feed-Forward Networks.

FLOPs Floating Point Operations.

FN True Negatives.

FP False Positives.

FPS Frames Per Second.

GEE Google Earth Engine.

GELAN Generalized Efficient Layer Aggregation Network.

GT Ground-Truth.

IoU Intersection over Union.

LF Late Fusion.

LSTM Long Short-Term Memory.

mAP mean Average Precision.

MHSA Multi-Head Self-Attention.

ML Machine Learning.

MLP Multilayer Perceptron.

MPA Mean Pixel Accuracy.

MSI Multispectral Instrument.

NIR Near-Infrared.

NLP Natural Language Processing.

OLI Operational Land Imager.

PA Pixel Accuracy.

PGI Programmable Gradient Information.

PRODES Programa de Monitoramento da Floresta Amazônica Brasileira por Satélite.

ReLU Rectified Linear Unit.

RS Remote Sensing.

SGD Stochastic Gradient Descent.

SiLU Sigmoid Linear Unit.

SN Siamese Network.

SPPF Spatial Pyramid Pooling Fast.

SWIR Shortwave Infrared.

TIRS Thermal Infrared Sensor.

TN False Negatives.

TP True Positives.

UAV Unmanned Aerial Vehicles.

ViT Vision Transformers.

WCE Weighted Cross-Entropy.

YOLO You Only Look Once.

CONTENTS

1 INTRODUCTION

Brazilian territory, the fifth largest in the world, encompasses vast regions characterized by diverse climatic and geological conditions. These areas are classified into six major biomes, Amazon, Atlantic Forest, Cerrado, Caatinga, Pampa, and Pantanal, each exhibiting remarkable diversity in vegetation and wildlife (Ellwanger; Nobre; Chies, 2023). This biological richness plays a fundamental role in shaping Brazil's natural heritage and cultural diversity.

Although the country is home to an extraordinary abundance of living organisms and a wide range of ecosystems, Brazilian biomes have experienced recurrent environmental crises over the past decades. These include the expansion of deforestation and wildfires which, among several factors, are largely driven by anthropogenic activities (Nobre et al., 2016; Pivello et al., 2021). Indeed, the impacts of deforestation and wildfires cause irreversible damage not only to Brazil's environment but also indirectly affect other countries worldwide, as well as the global economy itself.

For instance, to better contextualize the magnitude of the problem, data released by the Federal Government Portal (Secretaria de Comunicação Social – SECOM, 2025) indicate a $32.4\,\%$ reduction in deforested areas in Brazil between 2023 and 2024. The total area cleared in the Amazon biome remained substantial, reaching $377\,708\,\mathrm{ha}$, equivalent to approximately $528\,000$ soccer fields. The Atlantic Forest, in the meantime, recorded $13\,472\,\mathrm{ha}$ of deforestation. Although this area is significantly smaller, it was the only biome to show an increase in deforestation compared to the previous years.

In light of the intensification of these critical events and considering global initiatives aimed at forest preservation, such as the United Nations Sustainable Development Goal 15 (SDG 15), one of the objectives established by the United Nations in the 2030 Agenda for Sustainable Development (Nações Unidas Brasil, 2025), as well as the 30th United Nations Climate Change Conference (COP30) (Presidência da República, Casa Civil, 2025) in 2025, it becomes evident that there is an urgent need to develop technological tools, as well as prevention and mitigation strategies, capable of supporting the characterization, detection and monitoring of risk areas and environmental incidents across Brazilian biomes.

Through continuous monitoring of the country's diverse ecosystems, it becomes possible to track the emergence and expansion of deforested areas, revealing potential illegal activities within protected regions. In this context, quick assessment of the case enables timely alerts to public authorities and other responsible agencies, allowing measures to be implemented to mitigate damage, protect lives, and mobilize health and inspection teams in the field (Luemba; Kunzayila; Casaca, 2021).

To ensure the continuous monitoring of biomes, one of the key technologies that enables the multitemporal analysis of vast vegetated areas is Remote Sensing (RS) (Ban, 2016). Data acquired through different Earth observation sensors are generally georeferenced, when combined with appropriate computational techniques, allow for the identification of phenomena ranging from geophysical variables to environmental changes. These include mass movement rates (Negri et al., 2021), vegetation cover changes (Sapucci et al., 2021), land use and land cover dynamics (Negri; Silva; Casaca, 2018), landscape dynamics (Gino et al., 2023), flooded-area mapping (Negri et al., 2025), among others.

Once acquired, RS data can be combined and analyzed to, for example, characterize vegetation disturbance patterns (Rodrigues et al., 2024) or detect deforestation hotspots within biome areas. However, due to the nature of these images, which typically exhibit high dimensionality and complex patterns that are difficult to categorize, the adoption of advanced computational approaches becomes necessary. In this context, robust Pattern Recognition and Computer Vision techniques play a crucial role, being fundamental for the effective processing and interpretation of such data in environmental monitoring applications (Basso et al., 2021).

Given the challenges associated with processing large collections of RS imagery, including the joint manipulation of spatial, temporal, and spectral information with high levels of precision and detail (Yuan et al., 2020), Deep Learning (DL) models have gained increasing prominence in this field. This trend is largely supported by the massive volume of data generated by modern sensors, as well as by the flexibility of multilayer architectures employed in DL. Such architectures are capable of accurately approximating the complex nonlinear relationships among environmental, geometric, and temporal parameters, thereby enabling more robust modeling of intricate Earth observation phenomena (LeCun; Bengio; Hinton, 2015; Yuan et al., 2020).

Two DL architectures have been particularly prominent in this context: Convolutional Neural Network (CNN) and Transformer-based models (LeCun; Bengio; Hinton, 2015; Oubadriss; Laassiri; Makrani, 2024). CNNs have been widely adopted in RS applications, especially for semantic segmentation and change detection tasks, due to their ability to learn hierarchical spatial features from large volumes of annotated imagery (Yuan et al., 2020). More recently, the Transformer architecture, originally proposed for natural language processing tasks (Vaswani et al., 2017), has shown a remarkable ability to capture global contextual relationships within data and has been adapted to Computer Vision through the development of Vision Transformers (ViT) (Dosovitskiy et al., 2021). Despite the growing attention these models have received, their use in applications such as deforestation monitoring remains relatively limited (Benvenuto et al., 2025).

Considering the aforementioned challenges, this doctoral research aimed to explore and develop new methodologies based on the DL paradigm for multitemporal deforestation pattern detection, performing semantic segmentation, in biomes such as the Amazon Rainforest and the Atlantic Forest. Here, special attention was given to modeling temporal dependencies in RS data so that deforestation signals could be captured consistently over time, even in the presence of noise, cloud contamination and acquisition gaps. Beyond accurately identifying targets with temporal consistency, that is, preserving the historical coherence of the scene across frames captured at different time points, the proposed methodologies were designed to enable automatic target classification. As a result, the developed approaches constitute supervised methods capable of delivering accurate and temporally coherent deforestation mapping.

From a more technical perspective, we developed hybrid models that combine the well-established performance of CNNs with spatial attention mechanisms, and later with the representation and learning capabilities of ViT architectures. Indeed, our proposed designs are further enhanced by attention mechanisms and a composite loss function to effectively address the problem. More specifically, the first model adapts layers derived from the eighth version of the You Only Look Once (YOLO) architecture to the backbone of the network, while the second integrates these layers with a ViT module, followed by a convolutional upscaling strategy. For both, Spatial Attention mechanisms are applied at two distinct stages of the network, and skip connections are established between the downsampling and upsampling paths to preserve multiscale feature information. Finally, the learning scheme is guided by a composite loss function that combines Focal Loss and Dice Loss. This formulation is particularly suitable for handling class imbalance and contributes to improved model refinement and segmentation accuracy.

The proposed models were evaluated on three datasets: two single-temporal datasets from the Amazon Rainforest and the Atlantic Forest, and a third bi-temporal dataset from the city of Altamira, located in the Amazon biome. The third dataset is also an original product of this research, designed to represent real-world scenarios with imbalanced data and to incorporate temporal aspects, thereby enriching the evaluation set. The experimental setup involved a comparative analysis with a total of thirteen methods, considering qualitative, quantitative, and efficiency perspectives. Among the variations of the proposed architectures, the lightest version of the ViT-based model stood out, providing the best trade-off between segmentation accuracy and computational efficiency. These results indicate that the proposed approaches are suitable for application in real-world deforestation scenarios, as they achieve high accuracy even when trained on limited datasets. Their lightweight architectures enable fast training and inference, and they can be applied both to detect deforestation in a single period and to assess increases in deforestation over time. This supports continuous observation and monitoring of conservation areas, while enabling the early detection of deforestation at both small and large scales.

1.1 JUSTIFICATION AND OBJECTIVES

Over the past few years, DL has become a key resource for mapping deforested areas and other landscape disturbances. In this context, progress has been driven mainly by increasingly sophisticated CNN architectures and by the refinement of classic Machine Learning (ML) strategies for large-scale RS data.

Although recent literature reports techniques with relative effectiveness in specific supervised contexts, there are still numerous gaps. In particular, many approaches do not explore bi-temporal inputs, with results restricted to a single image, fail to discuss the computational cost associated with the proposed models, and do not test them in different domains, varying regions and vegetation (Benvenuto et al., 2025). Furthermore, although many studies have adopted DL algorithms to address different RS applications in recent years, newer techniques coupled with these networks, such as attention mechanisms and transformer-based modules, remain relatively underexplored. In this sense, the application of DL to these problems is still at the frontier of knowledge.

The need to develop innovative technological solutions that combine accuracy, efficiency, and automation for the classification and monitoring of environmental incidents in protected forest areas in Brazil is another central motivation for this research. Such solutions may create tangible opportunities for integration between academia and industry.

Therefore, in light of the previously discussed aspects, the identified motivations and the gaps observed in the literature, the main objective of this PhD research is to design and evaluate new DL-based methodologies for the multitemporal semantic segmentation of deforestation patterns in Brazilian biomes, with emphasis on the Amazon Rainforest and the Atlantic Forest. To accomplish this, the following specific objectives were established:

a) The investigation of classical and recent DL algorithms and techniques in the context of RS for deforestation detection and segmentation, analyzing their contributions and limitations.

b) The implementation, enhancement, and evaluation of different DL strategies and network architectures for the application addressed in this study.

c) The combination of DL-based algorithmic solutions in order to build more robust and accurate model for segment deforested areas.

d) The analysis of distinct data collections, including temporal series of multispectral images, as well as the investigation of preprocessing strategies required for their effective use within the implemented Artificial Intelligence (AI) computational frameworks.

e) The exploration and optimization of the training procedures for the developed models in order to improve their overall performance.

f) Lastly, an experimental evaluation of the impact of the developed computational model in practical RS applications, with a particular focus on the Brazilian forest biomes of the Amazon and the Atlantic Forest.

1.2 CONTRIBUTIONS

Considering the objectives established throughout this PhD thesis, the main contributions of this research can be summarized as follows:

- Addressing the first objective of this research, i.e., an extensive literature review was conducted to explore algorithms relevant to the problem under investigation. A critical analysis of the selected studies was also performed, while identifying representative models to be used as baselines in the comparative analysis.

- Embracing objectives b) and c), this work proposes two DL models. The first incorporates and adapts layers from a network widely known for real-time object detection, YOLOv8, integrating them into its encoder. The second replaces part of the YOLO architecture with a ViT module. Both models employ a U-shaped decoder guided by a spatial attention mechanism and skip connections. The proposed models are trained using a composite loss function that combines Dice Loss and Focal Loss. This combination encourages the models to achieve a high overlap with the reference data while reducing the impact of class imbalance.

- Considering objective d), a custom dataset was developed using images acquired by the Landsat-8 satellite with the support of the Google Earth Engine (GEE) platform, comprising bi-temporal data for monitoring the expansion of deforested areas. Within this objective, we also propose a preprocessing strategy for bi-temporal images, in which a fusion procedure based on logical operations is applied to generate maps that highlight areas of interest.

- Based on objective e), we conducted a comprehensive ablation and optimization study of the modules composing the proposed model, focusing in the ViT approach, aiming to evaluate the effectiveness of the adopted combinations and to establish a robust, consistent, and competitive architecture across different scenarios.

- Finally, according to the last objective, the proposed approaches was evaluated in three different scenarios and compared with well-established models from the literature. The results demonstrate its effectiveness across distinct experimental settings, outperforming competing approaches in terms of both accuracy and efficiency, while providing the best trade-off between segmentation quality and computational cost.

In summary, this research addresses the challenge of accurately and efficiently detecting deforestation using RS data and DL techniques, including the synergic combination of CNNs and ViTs. By combining convolutional architectures with transformer-based modules and attention mechanisms, the proposed approach seeks to improve the representation of spatial patterns associated with forest degradation while maintaining computational efficiency. Through extensive experimentation and comparative analyses, this work aims to contribute to the development of more robust and practical tools for environmental monitoring in Brazilian forest biomes.

In addition to these scientific and technological contributions, this doctoral research also resulted in a set of peer-reviewed publications, both directly related to the thesis and in broader, but complementary, topics.

The main publications directly related to this thesis are:

- (Benvenuto et al., 2025) "Precision Meets Speed: An Attention Encoder–Decoder Network for Deforestation Segmentation", *IEEE Geoscience and Remote Sensing Letters - Topic: 38h Conference on Graphics, Patterns and Images (SIBGRAPI 2025)* (**Qualis/CAPES: A1, Google h5-index:86**).
- (Benvenuto et al., 2025) "Assessing the Performance of Deep Learning Networks for Real-Time Deforestation Segmentation in the Amazon and Atlantic Biomes", *Apresentado em 52º Seminário Integrado de Software e Hardware (SEMISH)*, 2025 (**Qualis/CAPES: A4**)

Other publications produced during the PhD period, include:

- (Benvenuto; Colnago; Casaca, 2022) "Unsupervised Deep Learning Network for Deformable Fundus Image Registration", *47th IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP, 2022* (**Google h5-index: 137**).
- (Benvenuto et al., 2022) "A Fully Unsupervised Deep Learning Framework for Non-Rigid Fundus Image Registration", *Bioengineering*, 2022. (**Google h5-index: 71**).
- (Benvenuto; Casaca, 2023) "Retinal Images Registration via Unsupervised Deep Learning”, *Workshop de Teses e Dissertações - 36h Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2023. (**Qualis/CAPES: A3**).
- (Colnago et al., 2022) “Risk Factors Associated with Mortality in Hospitalized Patients with COVID-19 during the Omicron Wave in Brazil", *Bioengineering*, 2022. (**Google h5-index: 71**)
- (Benvenuto et al., 2024) "GECET: Garotas nas Engenharias, Ciências Exatas e Tecnologias: Projeto de Extensão voltado para a participação feminina nas áreas STEM" *Anais Estendidos do XXX Simpósio Brasileiro de Sistemas Multimídia e Web*, 2024. (**Qualis/CAPES: A3**).

1.3 TEXT ORGANIZATION

This manuscript is organized as follows. Chapter 2 presents the theoretical foundation of the study, introducing general concepts of RS, the fundamentals of DL, with emphasis on CNNs and Transformers, and the problem of semantic segmentation. Chapter 3 reviews related works that employ different DL strategies for detecting deforestation areas across various types of imagery, along with a comparative analysis of these approaches. Chapter 4 describes the proposed methodologies, highlighting its main components and the strategy adopted to address the targeted problem, as well as the evaluation frameworks. Chapter 5 presents the experimental development and discusses the results obtained from the multiple trained configurations. Finally, Chapter 6 summarizes the main findings of this study.

6 CONCLUSION

In this doctoral thesis, we addressed the problem of deforestation area segmentation, focusing on the Brazilian biomes of the Amazon and the Atlantic Forest. The literature review revealed that, although many studies have benefited from DL techniques, important gaps still remain, particularly regarding the use of single-temporal and bi-temporal inputs, as well as the evaluation of methods across different geographic regions and application scenarios. Furthermore, only a limited number of works systematically investigate the impact of computational complexity on model performance, especially in terms of scalability and practical deployment.

Therefore, this research contributes to scientific advancement by proposing two novel architectures. The first combines the efficiency of YOLOv8 with spatial attention mechanisms, while the second builds upon these initial findings and integrates them with the strong representation capability of ViT. Furthermore, the models are enhanced by an attention-based decoder employing spatial attention mechanisms, as well as a composite loss function that integrates the class-balancing properties of Focal Loss with the region-overlap optimization of Dice Loss.

Another important contribution of this work is the construction and organization of a dedicated dataset composed of bi-temporal image pairs from the Amazon region, specifically from the municipality of Altamira, using Landsat-8 satellite imagery acquired between 2016 and 2017. Furthermore, a temporal fusion strategy was proposed to enhance relevant changes between the image pairs through the application of logical operations, enabling the generation of maps that better highlight deforestation patterns. The dataset and the first developed model are publicly available on GitHub[1].

In addition, we introduced a new bi-temporal fusion strategy specifically designed to effectively handle pairs of multitemporal images. The proposed method undergone a comprehensive comparative evaluation across three distinct datasets and was benchmarked against thirteen well-established architectures from the literature, taking into account the particular characteristics of each model and dataset. A systematic ablation study was also conducted to identify the most effective configuration of the second proposed architecture, ensuring a well-founded and optimized final design.

Based on the conducted experiments, we conclude that the proposed models outperformed the state of the art, particularly in its ViT 128-variant configuration. This version, being the lightest, employs smaller patch dimensions in the Transformer module, resulting in a more efficient yet highly effective architecture. The comparative experiments demonstrated that the proposed method achieved improvements both quantitatively and qualitatively, showing strong adherence to ground truth boundaries and consistently attaining higher average accuracy values, especially in terms of F1-Score. Across all three evaluated datasets, the model delivered competitive or

[1] https://github.com/GiBenvenuto/Precision-Meets-Speed-Attention-Encoder-Decoder-Network-for-Deforestation-Segmentation

superior performance.

For Datasets 1 and 2, it outperformed classical architectures such as U-Net and ResUnet, as well as Transformer-based models like TransU-Net++ and attention-enhanced approaches such as Attention U-Net. For Dataset 3, the proposed method surpassed different YOLO variations and traditional architectures specifically designed for image pair analysis, such as Siamese-based networks. These results reinforce the robustness, adaptability, and effectiveness of the proposed hybrid framework across distinct scenarios and data characteristics.

Additionally, to achieving the highest accuracy results, the proposed 128 model stands out particularly in terms of computational efficiency, demonstrating remarkably fast performance. Even with the inclusion of Transformer modules, the architecture maintains a well-balanced number of parameters and GFLOPs. The lightest configuration requires less than 13 minutes for training, and, even more notably, achieves inference times ranging from 0.0034 to 0.0054 seconds, depending on the dataset. This represents a highly favorable trade-off between accuracy and efficiency, especially considering that competing methods with comparable accuracy often require training and inference times up to ten times longer.

The growing demand for sustainable and efficient models promotes the relevance of this work. In this context, the proposed methodologies contributes to the scientific community by demonstrating the effectiveness of hybrid architectures, specifically, the integration of highly optimized CNNs for real-time applications with Transformer-based modules. The results show that it is possible to achieve high accuracy while maintaining a balanced computational cost and reduced execution time, making such models suitable for real-world environmental monitoring tasks.

As highlighted previously, the proposed models achieved strong performance even when trained on limited datasets and without the use of data augmentation techniques. The hybrid ViT 128 model also demonstrated satisfactory results under domain shift conditions, reinforcing its generalization capability. In this way, the proposed solutions presents strong potential for operational and scalable deployment. Both can be applied in parallel processing frameworks to efficiently cover large territorial areas, enabling fast identification of deforestation hotspots and supporting timely decision-making processes.

The proposed models are based on a supervised learning approach and, as such, presents the inherent limitation of requiring labeled data for training. Although robust results even with limited data availability, the quality of the annotations directly influences the results. In particular, poorly constructed or imprecise GT data can induce the network to learn incorrect patterns, ultimately leading to prediction errors and penalizing overall performance. Therefore, careful dataset curation and high-quality reference labeling remain critical factors for achieving optimal results.

As future work, multiple improvements can be explored. One possible direction is to expand the experimental evaluation by testing the models on additional datasets and exploring different input compositions, particularly considering alternative spectral band combinations provided by satellite sensors. Or even to extended to related tasks, such as the detection of fire hotspots or active fire events, which share visual and contextual similarities with deforestation patterns.

Another promising avenue involves optimizing the network to further reduce its computational cost, including structural adjustments aimed at decreasing the number of required operations. In parallel, investigating self-supervised or semi-supervised learning strategies could help reduce the dependence on labeled data, thereby addressing the current limitations of the proposed approaches.

Finally, integrating the proposed approaches into operational systems represents a pivotal step toward practical deployment, enabling the effective use of the developed methodologies in real-world environmental monitoring applications.

BIBLIOGRAPHY

ADARME, M. O. et al. Evaluation of deep learning techniques for deforestation detection in the brazilian amazon and cerrado biomes from remote sensing imagery. **Remote Sensing**, v. 12, n. 6, 2020. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/12/6/910.

AKIBA, T. et al. Optuna: A next-generation hyperparameter optimization framework. In: **Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining**. New York, NY, USA: Association for Computing Machinery, 2019. (KDD '19), p. 2623–2631. ISBN 9781450362016. Disponível em: https://doi.org/10.1145/3292500.3330701.

ALKHATIB, A. A. A. A review on forest fire detection techniques. **International Journal of Distributed Sensor Networks**, v. 10, n. 3, p. 597368, 2014. Disponível em: https://doi.org/10.1155/2014/597368.

ALMEIDA, C. et al. **METODOLOGIA UTILIZADA NOS SISTEMAS PRODES E DETER -2 a EDIÇÃO (ATUALIZADA) INPE São José dos Campos 2022**. 2022.

ALSHEHRI, M.; OUADOU, A.; SCOTT, G. J. Deep transformer-based network deforestation detection in the brazilian amazon using sentinel-2 imagery. **IEEE Geoscience and Remote Sensing Letters**, v. 21, p. 1–5, 2024.

ALZU'BI, A.; AL-SMADI, L. Monitoring deforestation in jordan using deep semantic segmentation with satellite imagery. **Ecol. Informatics**, v. 70, p. 101745, 2022. Disponível em: https://api.semanticscholar.org/CorpusID:252384115.

ANDRADE, R. B. de; MOTA, G. L. A.; COSTA, G. A. O. P. da. Deforestation detection in the amazon using deeplabv3+ semantic segmentation model variants. **Remote Sensing**, v. 14, n. 19, 2022.

BAN, Y. Multitemporal remote sensing: Current status, trends and challenges. In: ____. **Multitemporal Remote Sensing: Methods and Applications**. Cham: Springer International Publishing, 2016. p. 1–18. ISBN 978-3-319-47037-5. Disponível em: https://doi.org/10.1007/978-3-319-47037-5_1.

BANDARA, W. G. C.; PATEL, V. M. A transformer-based siamese network for change detection. In: **IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium**. [S.l.: s.n.], 2022. p. 207–210.

BASSO, D. et al. Automatically detecting textual content in high-resolution images. In: **IEEE International Geoscience and Remote Sensing Symposium (IGARSS)**. [S.l.: s.n.], 2021. p. 4204–4207.

BEM, P. P. de et al. Change detection of deforestation in the brazilian amazon using landsat data and convolutional neural networks. **Remote Sensing**, v. 12, n. 6, 2020. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/12/6/901.

BENVENUTO, G. et al. Gecet: Garotas nas engenharias, ciências exatas e tecnologias: Projeto de extensão voltado para a participação feminina nas áreas stem. In: **Anais Estendidos do XXX Simpósio Brasileiro de Sistemas Multimídia e Web**. Porto Alegre, RS, Brasil: SBC, 2024. p.

211–216. ISSN 2596-1683. Disponível em: https://sol.sbc.org.br/index.php/webmedia_estendido/article/view/30498.

BENVENUTO, G.; CASACA, W. Retinal images registration via unsupervised deep learning. In: **Anais Estendidos da XXXVI Conference on Graphics, Patterns and Images**. Porto Alegre, RS, Brasil: SBC, 2023. p. 42–48. ISSN 0000-0000. Disponível em: https://sol.sbc.org.br/index.php/sibgrapi_estendido/article/view/27450.

BENVENUTO, G. et al. Assessing the performance of deep learning networks for real-time deforestation segmentation in the amazon and atlantic biomes. In: **Anais do LII Seminário Integrado de Software e Hardware**. Porto Alegre, RS, Brasil: SBC, 2025. p. 227–238. ISSN 2595-6205. Disponível em: https://sol.sbc.org.br/index.php/semish/article/view/36812.

BENVENUTO, G. A.; COLNAGO, M.; CASACA, W. Unsupervised deep learning network for deformable fundus image registration. In: **ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.: s.n.], 2022. p. 1281–1285.

BENVENUTO, G. A. et al. A fully unsupervised deep learning framework for non-rigid fundus image registration. **Bioengineering**, v. 9, n. 8, 2022. ISSN 2306-5354. Disponível em: https://www.mdpi.com/2306-5354/9/8/369.

BENVENUTO, G. A. et al. Precision meets speed: An attention encoder–decoder network for deforestation segmentation. **IEEE Geoscience and Remote Sensing Letters**, v. 22, p. 1–5, 2025.

BISHOP, C. M. **Neural Networks for Pattern Recognition**. [S.l.]: Oxford University Press, 1995.

BRAGAGNOLO, L.; da Silva, R.; GRZYBOWSKI, J. Amazon forest cover change mapping based on semantic segmentation by u-nets. **Ecological Informatics**, v. 62, p. 101279, 2021.

CAMPBELL, J.; WYNNE, R. **Introduction to Remote Sensing, Fifth Edition**. Guilford Publications, 2011. ISBN 9781609181772. Disponível em: https://books.google.com.br/books?id=NkLmDjSS8TsC.

CARELI, Á. S. et al. Deep learning segmentation models evaluation for deforestation monitoring embedded systems. In: **2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)**. [S.l.: s.n.], 2024. p. 274–278.

CHAVES, M. E. D.; PICOLI, M. C. A.; SANCHES, I. D. Recent applications of landsat 8/oli and sentinel-2/msi for land use and land cover mapping: A systematic review. **Remote Sensing**, v. 12, n. 18, 2020. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/12/18/3062.

CHEN, J. et al. Transunet: Transformers make strong encoders for medical image segmentation. **arXiv preprint arXiv:2102.04306**, 2021.

CHEN, J. et al. Dasnet: Dual attentive fully convolutional siamese networks for change detection in high-resolution satellite images. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 14, p. 1194–1206, 2021.

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, 2018.

CHEN, L.-C. et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: FERRARI, V. et al. (Ed.). **Computer Vision – ECCV 2018**. Cham: Springer International Publishing, 2018. p. 833–851. ISBN 978-3-030-01234-2.

CLAVERIE, M. et al. The harmonized landsat and sentinel-2 surface reflectance data set. **Remote Sensing of Environment**, v. 219, p. 145–161, 2018. ISSN 0034-4257. Disponível em: https://www.sciencedirect.com/science/article/pii/S0034425718304139.

COLNAGO, M. et al. Risk factors associated with mortality in hospitalized patients with covid-19 during the omicron wave in brazil. **Bioengineering**, v. 9, n. 10, 2022. ISSN 2306-5354. Disponível em: https://www.mdpi.com/2306-5354/9/10/584.

CORDTS, M. et al. The cityscapes dataset. In: **CVPR Workshop on The Future of Datasets in Vision**. [S.l.: s.n.], 2015.

DANTAS, C.; PASSOS, C.; HIROTA, R. **São Paulo x Altamira: uma batalha de gigantes**. 2025. https://infoamazonia.org/2025/09/24/sao-paulo-x-altamira-uma-batalha-de-gigantes/. InfoAmazonia. Publicado em 24 set. 2025. Acesso em 09 fev. 2026.

DAS, P. K. et al. A deforestation detection network using deep learning-based semantic segmentation. **IEEE Sensors Letters**, v. 8, n. 1, p. 1–4, 2024.

DAUDT, R. C.; SAUX, B. L.; BOULCH, A. Fully convolutional siamese networks for change detection. In: **2018 25th IEEE International Conference on Image Processing (ICIP)**. [S.l.: s.n.], 2018. p. 4063–4067.

DAUDT, R. C. et al. Urban change detection for multispectral earth observation using convolutional neural networks. In: **IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium**. [S.l.: s.n.], 2018. p. 2115–2118.

DOSOVITSKIY, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. In: **International Conference on Learning Representations**. [s.n.], 2021. Disponível em: https://openreview.net/forum?id=YicbFdNTTy.

Dumoulin, V.; Visin, F. A guide to convolution arithmetic for deep learning. **arXiv e-prints**, p. arXiv:1603.07285, mar. 2016.

ELLWANGER, J. H.; NOBRE, C. A.; CHIES, J. A. B. Brazilian biodiversity as a source of power and sustainable development: A neglected opportunity. **Sustainability**, v. 15, n. 1, 2023. ISSN 2071-1050. Disponível em: https://www.mdpi.com/2071-1050/15/1/482.

EVERINGHAM, M. et al. The pascal visual object classes (voc) challenge. **International Journal of Computer Vision**, Springer, v. 88, n. 2, p. 303–338, 2010.

FISCHER, W.; HEMPHILL, W.; KOVER, A. Progress in remote sensing (1972–1976). **Photogrammetria**, v. 32, n. 2, p. 33–72, 1976. ISSN 0031-8663. Disponível em: https://www.sciencedirect.com/science/article/pii/0031866376900132.

GARCIA-GARCIA, A. et al. A review on deep learning techniques applied to semantic segmentation. **ArXiv**, abs/1704.06857, 2017. Disponível em: https://api.semanticscholar.org/CorpusID:8574504.

GARNOT, V. S. F.; LANDRIEU, L. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. In: **2021 IEEE/CVF International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2021. p. 4852–4861.

GETREUER, P. Linear Methods for Image Interpolation. **Image Processing On Line**, v. 1, 2011.

GINO, V. L. et al. Integrating unsupervised machine intelligence and anomaly detection for spatio-temporal dynamic mapping using remote sensing image series. **Sustainability**, Mdpi, v. 15, n. 6, p. 4725, 2023.

GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. In: TEH, Y. W.; TITTERINGTON, M. (Ed.). **Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics**. Chia Laguna Resort, Sardinia, Italy: PMLR, 2010. (Proceedings of Machine Learning Research, v. 9), p. 249–256.

GONZALEZ, R.; WOODS, R. **Digital Image Processing**. Pearson, 2018. ISBN 9780133356724. Disponível em: https://books.google.com.br/books?id=0F05vgAACAAJ.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. http://www.deeplearningbook.org.

GUEDE-FERNáNDEZ, F. et al. A deep learning based object identification system for forest fire detection. **Fire**, v. 4, n. 4, 2021. ISSN 2571-6255. Disponível em: https://www.mdpi.com/2571-6255/4/4/75.

HANSEN, M. et al. High-resolution global maps of 21st-century forest cover change. **Science (New York, N.Y.)**, v. 342, p. 850–853, 11 2013.

HAO, S.; ZHOU, Y.; GUO, Y. A brief survey on semantic segmentation with deep learning. **Neurocomputing**, v. 406, p. 302–321, 2020. ISSN 0925-2312. Disponível em: https://www.sciencedirect.com/science/article/pii/S0925231220305476.

HARRIS, C. R. et al. Array programming with numpy. **Nature**, v. 585, p. 357–362, 2020.

HE, K. et al. Deep residual learning for image recognition. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 770–778.

HINTON, G. E.; ZEMEL, R. Autoencoders, minimum description length and helmholtz free energy. In: COWAN, J.; TESAURO, G.; ALSPECTOR, J. (Ed.). **Advances in Neural Information Processing Systems**. Morgan-Kaufmann, 1993. v. 6. Disponível em: https://proceedings.neurips.cc/paper_files/paper/1993/file/9e3cfc48eccf81a0d57663e129aef3cb-Paper.pdf.

HOWARD, A. G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. **CoRR**, abs/1704.04861, 2017. Disponível em: http://arxiv.org/abs/1704.04861.

HUANG, G. et al. Densely Connected Convolutional Networks . In: **2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Los Alamitos, CA, USA: IEEE Computer Society, 2017. p. 2261–2269. ISSN 1063-6919. Disponível em: https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.243.

HUANG, L. et al. Deep-learning-based semantic segmentation of remote sensing images: A survey. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 17, p. 8370–8396, 2024.

Iniciativa Amazônia+10. **Amazônia**. 2023. https://www.amazoniamaisdez.org.br/amazonia. Acesso em: 20 fev. 2026.

JADON, S. A survey of loss functions for semantic segmentation. In: **2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)**. [S.l.: s.n.], 2020. p. 1–7.

JAMALI, A. et al. Transu-net++: Rethinking attention gated transu-net for deforestation mapping. **Int. J. of Appl. Earth Obs. Geoinf.**, v. 120, p. 103332, 2023.

JELAS, I. M. et al. Deforestation detection using deep learning-based semantic segmentation techniques: a systematic review. **Frontiers in Forests and Global Change**, v. 7, 2024. ISSN 2624-893X. Disponível em: https://www.frontiersin.org/journals/forests-and-global-change/articles/10.3389/ffgc.2024.1300060.

JetBrains. **PyCharm Python IDE**. 2024. https://www.jetbrains.com/pycharm/.

JOCHER, G.; CHAURASIA, A.; QIU, J. **Ultralytics YOLOv8**. 2023. Disponível em: https://github.com/ultralytics/ultralytics.

JOCHER, G.; QIU, J. **Ultralytics YOLO11**. 2024. Disponível em: https://github.com/ultralytics/ultralytics.

JOHN, D.; ZHANG, C. An attention-based u-net for detecting deforestation within satellite sensor imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 107, p. 102685, 2022. ISSN 1569-8432. Disponível em: https://www.sciencedirect.com/science/article/pii/S0303243422000113.

KARAMAN, K.; GARNOT, V. S. F.; WEGNER, J. D. Deforestation detection in the amazon with sentinel-1 sar image time series. **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, X-1/W1-2023, p. 835–842, 2023. Disponível em: https://isprs-annals.copernicus.org/articles/X-1-W1-2023/835/2023/.

KASELIMI, M. et al. A vision transformer model for convolution-free multilabel classification of satellite imagery in deforestation monitoring. **IEEE Transactions on Neural Networks and Learning Systems**, v. 34, n. 7, p. 3299–3307, 2023.

KHORRAM, S. et al. **Remote Sensing**. Springer New York, 2012. (SpringerBriefs in Space Development). ISBN 9781461431039. Disponível em: https://books.google.com.br/books?id=QpdQut59MPsC.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Communications of the ACM**, ACM New York, NY, USA, v. 60, n. 6, p. 84–90, 2017.

KUMAR, R.; KUMAR, A.; SAIKIA, P. Deforestation and forests degradation impacts on the environment. In: ____. **Environmental Degradation: Challenges and Strategies for Mitigation**. Cham: Springer International Publishing, 2022. p. 19–46. ISBN 978-3-030-95542-7. Disponível em: https://doi.org/10.1007/978-3-030-95542-7_2.

KUNDU, S. et al. Real-time deforestation anomaly detection using yolo and langchain agents for sustainable environmental monitoring. **Scientific Reports**, v. 15, n. 1, p. 39961, nov. 2025. ISSN 2045-2322.

LAROCHELLE, H.; HINTON, G. E. Learning to combine foveal glimpses with a third-order boltzmann machine. In: LAFFERTY, J. et al. (Ed.). **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2010. v. 23. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2010/file/677e09724f0e2df9b6c000b75b5da10d-Paper.pdf.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

LEE, D.; CHOI, Y. A learning strategy for amazon deforestation estimations using multi-modal satellite imagery. **Remote Sensing**, v. 15, n. 21, 2023. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/15/21/5167.

LEE, S.-H. et al. Classification of landscape affected by deforestation using high-resolution remote sensing data and deep-learning techniques. **Remote Sensing**, v. 12, n. 20, 2020.

LIN, D. et al. Marta gans: Unsupervised representation learning for remote sensing image classification. **IEEE Geoscience and Remote Sensing Letters**, v. 14, n. 11, p. 2092–2096, 2017.

LIN, T.-Y. et al. Focal loss for dense object detection. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2017. p. 2980–2988.

LIN, T.-Y. et al. Microsoft coco: Common objects in context. In: FLEET, D. et al. (Ed.). **Computer Vision – ECCV 2014**. Cham: Springer International Publishing, 2014. p. 740–755. ISBN 978-3-319-10602-1.

LIU, P. et al. A lightweight object detection algorithm for remote sensing images based on attention mechanism and yolov5s. **Remote Sensing**, v. 15, n. 9, 2023. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/15/9/2429.

LIU, X. et al. Self-supervised learning: Generative or contrastive. **IEEE Transactions on Knowledge and Data Engineering**, v. 35, n. 1, p. 857–876, Jan 2023. ISSN 1558-2191.

LIU, Y. et al. Assessment and comparison of six machine learning models in estimating evapotranspiration over croplands using remote sensing and meteorological factors. **Remote Sensing**, v. 13, n. 19, 2021. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/13/19/3838.

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

LUEMBA, M.; KUNZAYILA, N.; CASACA, W. Exploratory analysis and visualization of brazilian forest data from the forest document system of the brazilian institute of the environment. In: **Computational Science and Its Applications**. [S.l.: s.n.], 2021. p. 145–159.

MA, L. et al. Deep learning in remote sensing applications: A meta-analysis and review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 152, p. 166–177, 2019. ISSN 0924-2716. Disponível em: https://www.sciencedirect.com/science/article/pii/S0924271619301108.

MARETTO, R. V. et al. Spatio-temporal deep learning approach to map deforestation in amazon rainforest. **IEEE Geoscience and Remote Sensing Letters**, v. 18, n. 5, p. 771–775, 2021.

MARTINS, V.; CARNEIRO, T. C. **O que acontece em Altamira, cidade que mais desmata no Brasil e tem mais boi que gente**. 2025. https://g1.globo.com/meio-ambiente/cop-30/noticia/2025/10/17/o-que-acontece-em-altamira.ghtml. G1 Pará. Atualizado em 17 out. 2025. Acesso em 09 fev. 2026.

MATOSAK, B. M. et al. Mapping deforestation in cerrado based on hybrid deep learning architecture and medium spatial resolution satellite time series. **Remote Sensing**, v. 14, n. 1, 2022. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/14/1/209.

MOURA, L. F. de et al. Enhancing deforestation detection through multi-domain adaptation with uncertainty estimation. **Forests**, v. 16, n. 5, 2025. ISSN 1999-4907. Disponível em: https://www.mdpi.com/1999-4907/16/5/742.

Nações Unidas Brasil. **Objetivo de Desenvolvimento Sustentável 15: Vida terrestre**. 2025. Acesso em 1 de julho de 2025. Disponível em: https://brasil.un.org/pt-br/sdgs/15.

NEGRI, R. G. et al. Assessing machine learning models on temporal and multi-sensor data for mapping flooded areas. **Transactions in GIS**, Wiley Online Library, v. 29, n. 2, p. e70028, 2025.

NEGRI, R. G. et al. Spectral–spatial-aware unsupervised change detection with stochastic distances and support vector machines. **IEEE Transactions on Geoscience and Remote Sensing**, v. 59, n. 4, p. 2863–2876, 2021.

Negri, R. G.; Silva, E.; Casaca, W. Inducing contextual classifications with kernel functions into support vector machines. **IEEE Geoscience and Remote Sensing Letters**, v. 15, n. 6, p. 962–966, 2018.

NOBRE, C. A. et al. Land-use and climate change risks in the amazon and the need of a novel sustainable development paradigm. **Proceedings of the National Academy of Sciences**, v. 113, n. 39, p. 10759–10768, 2016.

NWANKPA, C. et al. Activation functions: Comparison of trends in practice and research for deep learning. **ArXiv**, abs/1811.03378, 2018.

OLDEKOP, J. et al. Forest-linked livelihoods in a globalized world. **Nature Plants**, 11 2020.

OUBADRISS, A.; LAASSIRI, J.; MAKRANI, A. E. An overview comparison between convolutional neural networks and vision transformers. In: **Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security**. [S.l.: s.n.], 2024. p. 1–9.

PARMAR, N. et al. Image transformer. In: DY, J.; KRAUSE, A. (Ed.). **Proceedings of the 35th International Conference on Machine Learning**. PMLR, 2018. (Proceedings of Machine Learning Research, v. 80), p. 4055–4064. Disponível em: https://proceedings.mlr.press/v80/parmar18a.html.

PASZKE, A. et al. Pytorch: an imperative style, high-performance deep learning library. In: ____. **Proceedings of the 33rd International Conference on Neural Information Processing Systems**. Red Hook, NY, USA: Curran Associates Inc., 2019.

PIHUR, V.; DATTA, S.; DATTA, S. Weighted rank aggregation of cluster validation measures: a monte carlo cross-entropy approach. **Bioinformatics**, Oxford University Press, v. 23, n. 13, p. 1607–1615, 2007.

PIMPALKAR, S. et al. Evaluating afforestation and deforestation using attention-based semantic segmentation of sentinel-2 rgb-nir bands. **International Journal on Advanced Science, Engineering and Information Technology**, v. 15, p. 1309–1317, 08 2025.

PIVELLO, V. R. et al. Understanding brazil's catastrophic fires: Causes, consequences and policy needed to prevent future tragedies. **Perspectives in Ecology and Conservation**, v. 19, n. 3, p. 233–255, 2021.

Presidência da República, Casa Civil. **COP 30 no Brasil**. 2025. Página da Agenda Internacional – Missões Internacionais. Acesso em 1 de julho de 2025. Disponível em: https://www.gov.br/planalto/pt-br/agenda-internacional/missoes-internacionais/cop28/cop-30-no-brasil.

Python Software Foundation. **Python Programming Language**. 2024. https://www.python.org.

RODRIGUES, J. et al. A robust dual-mode machine learning framework for classifying deforestation patterns in amazon native lands. **Land**, MDPI, v. 13, n. 9, p. 1427, 2024.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: NAVAB, N. et al. (Ed.). **Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015**. [S.l.: s.n.], 2015.

ROS, G. et al. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 3234–3243.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, v. 65 6, p. 386–408, 1958. Disponível em: https://api.semanticscholar.org/CorpusID:12781225.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, out. 1986.

SALEHI, S. S. M.; ERDOGMUS, D.; GHOLIPOUR, A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In: SPRINGER. **International workshop on machine learning in medical imaging**. [S.l.], 2017. p. 379–387.

SAPKOTA, R. et al. Yolo advances to its genesis: a decadal and comprehensive review of the you only look once (yolo) series. **Artificial Intelligence Review**, v. 58, 06 2025.

SAPUCCI, G. R. et al. Analyzing spatio-temporal land cover dynamics in an atlantic forest portion using unsupervised change detection techniques. **Environmental Modeling and Assessment**, v. 26, p. 581–590, 2021.

SARINOVA, A. et al. Comparison of deep learning-based models for detection of diseased trees using an image compression algorithm. **Eastern-European Journal of Enterprise Technologies**, v. 5, n. 2 (131), p. 24–35, Oct. 2024. Disponível em: https://journals.uran.ua/eejet/article/view/314148.

Secretaria de Comunicação Social – SECOM. **Desmatamento no Brasil caiu 32,4% em 2024**. 2025. Site do Governo Federal. Acesso em 1 de julho de 2025. Disponível em: https://www.gov.br/secom/pt-br/assuntos/noticias/2025/05/desmatamento-no-brasil-caiu-32-4-em-2024.

SEYDI, S. T. et al. Fire-net: A deep learning framework for active forest fire detection. **Journal of Sensors**, v. 2022, p. 14, 2022.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. In: **International Conference on Learning Representations**. [S.l.: s.n.], 2015.

SINGH, P. et al. Hetconv: Heterogeneous kernel-based convolutions for deep cnns. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**. [S.l.: s.n.], 2019. p. 4835–4844.

SUDRE, C. H. et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: SPRINGER. **International Workshop on Deep Learning in Medical Image Analysis**. [S.l.], 2017. p. 240–248.

SZELISKI, R. **Computer Vision: Algorithms and Applications**. Springer London, 2010. (Texts in Computer Science). ISBN 9781848829350. Disponível em: https://books.google.com.br/books?id=bXzAlkODwa8C.

SZELISKI, R. **Computer Vision: Algorithms and Applications**. 2nd. ed. Cham, Switzerland: Springer International Publishing, 2022. ISBN 9783030343712. Disponível em: https://link.springer.com/book/10.1007/978-3-030-34372-9.

TORRES, D. L. et al. Deforestation detection with fully convolutional networks in the amazon forest from landsat-8 and sentinel-2 images. **Remote Sensing**, v. 13, n. 24, p. 5084, 2021.

TOVAR, P.; ADARME, M. O.; FEITOSA, R. Q. Deforestation detection in the amazon rainforest with spatial and channel attention mechanisms. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, XLIII-B3-2021, p. 851–858, 2021. Disponível em: https://isprs-archives.copernicus.org/articles/XLIII-B3-2021/851/2021/.

VASWANI, A. et al. Attention is all you need. In: **Proceedings of the 31st International Conference on Neural Information Processing Systems**. Red Hook, NY, USA: Curran Associates Inc., 2017. (NIPS'17), p. 6000–6010. ISBN 9781510860964.

VEGA, P. J. S. et al. Assessing the generalization capacity of convolutional neural networks and vision transformers for deforestation detection in tropical biomes. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, XLVIII-3-2024, p. 519–525, 2024. Disponível em: https://isprs-archives.copernicus.org/articles/XLVIII-3-2024/519/2024/.

Visin, F. et al. ReNet: A Recurrent Neural Network Based Alternative to Convolutional Networks. **arXiv e-prints**, p. arXiv:1505.00393, maio 2015.

VLăSCEANU, G. et al. Selecting the right metric: A detailed study on image segmentation evaluation. **BRAIN. Broad Research in Artificial Intelligence and Neuroscience**, v. 15, n. 4, p. 295–318, 2024. ISSN 2067-3957.

VOROTYNTSEV, P. et al. Satellite image segmentation using deep learning for deforestation detection. In: **IEEE 3rd Ukraine Conference on Electrical and Computer Engineering (UKRCON)**. [S.l.: s.n.], 2021. p. 226–231.

WANG, C.-Y.; YEH, I.-H.; LIAO, H.-Y. M. Yolov9: Learning what you want to learn using programmable gradient information. In: LEONARDIS, A. et al. (Ed.). **Computer Vision – ECCV 2024**. Cham: Springer Nature Switzerland, 2025. p. 1–21. ISBN 978-3-031-72751-1.

WANG, X. et al. A deep learning method for optimizing semantic segmentation accuracy of remote sensing images based on improved unet. **Scientific Reports**, v. 13, n. 1, p. 7600, 2023.

WOO, S. et al. Cbam: Convolutional block attention module. In: **Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VII**. Berlin, Heidelberg: Springer-Verlag, 2018. p. 3–19. ISBN 978-3-030-01233-5. Disponível em: https://doi.org/10.1007/978-3-030-01234-2_1.

XU, L. Research on the application of computer big data technology in satellite remote sensing map cartography and geographic information. In: **2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA)**. [S.l.: s.n.], 2022. p. 776–779.

YANG, Z. et al. Uav remote sensing applications in marine monitoring: Knowledge visualization and review. **Science of The Total Environment**, v. 838, p. 155939, 2022. ISSN 0048-9697. Disponível em: https://www.sciencedirect.com/science/article/pii/S0048969722030364.

YI-DE, M.; QING, L.; ZHI-BAI, Q. Automated image segmentation using improved pcnn model based on cross-entropy. In: IEEE. **Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.** [S.l.], 2004. p. 743–746.

YUAN, Q. et al. Deep learning in environmental remote sensing: Achievements and challenges. **Remote Sensing of Environment**, v. 241, p. 111716, 2020. ISSN 0034-4257. Disponível em: https://www.sciencedirect.com/science/article/pii/S0034425720300857.

ZHANG, B. et al. Progress and challenges in intelligent remote sensing satellite systems. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 15, p. 1814–1822, 2022.

ZHANG, Z.; LIU, Q.; WANG, Y. Road extraction by deep residual u-net. **IEEE Geoscience and Remote Sensing Letters**, v. 15, n. 5, p. 749–753, 2018.

ZHU, X. X. et al. Deep learning in remote sensing: A comprehensive review and list of resources. **IEEE Geoscience and Remote Sensing Magazine**, v. 5, n. 4, p. 8–36, 2017.