Deep Learning Segmentation Models Evaluation for Deforestation Monitoring Embedded Systems

1st Álvaro S. Careli
Cyber Security and IoT Laboratory, CS&I Lab.
National Institute of Telecommunication, Inatel
Santa Rita do Sapucaí, Brazil
alvaro.sampaio@ges.inatel.br

2nd Evandro C. Vilas Boas
Cyber Security and IoT Laboratory, CS&I Lab.
National Institute of Telecommunication, Inatel
Santa Rita do Sapucaí, Brazil
evandro.cesar@inatel.br

3rd Eduardo H. Teixeira
IoT Research Group Laboratory
National Institute of Telecommunication, Inatel
Santa Rita do Sapucaí, Brazil
eduardot@gea.inatel.br

4th Elaine C. C. Silva
Cyber Security and IoT Laboratory, CS&I Lab.
National Institute of Telecommunication, Inatel
Santa Rita do Sapucaí, Brazil
elaine.silva@inatel.br

5th Guilherme P. Aquino
Cyber Security and IoT Laboratory, CS&I Lab.
National Institute of Telecommunication, Inatel
Santa Rita do Sapucaí, Brazil
guilhermeaquino@inatel.br

6th Felipe A. P. Figueiredo
Wireless and Artificial Intelligence Laboratory, WAI Lab.
National Institute of Telecommunication, Inatel
Santa Rita do Sapucaí, Brazil
felipe.figueiredo@inatel.br

2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA) | 979-8-3315-0649-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICICYTA64807.2024.10913179

***Abstract*—This work evaluates deep learning segmentation models to propose a deforestation monitoring embedded system. The approach stands for environmental monitoring using remote sensing imagery, edge computing, and a deep learning segmentation model. Thus, the performance of you only look once architecture version 8 (YOLOv8) and Mask Region-based convolutional neural networks (Mask R-CNN) embedded in Raspberry Pi Model 4 regarding Intersection over Union (IoU), mean Average Precision (mAP), and time per image processing metrics is compared. The models are combined with a pixel-based algorithm that analyzes the temporal segmented images to define their forest area percentage for deforestation monitoring and detection. The results demonstrate YOLOv8x model achieved an IoU of 0.762, with a time per image of 0.4777 seconds, while Mask R-CNN R101 FPN 3x obtained an IoU of 0.763, with a time per image of 0.2669 seconds. The average times for YOLOv8 ranged from 0.0434 to 0.4777 seconds, and for Mask R-CNN from 0.1969 to 0.2669 seconds. Finally, this work proposes evaluating the model's performance when working with generative AI models Dall-e, Craiyon, and Tess-AI to create a synthetic dataset to augment the initial one with synthetic samples and improve the model's training with a large dataset. The Dall-e has been shown to outperform the others regarding the IoU metric, which was suggested to augment datasets with synthetic samples.**

***Index Terms*—Environmental Monitoring, Deforestation, Mask R-CNN, Segmentation, YOLOv8.**

I. Introduction

Forests are vital in maintaining ecological balance, acting as carbon sinks, climate regulators, and habitats for biodiversity. However, its preservation faces challenges due to the deforestation process imposed by human activity and natural events such as fires [1]–[3]. Deforest monitoring is crucial to provide up-to-date data for authorities and drive decision-making policies to combat illegal human actions or to implement reforestation campaigns. Conventional methods have been used to provide information for deforest monitoring and detecting tasks based on human observation or field collection. However, they are expensive and time-consuming for tracking temporal changes [1], [2].

Remote sensing imagery provided by satellite or unmanned aerial vehicle (UAV) emerges as a cost-effective and more reliable approach to tracking the advance of deforestation over time [1], [4]–[8]. The temporal collected images are analyzed and compared to identify changes by considering pixel processing techniques, such as algebra-based techniques, machine learning classifiers, and deep learning models. The latter relies on computer vision models that have been shown to outperform the other methods, considering high-resolution images. These methods evaluate the image pixels and label each based on the content information, comparing temporal images for a given area to track changes.

Related works on deep learning models relying on convolutional neural networks (CNNs) perform image segmentation that is further post-processed, aiming at detecting forest areas or trees. For instance, Region-based CNNs (R-CNNs) algorithms have been applied to identify trees in images obtained from satellites or UAVs. Forest health diagnosis based on a Mask R-CNN approach is proposed to segment dead trees from aerial images [9]. This model was trained based on a transfer learning scheme using an augmented dataset and a synthetic method. It allows for counting the number of dead trees in an image of a forest region to indicate its health, which fosters causal analysis of environmental changes and the predictive likelihood of forest fire. Additionally, Mask R-CNN has been applied to map tree species and crowns in a forest [10]–[12]. The experiments were conducted using images captured by UAVs due to their superior quality compared to satellite systems. Fast R-CNN was also used

This work was partially funded by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), National Council for Scientific and Technological Development (CNPq), Cyber Security and Internet of Things Laboratory (CS&I Lab.) and Inatel Cybersecurity Center (CxSC Telecom).

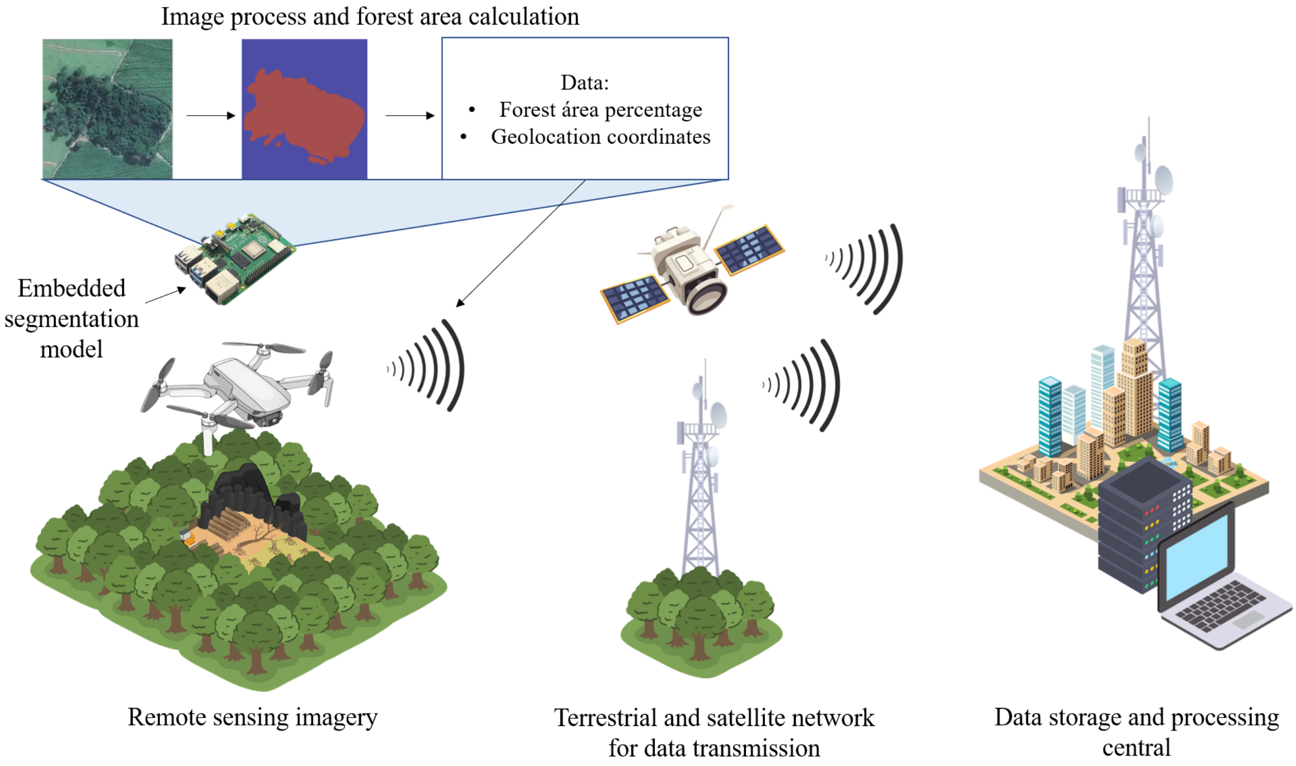

Fig. 1: Deforestation monitoring remote sensing embedded system.

to detect coconut trees in littoral areas, monitoring it due to economic value [13], [14]. Mask R-CNN was combined with other techniques to enhance segmentation performance, such as Vision Transformer [15]. Finally, some works elaborated on you only look once (YOLO) architecture to detect living tree or dead tree [16].

The CNN models, like the U-Net family, were extensively exploited to process satellite and UAV images and segment forest areas [5], [17], [18]. A CNN model object-based change detection that incorporates contextual information as spatial, spectral, and temporal relationships is proposed as an alternative to pixel-based approaches to evaluate satellite images [18]. Some works are devoted to evaluating deep learning techniques for deforestation detection [6], [7]. Adarme *et al.* proposed exploring the performance of the Early Fusion (EF), Siamese Network (SN), Convolutional Support Vector Machine (CSVM), and Support Vector Machine (SVM) [7]. Meanwhile, other authors have worked on CNN-based architecture, such as SharpMask, U-Net, and ResUnet [6]. Furthermore, a Deep Transformer-Based Network was trained using satellite images for deforestation detection as an alternative to the abovementioned deep learning models [1].

YOLO architecture has been widely applied to forest fire detection [19]–[21], while its potential for deforest monitoring through segmentation tasks is unveiled. Hence, this work contributes to the deforest monitoring field by evaluating and comparing the YOLOv8 model variants to task forest area segmentation with Mask R-CNN-based architecture. Image segmentation models are integrated with a pixel-based forest percentage area calculation algorithm. The algorithms are embedded into a Raspberry Pi Model 4 and evaluated based on Intersection over Union (IoU), mean Average Precision (mAP), and time per image process metrics, with the aim of image edge processing and transmission of the forest area percentage to storage for further analysis of deforestation monitoring. The dataset has been created and handcraft segmented by the authors, while available for future related work usage[1] based on a detection dataset[2] and web scraping. Finally, generative AI models such as Dall-e, Craiyon, and Tess-AI are evaluated for task dataset augmentation with synthetic samples. This approach is considered an alternative for augmenting small datasets and enhancing the model's training.

The remainder of this work is organized as follows. Section II discusses the methodology to compose a dataset for training the selected models and explains the pixel-based forest area calculation algorithm. Section III presents the experiment conducted to assess the models while proposing and evaluating the model's performance when working on synthetic images created by generative AI models for dataset augmentation purposes. Finally, Section IV presents conclusions and suggestions for future work.

II. Methodology

This section presents the dataset, the selected deep-learning-based segmentation models, and the performance evaluation methodology. The training process aims at deforestation monitoring for remote sensing by embedding the model into an IoT system, as seen in Fig. 1. Hence, the proposal considers a UAV to monitor forest areas using an embedded system capable of locally processing aerial capture images. The deep-learning model processes the image by applying a segmentation process to identify forest (red segmentation) and deforest areas (blue segmentation).

[1]https://universe.roboflow.com/inatel-rynhm/ic-i8d9e
[2]https://universe.roboflow.com/sliit-fi7f0/deforestation-p5ehc

Fig. 2: Dataset image samples with the ground truth masks.

Lately, a proposed algorithm counts the image's total pixel number and the respective number of pixels in each segmentation. These values define the forest percentage, which is further transmitted with geolocation coordinates to a central to track deforestation over time in a given area. Since the models work on images by obtaining multiple segmentations, OR logic was applied to combine all segmentations into a single mask.

A. Dataset and Evaluated Models

The dataset was built based on web scraping and images from other detection datasets. However, the authors manually labeled it to obtain the ground truth mask for the segmentation task using the Roboflow tools. This work used a dataset of 224 positive images with forested and deforested areas, as shown in Fig. 2. The set was divided into 70% images for training, 20% for validation, and 10% for testing. The segmentation models were trained and validated based on the handcrafted labels and tested based on the respective labels generated in the inference stage.

The YOLOv8 model architecture was selected considering its variants: XtraLarge (YOLOv8x), Large (YOLOv8l), Medium (YOLOv8m), Small (YOLOv8s), and Nano (YOLOv8n). The main difference among these sub-models is the number of parameters influencing the architecture's performance [22], [23]. Additionally, Mask R-CNN models were used for image segmentation due to their accuracy [24]. The Mask R-CNN R101 FPN 3x uses a ResNet-101 network with a Feature Pyramid Network (FPN), known for improving object detection at different scales. The Mask R-CNN R101 C4 3x and Mask R-CNN R101 DC5 3x use the ResNet-101 with different configurations affecting these model's depth and complexity. The Mask R-CNN R50 C4 3x and Mask R-CNN R50 DC5 1x models, which use the ResNet-50, are known for being lighter and faster than those using ResNet-101 while offering accuracy in object segmentation. These characteristics make Mask R-CNN models suitable for complex image segmentation.

B. Forest area calculation algorithm

Regarding an image, a given model can return k segmented masks (M^k) with a size ($m \times n$). The mask elements (M^k_{ij}) comprise an image pixel that equals 1 if it is into the segmentation indicating forest area or equals 0; otherwise. Hence, the k mask must be combined to obtain the final mask M^f applying a pixel-by-pixel OR boolean logic

$$M_f = M^1 \vee M^2 \vee \cdots \vee M^k, \tag{1}$$

where $\vee$ denotes the boolean OR logic. As a result, each element M^f_{ij} of the final mask is given by

$$M^f_{ij} = M^1_{ij} \vee M^2_{ij} \vee \cdots \vee M^k_{ij}. \tag{2}$$

The operation in (2) includes any M_{ij} pixel into the final mask if its value equals 1 in one of the k masks. Afterward, the forest percentage area (A_f) is obtained by summing up the M^f_{ij} elements of the matrix defined in (1) and dividing it by the image pixel numbers

$$A_f = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} M^f_{ij}. \tag{3}$$

The proposed strategy effectively integrates multiple segmentations, ensuring that all areas of interest are accounted for in the final mask while solving the intersection among the masks. It guarantees the measurement accuracy of segmented areas in complex images with a simple approach, aiming at low processing for embedding the solution.

C. IoU Models Performance

The model's accuracy performance upon image segmentation was evaluated based on the IoU metric, which allows for assessing how close the generated segmentations are to the handcrafted labels. It is defined as the intersection area between the model's predicted mask and the ground truth mask divided by the area of the union of these two masks.

Considering the ground truth mask (M^t) and a predicted mask (M^p) generated by a segmentation model, the IoU metric is defined as a ratio between the intersection and the union of the mask pixel

$$IoU = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (M^p_{ij} \wedge M^t_{ij})}{\sum_{i=1}^{m} \sum_{j=1}^{n} (M^p_{ij} \vee M^t_{ij})}. \tag{4}$$

The intersection between the masks is defined as the pixels in both mask segmentation, while the union relates to the pixels in at least one mask segmentation. The IoU value ranges from 0 to 1, whereas 1 indicates a perfect match between the prediction mask and the ground truth mask, and 0 accounts for no match. The mask used to calculate the IoU is the the M^f defined in (1), ensuring consistency in the evaluations. This quantitative approach allows for assessing the model accuracy for image segmentation tasks, providing a clear understanding of the model's performance in proximity to handcraft labels.

III. Experiments and Results

This section details the experiments conducted to evaluate the performance of forested area segmentation models YOLOv8 and Mask R-CNN. The results were obtained by assessing the trained models embedded on a Raspberry Pi Model 4 with 8GB, considering future implementation of the IoT deforestation system proposed in Fig. 1. Table I shows the

TABLE I: YOLOv8 and Mask R-CNN models performance evaluation.

YOLOv8 variant models			
Models	IoU	mAP	Time/image (s)
YOLOv8x	0.762	22.87	0.4777
YOLOv8l	0.780	22.02	0.3192
YOLOv8m	0.717	13.75	0.1723
YOLOv8s	0.757	22.18	0.0862
YOLOv8n	0.700	22.00	0.0434
Mask R-CNN variant models			
Models	IoU	mAP	Time/image (s)
Mask R-CNN R101 FPN 3x	0.763	19.15	0.2669
Mask R-CNN R101 C4 3x	0.757	21.67	0.2585
Mask R-CNN R101 DC5 3x	0.764	16.67	0.2562
Mask R-CNN R50 C4 3x	0.760	22.30	0.2477
Mask R-CNN R50 C4 1x	0.764	26.28	0.2123
Mask R-CNN R50 DC5 1x	0.761	20.47	0.1969

trained model performance comprising the following metrics: IoU, mAP, and time per image processing. The results were obtained from the dataset image test.

The IoU values demonstrate that YOLOv8 and Mask R-CNN models presented similar performance, with YOLOv8x reaching 0.762 and Mask R-CNN R101 FPN 3x achieving 0.763. However, neither model surpassed the 80% threshold, which can be related to the dataset size affecting the model generalizability. These results indicate that forest segmentation remains a significant challenge and a promising area for future research. Performance in terms of mAP for mask also varied, with the YOLOv8 models achieving a score between 13.75 and 22.87, while the Mask R-CNN models achieved between 16.67 and 26.28. The average processing time per image for YOLOv8 and Mask R-CNN models embedded on the Raspberry Pi 4 was similar. However, the variance in processing times among the Mask R-CNN models was lower, indicating consistent performance. For instance, YOLOv8n had a processing time of 0.0434 seconds per image, while YOLOv8x had 0.4777 seconds. Among the Mask R-CNN models, Mask R-CNN R50 DC5 1x had a time of 0.1969 seconds, and Mask R-CNN R101 FPN 3x took 0.2669 seconds.

Figure 3(a) and (b) present two images of the same area over time. These images were applied to model YOLOv8x, and the segmented masks were combined using the proposed algorithm to calculate the forest percentage area. Fig. 2(c) and (d) depict the algorithm output versions, already incorporating the combined segmentation models with area calculation. As mentioned, the blue segmentation corresponds to deforested areas, while the red segmentation represents forested areas. By processing these images based on (3), the results are forest percentage areas of 35.62% and 15.47%, showing the proposal's effectiveness in tracking the deforestation process over time.

Additionally, training and validation images were explored as inputs into generative AIs such as Dall-e, Craiyon, and Tess-AI to create a synthetic dataset. The generated images were applied to the trained models to evaluate their performance using the IoU metric. This allows us to understand

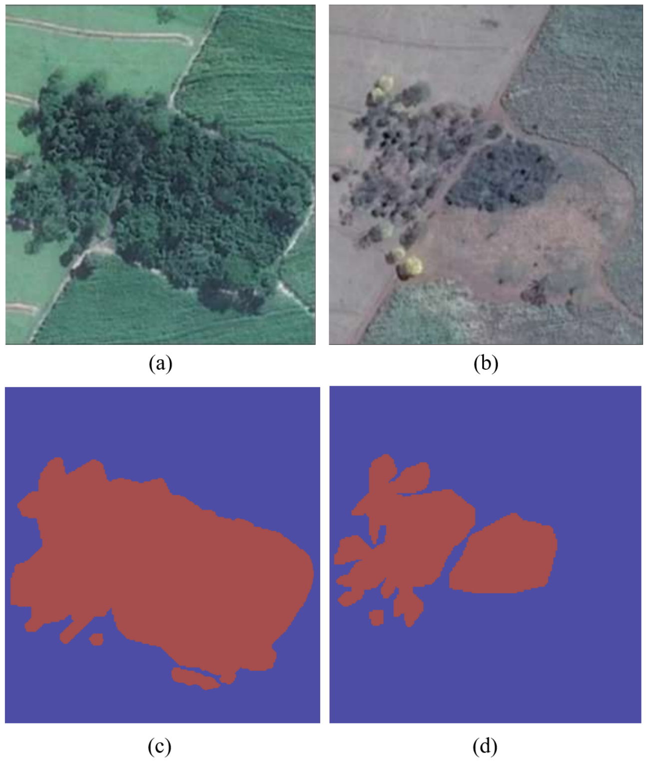

Fig. 3: Final segmentation mask obtained by the pixel-based algorithm.

TABLE II: Models IoU performance based on generative AI images.

Models	Crayon	Tess-AI	Dall-e
YOLOv8x	0.696	0.673	0.637
YOLOv8l	0.780	0.787	0.714
YOLOv8m	0.750	0.727	0.674
YOLOv8s	0.750	0.727	0.520
YOLOv8n	0.010	0.260	0.640
Mask R-CNN R101 FPN 3x	0.730	0.760	0.750
Mask R-CNN R101 C4 3x	0.740	0.690	0.756
Mask R-CNN R101 DC5 3x	0.730	0.700	0.750
Mask R-CNN R50 C4 3x	0.740	0.700	0.758
Mask R-CNN R50 C4 1x	0.743	0.700	0.751
Mask R-CNN R50 DC5 1x	0.727	0.682	0.730
IoU average value	0.672	0.673	0.698

how each model processes and interprets the nuances of images created through generative AIs. Furthermore, among the three generative AIs used to create new datasets from original images, Dall-e outperformed Crayon and Tess-AI, as seen in Table II. However, this advantage was mainly due to the performance of the YOLOv8n model. Comparing the labels generated by the models with true labels, the results suggest that any of the three generative AI models could be used to create synthetic datasets or expand existing datasets with synthetic samples, as the metrics obtained were similar to metrics with real images.

IV. Conclusion

This work evaluated the YOLOv8 and Mask R-CNN models for segmenting forest areas for deforestation monitoring for embedded systems. The models were compared based on IoU, mAP, and processing time per image while embedded in a Raspberry Pi Model 4. Results showed that the model's IoU metrics were below the 0.8 threshold, which can be related to the dataset size affecting the model's generalizability. The mAP performance varied slightly among the YOLOv8 and Mask R-CNN model variants. Despite similar average processing times per image, Mask R-CNN models exhibited lower variance, suggesting a consistent performance. The results have shown that dataset size must be improved as future work, and the models must be re-trained to evaluate their performance on a large dataset. A forward-thinking strategy for dataset augmentation has been proposed in this work using generative AI to augment datasets with synthetic samples. Notably, the performance of the three models was promising, with slight IoU disparities compared to the real labels. Hence, future work carries out on dataset augmentation and the architecture implementation of the proposed IoT system with real tests based on a drone.

References

[1] M. Alshehri, A. Ouadou, and G. J. Scott, "Deep transformer-based network deforestation detection in the brazilian amazon using sentinel-2 imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.

[2] M. Lu, E. Pebesma, A. Sanchez, and J. Verbesselt, "Spatio-temporal change detection from multidimensional arrays: Detecting deforestation from modis time series," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 227–236, 2016.

[3] B. Arteaga, M. Diaz, and M. Jojoa, "Deep learning applied to forest fire detection," in *2020 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2020, pp. 1–6.

[4] P. Vorotyntsev, Y. Gordienko, O. Alienin, O. Rokovyi, and S. Stirenko, "Satellite image segmentation using deep learning for deforestation detection," in *2021 IEEE 3rd Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 2021, pp. 226–231.

[5] J. Villalobos-Montiel, A. Aguilar-Gonzalez, L. Orona, and C. Lozoya, "Identifying deforested areas through convolutional neural network for drone reforesting," in *2023 IEEE Conference on Technologies for Sustainability (SusTech)*, 2023, pp. 138–143.

[6] P. P. De Bem, O. A. de Carvalho Junior, R. Fontes Guimarães, and R. A. Trancoso Gomes, "Change detection of deforestation in the brazilian amazon using landsat data and convolutional neural networks," *Remote Sensing*, vol. 12, no. 6, p. 901, 2020.

[7] M. Ortega Adarme, R. Queiroz Feitosa, P. Nigri Happ, C. Aparecido De Almeida, and A. Rodrigues Gomes, "Evaluation of deep learning techniques for deforestation detection in the brazilian amazon and cerrado biomes from remote sensing imagery," *Remote Sensing*, vol. 12, no. 6, p. 910, 2020.

[8] S. H. Khan, X. He, F. Porikli, and M. Bennamoun, "Forest change detection in incomplete satellite images with deep neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5407–5423, 2017.

[9] C.-Y. Chiang, C. Barnes, P. Angelov, and R. Jiang, "Deep learning-based automated forest health diagnosis from aerial images," *IEEE Access*, vol. 8, pp. 144 064–144 076, 2020.

[10] T. Yoshii, C. Lin, S. Tatsuhara, S. Suzuki, and T. Hiroshima, "Tree species mapping of a hemiboreal mixed forest using mask r-cnn," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 6228–6231.

[11] R. Hamzah and M. F. Md.Noor, "Visualization of tree species identification using mask rcnns for tropical forests in malaysia," in *2022 International Conference on Computer and Drone Applications (IConDA)*, 2022, pp. 55–60.

[12] Z. Sun, B. Xue, M. Zhang, and J. Schindler, "An improved mask r-cnn for instance segmentation of tree crowns in aerial imagery," in *2023 38th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2023, pp. 1–6.

[13] J. Zheng, W. Wu, L. Yu, and H. Fu, "Coconut trees detection on the tenarunga using high-resolution satellite images and deep learning," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 6512–6515.

[14] J. V. D. Prasad, M. K. Chandra, J. Akash, and K. Vivek, "Coconut tree detection in coastal areas with fast-rcnn using resnet-50," in *2023 Global Conference on Information Technologies and Communications (GCITC)*, 2023, pp. 1–4.

[15] Q. Liang, "Application of the vision transformer and mask r-cnn joint algorithm to assist forest decisions," in *2023 5th International Conference on Geoscience and Remote Sensing Mapping (GRSM)*, 2023, pp. 127–131.

[16] X. Wang, Q. Zhao, P. Jiang, Y. Zheng, L. Yuan, and P. Yuan, "Lds-yolo: A lightweight small object detection method for dead trees from shelter forest," *Computers and Electronics in Agriculture*, vol. 198, p. 107035, 2022.

[17] P. Vorotyntsev, Y. Gordienko, O. Alienin, O. Rokovyi, and S. Stirenko, "Satellite image segmentation using deep learning for deforestation detection," in *2021 IEEE 3rd Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 2021, pp. 226–231.

[18] S. H. Khan, X. He, F. Porikli, and M. Bennamoun, "Forest change detection in incomplete satellite images with deep neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5407–5423, 2017.

[19] S. Goyal, M. Shagill, A. Kaur, H. Vohra, and A. Singh, "A yolo based technique for early forest fire detection," *Int. J. Innov. Technol. Explor. Eng*, vol. 9, pp. 1357–1362, 2020.

[20] S. Wang, T. Chen, X. Lv, J. Zhao, X. Zou, X. Zhao, M. Xiao, and H. Wei, "Forest fire detection based on lightweight yolo," in *2021 33rd Chinese Control and Decision Conference (CCDC)*, 2021, pp. 1560–1565.

[21] S. Wu and L. Zhang, "Using popular object detection methods for real time forest fire detection," in *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 01, 2018, pp. 280–284.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[23] M. H. F. Afonso, E. H. Teixeira, M. R. Cruz., G. P. Aquino, and E. C. Vilas Boas, "Vehicle and plate detection for intelligent transport systems: Performance evaluation of models yolov5 and yolov8," in *2023 IEEE International Conference on Computing (ICOCO)*, 2023, pp. 328–333.

[24] E. Hassan, N. El-Rashidy, M. Talaa *et al.*, "mask r-cnn models," *Nile Journal of Communication and Computer Science*, vol. 3, no. 1, pp. 17–27, 2022.