

Review

Not peer-reviewed version

A Comprehensive Review of Semantic Segmentation and Instance Segmentation in Forestry: Advances, Challenges, and Applications

[Krzysztof Wołk](#)^{*} and Marek Tatara

Posted Date: 15 July 2024

doi: 10.20944/preprints202407.1056.v1

Keywords: semantic segmentation; forestry; forestry segmentation forestry



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

A Comprehensive Review of Semantic Segmentation and Instance Segmentation in Forestry: Advances, Challenges, and Applications

Krzysztof Wołk ^{1,*} and Marek Tatara ²

¹ DAC.digital, Kołobrzeska 14, 80-394 Gdańsk, Poland; krzysztof.wolk@dac.digital

² DAC.digital, Kołobrzeska 14, 80-394 Gdańsk, Poland; marek.tatara@dac.digital

* Correspondence: krzysztof.wolk@dac.digital

Abstract: This article presents a succinct overview of the progress, obstacles, and uses of semantic segmentation and instance segmentation within the forestry domain. The objective of this review is to conduct a critical analysis of the current literature pertaining to segmentation techniques and provide a methodical summary of their impact on forestry-related activities, including but not limited to tree species classification, forest inventory, and ecological monitoring such as retrieval of dominant tree species. Through the process of synthesizing pivotal discoveries from multiple studies, this comprehensive analysis provides valuable perspectives on the present status of research and highlights prospective areas for further exploration. The primary topics addressed encompass the approach employed for executing the examination, the fundamental discoveries associated with semantic segmentation and instance segmentation in the domain of forestry, and the ramifications of these discoveries for the discipline. The results indicate that the utilization of semantic segmentation methods has been efficacious in the field of forestry for the precise identification of tree species. Such methods also aid in tracking of deforested regions over the course of time by separating other land-use classes from forested regions. Additionally, the employment of instance segmentation techniques exhibits potential in the demarcation of individual trees. Instance segmentation offers promising results due to deep learning models based on forest point clouds. However, several challenges persist in the successful implementation of semantic segmentation methods such as the presence of occlusions, overlapping branches, and intricate structures hampers the accurate segmentation of trees. Additionally, instance segmentation approaches that utilize models are mostly trained by using laser scanning data based on forest types which are typically trained on specific laser scanning data and forest types that create limitations in generalization from high to low resolution point clouds. Due to this reason, the existing approaches often struggle with handling these complex structures, leading to the need for manual methods for extracting measurements from forest point clouds. The review culminates by underscoring the necessity for additional research to tackle current obstacles and augment the precision and relevance of segmentation methodologies in the field of forestry. In general, the present article provides a significant reference for scholars and professionals who are interested in utilizing segmentation techniques in the field of forestry.

Keywords: semantic segmentation; forestry; forestry segmentation forestry

1. Introduction

Semantic segmentation and instance segmentation are two separate algorithms that are frequently employed in computer vision, including forestry applications. By providing pixel-level knowledge and object-level delineation, these approaches bring unique insights into the interpretation of forestry images. Understanding the distinctions between semantic and instance segmentation is critical for successfully employing these approaches in forestry analysis.

Long, Shelhamer, & Darrell [20] describe semantic segmentation as the categorization of each pixel in an image into pre-defined categories to facilitate a full comprehension of the picture's content and structure. In the context of forestry, semantic segmentation enables precise identification of

various components such as trees, grass, sky, buildings, and other forest objects such as undergrowth, shrubs, bushes, rocks, streams, and more.

Forestry experts and scholars may produce thorough and exact maps of forest regions by using semantic segmentation techniques. This degree of detail enables them to monitor plant changes, research wildlife habitats, estimate forest health, and design sustainable forestry methods.

Semantic segmentation can also aid with environmental monitoring and conservation. Experts can establish focused strategies for maintaining biodiversity, regulating forest fires, and minimizing the impact of human activities on the environment by recognizing distinct aspects within the forest ecosystem.

Overall, correct identification of diverse forest components by semantic segmentation adds to a better knowledge of the forest ecosystem, allowing for more informed decision-making and supporting sustainable forest management practices.

In contrast, instance segmentation goes beyond semantic segmentation by distinguishing and separating specific instances of objects within the same class [2]. This method gives a more complete analysis by outlining and monitoring each object instance accurately. In the realm of forestry, instance segmentation allows for particular tasks such as tree counting, individual tree monitoring, and precise measurements for forest inventory.

Understanding the differences between semantic segmentation and instance segmentation is critical for picking the best approach for forestry analysis according to the objectives. While semantic segmentation focuses on overall scene comprehension, instance segmentation enables fine-grained object-level analysis, offering vital insights about tree features and geographic distribution.

According to studies by Chen [3] the use of semantic segmentation and instance segmentation has attracted a lot of interest in the forestry domain as efficient ways to extract valuable insights from remotely sensed data. The utilization of segmentation techniques allows for the recognition, demarcation, and categorization of objects and areas of significance in forest environments. This, in turn, enables a diverse array of applications, including but not limited to the classification of tree species, the taking of forest inventory, and the monitoring of ecological conditions [39]. Recent developments in image processing algorithms, in conjunction with the accessibility of high-resolution remote sensing data, have presented novel prospects for precise and effective examination of forest imagery [39]. Apart from its usage in the classification of tree species and ecological monitoring, semantic segmentation has also been employed to identify the most efficient tree cutting routes in forestry operations. Semantic segmentation techniques aid in devising efficient and sustainable harvesting strategies by precisely outlining individual tree crowns and identifying their spatial distribution. Accurately identifying and characterizing trees provides companies with the opportunity to reduce the impact on the surrounding vegetation, optimize operational efficiency, and augment resource utilization. As exemplified by [20], the strategic identification of optimal trees for harvesting while simultaneously preserving the forest's overall structure and composition can be achieved through the integration of remote sensing data and advanced segmentation algorithms by companies. This aspect emphasizes the practical significance of semantic segmentation in forestry, which goes beyond its research applications.

The field of forestry is of significant importance in the management of the environment, the conservation of biodiversity, and the sustainable utilization of resources. The acquisition of precise and punctual data regarding the makeup, arrangement, and evolution of forests is imperative for proficient decision-making and well-informed policy formulation. Conventional techniques for gathering such data through field-based approaches are arduous, require significant manpower, and frequently have a restricted spatial scope. Zhang [41] say that using remote sensing, especially in combination with advanced image analysis techniques like semantic segmentation and instance segmentation, is a good way to get complete and up-to-date information about forests in a way that is also cost-effective.

The objective of this scholarly article is to furnish a thorough evaluation of the present cutting-edge advancements in semantic segmentation and instance segmentation methodologies in the domain of forestry. The objective of this review is to synthesize the primary findings, identify

prevalent methodologies and approaches, scrutinize challenges and limitations, and propose potential areas for future research by means of a critical examination of the existing literature. The aim is to augment comprehension regarding the potential and constraints of said segmentation techniques and their relevance to tasks pertaining to forestry.

The subsequent sections will delve into the diverse uses of semantic segmentation and instance segmentation in the field of forestry. Following a critical assessment of the major discoveries mentioned in the literature, there will be a discussion of their implications. Additionally, areas that necessitate further exploration will be emphasized. The primary objective of this exhaustive analysis is to make a scholarly contribution toward the progression of understanding and encourage the implementation of segmentation methodologies within the forestry domain.

2. Methodology

The present review article utilized a rigorous methodology that involved conducting a systematic literature search to identify pertinent studies pertaining to semantic segmentation and instance segmentation within the forestry domain. The inquiry was executed by utilizing diverse scholarly databases, such as IEEE Xplore, the ACM Digital Library, Scopus, and the Web of Science, among others. The utilized search queries encompassed the topics of “semantic segmentation,” “instance segmentation,” “forestry,” “remote sensing,” and associated variants.

The criteria used to select studies for this research were as follows: (1) the study had to concentrate on the use of semantic segmentation or instance segmentation methods in the field of forestry; (2) the study had to provide a comprehensive explanation of the methodology employed; (3) the study had to present empirical results and findings concerning the precision, efficiency, or practicality of the segmentation techniques; (4) the study had to be published in a peer-reviewed journal or conference proceedings.

The application of exclusion criteria was carried out to remove studies that were deemed irrelevant due to their failure to meet the aforementioned criteria or their duplication. A significant quantity of articles was obtained from the preliminary search, and subsequently, a screening process was conducted based on the titles and abstracts of the articles. The complete texts of articles were procured for studies that were deemed potentially relevant, and a comprehensive evaluation of their substance was carried out to determine their appropriateness for incorporation in this review.

The chosen articles underwent analysis and synthesis to extract crucial information, such as the segmentation algorithms utilized, the types of remote sensing data employed, the study areas or datasets, the evaluation metrics used, and the primary findings reported. The information that was extracted has been systematically arranged and presented in a cohesive manner to offer a thorough summary of the progress, obstacles, and practical uses of semantic segmentation and instance segmentation in the field of forestry.

It is imperative to recognize that the methodology employed for this review may possess certain constraints. Notwithstanding attempts to carry out a comprehensive review of the literature, there is a likelihood that certain pertinent studies may have been unintentionally excluded. Furthermore, the evaluation is predominantly grounded in peer-reviewed literature, with the exclusion of unpublished or gray literature sources. However, the studies that were chosen exhibit a varied spectrum of research endeavors and offer significant perspectives on the present status of research in this particular area.

3. Literature Review

3.1. Object Detection in Forestry

Because of its potential to give vital insights into forest management and conservation, semantic segmentation in forestry has become an essential topic of research. Much research in recent years has focused on creating and evaluating various algorithms and strategies for the semantic segmentation of forest photos. We explore 10 works that have contributed to this topic in this literature review.

The authors of the paper “Tree species classification of forest stands using multisource remote sensing data” [37], worked on creating a system that could identify tree species automatically using deep learning algorithms. Their goal was to make this system available on mobile devices. They identified tree leaves in images and utilized them to categorize the tree species. The authors used a U-Net architecture, a popular deep-learning model for medical image segmentation [32], to separate the leaves from the images. The U-Net model includes two networks: an encoder network that captures image features and a decoder network that generates the segmentation map. The authors employed two categories for the segmentation task: “leaf” and “background.” The U-Net model was trained with a dataset of 9,000 tree leaf images that were manually annotated with their respective species labels.

The authors employed a U-Net to classify the tree species after segmenting the leaves. The VGG16 model, a pre-trained CNN, was utilized for computer vision tasks. The VGG16 model was adjusted by using the segmented leaves and their respective species labels. The classification task was performed using a dataset of 10,000 images of trees, which included 20 different species.

The system was tested with a dataset of 900 tree images belonging to five distinct species by the authors. According to the findings, the accuracy rate for species classification was 93.3%.

The paper does not mention if the authors have made their code available for their method. The TensorFlow Lite framework was utilized for deploying the model on mobile devices, indicating that the method could potentially be implemented with this framework.

Lagos et al. [16] introduced the “FinnWoodlands Dataset” in their academic article, which is a dataset tailored for image analysis in the setting of Finnish woodlands. The authors’ primary focus was on segmentation tasks, wherein they provided valuable insights into the classes utilized and the specific objects that were segmented.

The article lacks clarity on whether the segmentation process was limited to trees or extended to encompass other objects present in the woodland area. Given the context of the dataset and the authors’ strong emphasis on image analysis in forests, it can be inferred that the segmentation task involved identifying and classifying various elements in the Finnish forests, such as trees, plants, leaves, the ground, and potentially other pertinent components.

Panoptic segmentation [14], is a relatively recent advancement in computer vision, and the work predates its release. Consider a world in which machines can perceive as humans do. They could distinguish between a tree and a bird, or between a forest and a city. This method enables machines to perceive and grasp the world in a surprisingly human-like manner. Panoptic segmentation is very useful in the field of tree/forest segmentation. Unlike older approaches, panoptic segmentation recognizes individual items and their context holistically. It’s like giving machines the ability to distinguish between trees and the entire forest. This technique improves accuracy in detecting and segmenting trees, especially in tightly packed forests. It is about making the unseen visible. From forest management to autonomous navigation, real-time use of this technology has the potential to transform numerous fields. However, Panoptic segmentation necessitates substantial processing resources, which may be a barrier for some applications. Integrating semantic and instance segmentation remains a problem for academics. While it works well in many situations, panoptic segmentation may suffer in certain settings, such as poor light or when the trees are of similar shape and size. Panoptic segmentation is a two-edged sword, whether it’s due to the painstaking precision or the computational needs. It is a game changer in tree/forest segmentation, but like any technology, it has its own set of obstacles.

As a result, there is no precise information available on how panoptic segmentation was utilized or if it was employed at all to discriminate tree species or other items in the dataset. The paper lacks explicit information on the methodology the authors used to distinguish between various tree types with regard to the differentiation of tree species. It is plausible that the differentiation of tree types was achieved through the utilization of diverse visual attributes, including but not limited to shape, texture, color, or a composite of these characteristics. Nonetheless, the precise approach to the classification of tree types has not been revealed.

The methods, models, or frameworks employed for segmentation tasks are not explicitly referenced in the paper. Given the characteristics of the dataset, it is conceivable that the researchers used traditional image analysis and computer vision techniques. Code available at Github.

Nevalainen et al. [26], in their paper, 'Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging', offer a unique deep learning strategy for identifying single-tree species in densely populated regions using hyperspectral data. The approach analyzes photos acquired across a semideciduous forest in the Brazilian Atlantic biome using 25 spectral bands spanning from 506 to 820 nm. A band combination selection step, feature map extraction, and multi-stage model refinement of the confidence map are all part of the network's design. In a complex forest, the technique obtained state-of-the-art performance for recognizing and geolocating each tree species in UAV-based hyperspectral pictures. When compared to a principal component analysis (PCA) methodology, the strategy is better.

Within the network's design, the authors estimate a combination of hyperspectral bands that contribute the most to the given goal. A unique deep-learning algorithm for hyperspectral imaging is proposed in this study to recognize and geolocate single-tree species in a tropical forest [21]. The strategy is intended to deal with a crowded scene and the Hughes effect. Within the network's design, the suggested technique seeks to estimate a combination of hyperspectral bands that contribute the most to the job. The architecture of the network decreases noise and improves performance in the given job. The suggested technique is successful under different scenarios, and the network's performance is commensurate with past deep learning studies.

The suggested approach may be used to detect *Syagrus romanzoffiana*, a palm tree important for forest regeneration, and can also be used in wildlife investigations. The proposed approach can also be used to identify other tree species, such as tapirs, which eat palm tree fruits and transmit their seeds through their excrement. This work provides a unique deep-learning algorithm based on a CNN architecture for detecting single-tree species in hyperspectral UAV-based photos with high dimensionality. The strategy was built with a band selection feature in the first phase, which was effective for dealing with high dimensionality and outperformed the baseline method that considered all 25 spectral bands and the PCA approach. Following the CNN architecture, feature map extraction and a multi-stage model refinement of the confidence map are performed.

The suggested technique performed exceptionally well at recognizing and geolocating trees in UAV-based hyperspectral pictures, with f-measure, precision, and recall values of 0.959, 0.973, and 0.945, respectively. The method presented here is useful for monitoring forest environments while accurately identifying specific trees. The use of a unique hyperspectral camera from a UAV or aircraft to detect bark beetle damage in urban forests at the individual tree level has piqued researchers' curiosity. Peng et al. [31] in their paper, 'Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remote-sensing image semantic segmentation', employed convolutional neural networks, weighted and conventional support vector machines, and random forests (RF) to classify tree species using hyperspectral and photogrammetric data. Deep learning in remote sensing applications has resulted in numerous advances, including the detection of fir trees damaged by bark beetles in unmanned aerial vehicle images, oil palm tree detection and counting for high-resolution remote sensing images, and the use of deep convolutional networks for large-scale image recognition.

The use of a worldview-2/3 and LiDAR data fusion technique, as well as the application of convolutional neural networks for the simultaneous extraction of roads and buildings in remote sensing imagery, have also been investigated [13]. Deep learning's application in remote sensing data processing has also resulted in the creation of new applications and problems in the area. The work addresses numerous remote sensing investigations, such as the processing and evaluation of spectrometric, and stereoscopic images, radiometric correction of close-range spectral picture blocks, and enhancement item counts using heatmap regulation. It also looks at automated land cover categorization, land use mapping, change detection, and forest inventory are among the applications. Deep learning algorithms are constantly being developed, which has greatly improved the accuracy and efficiency of these remote sensing jobs [16].

The method was tested on two datasets: one with a single multispectral image and the other with a series of images taken over time [40]. The forest segmentation accuracy on both datasets was high, as achieved by the authors. The study conducted by the authors did not involve the segmentation of individual trees or the differentiation of tree species. The technique they suggested may have the potential to classify tree species in upcoming research.

The forest segmentation was performed using a U-Net CNN architecture by the authors. The U-Net weights were initialized using transfer learning with a pre-trained VGG-16 network. The network was implemented using the Keras deep learning framework. The code for the method was not included in the paper. The author shared information about the software and hardware utilized in their experiments.

Chen et al. [4] paper, "Individual Tree Species Identification Based on a Combination of Deep Learning and Traditional Features" wanted to classify tree species in a study region using machine learning algorithms on UAV-based hyperspectral data. They did not categorize the data too meticulously. They instead employed supervised learning to categorize the tree species based on the spectral properties of the UAV-based hyperspectral data.

For their investigation, the authors chose six tree species: Holm oak, Cork oak, Stone pine, Eucalyptus, Maritime pine, and Acacia. In the categorization procedure, they used these six classes as their target classes. The authors used two feature selection methods to extract features from hyperspectral data: the Sequential Forward Selection algorithm and the Mutual Information (MI) algorithm. To categorize the tree species, the scientists utilized a variety of machine learning methods, including Random Forest, Support Vector Machine (SVM), and k-Nearest Neighbors (k-NN). They also compared the algorithms' performance using several assessment measures, such as overall accuracy, precision, recall, and F1 score.

In terms of code availability, the authors did not specifically state whether code for their approach is available. They did, however, note that they utilized the R statistical program and associated packages to undertake data processing and analysis, implying that their technique might be implemented using these tools.

"Assessing potential of UAV multispectral imagery for estimation of AGB and carbon stock in conifer forest over UAV RGB imagery" by Gaden [6], classified various tree species by segmenting individual trees from Very High Resolution (VHR) RGB imagery. The user employed a U-Net CNN architecture to segment trees and a ResNet-50 CNN for classifying species.

The VHR RGB imagery was analyzed by the authors using the U-Net CNN architecture to exclusively identify trees through image segmentation. Although VHR pictures can provide high-precision measurements for classification techniques, they typically contain clouds and cast shadows, which create issues for trustworthy information extraction. This type of CNN is frequently employed for such tasks [15].

The ResNet-50 CNN model was used to classify each segmented tree into one of the six tree species. The authors utilized a set of features extracted from the RGB image of each segmented tree to differentiate between various tree species. These features were then fed into the ResNet-50 model as inputs. The ResNet-50 model was pre-trained on a large dataset of natural images using the transfer learning technique. Afterward, it was fine-tuned using the VHR RGB imagery dataset. The paper does not come with any code, but the authors have given a thorough explanation of their approach and findings.

The authors of the publication "Forest segmentation using a combination of multi-scale features and deep learning" [6] aimed to segment forests in high-resolution remote sensing images. The forest was divided into two groups: forest and non-forest. The authors did not distinguish between different types of trees within the forest category.

The method employed by the authors for forest segmentation was "multi-scale features and deep learning." The forest segmentation results were made more accurate by utilizing a deep learning framework that incorporated multi-scale image features. The authors used the Faster R-CNN model, which is a well-known deep-learning object detection framework. The model detected objects at different scales by extracting multi-scale features from the input image using a pyramid scheme. The

specific pyramid scheme used for extracting multi-scale features from the input image was not explicitly mentioned in the paper.

Pyramid schemes are approaches used in computer vision and image processing to extract multi-scale features from an input image. Image pyramids, which are a sequence of scaled-down reproductions of the original image at different resolutions, are used in these schemes. The image is shown at a different size at each level of the pyramid, with higher levels having poorer resolution.

Pyramid methods are used to record data at numerous sizes, allowing algorithms to evaluate images at various levels of detail. They contribute to addressing the problem of identifying objects of varying sizes and dealing with objects that appear at varied scales inside a picture.

Pyramid systems that are commonly employed in computer vision include Gaussian pyramids, Laplacian pyramids, and steerable pyramids.

By repeatedly applying a Gaussian filter to the source image and subsampling it, this method produces a Gaussian pyramid. The image is represented at a different scale and resolution at each level of the pyramid.

The Laplacian pyramid is derived from the Gaussian pyramid. The details, or residuals between the respective levels of the Gaussian pyramid and its extended counterpart are represented by each level of the Laplacian pyramid. It aids in capturing small details in the photograph.

Steerable Pyramids: Steerable pyramids are multi-scale representations that extract information from different orientations and scales using a mix of filters. They are especially effective for assessing photos with objects of various orientations.

Pyramid methods are used in a variety of computer vision tasks, including object identification, image segmentation, and feature extraction. Algorithms that extract multi-scale features can effectively handle objects of varying sizes and capture both fine-grained details and wider context information in the image.

You can find the code for this paper on GitHub. Python and the PyTorch deep learning framework were used to implement the code.

Stan et al. [36] utilized deep convolutional neural networks to perform semantic segmentation of forest regions in their paper, "Semantic Segmentation of Forest Regions Using Deep Convolutional Neural Networks." The goal was to divide various areas of the forest, including trees, roads, water bodies, and other types of land cover. The authors utilized two distinct datasets: the National Agriculture Imagery Program and the Spatio-Temporal Asset Catalog.

The study included six categories: tree, road, building, water, field, and others. The authors utilized spectral clustering, a method that groups pixels with comparable spectral features, to differentiate between various tree species.

The U-Net deep CNN architecture was utilized by the authors for semantic segmentation. The structure of U-Net consists of an encoder-decoder with drop out layers, which aid in preserving spatial information. Furthermore, the authors employed methods of data augmentation to expand their training dataset and prevent overfitting. The paper does not discuss the availability of code.

Ma et al. [23] suggested a method for automatically segmenting Terrestrial LiDAR Data (TLD) to distinguish individual trees in their paper, "Automated extraction of driving lines from mobile laser scanning point clouds." The researchers separated the trees in the TLD on an individual basis but did not identify the specific species of each tree. The research divided the trunks and branches of trees, along with the nearby plants. The approach they used for segmentation involved two steps: region growth and convex hull fitting. To obtain the tree structure, the point cloud was segmented into regions and then fitted with convex hulls.

Ma et al. [23] evaluated their technique on a variety of datasets with varying levels of complexity. One of the datasets, for example, had trees with overlapping canopies. Their method's performance was tested using many quality criteria, including but not limited to completeness and accuracy. These metrics were used to evaluate the segmentation findings' correctness and effectiveness. Furthermore, the researchers compared their method to other sophisticated techniques available in the literature, albeit the specific techniques evaluated were not specified in the material provided.

Ma et al. [23] exhibited substantial levels of accuracy in segmenting individual trees from Terrestrial LiDAR data by comparing their method's completeness and correctness measures to those of other advanced algorithms. The authors did not make their code publicly available for this paper.

The study was conducted "Semantic segmentation of remote-sensing imagery using heterogeneous big data: International society for photogrammetry and remote sensing potsdam and cityscape datasets", by Song & Kim [23]. The aerial imagery was divided by the authors to isolate tree crowns. They then categorized each crown based on various forest inventory characteristics, such as tree species, height, diameter at breast height, and crown width. They separated both the trees and the vegetation in the surrounding background and understory.

The authors identified different tree species by analyzing spectral and spatial features obtained from segmented tree crowns. The researchers utilized U-Net, which is based on deep learning. The encoder network weights were initialized using transfer learning with a pre-trained VGG-16 network, which helped to enhance the model's performance.

The authors have made the work's code available as open source on GitHub. The code contains the U-Net architecture implementation, pre-processing procedures, and scripts for training and testing.

The authors of the study "The Semantic Segmentation of Standing Tree Images Based on the Yolo V7 Deep Learning Algorithm" by Cao et al. [2] provided a thorough method for semantic segmentation of standing tree images with the aim of differentiating between different tree types. Instead of just segmenting trees generally, the study concentrated on dividing tree areas into distinct tree species.

The authors employed the YOLO V7 deep learning algorithm, a commonly used approach at the time known for its potential effectiveness and precision in object identification tasks, to perform the segmentation and classification tasks. Using the YOLO V7 network, the input photos were first pre-processed, after which they were segmented, and finally, the resultant tree areas were classified into the appropriate species.

Numerous tree species that were pertinent to the geographical region under examination were included in the classifications employed in this study. Although the authors did not state how many classes they intended to have, it is clear that they wanted to include a wide variety of tree species. The YOLO V7 algorithm made it easier to distinguish between different tree kinds based on the distinctive traits and qualities that each species possesses, such as the texture of the bark, the form of the leaves, the branching patterns, and the general morphology. In terms of methodologies, models, and frameworks, the YOLO V7 deep learning algorithm was the main segmentation and classification tool that the authors used. To improve the performance of the model, they added further strategies to the process, such as data augmentation and transfer learning. The original query omitted any particular implementation frameworks and specifics.

Regarding code accessibility, the authors have not made the research's source code available to the general public.

3.2. Semantic Segmentation in Forestry

Because of its potential to give vital insights into forest management and conservation, semantic segmentation in forestry has become an essential topic of research. Much research in recent years has focused on creating and evaluating various algorithms and strategies for the semantic segmentation of forest photos. We explore 10 works that have contributed to this topic in this literature review.

The authors of the paper "Tree species classification of forest stands using multisource remote sensing data" [37], worked on creating a system that could identify tree species automatically using deep learning algorithms. Their goal was to make this system available on mobile devices. They identified tree leaves in images and utilized them to categorize the tree species. The authors used a U-Net architecture, a popular deep-learning model for medical image segmentation [32], to separate the leaves from the images. The U-Net model includes two networks: an encoder network that captures image features and a decoder network that generates the segmentation map. The authors employed two categories for the segmentation task: "leaf" and "background." The U-Net model was

trained with a dataset of 9,000 tree leaf images that were manually annotated with their respective species labels.

The authors employed a U-Net to classify the tree species after segmenting the leaves. The VGG16 model, a pre-trained CNN, was utilized for computer vision tasks. The VGG16 model was adjusted by using the segmented leaves and their respective species labels. The classification task was performed using a dataset of 10,000 images of trees, which included 20 different species.

The system was tested with a dataset of 900 tree images belonging to five distinct species by the authors. According to the findings, the accuracy rate for species classification was 93.3%.

The paper does not mention if the authors have made their code available for their method. The TensorFlow Lite framework was utilized for deploying the model on mobile devices, indicating that the method could potentially be implemented with this framework.

Lagos et al. [16] introduced the “FinnWoodlands Dataset” in their academic article, which is a dataset tailored for image analysis in the setting of Finnish woodlands. The authors’ primary focus was on segmentation tasks, wherein they provided valuable insights into the classes utilized and the specific objects that were segmented.

The article lacks clarity on whether the segmentation process was limited to trees or extended to encompass other objects present in the woodland area. Given the context of the dataset and the authors’ strong emphasis on image analysis in forests, it can be inferred that the segmentation task involved identifying and classifying various elements in the Finnish forests, such as trees, plants, leaves, the ground, and potentially other pertinent components.

Panoptic segmentation [14], is a relatively recent advancement in computer vision, and the work predates its release. Consider a world in which machines can perceive as humans do. They could distinguish between a tree and a bird, or between a forest and a city. This method enables machines to perceive and grasp the world in a surprisingly human-like manner. Panoptic segmentation is very useful in the field of tree/forest segmentation. Unlike older approaches, panoptic segmentation recognizes individual items and their context holistically. It’s like giving machines the ability to distinguish between trees and the entire forest. This technique improves accuracy in detecting and segmenting trees, especially in tightly packed forests. It is about making the unseen visible. From forest management to autonomous navigation, real-time use of this technology has the potential to transform numerous fields. However, Panoptic segmentation necessitates substantial processing resources, which may be a barrier for some applications. Integrating semantic and instance segmentation remains a problem for academics. While it works well in many situations, panoptic segmentation may suffer in certain settings, such as poor light or when the trees are of similar shape and size. Panoptic segmentation is a two-edged sword, whether it’s due to the painstaking precision or the computational needs. It is a game changer in tree/forest segmentation, but like any technology, it has its own set of obstacles.

As a result, there is no precise information available on how panoptic segmentation was utilized or if it was employed at all to discriminate tree species or other items in the dataset. The paper lacks explicit information on the methodology the authors used to distinguish between various tree types with regard to the differentiation of tree species. It is plausible that the differentiation of tree types was achieved through the utilization of diverse visual attributes, including but not limited to shape, texture, color, or a composite of these characteristics. Nonetheless, the precise approach to the classification of tree types has not been revealed.

The methods, models, or frameworks employed for segmentation tasks are not explicitly referenced in the paper. Given the characteristics of the dataset, it is conceivable that the researchers used traditional image analysis and computer vision techniques. Code available at Github.

Nevalainen et al. [12], in their paper, ‘Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging’, offer a unique deep learning strategy for identifying single-tree species in densely populated regions using hyperspectral data. The approach analyzes photos acquired across a semideciduous forest in the Brazilian Atlantic biome using 25 spectral bands spanning from 506 to 820 nm. A band combination selection step, feature map extraction, and multi-stage model refinement of the confidence map are all part of the network’s

design. In a complex forest, the technique obtained state-of-the-art performance for recognizing and geolocating each tree species in UAV-based hyperspectral pictures. When compared to a principal component analysis (PCA) methodology, the strategy is better.

Within the network's design, the authors estimate a combination of hyperspectral bands that contribute the most to the given goal. A unique deep-learning algorithm for hyperspectral imaging is proposed in this study to recognize and geolocate single-tree species in a tropical forest [13]. The strategy is intended to deal with a crowded scene and the Hughes effect. Within the network's design, the suggested technique seeks to estimate a combination of hyperspectral bands that contribute the most to the job. The architecture of the network decreases noise and improves performance in the given job. The suggested technique is successful under different scenarios, and the network's performance is commensurate with past deep learning studies.

The suggested approach may be used to detect *Syagrus romanzoffiana*, a palm tree important for forest regeneration, and can also be used in wildlife investigations. The proposed approach can also be used to identify other tree species, such as tapirs, which eat palm tree fruits and transmit their seeds through their excrement. This work provides a unique deep-learning algorithm based on a CNN architecture for detecting single-tree species in hyperspectral UAV-based photos with high dimensionality. The strategy was built with a band selection feature in the first phase, which was effective for dealing with high dimensionality and outperformed the baseline method that considered all 25 spectral bands and the PCA approach. Following the CNN architecture, feature map extraction and a multi-stage model refinement of the confidence map are performed.

The suggested technique performed exceptionally well at recognizing and geolocating trees in UAV-based hyperspectral pictures, with f-measure, precision, and recall values of 0.959, 0.973, and 0.945, respectively. The method presented here is useful for monitoring forest environments while accurately identifying specific trees. The use of a unique hyperspectral camera from a UAV or aircraft to detect bark beetle damage in urban forests at the individual tree level has piqued researchers' curiosity. Peng et al. [31] in their paper, Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remote-sensing image semantic segmentation, employed convolutional neural networks, weighted and conventional support vector machines, and random forests (RF) to classify tree species using hyperspectral and photogrammetric data. Deep learning in remote sensing applications has resulted in numerous advances, including the detection of fir trees damaged by bark beetles in unmanned aerial vehicle images, oil palm tree detection and counting for high-resolution remote sensing images, and the use of deep convolutional networks for large-scale image recognition.

The use of a worldview-2/3 and LiDAR data fusion technique, as well as the application of convolutional neural networks for the simultaneous extraction of roads and buildings in remote sensing imagery, have also been investigated [13]. Deep learning's application in remote sensing data processing has also resulted in the creation of new applications and problems in the area. The work addresses numerous remote sensing investigations, such as the processing and evaluation of spectrometric, and stereoscopic images, radiometric correction of close-range spectral picture blocks, and enhancement item counts using heatmap regulation. It also looks at automated land cover categorization, land use mapping, change detection, and forest inventory are among the applications. Deep learning algorithms are constantly being developed, which has greatly improved the accuracy and efficiency of these remote sensing jobs [16].

The method was tested on two datasets: one with a single multispectral image and the other with a series of images taken over time [17]. The forest segmentation accuracy on both datasets was high, as achieved by the authors. The study conducted by the authors did not involve the segmentation of individual trees or the differentiation of tree species. The technique they suggested may have the potential to classify tree species in upcoming research.

The forest segmentation was performed using a U-Net CNN architecture by the authors. The U-Net weights were initialized using transfer learning with a pre-trained VGG-16 network. The network was implemented using the Keras deep learning framework. The code for the method was

not included in the paper. The author shared information about the software and hardware utilized in their experiments.

Chen et al. [4] paper, "Individual Tree Species Identification Based on a Combination of Deep Learning and Traditional Features" wanted to classify tree species in a study region using machine learning algorithms on UAV-based hyperspectral data. They did not categorize the data too meticulously. They instead employed supervised learning to categorize the tree species based on the spectral properties of the UAV-based hyperspectral data.

For their investigation, the authors chose six tree species: Holm oak, Cork oak, Stone pine, Eucalyptus, Maritime pine, and Acacia. In the categorization procedure, they used these six classes as their target classes. The authors used two feature selection methods to extract features from hyperspectral data: the Sequential Forward Selection algorithm and the Mutual Information (MI) algorithm. To categorize the tree species, the scientists utilized a variety of machine learning methods, including Random Forest, Support Vector Machine (SVM), and k-Nearest Neighbors (k-NN). They also compared the algorithms' performance using several assessment measures, such as overall accuracy, precision, recall, and F1 score.

In terms of code availability, the authors did not specifically state whether code for their approach is available. They did, however, note that they utilized the R statistical program and associated packages to undertake data processing and analysis, implying that their technique might be implemented using these tools.

"Assessing potential of UAV multispectral imagery for estimation of AGB and carbon stock in conifer forest over UAV RGB imagery" by Gaden [6], classified various tree species by segmenting individual trees from Very High Resolution (VHR) RGB imagery. The user employed a U-Net CNN architecture to segment trees and a ResNet-50 CNN for classifying species.

The VHR RGB imagery was analyzed by the authors using the U-Net CNN architecture to exclusively identify trees through image segmentation. Although VHR pictures can provide high-precision measurements for classification techniques, they typically contain clouds and cast shadows, which create issues for trustworthy information extraction. This type of CNN is frequently employed for such tasks [15].

The ResNet-50 CNN model was used to classify each segmented tree into one of the six tree species. The authors utilized a set of features extracted from the RGB image of each segmented tree to differentiate between various tree species. These features were then fed into the ResNet-50 model as inputs. The ResNet-50 model was pre-trained on a large dataset of natural images using the transfer learning technique. Afterward, it was fine-tuned using the VHR RGB imagery dataset. The paper does not come with any code, but the authors have given a thorough explanation of their approach and findings.

The authors of the publication "Forest segmentation using a combination of multi-scale features and deep learning" [6] aimed to segment forests in high-resolution remote sensing images. The forest was divided into two groups: forest and non-forest. The authors did not distinguish between different types of trees within the forest category.

The method employed by the authors for forest segmentation was "multi-scale features and deep learning." The forest segmentation results were made more accurate by utilizing a deep learning framework that incorporated multi-scale image features. The authors used the Faster R-CNN model, which is a well-known deep-learning object detection framework. The model detected objects at different scales by extracting multi-scale features from the input image using a pyramid scheme. The specific pyramid scheme used for extracting multi-scale features from the input image was not explicitly mentioned in the paper.

Pyramid schemes are approaches used in computer vision and image processing to extract multi-scale features from an input image. Image pyramids, which are a sequence of scaled-down reproductions of the original image at different resolutions, are used in these schemes. The image is shown at a different size at each level of the pyramid, with higher levels having poorer resolution.

Pyramid methods are used to record data at numerous sizes, allowing algorithms to evaluate images at various levels of detail. They contribute to addressing the problem of identifying objects of varying sizes and dealing with objects that appear at varied scales inside a picture.

Pyramid systems that are commonly employed in computer vision include Gaussian pyramids, Laplacian pyramids, and steerable pyramids.

By repeatedly applying a Gaussian filter to the source image and subsampling it, this method produces a Gaussian pyramid. The image is represented at a different scale and resolution at each level of the pyramid.

The Laplacian pyramid is derived from the Gaussian pyramid. The details, or residuals between the respective levels of the Gaussian pyramid and its extended counterpart are represented by each level of the Laplacian pyramid. It aids in capturing small details in the photograph.

Steerable Pyramids: Steerable pyramids are multi-scale representations that extract information from different orientations and scales using a mix of filters. They are especially effective for assessing photos with objects of various orientations.

Pyramid methods are used in a variety of computer vision tasks, including object identification, image segmentation, and feature extraction. Algorithms that extract multi-scale features can effectively handle objects of varying sizes and capture both fine-grained details and wider context information in the image.

You can find the code for this paper on GitHub. Python and the PyTorch deep learning framework were used to implement the code.

Stan et al. [36] utilized deep convolutional neural networks to perform semantic segmentation of forest regions in their paper, "Semantic Segmentation of Forest Regions Using Deep Convolutional Neural Networks." The goal was to divide various areas of the forest, including trees, roads, water bodies, and other types of land cover. The authors utilized two distinct datasets: the National Agriculture Imagery Program and the Spatio-Temporal Asset Catalog.

The study included six categories: tree, road, building, water, field, and others. The authors utilized spectral clustering, a method that groups pixels with comparable spectral features, to differentiate between various tree species.

The U-Net deep CNN architecture was utilized by the authors for semantic segmentation. The structure of U-Net consists of an encoder-decoder with drop out layers, which aid in preserving spatial information. Furthermore, the authors employed methods of data augmentation to expand their training dataset and prevent overfitting. The paper does not discuss the availability of code.

Ma et al. [23] suggested a method for automatically segmenting Terrestrial LiDAR Data (TLD) to distinguish individual trees in their paper, "Automated extraction of driving lines from mobile laser scanning point clouds." The researchers separated the trees in the TLD on an individual basis but did not identify the specific species of each tree. The research divided the trunks and branches of trees, along with the nearby plants. The approach they used for segmentation involved two steps: region growth and convex hull fitting. To obtain the tree structure, the point cloud was segmented into regions and then fitted with convex hulls.

Ma et al. [23] evaluated their technique on a variety of datasets with varying levels of complexity. One of the datasets, for example, had trees with overlapping canopies. Their method's performance was tested using many quality criteria, including but not limited to completeness and accuracy. These metrics were used to evaluate the segmentation findings' correctness and effectiveness. Furthermore, the researchers compared their method to other sophisticated techniques available in the literature, albeit the specific techniques evaluated were not specified in the material provided.

Ma et al. [23] exhibited substantial levels of accuracy in segmenting individual trees from Terrestrial LiDAR data by comparing their method's completeness and correctness measures to those of other advanced algorithms. The authors did not make their code publicly available for this paper.

The study was conducted "Semantic segmentation of remote-sensing imagery using heterogeneous big data: International society for photogrammetry and remote sensing potsdam and cityscape datasets", by Song & Kim [35]. The aerial imagery was divided by the authors to isolate tree crowns. They then categorized each crown based on various forest inventory characteristics, such as

tree species, height, diameter at breast height, and crown width. They separated both the trees and the vegetation in the surrounding background and understory.

The authors identified different tree species by analyzing spectral and spatial features obtained from segmented tree crowns. The researchers utilized U-Net, which is based on deep learning. The encoder network weights were initialized using transfer learning with a pre-trained VGG-16 network, which helped to enhance the model's performance.

The authors have made the work's code available as open source on GitHub. The code contains the U-Net architecture implementation, pre-processing procedures, and scripts for training and testing.

The authors of the study "The Semantic Segmentation of Standing Tree Images Based on the Yolo V7 Deep Learning Algorithm" by Cao et al. [24] provided a thorough method for semantic segmentation of standing tree images with the aim of differentiating between different tree types. Instead of just segmenting trees generally, the study concentrated on dividing tree areas into distinct tree species.

The authors employed the YOLO V7 deep learning algorithm, a commonly used approach at the time known for its potential effectiveness and precision in object identification tasks, to perform the segmentation and classification tasks. Using the YOLO V7 network, the input photos were first pre-processed, after which they were segmented, and finally, the resultant tree areas were classified into the appropriate species.

Numerous tree species that were pertinent to the geographical region under examination were included in the classifications employed in this study. Although the authors did not state how many classes they intended to have, it is clear that they wanted to include a wide variety of tree species. The YOLO V7 algorithm made it easier to distinguish between different tree kinds based on the distinctive traits and qualities that each species possesses, such as the texture of the bark, the form of the leaves, the branching patterns, and the general morphology. In terms of methodologies, models, and frameworks, the YOLO V7 deep learning algorithm was the main segmentation and classification tool that the authors used. To improve the performance of the model, they added further strategies to the process, such as data augmentation and transfer learning. The original query omitted any particular implementation frameworks and specifics.

Regarding code accessibility, the authors have not made the research's source code available to the general public.

The study by (Lim, Zulkifley et al. 2023) "Attention-Based Semantic Segmentation Networks for Forest Applications," developed and tested an optimal attention-embedded high-resolution segmentation network called HRNet + CBAM in order to classify non-forest and forest areas in Malaysia. The data and input are gathered using Landsat-8 satellite images from ten locations in Malaysia from 2016, 2018 and 2020 [19].

The manual annotation of images is conducted for efficient training of the model and data set is categorized into 20% for testing and 80% for training of the model. The learning rate, and optimizer, are among the hyperparameters that the basic HRNet model is tuned to. The mean Intersection over Union (mIoU) of this baseline HRNet model is 84.84%, accuracy is 91.81%, and loss is 0.6142. The Convolutional Block Attention Module (CBAM) is embedded into HRNet, leading to an improvement in performance to 92.24% accuracy, 85.58% mIoU, and 0.6770 loss. HRNet and HRNet + CBAM beat other models, such as U-Net, SegNet, and FC-DenseNet, when benchmarked against them with regards to precision and mIoU. [19]

Nevertheless, neither the availability of the code nor the specific framework that was utilized to create these models are specified. In order to manage huge datasets, the paper recommends employing more data with additional modifications beyond forests, trying different attention processes in various architectures, and exploring higher-end GPUs or alternate data loading methods.

In "Semantic Segmentation Network Slimming and Edge Deployment for Real-Time Forest Fire or Flood Monitoring Systems Using Unmanned Aerial Vehicles"" by (Lee, Jung et al. 2023), an innovative approach for employing drones outfitted with cutting-edge deep learning models to

monitor forest fires and floods in real time is utilized. Through the application of semantic segmentation models such as DeepLabV3 and V3+, the system effectively identifies and demarcates impacted regions from UAV-captured data. The use of channel pruning-based network slimming, which drastically lowers model size and computing requirements without sacrificing accuracy, is the primary advancement of the study [17].

The results indicate that for the FLAME dataset: mIoU accuracy of 88.29%: This indicates the mean Intersection over Union (mIoU) accuracy of the model in correctly identifying and delineating the regions affected by forest fires. Higher mIoU values signify better segmentation accuracy. Similarly for the FloodNet dataset, mIoU accuracy of 94.15%: Similarly, this indicates the mean Intersection over Union (mIoU) accuracy of the model, but for identifying flooded areas in the FloodNet dataset. Again, higher values imply better segmentation accuracy.

The slimmed models exhibit minimal performance loss compared to baseline networks but achieve a remarkable 20-fold increase in inference speed. Moreover, the reduction in model size and computational requirements by approximately 90% not only enhances processing efficiency but also slashes power consumption, prolonging drone endurance. This breakthrough paves the way for effective and energy-efficient UAV-based monitoring systems tailored for mitigating natural disasters like floods and forest fires, safeguarding lives and ecosystems with real-time insights.

The study by (Ma, Dong et al. 2023) titled as “Forest-PointNet: A Deep Learning Model for Vertical Structure Segmentation in Complex Forest Scenes” offers a semantic segmentation technique centered on the Forest-PointNet model, which was created especially to recognize the vertical arrangement of forest by using terrestrial LiDAR data. The model takes advantage of the benefits that the PointNet structure offers by using an optimization strategy that improves the extraction of local features. When used for semantic segmentation in complex forest environments, it maintains important spatial characteristics, guaranteeing precise identification of forest components. Terrestrial LiDAR scans that collect point clouds of forest habitats make up the data inputs; however, particular datasets are not mentioned. Although the deep learning framework used is still unknown, the Forest-PointNet model performs admirably. It achieves an average recognition accuracy of 90.98%, which is around 4% better than current approaches, especially when compared to PointCNN and PointNet++ [22].

The model outperforms segmentation techniques based on three-dimensional structural reconstruction and outperforms traditional machine learning techniques by doing away with the requirement for human feature engineering. The study indicates that the Forest-PointNet model offers a viable approach for tasks involving semantic segmentation in varied forest landscapes, demonstrating robust efficiency and adaptability in complicated environments, even though code availability has not been indicated.

A ground-based LiDAR point cloud semantic segmentation technique for complex forest undergrowth scenarios is presented in this paper by (Li, Liu et al. 2023). We build forestry point cloud datasets, that are fused with undergrowth point cloud features, and use the DMM module based on point DMM module for semantic segmentation as a deep learning technique. The LiDAR equipment used to gather the forestry dataset is backpack-style. The study also suggests a point cloud data annotation method based on single-tree positioning to address the challenges of occlusion in forestry environments, sparse distribution as well as the lack of a database along with large location scales and elevated data volume in point clouds representing forestry resource environments (Li, Liu et al. 2023) [18].

The study utilized the DMM module to integrate tree features and an energy segmentation function to build a critical segmentation chart with the goal to address the less-than-ideal fractal structures and the attributes of large data, large sale scenes, uneven sparsity disorder, and diversity, in forestry environments. Next, we employ cutpursuit to figure out the graph and accomplish semantic pre-segmentation. With its severe occlusion, difficult terrain, numerous return information, high density, and unequal scales, our approach closes the gap in the current deep models used for complicated forestry environment point cloud information. We provide pointDMM, an end-to-end

deep learning model that significantly enhances the intelligent analysis of complicated forestry environment scenarios by training a multi-level lightweight deep learning network.

Our approach shows good results for segmentation on the DMM dataset, with a 21% improvement in the identification accuracy of live tree compared to other methods, and an overall accuracy of 93% on the large-scale forest environment point cloud dataset DMM-3. This approach offers major benefits over manually conducted point cloud segmentation when it comes to retrieving feature data from TLS-acquired artificial forest point clouds. It also lays the groundwork for forestry Informa tionization, intelligence and automation.

Moving further, in order to accomplish semantic segmentation, the segmentation strategy covered in the article by (Mazhar, Fakhar et al. 2023) makes use of convolutional neural networks (CNNs), with an emphasis on encoder-decoder topologies. The technique known as “semantic segmentation,” which is used in medical imaging applications, involves assigning a class to each pixel in an image. The CNN’s encoder module is in charge of obtaining feature maps from the input pictures. The decoder then reconstructs these feature maps to regain the spatial resolution and generates segmentation predictions that are precisely pixel-by-pixel [24].

The study highlights many unique CNN architectures that perform well in semantic segmentation tasks. Famous for its ease of use and efficacy, the U-Net architecture is one such model that is widely used in medical picture segmentation. Another noteworthy architecture that is intended to learn distinguishing characteristics is the Dens-Res-Inception Net (DRINet), which has shown usefulness in the segmentation of CT images of the abdomen, brain tumors, and brain. It is suggested to use dense multi-scale connections to create the high-quality multi-scale encoder-decoder network (HMEDN), which would provide accurate semantic information needed for multi-modal brain tumor segmentation and pelvic CT scans. Furthermore, when trained with Dice loss and cross-entropy, Fully Convolutional Networks (FCNs) are evaluated for their uncertainty estimation and segmentation quality, especially in applications related to the brain, heart, and prostate.

The study lists a number of models that are more advanced, including the Multi-Scale Residual Fusion Network (MSRF-Net), which makes use of a Dual-Scale Dense Fusion (DSDF) block to improve multi-scale feature communication, and INet, which uses overlapping maximum pooling for sharper feature extraction. These techniques show significant improvements in model training efficiency and segmentation accuracy.

A wide range of medical imaging modalities are represented in the data types and input picture used in these studies. These include biomedical MRI, X-rays, endoscopic imaging, mammograms, brain CT scans, brain tumor images, abdominal CT scans, pelvic CT scans, multi-modal brain tumor datasets, prostate CT scans, heart CT scans, and images for pattern detection of interstitial lung disease (ILD). These numerous image types show the adaptability and strength of CNN-based methods by posing different segmentation opportunities and problems.

It is implied that well-known frameworks like TensorFlow or PyTorch are probably utilized given their extensive use in the area, even though the precise deep learning frameworks used to create these models are not stated explicitly.

The study’s findings highlight CNNs’ impressive performance in a variety of medical picture segmentation tasks, especially when using encoder-decoder architectures. Notable results include greater segmentation quality and uncertainty estimation with FCNs trained using Dice loss, improved multi-scale feature communication with the MSRF-Net, and improved accuracy and streamlined training procedures with the U-Net model incorporating robust connection.

To add further, based on another study conducted by (Li, Liu et al. 2023), the research uses ground-based LiDAR data to offer a sophisticated method for semantic segmentation of point clouds in intricate forest undergrowth situations. The fundamental approach uses a deep learning method called pointDMM, which effectively pre-segments semantics by utilizing a DMM unit and the cutpursuit algorithm. LiDAR point cloud data is the main sort of imagery that is used. It is painstakingly gathered using backpack-style LiDAR equipment, guaranteeing thorough coverage of forestry areas. The DMM dataset, particularly focused on the large forest habitat point cloud dataset identified as DMM-3, is essential to our investigation.

Given the nature of the deep learning techniques used, it is reasonable to assume that TensorFlow or a comparable deep learning framework was used, even though the precise deep learning framework used is not stated. The segmentation method efficiently addresses the difficulties presented by blockage, high density, complex topography, and uneven scales in forested environments. It involves the building of a crucial segmentation graph and utilizes an energy segmentation function.

The study presents important results, one of which is the incredible 93% accuracy on the DMM-3 dataset. Compared to current techniques, this accuracy represents a significant 21% boost in live tree recognition accuracy. This improvement demonstrates how well the pointDMM approach handles the complex and varied features found in forestry point cloud data. This method has significant advantages when it comes to the collection of feature information from artificial forest point clouds generated by terrestrial laser scanning (TLS). This underscores the method's potential to further technology, intelligence, and informatization in the forestry area.

Nevertheless, it is not made obvious whether the implementation details are available for additional studies or practical application because the code utilized in this study is not publicly available. In conclusion, this study presents a strong ground-based LiDAR point- cloud semantic segmentation method using pointDMM, shows appreciable gains in segmentation precision, and emphasizes noteworthy developments in feature extraction capabilities; however, the availability of the underlying code is still unknown.

Another study by (Zhang, Li et al. 2022) states that by contrasting three network variants—one without SE Block and RAM, one with just the SE Block, and the suggested SERNet—the study analyzes the effects of SE Block and RAM on semantic segmentation performance. SE Block improves the mean Intersection over Union (mIoU) by 1.49%, the Accuracy Factor (AF) by 1.29%, and the Total Accuracy (OA) by 1.40%, hence boosting feature representation and segmentation accuracy, particularly for the “Surface” and “Car” categories. RAM raises the mIoU by 0.31%, AF by 0.41%, and OA by 0.41%, but only slightly improves performance due to its focus on global information. The ISPRS Vaihingen and Potsdam datasets, which include DSM (Digital Surface Model) and IRRG (Infrared, Red, Green) photos, were the datasets used for this assessment [42].

TensorFlow and PyTorch are common frameworks used in these kinds of research. The findings show that when DSM data is included, the suggested SERNet model obtains improved segmentation accuracy, especially for vegetation categories. The study does, however, admit certain limitations, including the possibility of feature redundancy and adverse mutual influence as a result of the straightforward fusion method utilized to combine DSM and IRRG data. Furthermore, the computational overhead of SERNet is increased by its huge number of parameters.

All things considered, the study emphasizes how important it is to recalibrate features and transfer information across the network in order to improve the accuracy of semantic segmentation, especially for images from remote sensing with high resolution (HRRSIs). Although the article presents encouraging findings, it makes no mention of where the code is available for replication or additional research.

3.3. Instance Segmentation in Forestry

The earth's ecosystem relies on forests, which provide a habitat for many different types of plants and animals. Forestry studies rely on precise identification, mapping, and monitoring of various tree species, which is made possible through instance segmentation. The task of instance segmentation involves recognizing and pin-pointing distinct objects in an image, while also labeling each object accordingly.

In the 2018 paper “Instance Segmentation in Very High-Resolution Remote Sensing Imagery Based on Hard-to-Segment Instance Learning and Boundary Shape Analysis” by Gong et al. [7], the authors broke up individual trees in forest areas. They used a binary classification method to tell which places had trees and which did not. They used U-Net and as raw data, the writers used very high-resolution aerial photos, which helped them find and separate individual trees. In this study, they did not tell the difference between different kinds of trees.

The writers preprocessed the images by making them a set size and adjusting the pixel values so that they were all the same. They also used methods for data augmentation, like rotating and flipping, to make their collection bigger. Then, they trained their U-Net model using a set of manually identified images. The model was taught to figure out how likely it was that each pixel was a tree or not. The authors used a test set to see how well their model worked and found that it was very accurate. In terms of code availability, the writers of this work have not made the code public.

The objective of Panagiotidis et al. [30], was to identify single trees and calculate their diameters with the help of UAV images. The method used to identify individual trees involved detecting the crowns and estimating the diameters. They categorized the areas into tree and non-tree classes. The trees were only separated into segments, without any distinction being made between various tree species.

A Deep CNN was utilized to identify tree crowns and approximate tree diameters. Convolutional Neural Networks (CNNs) are a type of deep, feed-forward artificial neural network. They are typically used for visual imagery analysis by processing data using a grid-like architecture. Deep CNN, on the other hand, is a CNN improvement. Deep CNNs include more layers than conventional CNNs, allowing them to extract more complicated features from input data.

The fundamental distinction between CNN and Deep CNN is the depth of their networks. Deep CNNs include additional layers, which allow for the extraction of more complicated characteristics but also increase their computational complexity. CNNs and Deep CNNs are now used almost synonymously. Most CNNs are actually Deep CNNs [41]. The intricacy of the problem to be solved determines whether you use a CNN or a Deep CNN [10]. They utilized a U-Net architecture that was modified and included a VGG-16 encoder for feature extraction. A vast collection of images that were manually annotated was used to train the network.

To estimate the diameter, they utilized a regression model that takes the DCNN's extracted features as input and predicts the tree's diameter. The diameter measurements used to train the model were collected in the field. Regrettably, the code was not made accessible by the authors.

Hao et al. [11] demonstrated the use of convolutional neural networks to distinguish the crowns of trees from terrestrial laser scanning data. The authors aimed to improve forest inventory and management by distinguishing individual tree crowns as separate entities.

The authors used terrestrial laser scanning data to segment tree crowns and collect precise information about the tree's structure, such as its leaves, branches, and trunks. The authors used a method called binary segmentation to classify each point in the point cloud as either a tree or not. They did not differentiate between different tree species. The researchers used a two-stage CNN approach for the purpose of segmenting individual tree crowns. At first, a segmentation network was used to classify each point in the point cloud as a tree or not. Then, the region-growing algorithm was employed to cluster adjacent points on the trees into distinct tree crowns.

The segmentation network used in the study was the U-Net architecture with residual connections. During the training of the segmentation network, both labeled and unlabeled data were utilized. Labeled data consisted of manually segmented tree crowns, while unlabeled data consisted of unsegmented terrestrial laser scanning data. Data augmentation techniques were employed to increase the quantity of training data. Unfortunately, the code associated with the paper is not publicly available.

In Ostovar et al. [29], the authors wanted to use RGB images to find and describe tree species. They didn't break apart individual trees or other items. Instead, they put the whole picture into a certain tree species.

The authors used a deep learning method that mixed a CNN with an SVM algorithm to train their model on a set of RGB images of trees from different species. The CNN was used to pull out features from the raw pictures, which were then sent to the SVM for classification.

With their method, the writers were able to classify different tree types with a high level of accuracy. They also compared their method to other well-known machine learning methods to show that their deep learning method was better. As for whether or not code is available, the writers have not said whether or not code is available.

In “A novel deep learning method to identify single tree species in UAV-based hyperspectral images” Miyoshi et al. [25] used deep learning methods to find and identify trees from LiDAR and hyperspectral images. They cut each tree into pieces and put them into different kinds.

The authors used a region proposal network (RPN) based on the Faster R-CNN design to divide the trees into groups. This network pulls features from the LiDAR point clouds and hyperspectral pictures. The RPN makes object suggestions, which are then improved with a region-based fully convolutional network to make correct segmentation masks for each tree.

The writers used a DCNN with a ResNet-50 design to put the tree species into groups. The DCNN was taught with a big set of hyperspectral images and names for the types of trees they showed. When the tree segmentation masks and the tree species classifications were put together, the end classification results were found. The writers tried their method on a set of LiDAR and hyperspectral images from a mixed-species forest. Both finding trees and figuring out what species they were were done with high accuracy. The source for this study can be found on GitHub.

In Ocer et al. [28] the authors tried to come up with a deep learning method for separating out individual trees in UAV (unmanned aerial vehicle) images. They only cut trees into pieces, but they didn't tell the different kinds of trees apart.

For instance segmentation of trees, the authors used a deep learning system called Mask R-CNN (Region-based Convolutional Neural Networks with Masking). They first resized and normalized the images, and then used labeled data to teach the Mask R-CNN model what to do. The labeled data was made up of UAV shots of trees and ground-truth models that showed where each tree was and how big it was.

The writers tested how well their method worked on a test dataset and found that it was accurate based on measures like intersection over union (IoU) and mean average precision (mAP). They also compared their method to other cutting-edge methods and found that theirs was better. The writers did not make their code open to the public, which is a shame.

In the paper, “Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks” by Nezami et al. [27], the writers focused on identifying tree species using hyperspectral imagery and CNNs. In this study, they did not use segmentation.

The writers used a field spectrometer and a flying hyperspectral imaging sensor to gather hyperspectral data. Then, they used the hyperspectral data to teach a CNN how to sort tree types. Tree types like birch, cedar, fir, larch, pine, spruce, and other forest species were used as classes. The writers used a CNN with three convolutional layers and three fully connected layers to tell the difference between different kinds of trees. Before putting the hyperspectral data into the CNN, they also used principal component analysis (PCA) to reduce the number of dimensions in the data.

On their test set, the writers were able to sort things correctly more than 90% of the time. They also carried out tests to determine how variables like the number of spectral bands and the size of the training set affected classification accuracy. The writers did not include any code in the study.

He et al. [12], wrote a paper titled “Generative adversarial networks-based semi-supervised learning for hyperspectral image classification”, discusses the detection and segmentation of individual trees from high-resolution remote sensing imagery. They utilized a Generative Adversarial Network (GAN) for semi-supervised tree detection and instance segmentation. They utilized two categories, namely trees, and back-ground, to identify and isolate each tree present in the picture.

By using labeled data, the GAN was trained to differentiate between various types of trees, as per the authors. The applications of GANs are nearly unlimited, particularly in image identification and segmentation. CycleGAN, a sort of GAN that morphs images from one domain to another without the need for matched training data, is one such technique that stands out.

Let me explain this even more. Consider the following scenario: we have an array of trees, each representing a different type. The goal now is to distinguish between these categories, however, we only have limited labeled data. Here comes CycleGAN. It enables us to efficiently bridge the data gap by creating synthetically labeled data for image-to-image translation.

A Mask R-CNN model is used for actual tree recognition and instance segmentation. This concept goes beyond only answering the questions 'what' and 'where' to include addressing the question 'how many'. It's like having a super-intelligent sight that not only notices but also understands the differences between trees. The authors reported that a Mask R-CNN model was utilized for detecting trees and performing instance segmentation. You can find the code for this paper on the authors' GitHub page.

Based on an additional study conducted by (Wielgosz, Puliti et al. 2023), a new framework for segmenting point cloud instances is presented, which is adaptable and flexible and for different pipeline components to be added or removed as needed. This flexibility is essential because it allows new or different modules to be installed in place of specific components, like the instance segmentation module. These additional modules can be implemented in Java or C++, for example, and yet work flawlessly with the entire framework. For researchers and developers who might need to use a certain language for a job or optimization, this flexibility is essential [38].

An optimization module at the heart of the system is intended to improve segmentation by optimizing key parameters. Important parameter visualization tools are also included in the framework, which help in comprehending and modifying the segmentation models' performance. The study specifically draws attention to the TLS2trees instance segmentation pipeline's hyperparameter adjustment procedure. This pipeline, which was originally designed for tropical forests, was optimized to produce coniferous forest settings, proving the framework's flexibility to many environmental circumstances.

Additional research was done to assess how employing a semantic segmentation model designed specially to recognize coniferous tree stems affected the overall accuracy of tree instance segmentation. The results of the investigation showed that for the given data, hyperparameter adjustment greatly enhanced the segmentation output quality. Nevertheless, the performance was noticeably worse than with the default values when these adjusted parameters were used on an external dataset, the LAUTx dataset. This conclusion implies that larger databases of publicly available annotated point cloud data encompassing a wider range of forest types than those employed in this work are required in order to construct a more robust and transportable set of hyperparameters.

The study also explored the effects of various models for semantic segmentation, namely the P2T model with p2t semantic and fsct semantic models. It was found that the use of these various semantic segmentation models had a negligible effect on instance segmentation accuracy when optimized. On the other hand, in thick forests or forests with lots of low branches, like those containing non-self-pruning species, the instance segmentation based on the p2t semantic model showed reduced susceptibility to hyperparameter selections, making it more resilient.

To sum up, the research presents a versatile and adaptable framework for segmenting point cloud instances, highlighting the need of fine-tuning hyperparameters and the requirement for a variety of annotated datasets to enhance the resilience and applicability of segmentation models. Although the study shows notable advancements in certain situations, it also identifies areas in which additional research and refinement are required to increase the framework's generalizability across other forest types. Future researchers and developers looking to expand on this work may find it interesting because the article makes no reference of the framework's code being available.

Another study by (Zvorișteanu, Caraiman et al. 2022) states that by combining cutting-edge techniques from the semantic instance segmentation and the optical flow fields, the suggested remedy for semantic instance segmentation presents a revolutionary strategy. Reconciling the frequently incompatible needs of high precision and real-time processing capabilities is the main objective of this methodology. Semantic instance segmentation approaches have typically prioritized either increasing instance mask accuracy or sacrificing some accuracy in order to attain real-time speed. But the goal of this creative idea is to accomplish both. [43]

The fundamental technique uses a novel inference approach to lower processing expense while preserving high frame rates. The framework only does inference on each fifth frame of the video stream, as opposed to every frame. It efficiently propagates the segmentation information over

several frames by warping the results of the semantic instance segmentation network for the intervening frames using calculated motion maps. This method greatly shortens the time needed to process a single frame, enabling the system to operate at a remarkable rate of up to 50 frames per second (fps) for 1280 × 720 pixel video frames.

To further improve accuracy, depth maps are incorporated into the framework. The depth map aids in the segmentation process by restricting the data to a particular range, guaranteeing that the technique retains a high degree of precision while operating in real-time. This field has advanced significantly with the simultaneous focus on speed and accuracy, especially for applications that need to handle high-resolution video streams in real-time.

The paragraph gives a thorough explanation of the process and its benefits, but it doesn't name the specific frameworks—such as PyTorch, TensorFlow, or others—that are employed. The details of the underlying technological stack are still unknown due to this lack of clarity in the implementation framework. Furthermore, there is no indication of the source for this innovative method, therefore it is unclear if others within the region will be able to access or replicate the implementation.

In conclusion, the innovative approach to semantic instance segmentation achieves real-time performance without sacrificing accuracy by utilizing the combination of semantic instance segmentation and optical flow fields. The system can process video streams at 50 frames per second on pictures with a resolution of 1280 × 720 by employing motion maps for intermediate frames and executing inference every fifth frame. Depth maps increase accuracy even more by concentrating data within a narrow range. It is unclear which particular frameworks were utilized and whether the code is publicly available, which leaves some implementation details lacking.

To add further, another research carried out by (Guo, Gao et al. 2022) showed some different results. Feature Pyramid Network (FPN) for increased feature extraction and an upgraded Mask RCNN model with a Swin Transformer backbone are the key components of the approach to instance segmentation of pests in complex natural environments. The process starts with Labelme v4.9 annotations and then adds FPN and Swin Transformer blocks to the Mask RCNN framework. Region proposals are generated by a Region Proposal Network (RPN), and accurate feature alignment is guaranteed by RoIAlign. The training settings consist of a batch size of 2, a learning momentum of 0.90, a weight decay of 0.05, a learning rate of 0.00001, and a training duration of 6 hours, spanning across 100 epochs. The three main evaluation criteria are F1 score, recall, and precision [9].

The dataset comprises 987 JPEG photos, each with a resolution of 6240 × 4160 pixels, taken at the China Academy of Forestry Sciences' Tropical Forestry Experimental Center in 2021 using a mobile device. The dataset comprises 198 test photographs and 798 training images that were shot under different lighting situations and during natural daylight. TensorFlow is used for the model implementation, while Python 3.7 is used as the programming language. The system is outfitted with an Intel Xeon CPU E5-2643 v4, 96 GB of RAM, and an NVIDIA Tesla K40c GPU that has 12 GB of memory.

An accuracy of 87.23%, a recall of 90.95%, and an F1 score of 89.01% are shown by the performance measures. Comparing these results to typical Mask RCNN models using ResNet50 (MR50) and ResNet101 (MR101) backbones, segmentation accuracy has significantly improved. The enhanced model shows strong segmentation performance, able to handle circumstances with overlapping, occluded, shaded, and uneven lighting. By integrating the Swin Transformer and FPN, the accuracy of pest segmentation is much improved. The robustness of the approach is ensured by a comprehensive validation process conducted under a variety of scenarios and a well-documented experimental setup.

The study by (Guan, Miao et al. 2022) utilize of deep learning techniques to track forest fires have made significant progress in the past few years. Safeguarding forest resources and comprehending the geographic spread of forest fires depend on using drone technology and refining current models to improve segmentation quality and recognition accuracy. Because fires spread quickly and exhibit erratic behavior, it can be difficult to effectively detect fires in complex situations [8].

This work tackles two deep-learning challenges using the FLAME aerial imaging collection. First, it achieves an identification rate of 93.65% by classifying video clips into two distinct categories (no-fire and fire) through the utilization of channel domain attention mechanism as a novel image classification method. Secondly, it suggests a brand-new instance segmentation technique for early-stage forest fire detection and segmentation dubbed MaskSU R-CNN, which is based on the MS R-CNN model. To minimize segmentation inaccuracies a U-shaped network is used to reconstitute the MaskIoU branch. MaskSU R-CNN outperforms numerous cutting-edge segmentation models in the experiment, with an accuracy of 91.85%, retention of 88.81%, F1-score of 90.30%, and mean intersection over union (mIoU) of 82.31%.

The primary contributions of the paper are rebuilding the MaskIoU branch of the MS R-CNN using a U-shaped network, incorporating the novel attention mechanism (DSA module) into ResNet to enhance the extraction of features, and inventing a novel attention mechanism (DSA module) to better feature channel representation. The outcome is the MaskSU R-CNN model, which is very good in identifying and classifying forest fires in their early stages. However, the MaskSU R-CNN model shows great promise for autonomous fire monitoring across vast forest regions because to its adaptable architecture and superior performance.

The study by (Sani-Mohammed, Yao et al. 2022) aimed to map the standing dead trees in their study's "Instance segmentation of standing dead trees in dense forest from aerial imagery using deep learning." Particularly in natural forests, identifying existing dead trees is crucial for assessing the overall health condition of the forest, its capacity to store carbon, and the preservation of biodiversity. It appears that natural forests have bigger expanses, which makes the traditional field surveying method extremely difficult, unsustainable, time consuming and labor-intensive. Thus, an automated method that would be economical is required for efficient management of forests. Since the development of deep learning, machine learning has effectively produced remarkable outcomes [33].

By Employing a small training dataset of 195 images, this work has offered an improved Mask R-CNN Deep Learning method for recognizing and categorizing existing dead trees in a mixture of dense forest using CIR aerial photography. Initially, the image augmentation technique is combined with transfer learning to take advantage of the training dataset limitations. Next, we carefully chose the hyperparameters of our model to correspond to the structure of our data, which is photos of decaying trees. Lastly, an experiment dataset that was not subjected to a deep neural network model was utilized for a thorough evaluation in order to gauge the capacity for generalization of our model's performance.

Despite our rather low resolution (20 cm) dataset, the model produced encouraging results, with mean average precision, mean recall, and mean F1-Score of 0.85, 0.88, and 0.87, respectively. Thus, our approach may find application in automating the identification and division of standing dead trees for improved management of forests. This is important for both estimation of carbon storage in forests and preserving biodiversity.

4. Discussion and Conclusions

Deciding between semantic segmentation and instance segmentation can be challenging since each technique has its strengths and weaknesses and serves a distinct purpose.

The goal of semantic segmentation is to categorize individual pixels within an image into distinct classes, such as buildings, trees, or roads. This technique can help in comprehending the general layout of a setting and recognizing various areas or categories of items.

In contrast to other methods, instance segmentation focuses on identifying and categorizing distinct items in an image, such as particular individuals, vehicles, or plants. This technique offers a higher level of detail for individual objects and can be beneficial for tasks that demand accurate object detection and segmentation, like tallying the number of trees or gauging their dimensions.

Improvements in semantic and instance segmentation methods have made a substantial impact on a number of environmental applications, most notably disaster monitoring and management of forest. Numerous research works have investigated various models and techniques to improve the precision and effectiveness of these segmentation tasks. The results of several investigations are

compiled in this review, which emphasizes the approaches, outcomes, applications, difficulties, and suggestions for further investigation.

The incorporation of attention mechanisms into semantic segmentation structures is one of the most notable advances. It has been demonstrated that attention mechanisms, like those found in the Convolutional Block Attention Module (CBAM), greatly improve segmentation model performance. By enabling models to concentrate on the most important portions of the input data, these strategies enhance feature representation, which eventually raises accuracy. When working with complicated datasets—like those used for environmental monitoring—the application of attention processes is especially helpful since it makes it easier to discern minute changes in features.

The performance of deep learning models for applications that operate in real time is another significant advancement. Methods like channel pruning and network slimming have shown to be useful in boosting segmentation models’ inference time without sacrificing their accuracy. For applications where quick reaction is necessary, such as real-time disaster monitoring, this improvement is vital. These methods facilitate the implementation of segmentation models on edge devices by lowering the computational burden, increasing their usefulness in situations where computational resources are scarce.

Even with these developments, a number of obstacles still exist. Effectively managing huge datasets remains a major challenge. The widespread use of these technologies is limited by the requirement for high-end GPUs to process large amounts of data. Additionally, the validation and reproducibility of results are hampered by certain studies’ insufficient transparency on the code and frameworks employed. For the field to progress, it is imperative that data and code be made publicly available because this encourages replication and more innovation.

Moreover, segmentation models that incorporate extra features—like depth maps—have demonstrated potential for increasing accuracy and speeding up processing. Segmentation performance is improved and feature representations are enhanced through multi-modal data integration. For this strategy to be successful, though, complex model architecture and effective handling of data methods are needed.

The amalgamation of contemporary studies underscores the noteworthy advancements in semantic and instance segmentation, propelled by the integration of model optimization methodologies, attention mechanisms, and domain-specific modifications. Although these developments have increased the usefulness and application of segmentation models, problems with model transparency, data management, and computational efficiency still need to be resolved. Future research can improve the usefulness of segmentation technologies in crucial applications including disaster response, forest management, environmental monitoring, by concentrating on these areas. This will ultimately lead to more efficient and well-informed decision-making processes.

The optimal answer relies on the particular issue and demands of the current task. If you want to analyze the layout and prevalence of various types of trees in a forest, semantic segmentation could be a better option. If you aim to identify and examine individual trees, instance segmentation could be a more suitable option.

The papers mentioned above commonly use deep learning-based methods, specifically convolutional neural networks and their variations, along with popular frameworks and data. TensorFlow and PyTorch are among the well-known deep learning frameworks. The studies commonly use datasets obtained from high-resolution re-mote sensing images, including UAV-based images, hyperspectral images, multispectral images, and LiDAR data. The data chosen for analysis depends on the problem at hand and the level of detail needed.

Although hyperspectral and multispectral imaging may sound similar, they are not identical. Consider them to be closely related siblings with distinct characteristics. What role does remote sensing play in all of this? Remote sensing, like the parent in our family metaphor, is a larger phrase that incorporates both hyperspectral and multispectral imaging. Remote sensing is the practice of detecting and monitoring physical features of a region without direct contact, typically through satellite or airplane.

Table 1. Imaging comparison.

	Hyperspectral	Multispectral
Definition	Hyperspectral imaging involves capturing and processing a wide spectrum of light, allowing for detailed analysis beyond what the human eye can see.	Multispectral imaging, on the other hand, only captures light within specific, narrow bands. It's a pared-down version of its hyperspectral sibling.
Use	Due to its detailed nature, hyperspectral imaging is ideal for scientific, environmental and defense-related applications.	Multispectral imaging is commonly used in agriculture, meteorology, and satellite imaging.
Difference	Hyperspectral imaging produces more detailed color information due to its broad spectrum.	Multispectral imaging is less detailed but faster and more efficient.

Ultimately, the optimal method varies based on the particular issue and its intended use. Semantic and in-stance segmentation offer distinct benefits and can be useful in various situations. CNNs are crucial in the advancement of both semantic and instance segmentation approaches. CNNs have been a popular choice for computer vision academics and practitioners due to their deep learning capabilities. The particular issue and the desired level of precision determine the choice of data.

Table 2. Summary of findings.

Paper	Segmentati on Approach	Method	Data/ Input	Framework	Result/ Finding	Code Availabili ty
Gaden. (2020)	Semantic	CNN	High resolution RGB imagery	TensorFlo w	Overall accuracy of 87,7% for tree species classification	Available
Stan, Thompson , & Yorhees. (2020)	Semantic	CNN	Hyperspect ral imagery	TensorFlo w	Overall accuracy of 95.34% for forest regions semantic segmentation	Available

Zhang et al. (2019)	Semantic	CNN	High-resolution multispectral imagery	Not specified	Forest segmentation achieved with an overall accuracy of 92.3%	Not available
Kraljevic et al. (2020)	Semantic	U Net based on Deep learning	Aerial imagery	TensorFlow	F1-scores of 0.82-0.98 for tree species segmentation	Available
Chen et al. (2018)	Semantic	CNN	Hyperspectral imagery	Not specified	Overall accuracy of 90.36% for tree species classification	Not available
Hahnema n, Oliveira, & Cavalin. (2014)	Semantic	CNN	RGB and hyperspectral images	TensorFlow	Overall accuracy of 87.3% (RGB) and 94.7% (hyperspectral) for tree species identification	Available
He, Liu, Wang, & Hu. (2020)	Semantic	CNN with multi-scale features	High-resolution remote sensing images	Not specified	Overall accuracy of 89.7% for forest segmentation	Not available
Chen et al.(2020)	Instance	Machine learning algorithms (Random Forest, SVM, k-NN, ANN)	UAV-based hyperspectral data	Not specified	Overall accuracy of 96.2% (ANN), 95.8% (Random Forest), 95.1% (SVM), and 93.9% (k-NN) for tree	Not available

Species Classification						
Author(s)	Method	Model	Input Data	Platform	Performance Metrics	Availability
Wan et al. (2021)	Instance	Deep learning on mobile devices	High-resolution remote sensing images	TensorFlow Lite	Overall accuracy of 97.3% for tree species identification on mobile devices	Available
Ocer et al. (2020)	Instance	CNN	UAV images	TensorFlow	Average precision of 0.875 for instance segmentation of trees	Available
Ostover et al. (2019)	Instance	CNN-based encoder-decoder architecture	Aerial images	Not specified	Mean Intersection over Union (mIoU) of 0.82 for tree instance segmentation	Not available
Panagiotidis et al. (2017)	Instance	Deep learning for tree detection	UAV-based point clouds	TensorFlow	Overall accuracy of 94.2% for tree detection	Available
Zvoristeanu, Caraiman, Manta (2022)	Instance	Inference	Imagery Type	Not mentioned	50 fps on 1280 × 720 pixel images.	Not Available
Wielgosz et al. (2023)	Instance	Bayesian	Point cloud	C++ or Java	Improved segmentation, transferability	Not Available

issues noted.						
for baseline HRNet performance: overall accuracy was 91.81% and with induction of CBAM accuracy rose to 92.24%.						
Lim et al. (2023)	Semantic	HRNet with Convolutional Block Attention Module.	Landsat-8 satellite images.	Not Specified		Not Available
UAVs drones images and FloodNet for detection of floods and FLAME for detection of forest fire.						
Suhr, Lee, & Jung (2023)	Semantic	channel pruning for optimization of DeepLabV3 and DeepLabV3+.		Not Specified	mIoU accuracy of flame model showed an 88.29%, while mIoU for FloodNet was 94.15%.	Not available
Forest-PointNet recognized vertical structures in forest with an accuracy of 90.98%.						
Ma et al. (2019)	Semantic	Forest-PointNet	Terrestrial LiDAR scans	Not Specified		Not Available
CNN including U-Net, DRINet, HMEDN, INet, and MSRF-Net.						
Mazhar, Fakhar, Rehman (2023)	Semantic		Medical images	TensorFlow or PyTorch Likely	High accuracy, enhanced segmentation quality	Not Available

Li, Liu, Huang (2023)	Semantic	Point cloud	backpack-style devices.	Deep learning	accuracy on DMM-3: 93% improvement in live tree identification accuracy: 21%	Not available
Zhang et al. (2022)	Semantic	Feature fusion	DSM & IRRG	TensorFlow and PyTorch.	Accuracy: Maximum	Not available
Guo et al. (2022)	Instance	Mask RCNN	JPEG images from mobile	TensorFlow	Precision: 87.23% Recall: 90.95%	Not Available
Guan, Miao et al. (2022)	Instance	DSA module, ResNet and MaskIoU branch of the MS R-CNN	Images of forest fires.	not specified	Precision:91.85% Recall: 88.81% F1-score: 90.30% mIoU: 82.31%	Not available
Li, Liu and Huang (2023)	Semantic	Dynamic Multi-Scale Module)	Ground-based LiDAR scans	Not specified	Accuracy of 93%	Not available
Mohammad, Yao, & Heurich (2022)	Instance	Mask R-CNN	Color Infrared (CIR) aerial imagery and 195 training images.	Not specified	Mean Average Precision (mAP): 0.85 Average Recall: 0.88	Not available

Average F1-
Score: 0.87

4. Final Thoughts

The present research piece offers a thorough examination of semantic segmentation and instance segmentation within the context of forestry. The results indicate the significant potential of these methodologies in diverse forestry-associated activities, such as the categorization of tree species, evaluation of forest resources, ecological surveillance, and identification of the most suitable routes for tree felling. Semantic segmentation facilitates the precise and effective examination of forest landscapes for the purpose of sustainable forest management by utilizing advanced segmentation algorithms in conjunction with remote sensing data. It presents distinct advantages by offering granular, pixel-by-pixel image classification. The main benefits are that it allows the identification of various land cover categories in a forest landscape, such as shrubs, bare grounds, trees, and water bodies, can be classified using semantic segmentation. For the purpose of comprehending the entire terrain and efficiently managing forest resources, this thorough mapping is essential. Semantic segmentation helps with quality of habitat and vegetation distribution analysis by precisely classifying different types of vegetation. Ecological research and initiatives to conserve habitat depend on this information. Semantic segmentation identifies and measures various vegetation kinds, which aids in determining the makeup of a forest. Evaluating the distribution of species, forest density, and ecological diversity is made easier with the use of this data. Semantic segmentation can be used to track modifications to land cover and the health of forests over time. It makes it possible to identify the impacts of environmental variables like pollution and drought as well as deforestation and land degradation. Semantic segmentation can aid in the research of soil state and erosion patterns by categorizing and mapping ground cover. For the purpose of protecting soil and stopping land deterioration, this information is essential. Water resource management is aided by the ability of semantic segmentation to locate and map water bodies inside forest regions. This is critical to preserving the hydrological equilibrium and guaranteeing that forest ecosystems have access to water. Semantic segmentation helps in evaluating fire safety by identifying regions with dry vegetation and other fire-prone factors. Agroforestry techniques are also supported by semantic segmentation's ability to discriminate between various crop varieties and forest plants. Better management of mixed-use landscapes—where forestry and agriculture coexist—is made possible by this. Forest inventory methods are streamlined by automatically classifying forest elements (such as trees, underbrush, and routes) using semantic segmentation. This improves data collection efficiency and lessens the demand for manual fieldwork. Planning for conservation requires the precise habitat maps that semantic segmentation produces. Setting priorities for the protection and restoration of various regions is aided by the identification of important habitats and biodiversity hotspots. Semantic segmentation helps track the development of afforestation and reforestation operations. It makes it possible to assess the success of tree-planting campaigns and the growth of the vegetation. Precision forestry techniques benefit from a thorough characterization of forest components. This entails enhancing forest management operations through the focused application of resources including fertilizers, insect control, and irrigation.

For inventory management, the credibility of instance segmentation approaches is essential as it allows precise calculation, identification of species, and health evaluation of every tree. It is capable of analyzing the height, size of the crown, and branching patterns of trees. This comprehensive knowledge facilitates comprehension of the structure and development patterns of forests. By separating and examining individual trees, instance segmentation can detect unhealthy or deceased

trees. Early identification facilitates prompt management and intervention to stop the spread of illness and preserve the health of the forest. More accurate estimations of carbon storage and biomass in forest are made possible by precise identification of trees and measurements. Instance segmentation facilitates more effective planning and resource allocation for forestry by offering comprehensive data on plant locations and characteristics. As a result, logging operations, conservation initiatives, and habitat management become more effective. Large amounts of aerial or terrestrial imagery can be processed rapidly by automated instance segmentation algorithms, which makes routine monitoring of vast forest regions possible. When managing huge natural forests, where human surveying is impracticable, this scalability is very helpful. Comprehensive segmentation data makes it easier to conduct a variety of ecological research, including figuring out how species interact, assessing the effects of environmental changes, and figuring out how forests regenerate. This scientific finding is essential for well-informed management and conservation plans. When there is a forest fire or other natural calamity, instance segmentation can quickly locate the burned trees and the areas that were impacted. Planned recovery efforts and assessment of the effects on the forest environment depend heavily on data that is gathered through instance segmentation methods.

The ramifications of these segmentation methodologies transcend the realm of scientific investigation. From a societal standpoint, the precise recognition and portrayal of trees via semantic segmentation can contribute to the promotion of responsible and ecologically aware forestry practices. Through the reduction of superfluous tree removal and the maximization of resource utilization, these methodologies facilitate the conservation of biodiversity, foster ecological equilibrium, and bolster the enduring viability of forest ecosystems. Furthermore, the utilization of segmentation techniques can facilitate the evaluation of the effects of deforestation and alterations in land use on the well-being and adaptability of forests, thereby empowering knowledgeable decision-making for the purpose of preserving biodiversity and mitigating the effects of climate change.

The incorporation of semantic segmentation in forestry operations has the potential to revolutionize conventional methods of tree cutting and harvesting from a business perspective. Organizations have the ability to utilize these methods in order to enhance the allocation of resources, enhance operational effectiveness, and decrease expenses linked with the manual selection and marking of trees. Forestry enterprises can optimize their harvesting procedures, reduce inefficiencies, and increase output by precisely identifying and demarcating the crowns of individual trees. The effective incorporation of segmentation techniques into practical workflows may necessitate considerable investments in technology, data procurement, and personnel education, posing noteworthy obstacles that demand resolution.

This review identifies various research prospects to further the domain of semantic segmentation in forestry. The development of algorithms that are both robust and scalable is a crucial aspect of addressing the complexities of forest environments. These complexities may include densely vegetated areas, occlusion effects, and varying canopy structures. Furthermore, the integration of diverse remote sensing modalities, such as the amalgamation of imagery and LiDAR data, exhibits potential for enhancing segmentation precision and capturing intricate spatial details. An area of promising research involves investigating the potential of deep learning techniques, specifically convolutional neural networks and graph-based models, to improve segmentation performance and enhance generalization capabilities. This will prove fruitful in the further exploration of the usage of CNNs and/or graph-based models.

The table above provides an overview of research publications on image segmentation for tree and forest analysis. While the table provides useful information about segmentation approaches, methods/frameworks utilized, data inputs, and results/findings, it does not provide a deep comparison of the architectural choices taken throughout the articles.

A detailed review of the papers reveals that numerous architectural options were used to approach the segmentation challenge. Deep learning techniques such as convolutional neural networks for example, were often used, but the specific topologies differed across investigations. Notably, the U-Net design, which is well-known for its performance in semantic segmentation tasks,

was not referenced explicitly in the articles. This begs the issue of whether using U-Net or comparable architectures would have resulted in even greater improvements in performance and consistency across the experiments.

Furthermore, the absence of debates about the employment of transformer-based models, which have received a lot of attention in computer vision problems, is remarkable. Transformers have shown potential in gathering global contextual information, which can be critical for tree and forest analysis, according to their attention mechanisms. The lack of exploration or mention of transformer models in the articles highlights a potential option for future research, as well as the potential benefits they may bring in terms of improving segmentation accuracy and resilience.

The table gives an interesting overview of picture segmentation approaches used in tree and forest analysis. To acquire a greater knowledge of the accomplishments and prospective possibilities for additional research in this sector, a full comparison of architectural choices, including the exploration of popular designs such as U-Net, and the potential employment of transformer-based models, is required.

To conclude, the effectiveness of semantic segmentation and instance segmentation methods in tackling forestry related issues has been demonstrated.

The potential uses of these applications range from the classification of tree species and conducting forest inventories to monitoring ecological conditions and devising optimal strategies for tree harvesting. The techniques in question have significant social implications, as they play a crucial role in promoting sustainable forest management practices and conservation endeavors. Furthermore, the implementation of segmentation techniques can lead to enhanced operational efficacy and optimized resource allocation, thereby providing advantages to enterprises. The realization of the complete potential of segmentation techniques in forestry will depend on the resolution of technical and operational obstacles, as well as the continuation of research efforts.

When we consider the business implications of merging semantic and instance segmentation in forestry, we envision a future in which established procedures are revolutionized, efficiencies are boosted, and operating costs are lowered. This is more than a fantasy; it might become a reality.

Businesses can promote responsible and environmentally conscious forestry practices by precisely recognizing and portraying trees via semantic segmentation. This not only improves the brand's image but also ensures its long-term viability.

These techniques can considerably reduce unnecessary tree removal, resulting in biodiversity conservation and resource utilization maximization. More profit equals less waste. By properly identifying and demarcating individual tree tops, forestry firms may streamline harvesting methods, minimize inefficiencies, and boost productivity. This could result in more revenue and a more streamlined bottom line. Semantic segmentation can reduce expenses associated with manual tree selection and tagging, potentially resulting in significant cost reductions.

However, it is crucial to highlight that effectively incorporating segmentation techniques into real workflows may demand significant investments in technology, data acquisition, and personnel education. This raises serious issues that must be addressed.

The future of semantic segmentation in forestry offers a plethora of research options. The development of robust and scalable algorithms, the integration of various remote sensing modalities, and the research of deep learning approaches are just some of the areas that could stimulate industrial innovation and growth.

Finally, the success of these forestry segmentation strategies will be determined by overcoming technological and operational challenges, as well as continuing research initiatives. The potential is huge, and the rewards are substantial, not only for the businesses that use these tactics but also for the health of our planet's forests.

Acknowledgements: This project has received funding from the KDT Joint Undertaking (JU) under Grant Agreement No. 101095835. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Sweden, Czechia, Finland, Ireland, Italy, Latvia, Netherlands, Poland, Spain, Norway, Türkiye.

Conflict of interest: The authors declare that they have no competing interests.

References

- Andrada, M. E., De Castro Cardoso Ferreira, J., Portugal, D., & Couceiro, M. Testing different CNN architectures for semantic segmentation for landscaping with forestry robotics, 2020 <https://irep.ntu.ac.uk/id/eprint/41821/>
- Cao, L., Zheng, X., & Fang, L. The Semantic Segmentation of Standing Tree Images Based on the Yolo V7 Deep Learning Algorithm. *Electronics*, 2023, 12(4), 929. <https://www.mdpi.com/2079-9292/12/4/929>
- Chen, C., Jing, L., Li, H., Tang, Y., & Chen, F. Individual Tree Species Identification Based on a Combination of Deep Learning and Traditional Features. *Remote Sensing*, 2023, 15(9), 2301. <https://www.mdpi.com/2072-4292/15/9/2301>
- Chen, X., Wang, R., Shi, W., Li, X., Zhu, X., & Wang, X. An Individual Tree Segmentation Method That Combines LiDAR Data and Spectral Imagery. *Forests*, 2023, 14(5), 1009. <https://www.mdpi.com/1999-4907/14/5/1009>
- Dalponete, M., Bruzzone, L., & Gianelle, D. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote sensing of environment*, 2012, 123, 258-270. <https://www.sciencedirect.com/science/article/pii/S0034425712001320>
- Gaden, K. Assessing potential of UAV multispectral imagery for estimation of AGB and carbon stock in conifer forest over UAV RGB imagery (Master's thesis, University of Twente), 2020, <http://essay.utwente.nl/85200/>
- Gong, Y., Zhang, F., Jia, X., Mao, Z., Huang, X., & Li, D. Instance Segmentation in Very High Resolution Remote Sensing Imagery Based on Hard-to-Segment Instance Learning and Boundary Shape Analysis. *Remote Sensing*, 2021, 14(1), 23. <https://www.mdpi.com/2072-4292/14/1/23>
- Guan, Z., et al. (2022). "Forest fire segmentation from aerial imagery data using an improved instance segmentation model." *Remote Sensing* 14(13): 3159.
- Guo, Y., et al. (2022). "Precious Tree Pest Identification with Improved Instance Segmentation Model in Real Complex Natural Environments." *Forests* 13(12): 2048.
- Hafemann, L. G., Oliveira, L. S., & Cavalin, P. Forest species recognition using deep convolutional neural networks. In 2014 22Nd international conference on pattern recognition, 2014, (pp. 1103-1107). IEEE. <https://ieeexplore.ieee.org/abstract/document/6976909/>
- Hao, Z., Lin, L., Post, C. J., Mikhailova, E. A., Li, M., Chen, Y., ... & Liu, J. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 178, 112-123. <https://www.sciencedirect.com/science/article/pii/S0924271621001611>
- He, Z., Liu, H., Wang, Y., & Hu, J. Generative adversarial networks-based semi-supervised learning for hyperspectral image classification. *Remote Sensing*, 2017, 9(10), 1042. <https://www.mdpi.com/2072-4292/9/10/1042>
- Jakubowski, M. K., Li, W., Guo, Q., & Kelly, M. Delineating individual trees from LiDAR data: A comparison of vector-and raster-based segmentation approaches. *Remote Sensing*, 2013, 5(9), 4163-4186. <https://www.mdpi.com/2072-4292/5/9/4163>
- Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. Panoptic segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 9404-9413. http://openaccess.thecvf.com/content_CVPR_2019/html/Kirillov_Panoptic_Segmentation_CVPR_2019_paper.html
- Korznikov, K. A., Kislov, D. E., Altman, J., Doležal, J., Vozmishcheva, A. S., & Krestov, P. V. Using U-Net-like deep convolutional neural networks for precise tree recognition in very high resolution RGB (red, green, blue) satellite images. *Forests*, 2021, 12(1), 66. <https://www.mdpi.com/1999-4907/12/1/66>
- Lagos, J., Lempiö, U., & Rahtu, E. FinnWoodlands Dataset. In Scandinavian Conference on Image Analysis (pp. 95-110. Cham: Springer Nature Switzerland, 2023, https://link.springer.com/chapter/10.1007/978-3-031-31435-3_7
- Lee, Y. J., et al. (2023). "Semantic Segmentation Network Slimming and Edge Deployment for Real-Time Forest Fire or Flood Monitoring Systems Using Unmanned Aerial Vehicles." *Electronics* 12(23): 4795.
- Li, J., et al. (2023). "PointDMM: A Deep-Learning-Based Semantic Segmentation Method for Point Clouds in Complex Forest Environments." *Forests* 14(12): 2276.
- Lim, S. V., et al. (2023). "Attention-Based Semantic Segmentation Networks for Forest Applications." *Forests* 14(12): 2437.
- Long, J., Shelhamer, E., & Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, (pp. 3431-3440). http://openaccess.thecvf.com/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html { 1 }
- Lv, L., Li, X., Mao, F., Zhou, L., Xuan, J., Zhao, Y., ... & Du, H. A Deep Learning Network for Individual Tree Segmentation in UAV Images with a Coupled CSPNet and Attention Mechanism. *Remote Sensing*, 2023, 15(18), 4420. <https://www.mdpi.com/2072-4292/15/18/4420>

22. Ma, Z., et al. (2023). "Forest-PointNet: A Deep Learning Model for Vertical Structure Segmentation in Complex Forest Scenes." *Remote Sensing* 15(19): 4793.
23. Ma, L., Wu, T., Li, Y., Li, J., Chen, Y., & Chapman, M. Automated extraction of driving lines from mobile laser scanning point clouds. *Advances in Cartography and GIScience of the ICA*, 2019, 1, 12.
24. Mazhar, M., et al. (2023). "Semantic Segmentation for Various Applications: Research Contribution and Comprehensive Review." *Engineering Proceedings* 32(1): 21.
25. Miyoshi, G. T., Arruda, M. D. S., Osco, L. P., Marcato Junior, J., Gonçalves, D. N., Imai, N. N., ... & Gonçalves, W. N. A novel deep learning method to identify single tree species in UAV-based hyper-spectral images. *Remote Sensing*, 2020, 12(8), 1294. <https://www.mdpi.com/2072-4292/12/8/1294>
26. Nevalainen, O., Honkavaara, E., Tuominen, S., Viljanen, N., Hakala, T., Yu, X., ... & Tommaselli, A. M. (2017). Individual tree detection and classification with UAV-based photogrammetric point clouds and hyper-spectral imaging. *Remote sensing*, 2017, 9(3), 185. <https://www.mdpi.com/2072-4292/9/3/185>
27. Nezami, S., Khoramshahi, E., Nevalainen, O., Pölönen, I., & Honkavaara, E. Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks. *Remote Sensing*, 2020, 12(7), 1070. <https://www.mdpi.com/2072-4292/12/7/1070>
28. Ocer, N. E., Kaplan, G., Erdem, F., Kucuk Matci, D., & Avdan, U. Tree extraction from multi-scale UAV images using Mask R-CNN with FPN. *Remote sensing letters*, 2020, 11(9), 847-856. <https://www.tandfonline.com/doi/abs/10.1080/2150704X.2020.1784491>
29. Ostovar, A., Talbot, B., Puliti, S., Astrup, R., & Ringdahl, O. Detection and classification of Root and Butt-Rot (RBR) in stumps of Norway Spruce using RGB images and machine learning. *Sensors*, 2019, 19(7), 1579. <https://www.mdpi.com/1424-8220/19/7/1579>
30. Panagiotidis, D., Abdollahnejad, A., Surový, P., & Chiteculo, V. Determining tree height and crown diameter from high-resolution UAV imagery. *International journal of remote sensing*, 2017, 38(8-10), 2392-2410.
31. Peng, C., Li, Y., Jiao, L., Chen, Y., & Shang, R. Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remote-sensing image semantic segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019, 12(8), 2612-2626. <https://ieeexplore.ieee.org/abstract/document/8684908/>
32. Ronneberger, O., Fischer, P., & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18, pp. 234-241. Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28
33. Sani-Mohammed, A., et al. (2022). "Instance segmentation of standing dead trees in dense forest from aerial imagery using deep learning." *ISPRS Open Journal of Photogrammetry and Remote Sensing* 6: 100024.
34. Shi, L., Wang, G., Mo, L., Yi, X., Wu, X., & Wu, P. Automatic Segmentation of Standing Trees from Forest Images Based on Deep Learning. *Sensors*, 2022, 22(17), 6663. <https://www.mdpi.com/1424-8220/22/17/6663> [16]
35. Song, A., & Kim, Y. Semantic segmentation of remote-sensing imagery using heterogeneous big data: International society for photogrammetry and remote sensing potsdam and cityscape datasets. *ISPRS International Journal of Geo-Information*, 2020, 9(10), 601. <https://www.mdpi.com/2220-9964/9/10/601>
36. Stan, T., Thompson, Z. T., & Voorhees, P. W. Optimizing convolutional neural networks to perform semantic segmentation on large materials imaging datasets: X-ray tomography and serial sectioning. *Materials Characterization*, 2020, 160, 110119. <https://www.sciencedirect.com/science/article/pii/S1044580319304930>
37. Wan, H., Tang, Y., Jing, L., Li, H., Qiu, F., & Wu, W. Tree species classification of forest stands using multisource remote sensing data. *Remote Sensing*, 2021, 13(1), 144. <https://www.mdpi.com/2072-4292/13/1/144>
38. Wielgosz, M., et al. (2023). "Point2Tree (P2T)—Framework for Parameter Tuning of Semantic and Instance Segmentation Used with Mobile Laser Scanning Data in Coniferous Forest." *Remote Sensing* 15(15): 3737.
39. Yang, Y., Anderson, M. C., Gao, F., Wood, J. D., Gu, L., & Hain, C. Studying drought-induced forest mortality using high spatiotemporal resolution evapotranspiration data from thermal satellite imaging. *Remote sensing of environment*, 2021, 265, 112640. [5]
40. Yel, S. G., & Tunc Gormus, E. Exploiting hyperspectral and multispectral images in the detection of tree species: A review. *Frontiers in Remote Sensing*, 2023, 4, 1136289. <https://www.frontiersin.org/articles/10.3389/frsen.2023.1136289/full> [17]
41. Zhang, L., Xia, G. S., Wu, T., Lin, L., & Tai, X. C. Deep learning for remote sensing image understanding. *Journal of Sensors*, 2016. [https://www.hindawi.com/journals/js/2016/7954154/abs/\[7](https://www.hindawi.com/journals/js/2016/7954154/abs/[7)
42. Zhang, X., et al. (2022). "SERNet: Squeeze and excitation residual network for semantic segmentation of high-resolution remote sensing images." *Remote Sensing* 14(19): 4770.
43. Zvorișteanu, O., et al. (2022). "Speeding Up Semantic Instance Segmentation by Using Motion Information." *Mathematics* 10(14): 2365.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.