Technical Note

Unveiling the Potential of Drone-Borne Optical Imagery in Forest Ecology: A Study on the Recognition and Mapping of Two Evergreen Coniferous Species

Kirill Korznikov [1,2], Dmitriy Kislov [1], Tatyana Petrenko [1], Violetta Dzizyurova [1,3], Jiří Doležal [2,4], Pavel Krestov [1] and Jan Altman [2,5,*]

1   Botanical Garden-Institute FEB RAS, 690024 Vladivostok, Russia
2   Institute of Botany of the CAS, 379 01 Třeboň, Czech Republic
3   Faculty of Biology, Lomonosov Moscow State University, 119991 Moscow, Russia
4   Faculty of Science, University of South Bohemia, 370 05 České Budějovice, Czech Republic
5   Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague,
     165 21 Prague, Czech Republic
*   Correspondence: altman.jan@gmail.com

Abstract: The use of drone-borne imagery for tree recognition holds high potential in forestry and ecological studies. Accurate species identification and crown delineation are essential for tasks such as species mapping and ecological assessments. In this study, we compared the results of tree crown recognition across three neural networks using high-resolution optical imagery captured by an affordable drone with an RGB camera. The tasks included the detection of two evergreen coniferous tree species using the YOLOv8 neural network, the semantic segmentation of tree crowns using the U-Net neural network, and the instance segmentation of individual tree crowns using the Mask R-CNN neural network. The evaluation highlighted the strengths and limitations of each method. YOLOv8 demonstrated effective multiple-object detection (F1-score—0.990, overall accuracy (OA)—0.981), enabling detailed analysis of species distribution. U-Net achieved less accurate pixel-level segmentation for both species (F1-score—0.981, OA—0.963). Mask R-CNN provided precise instance-level segmentation, but with lower accuracy (F1-score—0.902, OA—0.822). The choice of a tree crown recognition method should align with the specific research goals. Although YOLOv8 and U-Net are suitable for mapping and species distribution assessments, Mask R-CNN offers more detailed information regarding individual tree crowns. Researchers should carefully consider their objectives and the required level of accuracy when selecting a recognition method. Solving practical problems related to tree recognition requires a multi-step process involving collaboration among experts with diverse skills and experiences, adopting a biology- and landscape-oriented approach when applying remote sensing methods to enhance recognition results. We recommend capturing images in cloudy weather to increase species recognition accuracy. Additionally, it is advisable to consider phenological features when selecting optimal seasons, such as early spring or late autumn, for distinguishing evergreen conifers in boreal or temperate zones.

Keywords: tree crown recognition; multiple-object detection; semantic segmentation; instance segmentation; unmanned aerial vehicle

1. Introduction

In recent years, significant advancements have been made in object recognition and image segmentation using neural networks, offering valuable opportunities for the utilization of airborne optical imagery in forestry and forest vegetation monitoring [1,2]. Numerous studies have focused on detecting, identifying, and enumerating trees in forest stands, as well as localizing rare, endangered, or invasive species [3–6]. Acquiring knowledge about the spatial distribution of various tree species plays a pivotal role in forest management and

scientific research, enabling a more comprehensive understanding of how different species are distributed across the landscape [7]. In a broader context, beyond nature protection or scientific research, tree recognition holds potential for applied forest management, such as in assessing the stocks of commercially exploited tree species.

The identification of individual trees in sparse forests, including those with distinctive crown shapes and colors, has been successfully accomplished using neural networks [8–11]. However, the task becomes more challenging when identifying outwardly similar trees, such as evergreen conifers, within dense natural forest stands [12–14]. Very few studies have explored the feasibility of distinguishing between various species of coniferous trees in boreal or temperate forests. Unlike broadleaf species, coniferous trees exhibit limited variability in crown branching and needle-shaped leaves, making species identification challenging. Additionally, the evergreen nature of most conifers, with consistent color throughout the year, further complicates the recognition process. Natesan et al. [15] applied the DenseNet semantic segmentation neural network to identify five coniferous tree species using RGB multi-temporal images captured from a drone, achieving a classification accuracy of 84%. Beloiu et al. [16] trained Faster R-CNN object detection neural networks to recognize three coniferous species (spruce, fir, and pine) and one broadleaf species (beech). Although single-species models effectively identified spruce and fir, multi-species models encountered increased false positive and false negative results for coniferous species.

Conifers play a crucial role in boreal and temperate forests and hold significant economic importance. Late-successional conifer species are prominent features that characterize old-growth forests. We address the issue of distinguishing between two evergreen coniferous species, Korean pine (*Pinus koraiensis* Siebold et Zucc.) and Manchurian fir (*Abies holophylla* Maxim.), using high-resolution RGB images taken from an inexpensive drone equipped with a conventional optical camera. These two species are the dominant species in the primary forests of the southern Ussuri taiga ecoregion [17,18]. Determining the spatial distribution of dominant coniferous trees is critical for understanding the structure of the tree layer and searching for the most valuable pristine habitats.

We utilized three different types of neural networks on the same dataset of training images—object detection, semantic segmentation, and instance segmentation methods. Object detection involves identifying the position of multiple objects in an image, semantic segmentation predicts the class of each pixel of the image, and instance segmentation generates a pixel-wise mask for every object in the image (see examples in [19]). We selected well-established and popular neural networks that represented each type. Our aim in this study was not to directly compare the performance of neural networks belonging to one class of neural networks, such as object detection or segmentation. Instead, our focus was on illustrating the outcomes achieved by employing neural networks from three distinct classes to solve the specific problem of recognizing the crowns of two coniferous evergreen tree species.

We conducted three different recognition tasks to simulate various research goals: multiple-object tree species detection using the YOLOv8 neural network, which aimed to map tree species (1); semantic segmentation of tree crowns using the U-Net convolutional neural network, which aimed to map pixel-level segmentation masks for each tree species (2); and instance segmentation of individual tree crowns using the Mask R-CNN neural network, which aimed to make pixel-level segmentation masks for each instance (3).

We utilized orthophotos instead of individual drone photos, which are better suited for real research tasks as orthophotos covering areas of tens to hundreds of hectares are necessary for identifying spatial or landscape patterns of tree distribution. To compare the results, we adopted a straightforward criterion to evaluate accuracy based on the number of correctly and incorrectly recognized objects, rather than calculating pixel-by-pixel accuracy.

2. Materials and Methods

2.1. Data Collection

Drone surveys were conducted to observe the canopy of old-growth forests in six locations in the southern part of the Ussuri taiga ecoregion in the Far East (Figure 1).
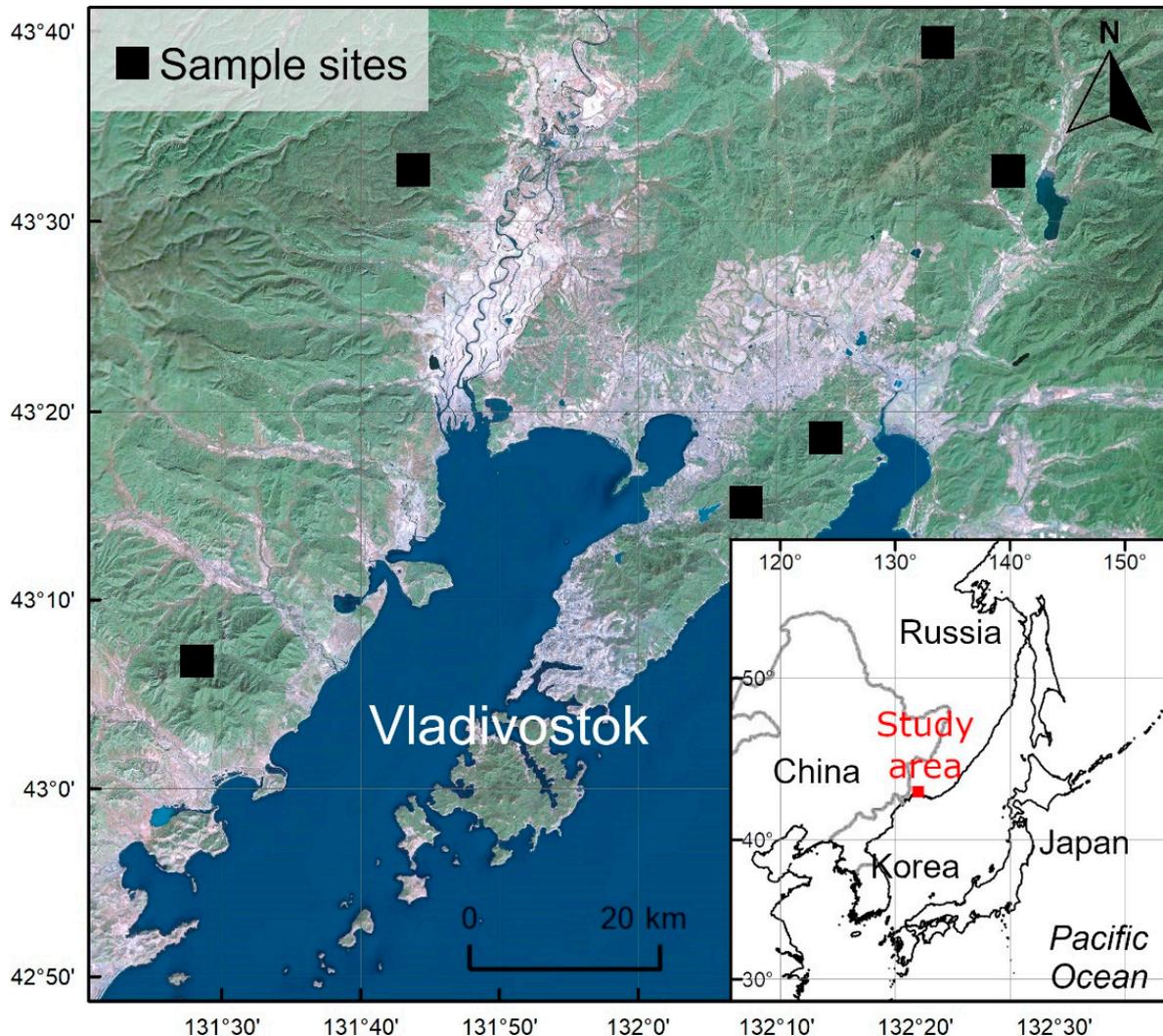

Figure 1. Location of the study area.

The survey utilized the DJI Mavic 2 Pro drone equipped with the Hasselblad L1D-20c aerial camera to capture high-resolution RGB images. The flight missions took place in March, April, and November 2021, covering the spring period before the broadleaf trees' leaf flush and the autumn period after leaf fall. It is worth noting that only Manchurian fir and Korean pine retained their leaves during this time. To ensure optimal image quality, data acquisition was scheduled on cloudy days to minimize the impact of direct sunlight. The flight operations were conducted using the Map Pilot Pro software for iOS [20], maintaining a flight height of approximately 300 m above ground level while following a grid-like trajectory over the area of interest (single grid flights). This allowed us to capture images at regular spatial intervals with the same ground sample distance. The resulting photo images had a resolution of 5472 × 3648 pixels, with overlays along and across the flight path set at 95% to ensure comprehensive coverage. The photo datasets were then processed using OpendDroneMap (ODM) software [21] with the default settings, producing orthophotos with a final resolution of ~10 cm pixel^{-1} (Figure 2).
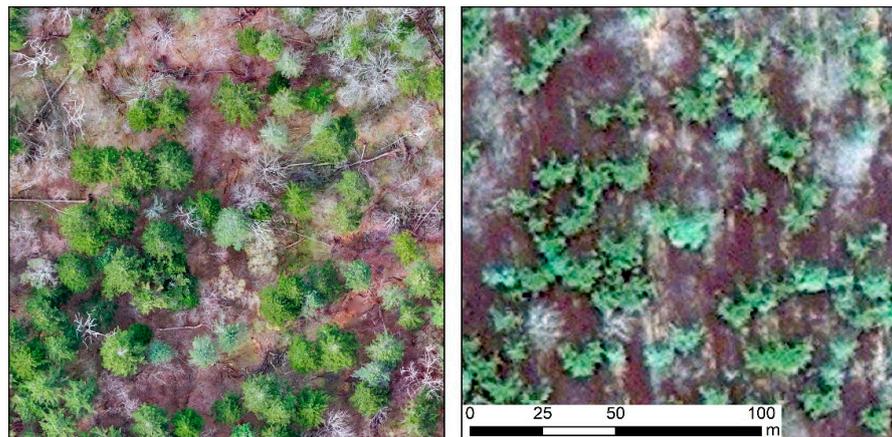
Figure 2. Comparison of the obtained orthophoto image (**left**) with a resolution of 10 cm pixel^{-1} and the Pleiades-1B satellite image (**right**) with a resolution of 50 cm pixel^{-1} (14 November 2017, id DS_PHR1B_201711140212170_FR1_PX_E131N43_0702_02902).

To delineate the tree crowns of Manchurian fir and Korean pine, we obtained 38 separate images from the orthophotos, each with a resolution of 1024 × 1024 pixels. The neural networks were trained using a dataset that included 32 images, and 6 images were set aside as a separate validation sample to assess the performance of the trained models. It is worth mentioning that the orthophoto images contained minor artifacts resulting from the stitching process of the original optical images. During the training phase, we intentionally included images of tree crowns with distortions and artifacts. This decision was made because our objective was to utilize neural networks to recognize tree crowns in orthophoto images rather than the original photos, which are generally of higher quality. A total of 249 Korean pine crowns were manually delineated, with a median of 6 crowns per image, and 421 Manchurian fir crowns were delineated, with a median of 10 crowns per image (Figure 3). We purposely excluded the marking of crowns belonging to young undergrowth trees, as our primary focus was on identifying mature trees that constitute the canopy. Additionally, understory trees do not provide sufficient information for species identification.
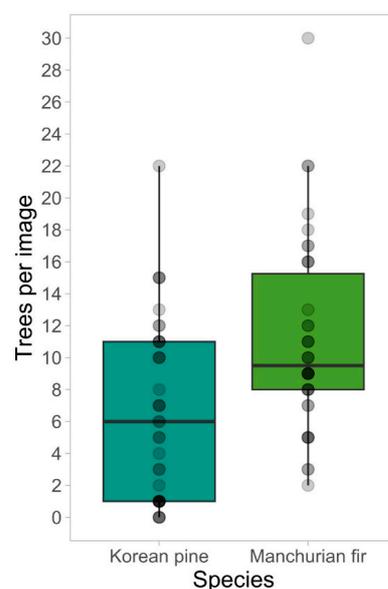

Figure 3. Distribution of the manually delineated tree crowns in the set of images used to train the neural networks.

Furthermore, to evaluate the object-based performance of the trained neural networks, a single high-resolution image measuring 4096 × 4096 pixels (covering an area of over 16 ha) was used. This test sample contained more than 300 tree crowns, which were expertly identified without their boundary delineations.

2.2. Neural Networks

The neural networks were trained using a GTX 1060 TI 6 GB GPU for up to 10 h. During training, the models were evaluated using the validation dataset, and training continued until there were no further improvements in scores for at least 1000 consecutive iterations. To ensure the models fit within the GPU memory, we followed the principle of making minimal changes to the default values of the configuration parameters. However, if necessary, we adjusted these accordingly.

For the U-Net model [22], the following parameter values were chosen for training: a batch size of 3 and an image size of 512 × 512 × 3. For the YOLOv8 model, a batch size of 8 was used. The other parameters were set to their default values, according to the Ultralytics repository [23]. For the Mask R-CNN model [24], we utilized the default configurations and a batch size of 3. During training, the Adam optimizer, with an initial learning rate of $1.e^{-4}$, was employed to optimize the models and update the weights.

We used the package [25,26] for augmentation and applied the following sequence of transformations: 1—random vertical and horizontal flips; 2—random rotation by 90 degrees; 3—ElasticTransform with parameters (alpha = 120, sigma = 6, alpha_affine = 3.6, *p* = 0.5); and 4—random brightness, contrast, and gamma adjustments with a probability of 0.8.

Variations in brightness and contrast throughout the augmentation process significantly improved the overall accuracy of crown recognition. This effect is likely due to the similarity in variability among the original images taken under various lighting conditions. To illustrate the significance of brightness and contrast augmentation, we conducted neural network training without employing this technique. Subsequently, we evaluated the performance of the models trained with and without brightness and contrast augmentation using a separate test image (2048 × 2048 pixels) that included the presence of minor fog as atmospheric distortion.

Loss function graphs for the U-Net and the Mask R-CNN models are presented in Figure A1. Graphs describing the learning process of YOLOv8 are presented in Figure A2.

To evaluate the accuracy of the models, we calculated several metrics. These included the number of true positive (TP), false positive (FP), and false negative (FN) results of tree crown detection. Additionally, we computed the following accuracy metrics: precision (1), which represents the percentage of correctly detected crowns among all the objects identified by the model; recall (2), also known as sensitivity or the true positive rate, indicating the percentage of correctly detected trees among all the reference trees; F1-score (3), combining both precision and recall, providing a comprehensive assessment of the model's performance; and overall accuracy (OA) (4), which provides a measure of the overall performance of the trained neural networks.

$$\text{PRECISION} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (1)$$

$$\text{RECALL} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (2)$$

$$\text{F1} = 2 \times \frac{\text{PRECISION} \times \text{RECALL}}{\text{PRECISION} + \text{RECALL}} \qquad (3)$$

$$\text{OA} = 2 \times \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \qquad (4)$$

3. Results

Among the algorithms assessed for tree crown recognition, YOLOv8 demonstrated the highest level of accuracy, achieving superior performance in accurately identifying species and localizing tree crowns within the image dataset. U-Net closely followed, attaining the second-highest accuracy. Mask R-CNN exhibited relatively lower accuracy scores compared to the other methods, indicating its comparatively lower effectiveness in accurately mapping tree crowns (Table 1).

Table 1. Achieved accuracy scores.

Object	TP	FP	FN	Precision	Recall	F1-Score	OA
Object-detection by YOLOv8							
Korean pine *Pinus koraiensis*	174	0	2	1	0.989	0.994	0.989
Manchurian fir *Abies holophylla*	134	3	1	0.98	0.993	0.985	0.971
Both species	308	3	3	0.99	0.990	0.990	0.981
Semantic segmentation by U-Net							
Korean pine *Pinus koraiensis*	174	5	1	0.97	0.994	0.983	0.967
Manchurian fir *Abies holophylla*	134	5	1	0.96	0.993	0.978	0.957
Both species	308	10	2	0.97	0.994	0.981	0.963
Instance segmentation by Mask R-CNN							
Korean pine *Pinus koraiensis*	149	5	26	0.97	0.851	0.906	0.828
Manchurian fir *Abies holophylla*	123	16	12	0.88	0.911	0.898	0.815
Both species	272	21	38	0.93	0.877	0.902	0.822

Abbreviations: TP—true positive, FP—false positive, FN—false negative, OA—overall accuracy.

Tree crown detection using the YOLOv8 model yielded promising results for Korean pine, with no false positive identifications and only two false negative identifications (Figure 4a). However, for Manchurian fir, there were three instances of false positive identifications. The flowering deciduous broadleaf trees of Korean poplar (*Populus suaveolens* Fisch.) were misclassified as Manchurian fir, with probabilities ranging from 0.30 to 0.60 (Figure 4b). Additionally, there was one false negative case where a small fir tree located at the edge of the image was not recognized. Despite these challenges, the YOLOv8 model produced accurate results for most of the tree crowns, even in scenarios where the crowns of different species overlapped (Figure 4c), the image quality was compromised due to artifacts in the orthophoto generation process, or when the crowns at the image boundaries were incomplete (Figure 4d).

Both the Mask R-CNN and the U-Net neural networks exhibited lower accuracy in object identification compared to the YOLOv8 neural network. However, the U-Net showed higher accuracy between the two segmentation methods. The occurrence of false positive and false negative results can be attributed to the misclassification of Manchurian fir as Korean pine and vice versa. It is worth noting that when YOLOv8 produced false positive identifications for images of flowering broadleaf Korean poplar trees, neither the Mask R-CNN nor the U-Net models generated such false positive results. However, false negative results for Korean pine were also observed. The false negative result produced by the YOLOv8 model for the small fir was also produced by the U-Net model, whereas the Mask R-CNN model correctly identified the crown of this tree. Overall, the Mask R-CNN model had a higher tendency for false negatives, whereas the U-Net had a higher incidence of false positives.
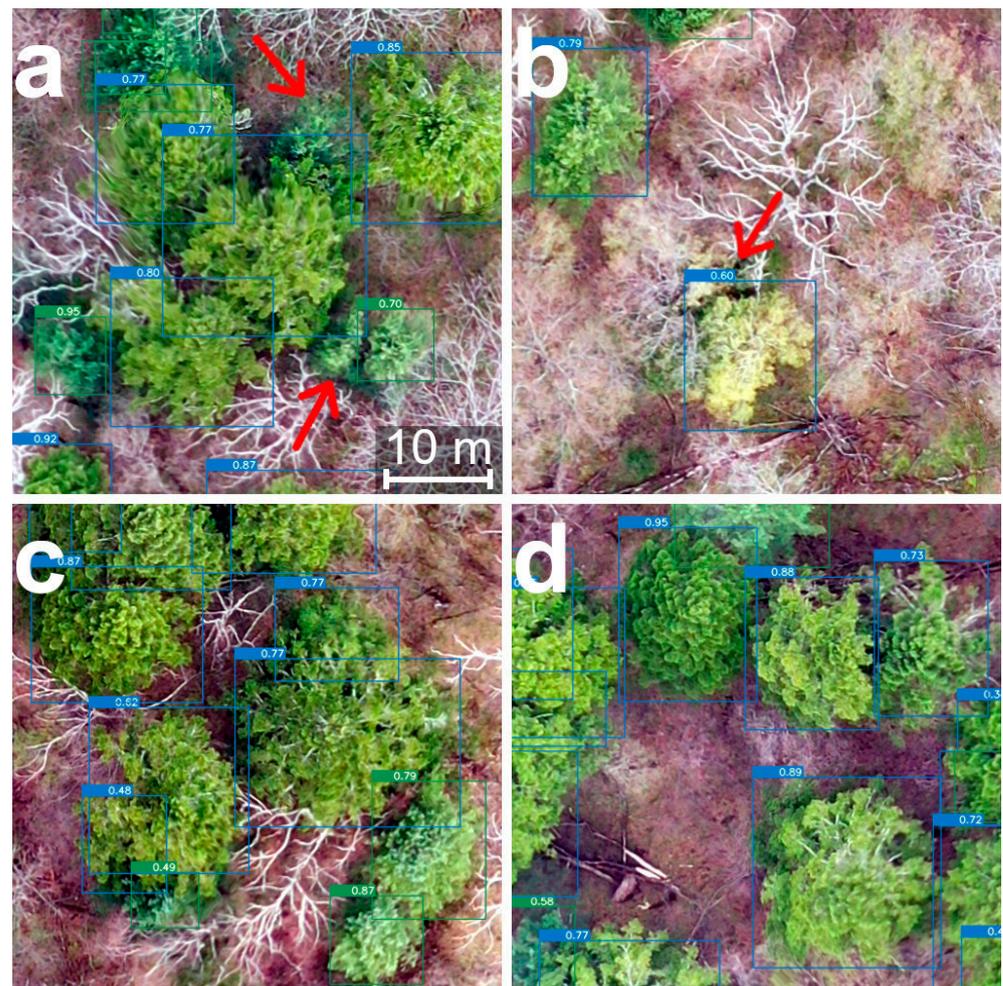
Figure 4. Results of tree crown detection using YOLOv8. The blue boxes represent the crowns of Manchurian fir (*Abies holophylla*) and the green boxes represent the crowns of Korean pine (*Pinus koraiensis*). (**a**) false negative results for *P. koraiensis* are indicated by red arrows; (**b**) false positive result for *A. holophylla* is indicated by red arrow; (**c**) correct results for partially overlapped crowns; (**d**) correct results for crowns located on an edge of the image (right side border).

The omission of brightness, contrast, and gamma adjustment augmentation in the training process for all three neural networks considerably reduced the recognition accuracy for images with atmospheric interference, such as fog. The absence of augmentation had a particularly severe impact on the YOLOv8 model. In the test image containing more than 70 crowns of coniferous trees partially covered by fog, the YOLOv8 model was able to detect only two trees of Manchurian fir and failed to identify any crowns of Korean pine (Figure A3). Similarly, there was a noticeable decline in recognition quality when brightness augmentation was not used for both the U-Net and the Mask R-CNN models (Figures A4 and A5).

Our results demonstrate that drone-based optical imagery, collected from an approximate height of 300 m and stitched into orthophoto images with a final resolution of approximately 10 cm per pixel, is capable of accurately identifying and mapping two coniferous species in dense mixedwood forests (Figure A6). However, the results obtained from the neural network approach for semantic segmentation and instance segmentation, which aim to map the two tree types onto orthophotos, exhibited lower accuracy compared to the YOLOv8 model, which is designed for object detection.

4. Discussion

In many practical forestry and forest vegetation monitoring tasks, the primary objective is often to map the spatial distribution of tree species [6,14,27–29]. Neural networks for object detection, such as the YOLOv8 model, are optimal tools for these tasks. Although there has been significant research interest in using neural networks for semantic and instance segmentation in tree canopy delineation, their practical application is not as widespread as simple mapping. The semantic segmentation of tree crowns can indeed be valuable for studying the dimensional characteristics of trees, determining height-to-crown size ratios, or quantifying the specific leaf area of trees belonging to different species. These tasks are crucial in the field of functional plant ecology and in the wood industry [13,29–32].

The segmentation algorithms used in our study had some limitations regarding false positive and false negative crown detection. Although the U-Net model, which is used for semantic segmentation, showed higher accuracy compared to the instance segmentation model Mask R-CNN, it is important to note that the results obtained through semantic segmentation do not apply to the detection of individual tree crowns. This limitation arises from the inherent nature of semantic segmentation methods, which do not delineate individual objects, making it challenging to accurately determine the number of trees in a given area (Figure A7). However, semantic segmentation methods still hold value in identifying and characterizing large-scale objects, such as windthrows or forested areas affected by forest pests [33–36]. They can also be useful in cases where tree species or other mapped plants are exclusively represented by single objects.

Instance segmentation methods, such as the Mask R-CNN model, allow for the recognition of individual tree crowns, enabling the mapping of individual trees and the calculation of crown sizes. It is important to note that the Mask R-CNN model exhibited the lowest recognition accuracy among the evaluated methods in our study. A common issue we encountered was the occurrence of tree crowns that simultaneously belonged to both recognized object classes (Figure A8). This ambiguity further complicated the precise delineation of individual tree crowns, resulting in a decrease in recognition accuracy. This finding highlights the challenge of accurately separating tree species with similar visual characteristics, even when employing advanced segmentation techniques. Hence, although the Mask R-CNN model serves as an instance segmentation method enabling the identification of individual tree crowns, in our research, its relatively lower accuracy does not render it suitable for fully automated tree crown detection tasks. Therefore, when utilizing the results of the Mask R-CNN model, expert supervision is necessary to ensure reliability of the findings.

Several studies have produced promising results for object detection and instance-based segmentation methods in simpler tasks, such as identifying isolated tree crowns of one species or without species identification [29,30,36,37]. Our research sheds light on the challenges that arise when dealing with more complex scenarios. Specifically, the intricate task of distinguishing between two or more similar tree species presents a more complex challenge [12,14–16]. In this context, we observed a decrease in the accuracy of segmentation methods. This highlights the need for further research and the development of segmentation techniques that can handle complex scenarios involving visually similar tree species.

Given the identified limitations and trade-offs, it is crucial to carefully choose the appropriate segmentation method based on the specific objectives. Semantic segmentation methods are valuable for characterizing large-scale objects, whereas instance segmentation methods provide more detailed information about individual tree crowns. Selection of the methodology should be guided by the desired outcome and the level of accuracy required for the intended applications. It is important to consider the limitations and challenges associated with each method and to weigh them against the specific goals of the study in order to make an informed decision.

Tree species recognition in natural or semi-natural forests presents particular challenges [16,30] compared to forest plantations or other types of homogeneous forest

stands [9–11,32,37]. Simply selecting and learning neural networks is not enough for achieving successful tree crown recognition. The process of capturing images with drones and creating a comprehensive training dataset with manually delineated trees is equally important. Involving experts with the capability to differentiate similar tree species visually and conducting ground-based verification are crucial steps to ensure the creation of high-quality training data [38].

We used drone images captured from a constant flight height of ~300 m above ground level to generate an orthophoto with a resolution of ~10 cm pixel^{-1}. This resolution proved sufficient for distinguishing between two similar evergreen tree species. Altering the flight altitude to achieve higher-resolution orthophotos could potentially improve recognition accuracy, particularly for segmentation algorithms. However, this approach comes with trade-offs. Lowering the flight altitude would necessitate capturing more photos of the same area and additional resources for orthophoto generation. Furthermore, higher-resolution orthophotos may introduce more artifacts due to minor changes in the tree crowns caused by wind movement, leading to difficulties in proper stitching. Conversely, increasing the flight altitude and obtaining lower-resolution orthophotos could result in lower recognition accuracy. The spatial resolution of very-high-resolution satellite systems of ~50 cm pixel^{-1} is inadequate for distinguishing between two selected evergreen coniferous tree species, as demonstrated earlier [33]. The choice of flight altitude and final image resolution should be carefully considered and tailored to the specific study and recognition objects. Ideally, researchers should have the opportunity to experimentally determine the optimal flight parameters based on empirical experiences [39,40].

Furthermore, the use of multispectral or hyperspectral data [41–43] can enhance recognition quality. Nevertheless, our study demonstrates that even with inexpensive drones equipped with a default RGB camera, satisfactory recognition results can be achieved.

The augmentation process during neural network training is another crucial aspect of achieving reliable recognition results [44]. For tree crown recognition using drone-borne aerial orthophoto images, it is important to include lighting changing transforms that emulate the variability of real atmospheric conditions. This includes variations in brightness associated with the time of day and brightness changes caused by aerosols in the atmosphere, such as dust, clouds, or fog (Figures A4–A6).

It is important to consider the influence of environmental conditions on the results of tree species recognition. In our study, we intentionally conducted drone surveys on cloudy days to avoid issues caused by uneven illumination from direct sunlight. This approach was necessary to minimize the presence of bright spots and shadows in the tree crowns, which could pose challenges in accurately distinguishing between the two conifer species [45,46]. Based on our experience, we recommend capturing images only in cloudy weather to increase the accuracy of species recognition. Taking this option helps to mitigate the potential impact of lighting conditions and to improve the overall reliability of the results.

Another factor that significantly contributed to the high recognition accuracy was the careful selection of the phenological period for the drone survey [47,48]. When addressing specific research objectives, it is crucial to consider the biological characteristics of the study objects and to select the appropriate seasons when the objects of interest are most distinguishable from other elements in the landscape. This requires a comprehensive understanding of phenology, including the phenological stages of the target objects and the overall appearance of the forest canopy [49]. By carefully choosing the right seasons, researchers can ensure that the objects they are studying are visually distinct and easily recognizable, thereby improving recognition accuracy.

For evergreen conifers in the boreal or temperate zones, the optimal seasons are early spring or late autumn. Although winter conditions with snow cover may initially seem suitable for the flight missions, there are several potential limitations to consider. Firstly, the presence of snow on the tree crowns can significantly hinder recognition accuracy, and the results obtained during this period may not be applicable to orthophotos captured

during snowless periods. Secondly, negative air temperatures during winter can greatly restrict the operational capabilities of many drones, as the performance of lithium batteries decreases significantly in low-temperature environments.

5. Conclusions

For tasks involving tree crown recognition for counting or mapping multiple tree species, dedicated neural networks designed for object detection and counting, such as the YOLOv8 model, are more suitable and reliable. Although more complex image segmentation algorithms can also yield satisfactory results for mapping, their accuracy may be lower, and the learning process may be longer and computationally intensive. Instance segmentation neural networks are primarily recommended for tasks involving the assessment of separate tree crowns, with results requiring careful expert validation.

We stress to carefully consider the specific research task and the complexity of object classification when selecting segmentation methods. More complex tasks, such as differentiating between visually similar tree species, may necessitate additional strategies or modifications to existing segmentation algorithms to enhance accuracy. The continuous development of robust and accurate segmentation methods for such intricate tasks is an ongoing focus of research in the fields of remote sensing and computer vision.

Solving practical problems related to tree recognition requires a multi-step process that involves collaboration among experts with different skills and experiences. It is essential to adopt biology- and landscape-oriented approaches when applying remote sensing methods, which requires proficiency not only in remote sensing and deep learning techniques but also in understanding the biological aspects of forest ecosystems. This approach will not only aid in collecting primary remote data but will also significantly enhance the quality of the final recognition results.

Author Contributions: Conceptualization, K.K. and D.K.; methodology, D.K.; software, D.K.; formal analysis, K.K., T.P. and V.D.; investigation, T.P. and V.D.; writing—original draft preparation, K.K. and J.A.; writing—review and editing, D.K., J.D. and J.A.; visualization, K.K., T.P. and V.D.; supervision, P.K.; funding acquisition, P.K. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.
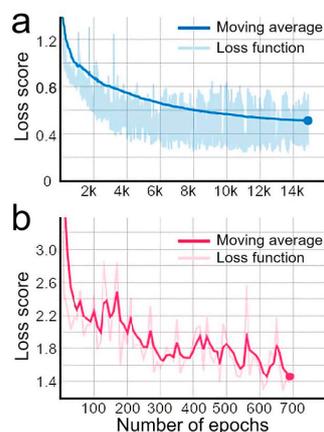
Appendix A

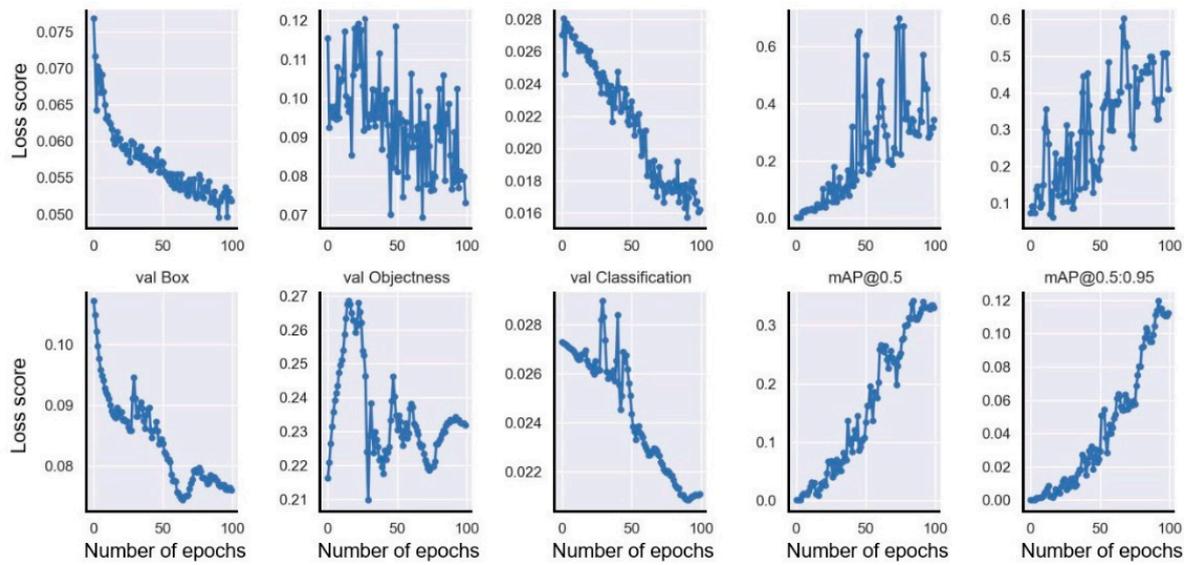**Figure A1.** Loss functions of U-Net (**a**) and Mask R-CNN (**b**) learning.

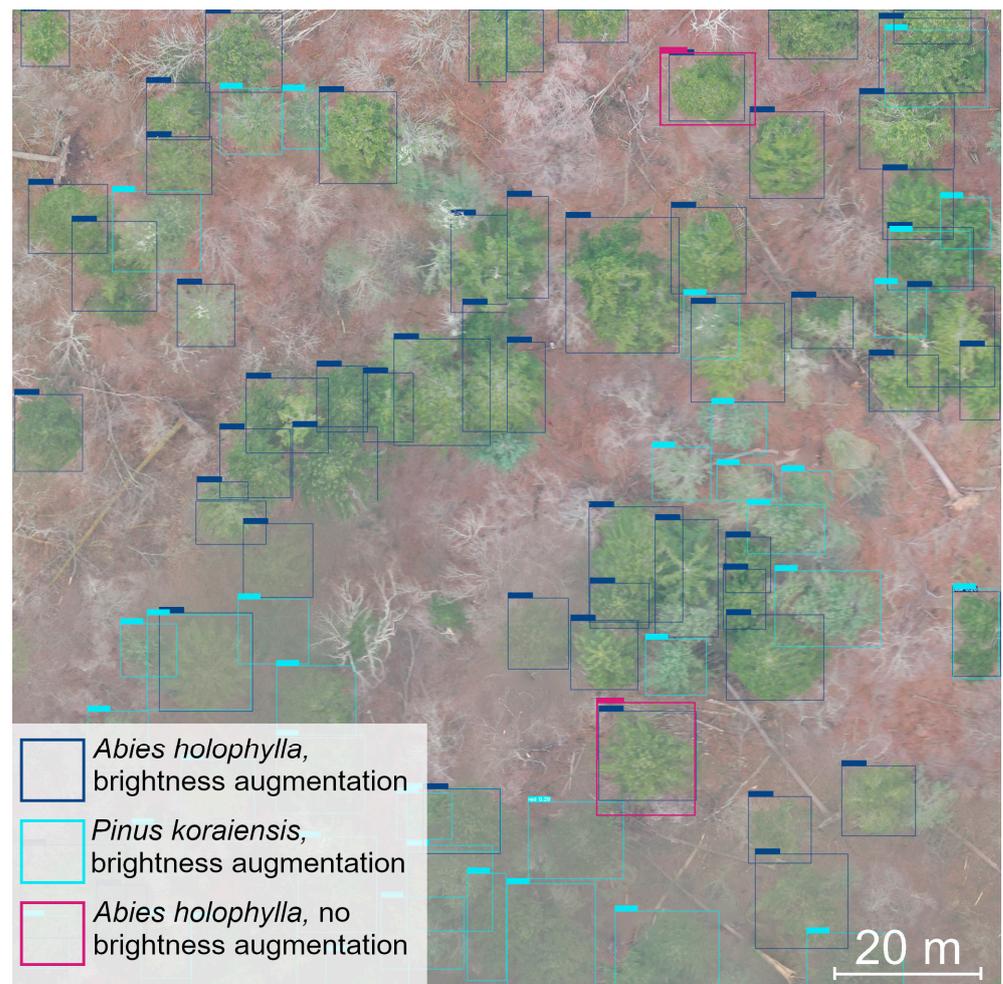**Figure A2.** Loss functions of YOLOv8 learning.

Figure A3. Recognition results of Korean pine (*Pinus koraiensis*) by YOLOv8 on a partially fogged fragment of the orthophoto, with and without brightness augmentation. The original colors of the orthophoto image have been retained.
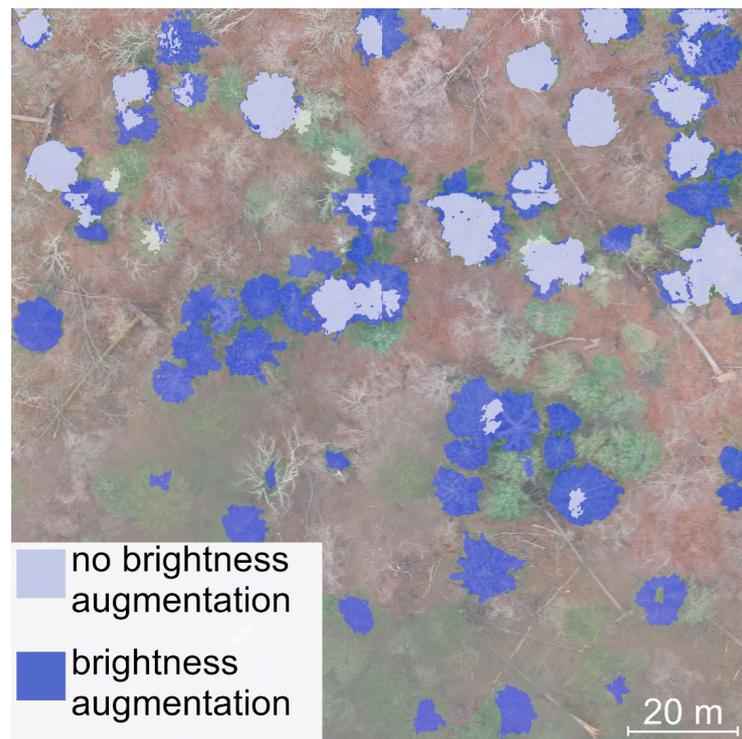
Figure A4. Recognition results of Korean pine (*Pinus koraiensis*) by U-Net on a partially fogged fragment of the orthophoto, with and without brightness augmentation. The original colors of the orthophoto image have been retained.
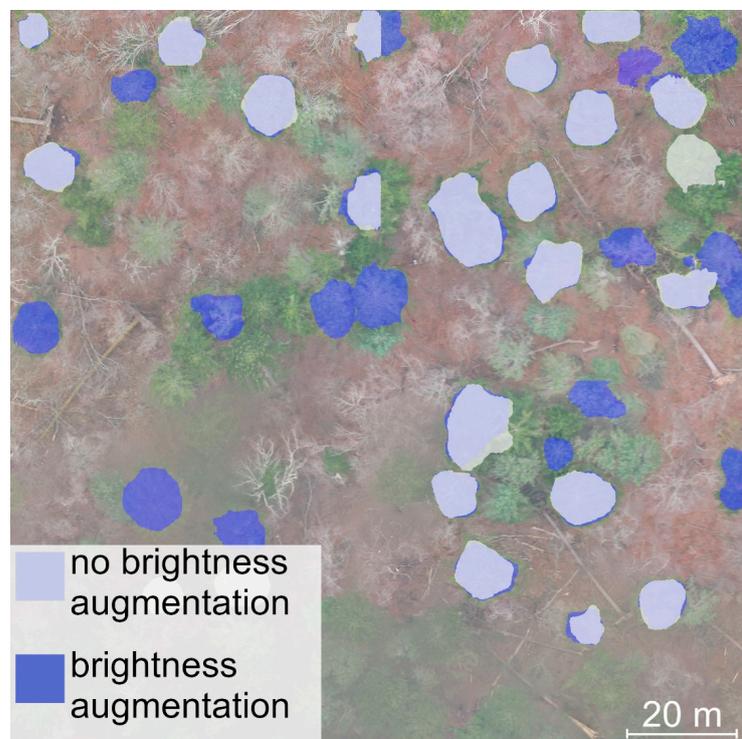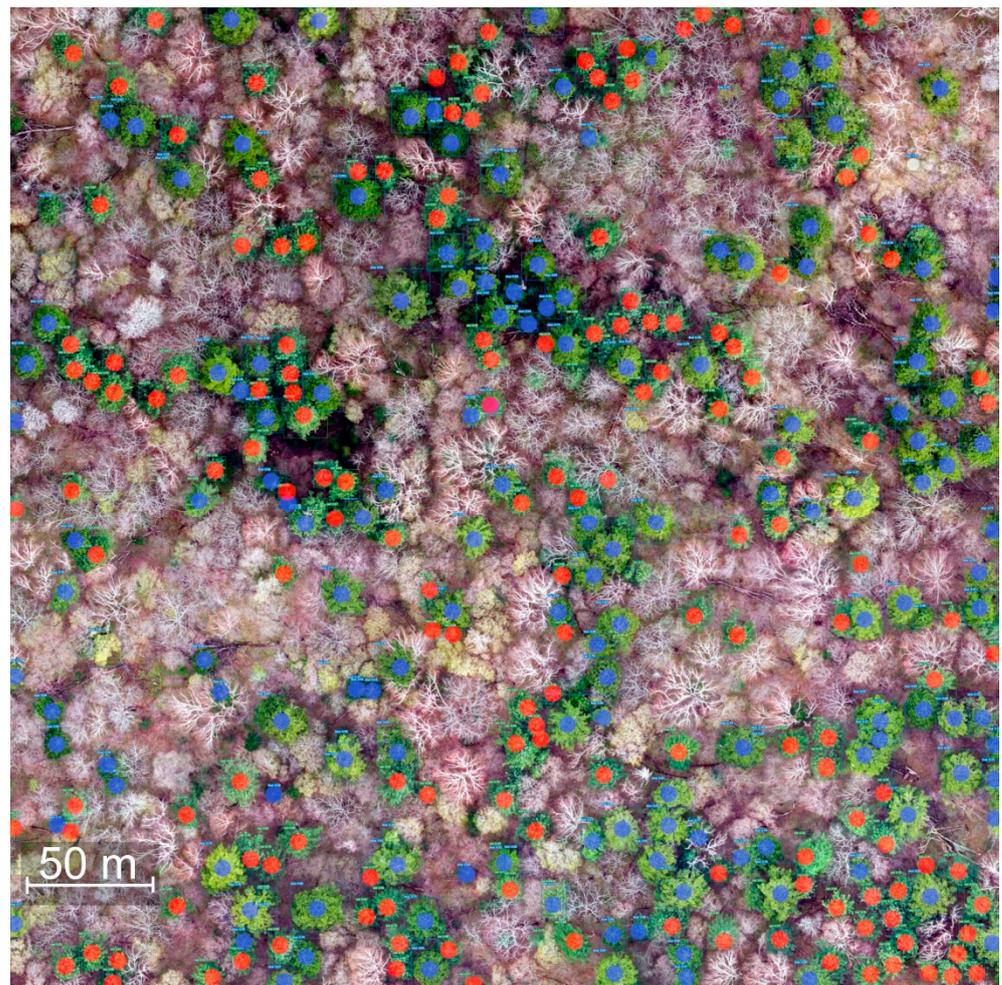
Figure A5. Recognition results of Korean pine (*Pinus koraiensis*) by Mask R-CNN on a partially fogged fragment of the orthophoto, with and without brightness augmentation. The original colors of the orthophoto image have been retained.

● Manchurian fir *Abies holophylla* ● Korean pine *Pinus koraiensis*

Figure A6. Recognition results by YOLOv8 on a mixedwood forest area of ~16 ha (the 4096×4096 pixels testing image), excluding a few false positive cases for both species.
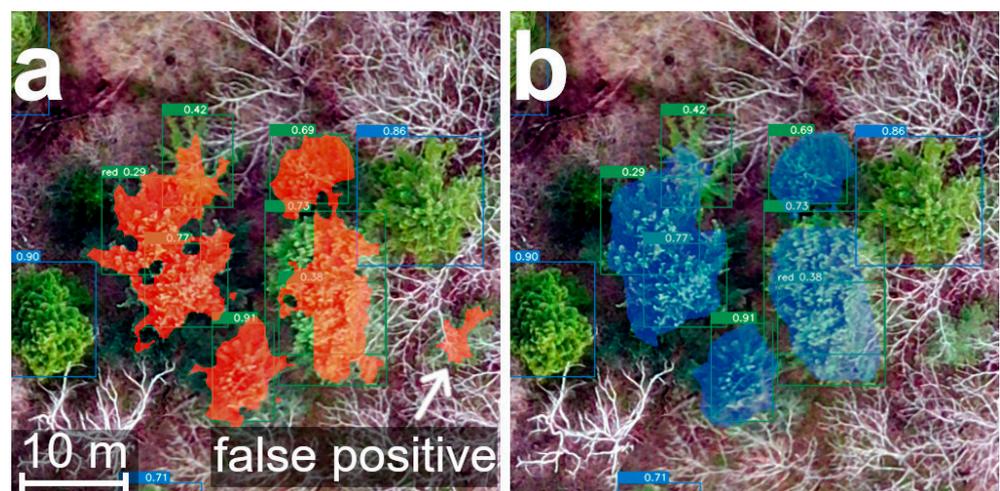

Figure A7. Korean pine (*Pinus koraiensis*) delineation results by U-Net (**a**) and Mask R-CNN (**b**). The green boxes represent Korean pine crown detections by YOLOv8, whereas the blue boxes represent the detection of Manchurian fir (*Abies holophylla*). Each delineated crown detected by Mask R-CNN is treated as a separate object, enabling the estimation of single crown areas.
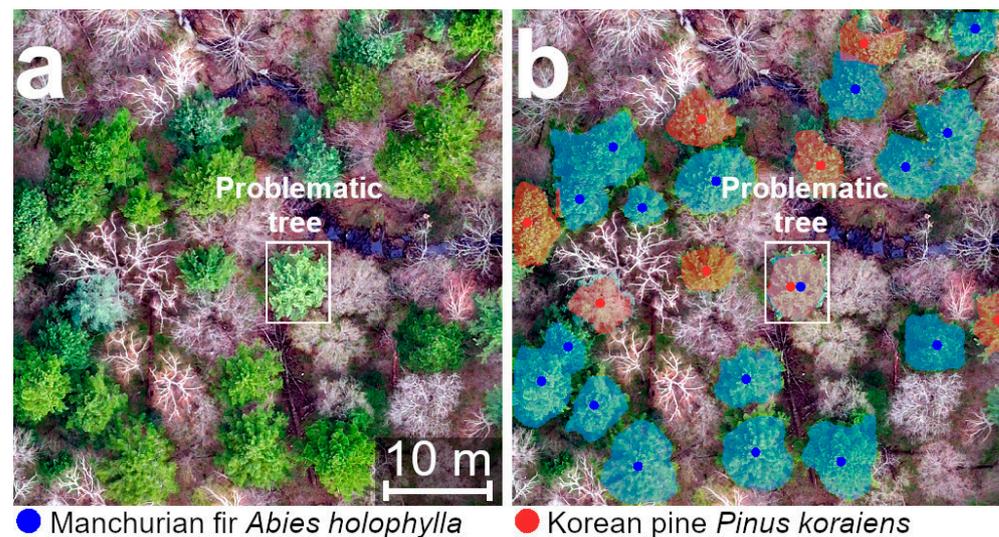
Figure A8. Recognition result by Mask R-CNN with the double recognition of a single tree of Manchurian fir (*Abies holophylla*). (**a**) original image; (**b**) image with segmented tree crowns.

References

1. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in Vegetation Remote Sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [CrossRef]
2. Komárek, J.; Klouček, T.; Prošek, J. The Potential of Unmanned Aerial Systems: A Tool towards Precision Classification of Hard-to-Distinguish Vegetation Types? *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *71*, 9–19. [CrossRef]
3. Tarantino, C.; Casella, F.; Adamo, M.; Lucas, R.; Beierkuhnlein, C.; Blonda, P. Ailanthus Altissima Mapping from Multi-Temporal Very High Resolution Satellite Images. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 90–103. [CrossRef]
4. Ball, J.G.C.; Hickman, S.H.M.; Jackson, T.D.; Koay, X.J.; Hirst, J.; Jay, W.; Archer, M.; Aubry-Kientz, M.; Vincent, G.; Coomes, D.A. Accurate Delineation of Individual Tree Crowns in Tropical Forests from Aerial RGB Imagery Using Mask R-CNN. *Remote Sens. Ecol. Conserv.* **2023**. [CrossRef]
5. Braga, G.J.R.; Peripato, V.; Dalagnol, R.; Ferreira, M.P.; Tarabalka, Y.; OC Aragão, L.E.; de Campos Velho, H.F.; Shiguemori, E.H.; Wagner, F.H. Tree Crown Delineation Algorithm Based on a Convolutional Neural Network. *Remote Sens.* **2020**, *12*, 1288. [CrossRef]
6. Albuquerque, R.W.; Vieira, D.L.M.; Ferreira, M.E.; Soares, L.P.; Olsen, S.I.; Araujo, L.S.; Vicente, L.E.; Tymus, J.R.C.; Balieiro, C.P.; Matsumoto, M.H.; et al. Mapping Key Indicators of Forest Restoration in the Amazon Using a Low-Cost Drone and Artificial Intelligence. *Remote Sens.* **2022**, *14*, 830. [CrossRef]
7. Zhang, J.; Hu, J.; Lian, J.; Fan, Z.; Ouyang, X.; Ye, W. Seeing the Forest from Drones: Testing the Potential of Lightweight Drones as a Tool for Long-Term Forest Monitoring. *Biol. Conserv.* **2016**, *198*, 60–69. [CrossRef]
8. Gibril, M.B.A.; Shafri, H.Z.M.; Al-Ruzouq, R.; Shanableh, A.; Nahas, F.; Al Mansoori, S. Large-Scale Date Palm Tree Segmentation from Multiscale UAV-Based and Aerial Images Using Deep Vision Transformers. *Drones* **2023**, *7*, 93. [CrossRef]
9. Zhu, Y.; Zhou, J.; Yang, Y.; Liu, L.; Liu, F.; Kong, W. Rapid Target Detection of Fruit Trees Using UAV Imaging and Improved Light YOLOv4 Algorithm. *Remote Sens.* **2022**, *14*, 4324. [CrossRef]
10. Guo, X.; Liu, Q.; Sharma, R.P.; Chen, Q.; Ye, Q.; Tang, S.; Fu, L. Tree Recognition on the Plantation Using UAV Images with Ultrahigh Spatial Resolution in a Complex Environment. *Remote Sens.* **2021**, *13*, 4122. [CrossRef]
11. Donmez, C.; Villi, O.; Berberoglu, S.; Cilek, A. Computer Vision-Based Citrus Tree Detection in a Cultivated Environment Using UAV Imagery. *Comput. Electron. Agric.* **2021**, *187*, 106273. [CrossRef]
12. Onishi, M.; Ise, T. Explainable Identification and Mapping of Trees Using UAV RGB Image and Deep Learning. *Sci. Rep.* **2021**, *11*, 903. [CrossRef]
13. Miraki, M.; Sohrabi, H.; Fatehi, P.; Kneubuehler, M. Individual Tree Crown Delineation from High-Resolution UAV Images in Broadleaf Forest. *Ecol. Inform.* **2021**, *61*, 101207. [CrossRef]
14. Weinstein, B.G.; Marconi, S.; Graves, S.J.; Zare, A.; Singh, A.; Bohlman, S.A.; Magee, L.; Johnson, D.J.; Townsend, P.A.; White, E.P. Capturing Long-Tailed Individual Tree Diversity Using an Airborne Imaging and a Multi-Temporal Hierarchical Model. *Remote Sens. Ecol. Conserv.* **2023**. [CrossRef]
15. Natesan, S.; Armenakis, C.; Vepakomma, U. Individual Tree Species Identification Using Dense Convolutional Network (DenseNet) on Multitemporal RGB Images from UAV. *J. Unmanned Veh. Sys.* **2020**, *8*, 310–333. [CrossRef]
16. Beloiu, M.; Heinzmann, L.; Rehush, N.; Gessler, A.; Griess, V.C. Individual Tree-Crown Detection and Species Identification in Heterogeneous Forests Using Aerial RGB Imagery and Deep Learning. *Remote Sens.* **2023**, *15*, 1463. [CrossRef]

17. Krestov, P.V. Forest Vegetation of Easternmost Russia (Russian Far East). In *Forest Vegetation of Northeast Asia*; Kolbek, J., Šrůtek, M., Box, E.O., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 2003; pp. 93–180, ISBN 978-94-017-0143-3.

18. Dinerstein, E.; Olson, D.; Joshi, A.; Vynne, C.; Burgess, N.D.; Wikramanayake, E.; Hahn, N.; Palminteri, S.; Hedao, P.; Noss, R.; et al. An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience* **2017**, *67*, 534–545. [CrossRef] [PubMed]

19. Casado-García, Á.; Domínguez, C.; García-Domínguez, M.; Heras, J.; Inés, A.; Mata, E.; Pascual, V. CLoDSA: A Tool for Augmentation in Classification, Localization, Detection, Semantic Segmentation and Instance Segmentation Tasks. *BMC Bioinform.* **2019**, *20*, 323. [CrossRef] [PubMed]

20. Map Pilot Pro. Available online: https://www.mapsmadeeasy.com/map_pilot/ (accessed on 19 June 2023).

21. OpenDroneMap/ODM. Available online: https://github.com/OpenDroneMap/ODM (accessed on 19 June 2023).

22. U-Net: Semantic Segmentation with PyTorch. Available online: https://github.com/milesial/Pytorch-UNet (accessed on 19 June 2023).

23. YOLO by Ultralytics. Available online: https://github.com/ultralytics/ultralytics (accessed on 19 June 2023).

24. GitHub-Matterport/Mask_RCNN: Mask R-CNN for Object Detection and Instance Segmentation on Keras and TensorFlow. Available online: https://github.com/matterport/Mask_RCNN (accessed on 19 June 2023).

25. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and Flexible Image Augmentations. *Information* **2020**, *11*, 125. [CrossRef]

26. Albumentations. Available online: https://github.com/albumentations-team/albumentations (accessed on 19 June 2023).

27. Sivanandam, P.; Lucieer, A. Tree Detection and Species Classification in a Mixed Species Forest Using Unoccupied Aircraft System (UAS) RGB and Multispectral Imagery. *Remote Sens.* **2022**, *14*, 4963. [CrossRef]

28. Sun, Y.; Li, Z.; He, H.; Guo, L.; Zhang, X.; Xin, Q. Counting Trees in a Subtropical Mega City Using the Instance Segmentation Method. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *106*, 102662. [CrossRef]

29. Yang, M.; Mou, Y.; Liu, S.; Meng, Y.; Liu, Z.; Li, P.; Xiang, W.; Zhou, X.; Peng, C. Detecting and Mapping Tree Crowns Based on Convolutional Neural Network and Google Earth Images. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102764. [CrossRef]

30. Gan, Y.; Wang, Q.; Iio, A. Tree Crown Detection and Delineation in a Temperate Deciduous Forest from UAV RGB Imagery Using Deep Learning Approaches: Effects of Spatial Resolution and Species Characteristics. *Remote Sens.* **2023**, *15*, 778. [CrossRef]

31. Nasiri, V.; Darvishsefat, A.A.; Arefi, H.; Pierrot-Deseilligny, M.; Namiranian, M.; Le Bris, A. Unmanned Aerial Vehicles (UAV)-Based Canopy Height Modeling under Leaf-on and Leaf-off Conditions for Determining Tree Height and Crown Diameter (Case Study: Hyrcanian Mixed Forest). *Can. J. For. Res.* **2021**, *51*, 962–971. [CrossRef]

32. Lou, X.; Huang, Y.; Fang, L.; Huang, S.; Gao, H.; Yang, L.; Weng, Y.; Hung, I.-K. uai Measuring Loblolly Pine Crowns with Drone Imagery through Deep Learning. *J. For. Res.* **2022**, *33*, 227–238. [CrossRef]

33. 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*, 66. [CrossRef]

34. Kislov, D.E.; Korznikov, K.A.; Altman, J.; Vozmishcheva, A.S.; Krestov, P.V. Extending Deep Learning Approaches for Forest Disturbance Segmentation on Very High-Resolution Satellite Images. *Remote Sens. Ecol. Conserv.* **2021**, *7*, 355–368. [CrossRef]

35. Safonova, A.; Tabik, S.; Alcaraz-Segura, D.; Rubtsov, A.; Maglinets, Y.; Herrera, F. Detection of Fir Trees (Abies Sibirica) Damaged by the Bark Beetle in Unmanned Aerial Vehicle Images with Deep Learning. *Remote Sens.* **2019**, *11*, 643. [CrossRef]

36. Hu, G.; Wang, T.; Wan, M.; Bao, W.; Zeng, W. UAV Remote Sensing Monitoring of Pine Forest Diseases Based on Improved Mask R-CNN. *Int. J. Remote Sens.* **2022**, *43*, 1274–1305. [CrossRef]

37. Zhang, C.; Zhou, J.; Wang, H.; Tan, T.; Cui, M.; Huang, Z.; Wang, P.; Zhang, L. Multi-Species Individual Tree Segmentation and Identification Based on Improved Mask R-CNN and UAV Imagery in Mixed Forests. *Remote Sens.* **2022**, *14*, 874. [CrossRef]

38. Jansen, A.J.; Nicholson, J.D.; Esparon, A.; Whiteside, T.; Welch, M.; Tunstill, M.; Paramjyothi, H.; Gadhiraju, V.; van Bodegraven, S.; Bartolo, R.E. Deep Learning with Northern Australian Savanna Tree Species: A Novel Dataset. *Data* **2023**, *8*, 44. [CrossRef]

39. Moreira, B.M.; Goyanes, G.; Pina, P.; Vassilev, O.; Heleno, S. Assessment of the Influence of Survey Design and Processing Choices on the Accuracy of Tree Diameter at Breast Height (DBH) Measurements Using UAV-Based Photogrammetry. *Drones* **2021**, *5*, 43. [CrossRef]

40. Perroy, R.L.; Sullivan, T.; Stephenson, N. Assessing the Impacts of Canopy Openness and Flight Parameters on Detecting a Sub-Canopy Tropical Invasive Plant Using a Small Unmanned Aerial System. *ISPRS J. Photogramm. Remote Sens.* **2017**, *125*, 174–183. [CrossRef]

41. Zhang, B.; Zhao, L.; Zhang, X. Three-Dimensional Convolutional Neural Network Model for Tree Species Classification Using Airborne Hyperspectral Images. *Remote Sens. Environ.* **2020**, *247*, 111938. [CrossRef]

42. Abdollahnejad, A.; Panagiotidis, D. Tree Species Classification and Health Status Assessment for a Mixed Broadleaf-Conifer Forest with UAS Multispectral Imaging. *Remote Sens.* **2020**, *12*, 3722. [CrossRef]

43. Wang, X.; Wang, Y.; Zhou, C.; Yin, L.; Feng, X. Urban Forest Monitoring Based on Multiple Features at the Single Tree Scale by UAV. *Urban For. Urban Green.* **2021**, *58*, 126958. [CrossRef]

44. Wang, J.; Mall, S.; Perez, L. The Effectiveness of Data Augmentation in Image Classification Using Deep Learning. *Convolutional Neural Netw. Vis. Recognit* **2017**, *11*, 1–8.

45. Milas, A.S.; Arend, K.; Mayer, C.; Simonson, M.A.; Mackey, S. Different Colours of Shadows: Classification of UAV Images. *Int. J. Remote Sens.* **2017**, *38*, 3084–3100. [CrossRef]

46. Kattenborn, T.; Eichel, J.; Fassnacht, F.E. Convolutional Neural Networks Enable Efficient, Accurate and Fine-Grained Segmentation of Plant Species and Communities from High-Resolution UAV Imagery. *Sci. Rep.* **2019**, *9*, 17656. [CrossRef] [PubMed]
47. Berra, E.F. Individual Tree Crown Detection and Delineation across a Woodland Using Leaf-on and Leaf-off Imagery from a UAV Consumer-Grade Camera. *JARS* **2020**, *14*, 034501. [CrossRef]
48. Lee, C.K.F.; Song, G.; Muller-Landau, H.C.; Wu, S.; Wright, S.J.; Cushman, K.C.; Araujo, R.F.; Bohlman, S.; Zhao, Y.; Lin, Z.; et al. Cost-Effective and Accurate Monitoring of Flowering across Multiple Tropical Tree Species over Two Years with a Time Series of High-Resolution Drone Imagery and Deep Learning. *ISPRS J. Photogramm. Remote Sens.* **2023**, *201*, 92–103. [CrossRef]
49. Klosterman, S.; Melaas, E.; Wang, J.A.; Martinez, A.; Frederick, S.; O'Keefe, J.; Orwig, D.A.; Wang, Z.; Sun, Q.; Schaaf, C.; et al. Fine-Scale Perspectives on Landscape Phenology from Unmanned Aerial Vehicle (UAV) Photography. *Agric. For. Meteorol.* **2018**, *248*, 397–407. [CrossRef]