

Article

Target Detection-Based Tree Recognition in a Spruce Forest Area with a High Tree Density—Implications for Estimating Tree Numbers

Mirzat Emin ^{1,2,3} , Erpan Anwar ^{1,2,3}, Suhong Liu ^{4,*}, Bilal Emin ^{1,2,3}, Maryam Mamut ^{1,2,3},
Abduwali Abdukeram ^{1,2,3} and Ting Liu ^{1,2,3}

¹ College of Resources and Environmental Science, Xinjiang University, Urumqi 830046, China; mirzat_edu@163.com (M.E.); erpan_edu@163.com (E.A.); bilalemin@stu.xju.edu.cn (B.E.); mlym18195859060@163.com (M.M.); a2511550848@163.com (A.A.); lting_liu@163.com (T.L.)

² Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi 830046, China

³ Institute of Arid Ecology and Environment, Xinjiang University, Urumqi 830046, China

⁴ Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

* Correspondence: liush@nbu.edu.cn

Abstract: Here, unmanned aerial vehicle (UAV) remote sensing and machine vision were used to automatically, accurately, and efficiently count Tianshan spruce and improve the efficiency of scientific forest management, focusing on a typical Tianshan spruce forest on Tianshan Mountain, middle Asia. First, the UAV in the sampling area was cropped from the image, and a target-labeling tool was used. The Tianshan spruce trees were annotated to construct a data set, and four models were used to identify and verify them in three different areas (low, medium, and high canopy closures). Finally, the combined number of trees was calculated. The average accuracy of the detection frame, mean accuracy and precision (mAP), was used to determine the target detection accuracy. The Faster Region Convolutional Neural Network (Faster-RCNN) model achieved the highest accuracies (96.36%, 96.32%, and 95.54% under low, medium, and high canopy closures, respectively) and the highest mAP (85%). Canopy closure affected the detection and recognition accuracy; YOLOv3, YOLOv4, and Faster-RCNN all showed varying spruce recognition accuracies at different densities. The accuracy of the Faster-RCNN model decreased by at least 0.82%. Combining UAV remote sensing with target detection networks can identify and quantify statistics regarding Tianshan spruce. This solves the shortcomings of traditional monitoring methods and is significant for understanding and monitoring forest ecosystems.

Keywords: Tianshan spruce; target detection; UAV; forest inventory



Citation: Emin, M.; Anwar, E.; Liu, S.; Emin, B.; Mamut, M.; Abdukeram, A.; Liu, T. Target Detection-Based Tree Recognition in a Spruce Forest Area with a High Tree Density—Implications for Estimating Tree Numbers. *Sustainability* **2021**, *13*, 3279. <https://doi.org/10.3390/su13063279>

Academic Editors: Marc A. Rosen and C. Ronald Carroll

Received: 30 January 2021

Accepted: 10 March 2021

Published: 16 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tianshan spruce forests play an irreplaceable role in water conservation [1], oxygen supply and carbon fixation [2], climate regulation, air purification, nutrient pattern [3], and biodiversity conservation. They also have an important impact on the ecological environment and climate regulation in Xinjiang [4,5]. With the rapid development of tourism, the Tianshan spruce forest at low latitudes on the northern slopes of Tianshan Mountain is vulnerable to anthropogenic deforestation, which could reduce the biomass of Tianshan spruce. Hence, the rapid identification and determination of quantitative statistics are of great significance for understanding and monitoring forest ecosystems.

Field measurements and remote sensing estimations are the two main methods used to quantify forest information [6]. The sample plot survey, a traditional method for determining forest quantity statistics, requires a heavy workload and a long cycle. Furthermore, it has low efficiency, especially in mountainous areas with complex terrain. It is difficult to determine statistics in this way, which cannot meet the requirements of modern forestry. With the rapid development of remote sensing technology and the improvement of image

resolution, satellite remote sensing technology has gradually become an important method for extracting and estimating the number of trees in forest canopies. Liu et al. [7] used the crown vertex detection method of multi-scale spot detection to extract the crown information for Tianshan spruce using WorldView-2 images. They achieved good extraction effects on crown information of Tianshan spruce under high, medium, and low canopy densities. Fabien et al. [8] proposed a new method for automatically dividing tree canopies based on very high-resolution WorldView-2 images and applied it to an Atlantic rainforest area covered by a highly heterogeneous tropical canopy (the San Gebra Forest in Brazil). This method can detect with 80% accuracy. Dilek et al. [9] used unmanned aerial vehicle (UAV) multispectral images (MSIs) and digital surface models (DSMs) to extract citrus trees using sequence threshold, canny edge detection, and circular Hough transform algorithms. Their proposed method was successful, and achieved a delineation accuracy of over 80%. However, most studies on crown extraction for single trees have focused on simple regular or low-density stands. The spatial structure of the Tianshan spruce forest is more complex and changeable, so there are still some difficulties regarding the accurate extraction and statistical quantification of single tree crowns [10]. These difficulties are especially prevalent in high canopy density forests, owing to overlapping and connection between tree crowns, and due to the occlusion of young trees. Most tree number acquisition methods obtain the number of trees by extracting the crown information of individual trees and then use the marked watershed segmentation method [11], but this approach is not ideal for high-density areas.

Developments in computer science and technology have provided feature extraction technologies based on deep convolutional neural networks, which are now widely used in the field of computer vision. Deep learning methods based on convolutional neural networks have gradually become a research hotspot in the field of image processing. Therefore, some studies have examined tree detection using the deep learning method. This method combines two technologies, and its progress offers great potential for tree counting as it combines UAVs and depth of learning [12]. Deng et al. [13] used drone images to improve the Fast Region Convolutional Neural Network (Fast-RCNN) deep learning framework, thereby establishing a detection model for dead pine trees. They established a geographic information output module to output the specific geographic locations of diseased trees and to accurately locate pine blight and dead trees. Felix et al. [14] used a semantic segmentation method (U-net, Convolutional Networks for Biomedical Image Segmentation) that simultaneously segmented and classified tree species from images and evaluated high-resolution red-green-blue (RGB) images provided by traditional convolutional neural networks (CNNs) and drones for rendering. In this way, they mapped the potential environmental benefits of temperate forest species. Ding et al. [15] adopted a PCB micro-defect detection method based on Fast-RCNN, which solved the shortcomings of deep convolutional networks in detecting small defective areas. They achieved good experimental results by opening the printed circuit board (PCB) defect database (a kind of database).

In recent years, the rapid development of UAV remote sensing technology has highlighted the advantages of high resolution and simple data acquisition—it is low cost, rapid, and low risk [16,17]. These advantages make up for the shortcomings of traditional satellite remote sensing. UAV remote sensing technology has been widely used in land use classification, agricultural resources surveys [18], forestry resources investigations, forest pest control, and fire prevention [19]. Yi et al. [20] used a local maximum and multi-scale algorithm to extract the number of trees in subtropical forests based on UAV remote sensing. Hernandez et al. [21] used mixed pixel- and region-based algorithms to segment images, thereby automatically extracting individual trees in plantations. In this way, they were able to estimate the heights and crown diameters of individual trees. However, there have been few studies carried out on spruce detection and forest management for Tianshan spruce, based on deep learning and UAV remote sensing.

This study combined UAV high-resolution image data with a target detection algorithm to identify and count Tianshan spruce in Yangjuangou, on the northern slopes of Tianshan Mountain. Experiments on citrus trees, an important agricultural product in the Mediterranean region, were performed in three areas with different tree densities but similar land areas and background coverage. This provided technical support for the extraction of biomass information from the Tianshan spruce forest to accurately estimate the biomass of spruce forest, predict the development and succession direction of the community, and provide a reference for the efficient management and protection of Tianshan spruce.

2. Materials and Methods

2.1. Introduction to Research Objectives

Tianshan spruce is a main tree species in the forest community of the Tianshan Mountains. It is tall and narrow but has a long crown that is high in the center and low around the edges. In remote sensing images, the peak of the crown appears bright, whereas the edges of the crown are dark. Tianshan spruce is mainly distributed across the northern slopes of Tianshan Mountain, at an altitude of 1500–2700 m; it belongs to the northern mountain coniferous forest system [22], which accounts for 44.9% of the total forested area in Xinjiang (with an area of 5.28×10^5 hm²).

2.2. Study Area

Urumqi County is located in the northwest of China, in the central part of Xinjiang. It lies at the northern foot of the Tianshan Mountains, south of the Dzungarian Basin (Figure 1). The area has a continental arid climate and lies in the middle temperate zone; it has an annual precipitation of 200 mm. The terrain mainly comprises mountainous areas, basins, and plains. The terrain is high in the south and low in the north. The vegetation and soil in the basin exhibit significant vertical zonality. The soil is dominantly grayish-brown forest soil, and the annual average temperature is 2–3 °C. Yangjuangou is located in the central part of Urumqi County, with a geographical location of 87°27'40" E–87°29'5" E, 43°24'30" N–43°26' N. More than 90% of the trees in the forest area are Tianshan spruce, with an average height of 15 m. The study area represents a typical distribution area of Tianshan spruce in Xinjiang. Figure 1 is the map of study area.

2.3. Data Acquisition and Preprocessing

The data used in this study comprised of a low-altitude remote sensing image obtained from a UAV (Dajiang spirit 4prov2.0, equipped with a CMOS camera, SZ DJI Technology Co., Ltd, Shenzhen, China). The image, which had 20 million effective pixels, was taken on 8 November 2019 (there is a layer of low green grass under the Tianshan spruce forest in summer. In November, the grass on the ground will wither and turn yellow, but Tianshan spruce is a green plant in all seasons. Therefore, the data collection in November is conducive to the detection effect). The conditions were sunny, and the wind speed was 3 m/s; the flight altitude was 175 m, the imaging area was 1.08 km², the image resolution was 4.38 cm, the band was the visible light band (RGB), and 1520 aerial photos were taken.

In this study, Agisoft Photoscan1.2.6 software (Agisoft LLC11 Degtyarniy per., St. Petersburg, Russia) was used to quickly splice UAV images. A data set of Tianshan spruce was built and divided into a training set and a verification set, according to the ratio of 7:3. The image for the training area was enhanced. In order to better distinguish shadows on the ground and trees, the band combination of the UAV image was split into Blue, Green, and Red when training samples. After random clipping, rotation, and noise addition, 3300 datasets were generated for model training. Data enhancement can reduce the model over the fitting phenomenon and can enhance the model generalization performance. Three areas with low, medium, and high canopy densities were selected for verification, each with a size of 1600 × 1600 pixels and an actual area of 4911 m².

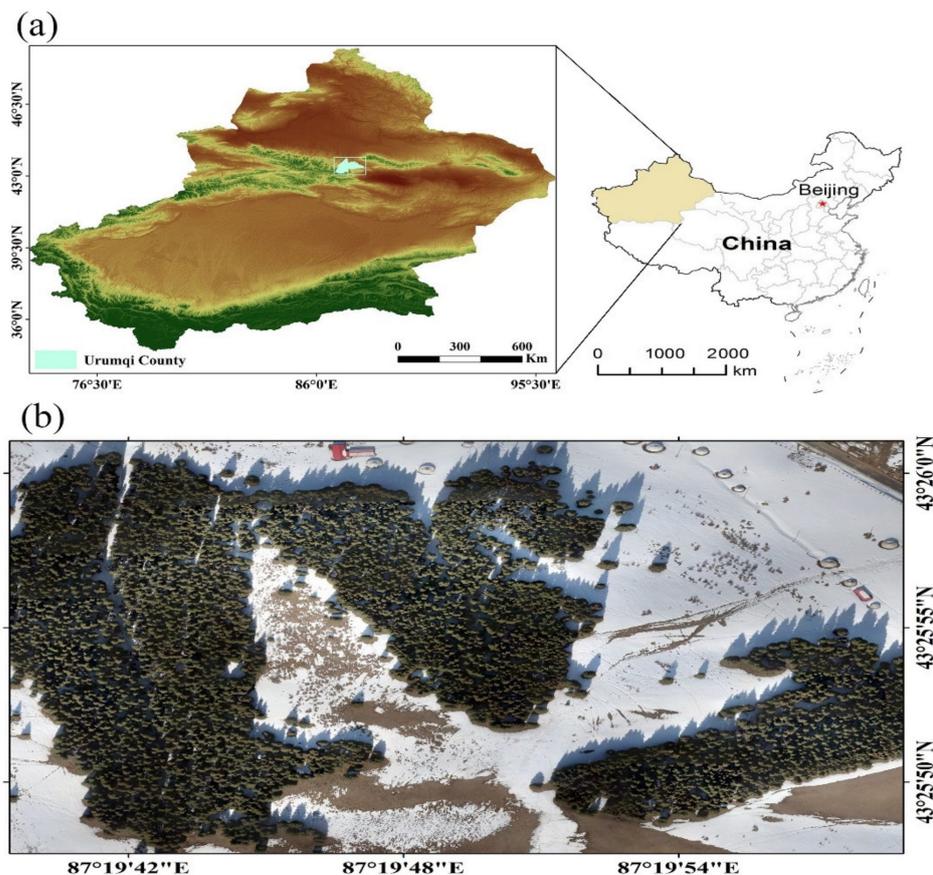


Figure 1. (a) Geographical location of Urumqi County and (b) sampling area in Yangjuangou, Urumqi County.

2.4. Methods

Deep learning has derived a variety of applications in scientific research, such as target detection [23], motion tracking [24], and action recognition [25]. Target detection is a challenging basic project in computer vision. In recent years, a large number of target detection models have emerged in the framework of Convolutional Neural Networks (CNNs), which have good performance in terms of speed and detection accuracy [26]. Generally, there are two kinds of frequently used algorithms. The first type is two-stage regions with CNN features (RCNN) [27] algorithms (RCNN, Fast-CNN, Faster-RCNN, etc.), which first need to generate the target position, and then classify and regress the candidate frames. During training, the first training area is the region proposal network (RPN), which is followed by training for the target area network detection. Thus, this model is highly accurate but slow. The other type is single-stage algorithms, such as You Only Look Once (YOLO) [28] and Single Shot MultiBox Detector (SSD) [29]. These use only one CNN to directly predict the categories and positions of different targets. These models are much faster than two-stage models, but they have lower accuracy. Regarding speed, single-stage models are more suitable for industrial applications than two-stage models.

Unlike traditional image classification methods, target detection does not need to extract specific artificial features for particular projects. Instead, training algorithms and different network models are used to learn the advanced semantic information of the image and complete the classification and positioning of the target, thus improving efficiency and accuracy. In this study, SSD, YOLOv3, YOLOv4, and Faster-RCNN were used to detect Tianshan spruce in Yangjuangou in Urumqi County. The most suitable model for spruce detection was selected by comparing their detection effects. Figure 2 is the overall technical route.

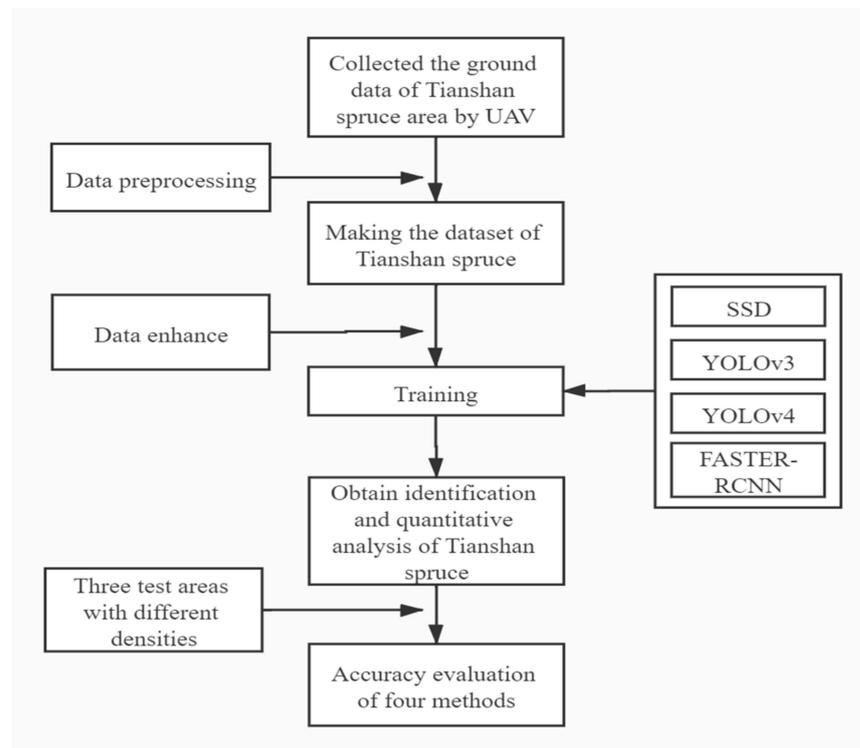


Figure 2. The overall technical route.

2.4.1. SSD

The SSD [29] model (Figure 3) is an end-to-end deep learning model. In terms of structure, the SSD model adds a new network structure to the Very Deep Convolutional Networks for Large-Scale Image Recognition (VGG—small filters, deep networks) basic network, converts the full connection layer into a convolution layer, adds convolution layers, and outputs the prediction result as a target detection frame by merging convolutions conv4_3, Conv7, Conv8_2, Conv9_2, Conv10_2, and Conv11_2. Two 3×3 convolution kernels were used for prediction: one for classification and the other for position regression.

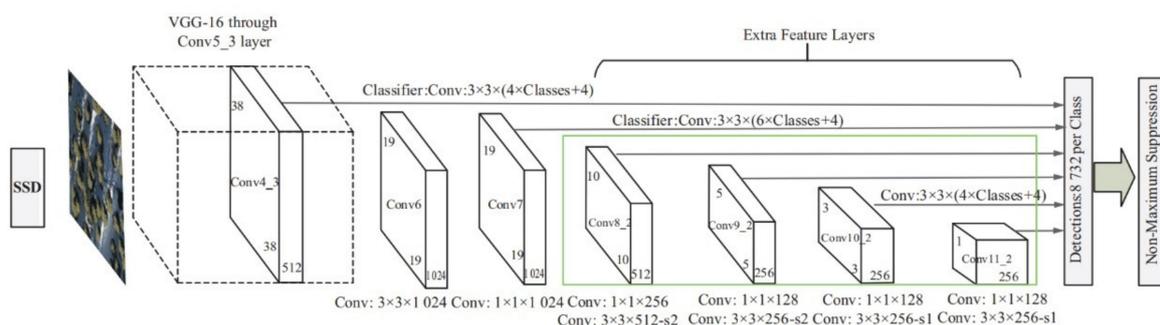


Figure 3. Single Shot MultiBox Detector (SSD) network structure.

2.4.2. YOLOv3

The YOLOv3 [30] (Figure 4) algorithm is an improved algorithm based on YOLOv2. It extends and improves the basic network structure of YOLOv2 from darknet19 (Darknet is an open-source neural network framework). It is fast, easy to install, and supports CPU and GPU computation and adopts the darknet53 network structure. This includes 53 volume layers for image feature extraction. When the deepest feature is extracted, the output and sampling are performed at the same time to fuse with the features at another scale. YOLOv3 combines the construction idea of the Deep Residual Learning for Image

Recognition (RESNET) network and adopts a large number of residual blocks, which allows the network structure to be set deeper. This means it has a strong feature extraction ability.

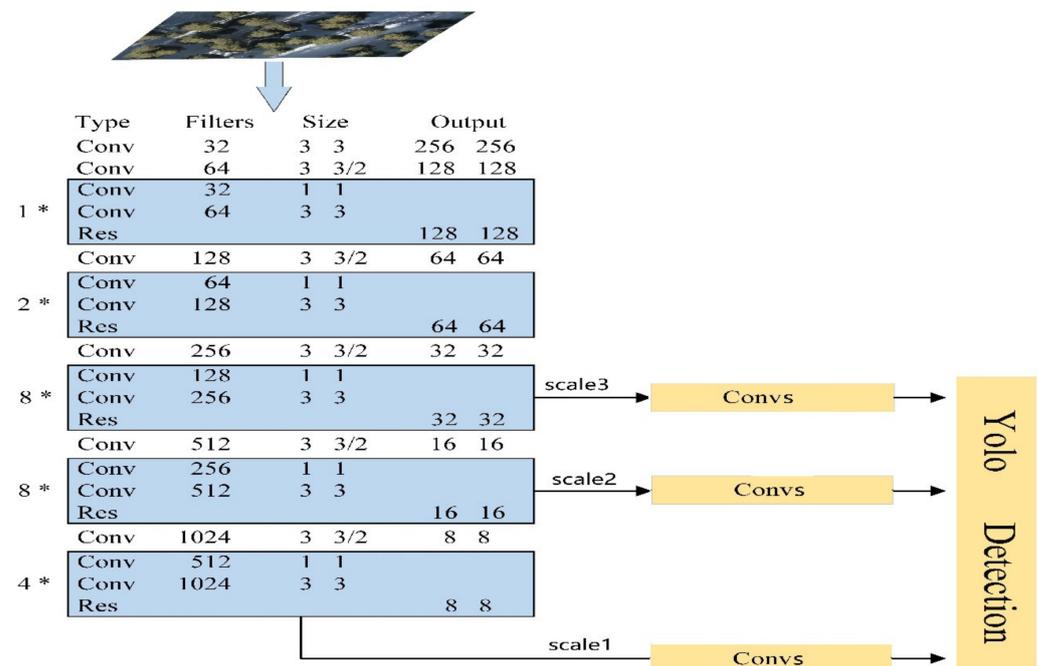


Figure 4. You Only Look Once (YOLO)v3 network structure.

2.4.3. YOLOv4

For YOLOv4 [31] (Figure 5), the target detection algorithm is more complex than that of YOLOv3 in the network structure, and the accuracy of the neural network is improved through more training skills. The network structure of YOLOv4 consists of four parts: input, backbone, neck, and prediction [32]. The input introduces a mosaic data enhancement method and self-confrontation training of SAT. The backbone main network combines CSPDarknet53, Mish activation function, and Dropblock. In the neck section, SPP modules are inserted between the backbone and the final output layer. In the prediction section, the anchor frame mechanism of the output layer is the same as that of YOLOv3. The main improvement is the CIOU-Loss [33] during training; the NMS [34] is filtered by the prediction box to become Diou NMS.

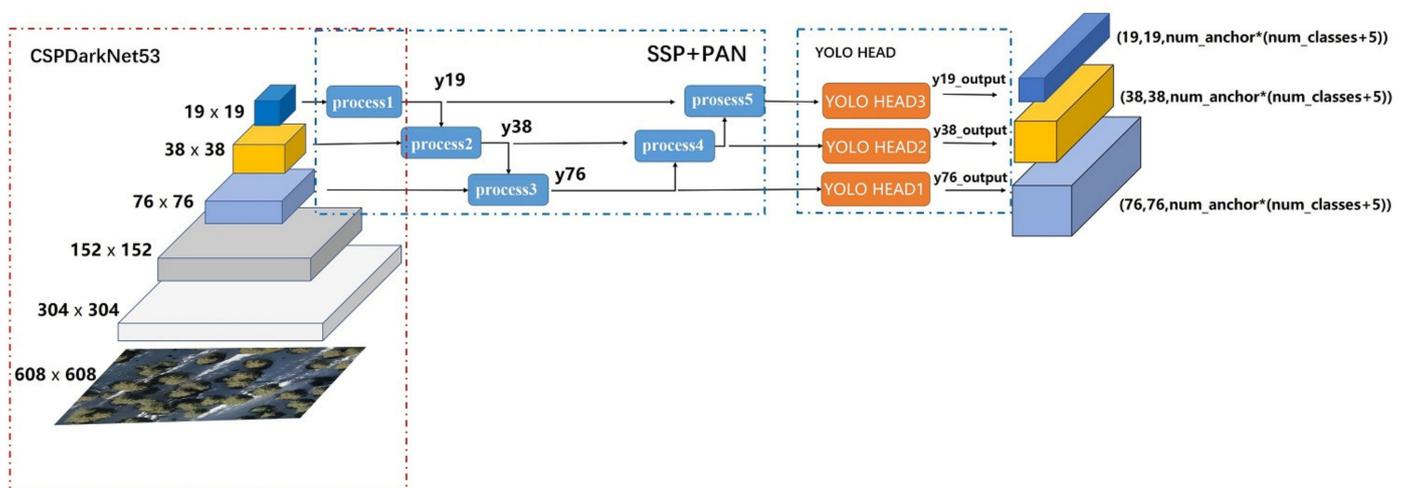


Figure 5. YOLOv4 network structure.

2.4.4. Faster-RCNN

Faster-RCNN [35] (Figure 6) is a two-stage convolution neural network algorithm. It consists of a feature extraction network (CNN), a target detection network (RCNN), and a region recommendation network (RPN) [36]. RPN is a full convolution-based network that can simultaneously predict the probability values of target area frames and the probability value of the true target. RPN is an end-to-end network training that aims to generate a high-quality regional suggestion box for Fast-RCNN classification and detection. Through a simple alternative operation optimization method, RPN and Fast-RCNN can share convolution features during training. Hence, the overall structure of Fast-RCNN can be considered as an RPN + Faster-RCNN.

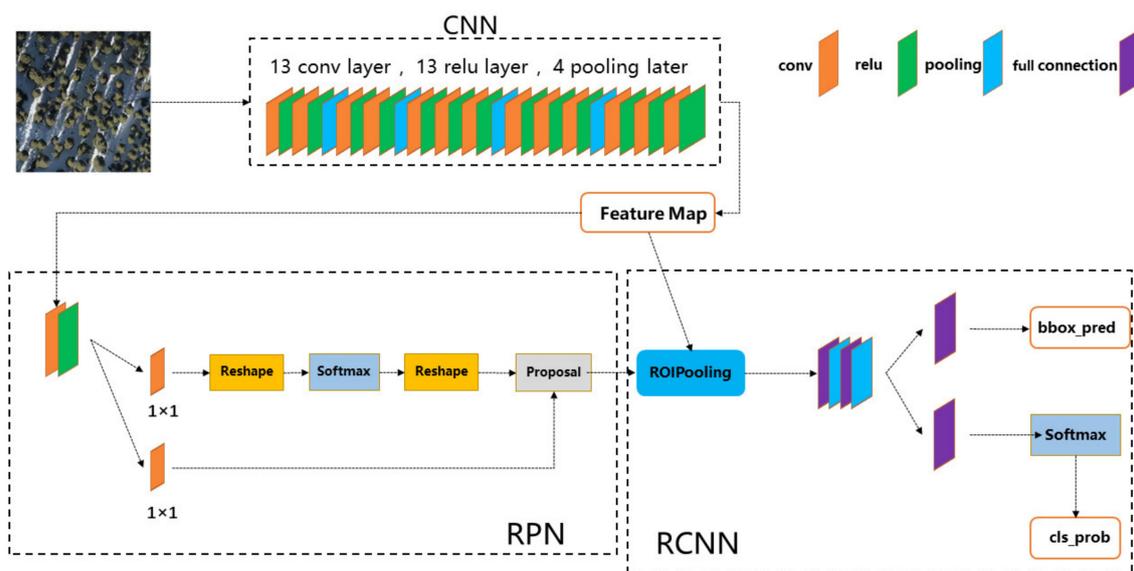


Figure 6. Faster Region Convolutional Neural Network (Faster-RCNN) network model.

3. Results

In this study, 3300 training samples were used to train four types of target detection models: SSD, YOLOv3, YOLOv4, and Faster-RCNN. Finally, three different density areas of spruce were identified, and the numbers of trees were counted. In order to improve the accuracy of spruce recognition, the three validation areas were each split into 16 small pieces using a regular grid, and then they were identified. Figure 7 shows the detection effects for Tianshan spruce in three different density regions under the four different models. The same training sample was selected as the data set, and the different network structures exhibited obvious differences regarding their recognition of the Tianshan spruce.

Accuracy Assessment

In this study, the actual number of Tianshan spruce in each validation area was obtained by visual interpretation. The objective detection algorithm was used to extract the number of spruce trees, and the overall accuracy (OA) and overlap error (OE) were estimated using Equation (1):

$$OA = |N_d - N_o| / N_v \times 100\% \quad (1)$$

where N_v is the total number of Tianshan spruce in the visual interpretation sample plot, N_d is the total number of spruce detected by the target detection model, and N_o is the number of repeatedly detected spruce caused by image rule clipping.

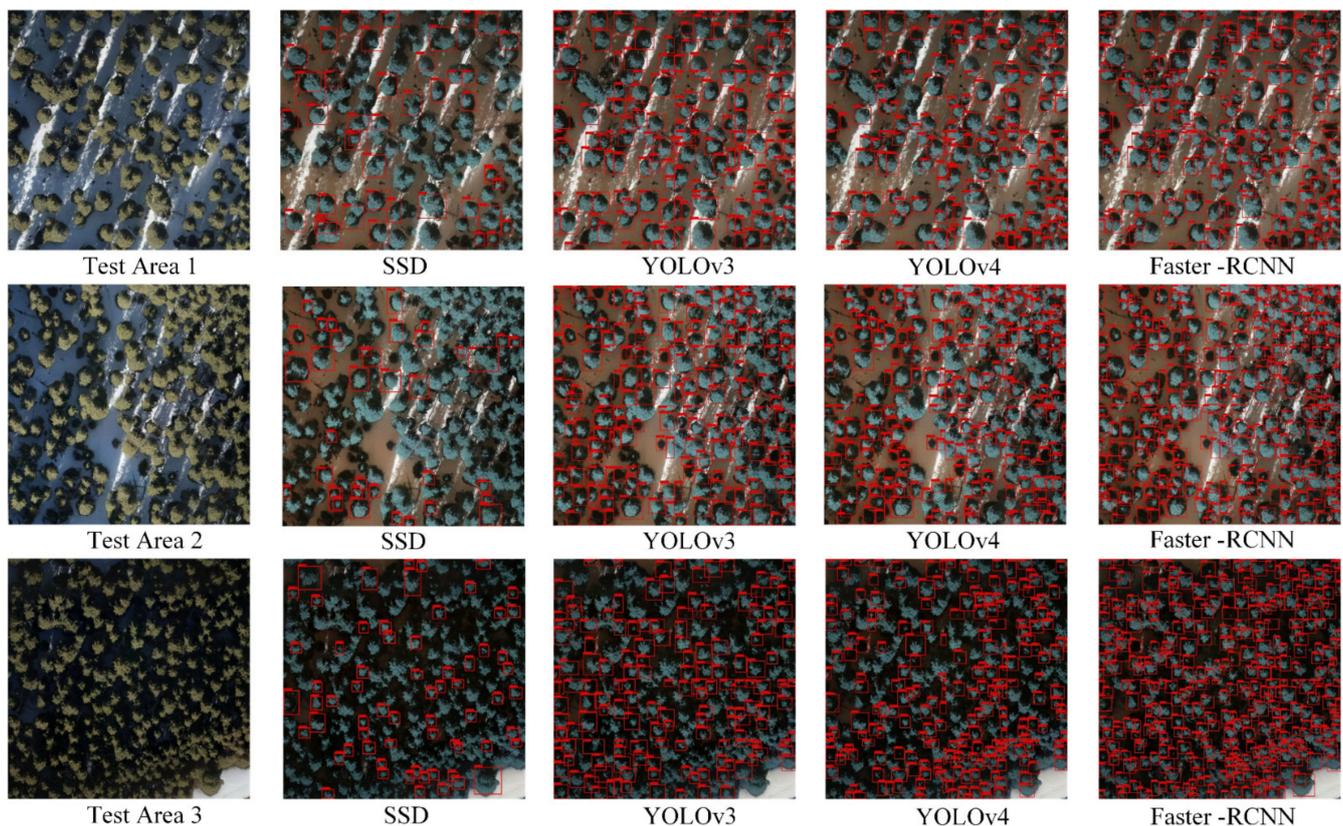


Figure 7. Test results of the four network structures.

Through visual interpretation, the actual quantity of Tianshan spruce in the three validation areas were 165, 245, and 359. Table 1 shows that the average accuracy of the SSD target detection algorithm, based on the VGG16 (Very Deep Convolutional Networks with 16 convolutions) network, was lower than that of the other algorithms. The accuracies of the three different density regions were 24.85%, 14.69%, and 18.38%, respectively. In contrast, the detection accuracy of SSD improved significantly. The accuracies of the three different density areas were 74.55%, 65.31%, and 41.23%, respectively. The detection accuracy of YOLOv4 exceeded that of YOLOv3, and the accuracy of the three validation areas were 82.42%, 82.04%, and 56.27%, respectively. The Faster-RCNN algorithm takes ResNet101-FPN as the backbone network model. Although the training speed of ResNet101-FPN was not as fast as that of the YOLO algorithm based on VGG16, the accuracies of the Faster-RCNN algorithm were 96.36%, 96.32%, and 95.54%, which are, respectively, higher than those of the other three algorithms.

This study also used accuracy (P), recall (R), and average precision (AP), which are commonly used as references when evaluating the accuracy of different detection models. AP is the area under the curve of accuracy and recall rate; it is an intuitive evaluation standard used to assess the accuracy of detection models and can be used to analyze the detection effect of a single category. The accuracy and recall were defined as shown in Equations (1) and (2):

$$R = TP / TP + FP \quad (2)$$

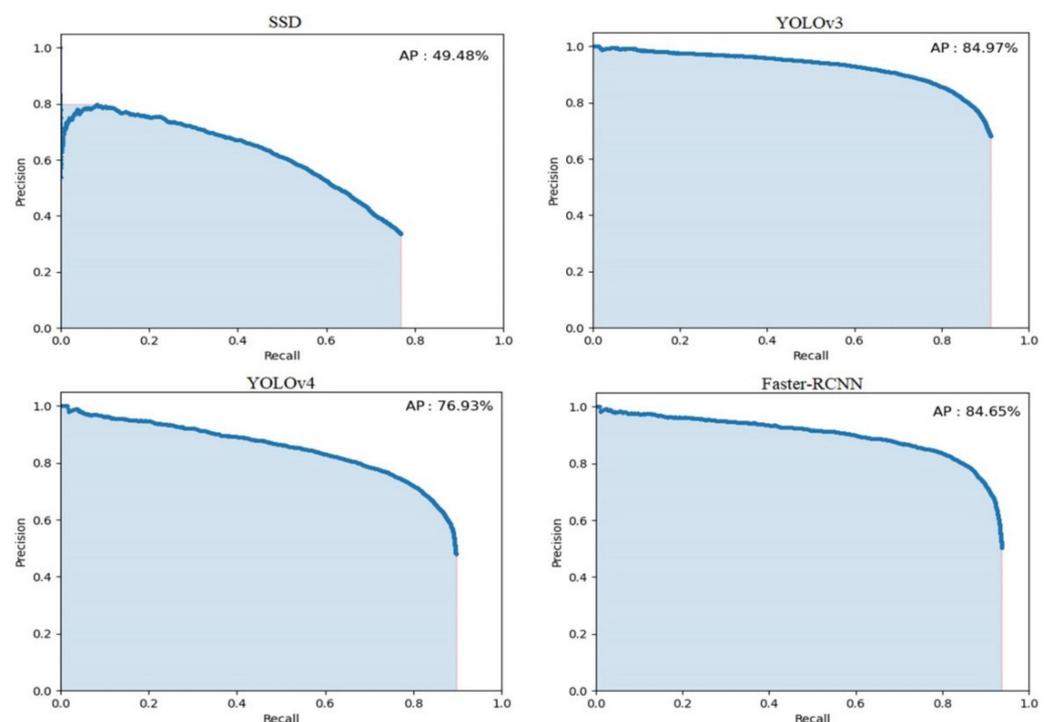
$$R = TP / TP + FN \quad (3)$$

where *TP* means that the test result is correct—that is, the category of the detection box and the label box are the same and the IOU is >0.5. *FP* signifies a target that has been wrongly identified, and *FN* represents a target that has not been detected.

Table 1. Accuracy evaluations of the four models.

	Test Area1, Nv = 165			Test Area2, Nv = 245			Test Area3, Nv = 359		
	N _d	N _o	OA (%)	N _d	N _o	OA (%)	N _d	N _o	OA (%)
SSD	44	3	24.85	37	1	14.69	66	0	18.38
YOLOv3	135	12	74.55	167	7	65.31	150	2	41.23
YOLOv4	146	10	82.42	211	10	82.04	209	7	56.27
Faster-RCNN	191	32	96.36	280	44	96.32	362	19	95.54

Precision refers to the proportion of the number of correctly identified targets compared to the total number of targets; it measures the degree of agreement between the predicted box and the real box. The curve drawn by different precision and recall points and the area AP formed by the X and Y axes were used to evaluate the prediction effect of each model. The larger the value, the better the model effect. Figure 8 shows that Faster-RCNN and YOLOv3 had the highest mAP (mean accuracy and precision) values (84.65% and 84.97%); for YOLOv4 mAP, it was 76%, and for SSD mAP, it was 49%. Therefore, the network model of Faster-RCNN is better than other models in terms of total accuracy and mAP accuracy, which is thereby more suitable for spruce identification and detection.

**Figure 8.** Precision–recall curves of the four network models.

4. Discussion

Traditional forest protection and management methods are mainly conducted through manual field investigations, which entail a heavy workload, low efficiency, and strong subjectivity. Furthermore, it is difficult to monitor using these methods at high altitudes and on steep slopes. The method proposed in this paper can establish a detection model for Tianshan spruce using UAV remote sensing and a target detection algorithm, thereby overcoming the above shortcomings. The identification effects of SSD, YOLOv3, YOLOv4, and Faster-RCNN models were compared through the identification of Tianshan spruce, which provided a method reference for quick detection and tree number estimation.

Although Faster-RCNN showed higher precision, it requires a longer training time than the other methods. A total of 5000 training samples were used in this study, and the training time of Faster-RCNN was 10 h more than the other three methods. There is still

more space for improvement in the structures of the YOLO and SSD models regarding parameter improvement and data selection [37]. Yu et al. [38] improved YOLOv4-FPM to achieve a real-time detection approach for bridge cracks. Hence, in the model architecture, corresponding residual units could be added to the residual block to obtain more target feature information. Regarding parameter improvement, attempts could be made to combine multiple parameters to obtain a better solution and therefore improve the usability of the model. Regarding data selection, a small sample data model could be used to avoid expanding the labor cost; the small sample data model could be used to expand the data set and thus improve the availability of the model.

Compared with traditional extraction methods for obtaining the number of trees, the target detection method is both rapid and precise. Considering real-time monitoring and model generality, combining UAV images with deep learning is better than the traditional image classification method for monitoring the distribution of Tianshan spruce. This approach can provide better technical support for studying Tianshan spruce in the typical watershed of the gentle slopes of Tianshan Mountain. Supervised learning based on the CNN model is limited by the large amount of labeled data needed as training samples for model training and feature learning [39]. Here, the training samples were optimized to reduce the over-fitting phenomenon of the model and to enhance it through image enhancement methods such as rotation, color change, and edge noise. Finally, a fast and high-precision detection effect was achieved for Tianshan spruce.

5. Conclusions

For forest planning and yield estimation, extracting the locations, numbers, and diameters of trees are important albeit difficult, and demanding. In this study, four detection models based on the target detection method were proposed (SSD, YOLOv3, YOLOv4, and Faster-RCNN), and the proposed approaches were tested in three different density areas. The most suitable model for detecting Tianshan spruce was determined by evaluating each model's quantity, pixel accuracy, and ability to identify and generate statistics for Tianshan spruce.

The SSD model had the lowest detection accuracy; its average accuracy over the three different density regions was below 20%. YOLOv3 achieved accuracies of 74.55% and 65.31%. YOLOv4 achieved accuracies of 82.42% and 82.04% in low and medium density regions but only produced accuracies of 41% and 56%, respectively, in the high-density region. The Faster RCNN model had the highest accuracy (96.36%, 96.32%, and 95.54%, respectively). The performances of these algorithms were similar, and their detection accuracies depended on the density of the trees. The accuracies of three algorithms (YOLOv3, YOLOv4, and Faster-RCNN) decreased with the increasing density; however, Faster-RCNN showed the smallest change.

Combining machine vision and UAV remote sensing can improve efficiency, ensure accuracy, and hence overcome the shortcomings of traditional survey methods. Therefore, the Faster-RCNN network model can be used to identify Tianshan spruce and has important practical significance for understanding and monitoring forest ecosystems. In order to carry out further studies, the output module could be used for the network by adding a coordinate system transformation function and geographic location information in the later stage. Thereby, the latitude and longitude information of each target detection result could be extracted and generated. Further research and discussions regarding the detection of dead spruce trees could then be explored.

Author Contributions: Conceptualization, M.E.; data curation, M.E.; formal analysis, E.A. and B.E.; funding acquisition, S.L.; investigation, A.A. and T.L.; methodology, E.A. and M.M.; Software, E.A.; supervision, S.L.; Writing—original draft, M.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the National Natural Science Foundation of China (41861053).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Acknowledgments: I would like to thank the anonymous referees for their helpful comments.

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

References

- Ding, Y.; Zang, R.G.; Huang, J.H.; Xu, Y.; Lu, X.H.; Guo, Z.J.; Ren, W. Intraspecific trait variation and neighborhood competition drive community dynamics in an old-growth spruce forest in northwest China. *Sci. Total. Environ.* **2019**, *678*, 25–532. [\[CrossRef\]](#)
- Jiao, L.; Jiang, Y.; Wang, M.C.; Kang, X.Y.; Zhang, W.T.; Zhang, L.N.; Zhao, S.D. Responses to climate change in radial growth of *Picea schrenkiana* along elevations of the eastern Tianshan Mountains, northwest China. *Dendrochronologia* **2016**, *40*, 117–127. [\[CrossRef\]](#)
- Sullivan, B.W.; Alvarez-Clare, S.; Castle, S.C.; Porder, S.; Reed, S.C.; Schreeg, L.; Townsend, A.R.; Cleveland, C.C. Assessing nutrient limitation in complex forested ecosystems: Alternatives to large-scale fertilization experiments. *Ecology* **2014**, *95*, 668–681. [\[CrossRef\]](#)
- Van der Sande, M.T.; Peña-Claros, M.; Ascarrunz, N.; Arets, E.J.M.M.; Licona, J.C.; Toledo, M.; Poorter, L. Abiotic and biotic drivers of biomass change in a Neotropical forest. *J. Ecol.* **2017**, *105*, 1223–1234. [\[CrossRef\]](#)
- Clark, J.S.; Bell, D.M.; Hersh, M.H.; Nichols, L. Climate change vulnerability of forest biodiversity: Climate and competition tracking of demographic rates. *Glob. Chang. Biol.* **2011**, *17*, 1834–1849. [\[CrossRef\]](#)
- Ozdemir, I.; Karnieli, A. Predicting forest structural parameters using the image texture derived from WorldView-2 multispectral imagery in a dryland forest, Israel. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 701–710. [\[CrossRef\]](#)
- Zhang, N.; Zhang, X.L.; Ye, L. Tree crown extraction based on segmentation of high-resolution remote sensing image improved peak-climbing algorithm. *Trans. Chin. Soc. Agric. Eng.* **2014**, *45*, 294–300.
- Wagner, F.H.; Ferreira, M.P.; Sanchez, A.; Hirye, M.C.; Zortea, M.; Gloor, E.; Phillips, O.L.; de Souza Filho, C.R.; Shimabukuro, Y.E.; Aragão, L.E. Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 362–377. [\[CrossRef\]](#)
- Koc-San, D.; Selim, S.; Aslan, N.; San, B.T. Automatic citrus tree extraction from UAV images and digital surface models using circular Hough transform. *Comput. Electron. Agric.* **2018**, *150*, 289–301. [\[CrossRef\]](#)
- Aubry-Kientz, M.; Dutrieux, R.; Ferraz, A.; Saatchi, S.; Hamraz, H.; Williams, J.; Coomes, D.; Piboule, A.; Vincent, G. A comparative assessment of the performance of individual tree crowns delineation algorithms from ALS data in tropical forests. *Remote Sens.* **2019**, *11*, 1086. [\[CrossRef\]](#)
- Duncanson, L.; Cook, B.; Hurtt, G.; Dubayah, R. An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. *Remote Sens. Environ.* **2014**, *154*, 378–386. [\[CrossRef\]](#)
- Gini, R.; Passoni, D.; Pinto, L.; Sona, G. Use of unmanned aerial systems for multispectral survey and tree classification: A test in a park area of northern Italy. *Eur. J. Remote Sens.* **2014**, *47*, 251–269. [\[CrossRef\]](#)
- Deng, X.; Tong, Z.; Lan, Y.; Huang, Z. Detection and location of dead trees with pine wilt disease based on deep learning and UAV remote sensing. *AgriEngineering* **2020**, *2*, 294–307. [\[CrossRef\]](#)
- Schiefer, F.; Kattenborn, T.; Frick, A.; Frey, J.; Schall, P.; Koch, B.; Schmidlein, S. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2020**, *170*, 205–215. [\[CrossRef\]](#)
- Ding, R.; Dai, L.; Li, G.; Liu, H. TDD-Net: A tiny defect detection network for printed circuit boards. *CAAI Trans. Intell. Technol.* **2019**, *4*, 110–116. [\[CrossRef\]](#)
- Lan, Y.; Zhu, Z.; Deng, X.; Lian, B.; Huang, J.; Huang, Z.; Hu, J. Monitoring and classification of citrus Huanglongbing based on UAV hyperspectral remote sensing. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 92–100.
- Tang, L.; Shao, G. Drone remote sensing for forestry research and practices. *J. For. Res.* **2015**, *26*, 791–797. [\[CrossRef\]](#)
- Lebourgeois, V.; Bégue, A.; Labbé, S.; Mallavan, B.; Prévot, L.; Roux, B. Can commercial digital cameras be used as multispectral sensors? A crop monitoring test. *Sensors* **2008**, *8*, 7300–7322. [\[CrossRef\]](#)
- Komarek, J. The perspective of unmanned aerial systems in forest management. Do we really need such details? *Appl. Veg. Sci.* **2020**, *23*, 718–721. [\[CrossRef\]](#)
- He, Y.; Zhou, X.C.; Huang, H.Y.; Xu, X.Q. Counting tree number in subtropical forest districts based on UAV remote sensing images. *Remote Sens. Technol. Appl.* **2018**, *33*, 168–176.
- Hernandez, J.G.; Ferreira, E.G.; Sarmiento, A.; Silva, J.; Nunes, A.; Correia, A.C.; Fontes, L.; de Brito Tavares, M.M.B.; Varela, R.A.D. Using high resolution UAV imagery to estimate tree variables in *Pinus pinea* plantation in Portugal. *For. Syst.* **2016**, *25*, 16. [\[CrossRef\]](#)
- Li, M.H.; He, F.H.; Liu, Y.; Pan, C.D. Spatial distribution pattern of tree individuals in the Schrenk spruce forest, northwest China. *Acta Ecol. Sin.* **2005**, *25*, 1000–1006. (In Chinese with English Abstract)

23. Zhang, H.; Xu, M.; Zhuo, L.; Havyarimana, V. A novel optimization framework for salient object detection. *Visual Comput.* **2016**, *32*, 31–41. [[CrossRef](#)]
24. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
25. Guan, H.; Cheng, B. How do deep convolutional features affect tracking performance: An experimental study. *Visual Comput.* **2018**, *34*, 1701–1711. [[CrossRef](#)]
26. Li, T.; Ye, M.; Ding, J. Discriminative Hough context model for object detection. *Visual Comput.* **2014**, *30*, 59–69. [[CrossRef](#)]
27. Girshick, R.; Donahue, J.; Darrell, T. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
28. Brill, M.H. Computer vision and pattern recognition: CVPR 92. *Color Res. Appl.* **2010**, *17*, 426–427. [[CrossRef](#)]
29. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot Multi-Box Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; 9905, pp. 21–37.
30. Redmon, J.; Farhadi, A. Yolo3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
31. Zhu, Q.; Zheng, H.; Wang, Y.; Cao, Y.; Guo, S. Study on the evaluation method of sound phase cloud maps based on an improved YOLOv4 algorithm. *Sensors* **2020**, *20*, 4314. [[CrossRef](#)]
32. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
33. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [[CrossRef](#)] [[PubMed](#)]
34. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS improving object detection with one line of code. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5562–5570.
35. Bao, J.; Wei, S.; Lv, J.; Zhang, W. Optimized faster-RCNN in real-time facial expression classification. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *790*, 012148. [[CrossRef](#)]
36. Fattal, A.K.; Karg, M.; Scharfenberger, C.; Adamy, J. Saliency-guided region proposal network for CNN based object detection. In Proceedings of the Saliency-Guided Region Proposal Network for CNN Based Object Detection, Yokohama, Japan, 16–19 October 2017; pp. 1–8.
37. Redmon, J.; Farhadi, A. Yolo9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
38. Yu, Z.W.; Shen, Y.G.; Shen, C.K. A real-time detection approach for bridge cracks based on YOLOv4-FPM. *Automat. Construct.* **2021**, *122*, 103514. [[CrossRef](#)]
39. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer: Cham, Switzerland, 2015; pp. 234–241.