

Original Russian Text © 2024 A. D. Nikitina published in Forest Science Issues Vol. 7, No 2, Article 146.

DOI: 10.31509/2658-607x-202473-150

## AUTOMATIC SEGMENTATION OF TREE CROWNS IN PINE FORESTS USING MASK R-CNN ON RGB IMAGERY FROM UAVS

© 2024

A. D. Nikitina

*Center for Forest Ecology and Productivity of the RAS  
Profsoyuznaya st. 84/32 bldg. 14, Moscow, 117997, Russia*

E-mail: nikitina.al.dm@gmail.com

Received: 18 May 2024

Revised: 05 June 2024

Accepted: 22 June 2024

This article presents the results of applying an improved method for automatic segmentation of RGB imagery (orthophotos) obtained using consumer-grade unmanned aerial vehicles (UAVs) based on the Mask R-CNN neural network. Preparation and post-processing blocks for raster and vector files were developed for geospatial data processing. The model was trained on 7.000 tree crowns identified in pine forest of drained habitats in the mixed coniferous-broadleaved forest subzone. Training was carried out using cross-validation. Additional data on 1.337 crowns were used for verification. Sequential filtering by area, score, and duplicate segments improved the quality of the final segmentation results for all age groups of pine forests. The model produced an average precision of 0.87, a recall of 0.81, and an F1-score of 0.83. The obtained results demonstrate the high efficiency of the filtering algorithm in reducing segment redundancy and increasing data reliability. The Mask R-CNN automatic segmentation method is an effective tool for studying the characteristics of pine forest stands based on RGB orthophotos obtained during UAV surveys; using this method, it is possible to reproduce the results of visual interpretation with high precision. This method is particularly effective for scaling studies to large areas, where manual interpretation would be labor-intensive.

**Keywords:** *Mask R-CNN, automatic segmentation, tree detection, pine forest, RGB imagery, UAVs, environmental monitoring, remote sensing*

In the context of global climate change, the issues of carbon balance and carbon stocks in forest ecosystems are becoming increasingly relevant. Effective monitoring of these parameters in forest ecosystems and forest resource management require obtaining detailed and accurate information about forest structure and condition (Espíndola, 2023). Within this context, the use of highly detailed GPS survey opens up new opportunities for environmental research and increases the efficiency of data collection and analysis. Forests with a predominance of Scotch pine (*Pinus sylvestris* L.) are common in the temperate latitudes of the Northern Hemisphere. This species is adaptable to diverse growing conditions and highly resistant to environmental stresses such as droughts and fires. Pine forests with their high ecological tolerance and significant contribution to the carbon cycle are an important object of environmental studies.

Existing studies (Medvedev et al., 2020; Tuominen et al., 2017; Nevalainen et al., 2017; Puliti et al., 2017; Ocer et al., 2020; Diez et al., 2021; Ball et al., 2023; Zhou et al., 2023) outline the prospects of UAV and neural network applications in image

processing for accurate and efficient study of forests. However, there are relatively few studies on forests of complex structure with a closed canopy, hence the need for further studies and higher accuracy of methods used in these conditions. It is also essential to take into account the potential for use of mass-market UAVs, since their availability and widespread use make these methods applicable in applied and research practice for a wider range of users. The *purpose* of this study is to evaluate the effectiveness of the automatic segmentation method using the Mask R-CNN neural network to identify individual trees in pine forest of different structures based on RGB orthophotos obtained using UAVs. In this study, segmentation of pine forests primarily focuses on the upper canopy layer, where the crowns of pine trees form the dominant component.

## MATERIALS AND METHODS

**Objects of research.** The objects of research are pine forests in the drained habitats of the coniferous-broadleaved forest subzone of the western part of the Russian Plain in the following protected areas: the *Kurshskaya Kosa* National Park (NP), the *Smolenskoye Poozerye* NP, the

*Bryansky Les* State Natural Biosphere Reserve (SNBR). Three age groups were identified in the forests under study: young (10–40 years old), middle-aged (40–80 years old) and old (over 80 years old). Some studies (Nezami et al., 2020; Diez et al., 2021) show that segmentation algorithm demonstrate the highest learning efficiency on single-species and structurally simple forests. With the data obtained in the *Kurshskaya Kosa* NP, it is possible to better adjust the model, as the studied pine forests of this park mainly have a single-species composition with a 10C stand formula. In accordance with the Guide to Identifying Forest Types in the European Russia (<http://cepl.rssi.ru/bio/forest/index.htm>), the young pine forests fall into the group of xerophytic green-moss and green-moss-lichen pine forest types. Xerophytic green-moss pine forests are predominant in both middle-aged and old forests. Data obtained in the pine forests of the *Smolenskoye Poozerye* and *Bryansky Les* improve the quality of annotation for segmentation algorithms applied to forest stands of mixed composition and complex structure, making these algorithms more versatile. In the *Smolenskoye Poozerye* NP, the studied areas in middle-aged and old forests are domi-

nated by low shrub-green-moss pine forests, whereas the young forests are mainly represented by small-grass-green-moss pine forests, with a closeness of 40–90%. Within the studied areas of the *Bryansky Les* Biosphere Reserve, middle-aged low shrub-green-moss pine forests and complex old pine forests with linden and oak are predominant, with a closeness of 70–80%.

**Aerial survey using UAVs.** In this study, the results of survey carried out using Phantom 3 Advanced and Mavic Pro UAVs by DJI were used. These devices are affordable and equipped with RGB cameras. Flight tasks were simulated using DroneDeploy software. The flights were performed at an altitude of 100–200 m depending on the complexity of the terrain and the height of the forest canopy with a longitudinal and transverse overlap of 90% with constant calm weather (wind velocity up to 10 m/s) at about 11:00 to 16:00 local time. The survey area using such parameters is ~15 hectares per battery.

**Aerial images processing** was carried out using Agisoft Metashape software and included the following main stages: uploading images; image alignment; building a dense point cloud; building a digital terrain model (DTM); building an

orthophoto; rasterizing the DTMs and the orthophotos. The calculated spatial resolution of the DTMs varied from 15 to 32 cm/pixel depending on the altitude of the flight, whereas that of orthophotos was from 2 to 8 cm/pixel. The total number of initial images used to create orthophotos exceeded 20000, with an average orthophoto resolution of 5.9 cm/pixel and DTM resolution of 23.6 cm/pixel. The total number of orthophotos was 55.

**Visual interpretation.** During interpretation, the boundaries of the crowns of individual trees were delineated manually using the QGIS software. Annotation is necessary to create training and validation datasets. It is carried out for the central sections of orthophotos ranging in size from 20×20 m (for young trees) to 100×100 m (for middle-aged and old forests). As a result, files with spatial vector data (.shp) with the crowns of individual trees in pine forest, as well as in additional key areas of diverse composition, were obtained to expand the training set during further segmentation. The final set of visual annotation data included ~8.300 individual crowns and covered 55 key sites. The most crowns were identified in pine forest — 6.799, including 3.330 in middle-aged

forests, 2.325 in old forests, and 1.144 in young forests. Auxiliary areas with other species make up a smaller part (broadleaved forests — 562, spruce forests — 823, small-leaved forests — 141).

**Automatic segmentation using the Mask R-CNN neural network.** The ultra-precision Mask R-CNN neural network developed in 2017 (He et al., 2017) is used to identify individual tree crowns on aerial images. It expands the capabilities of the Faster R-CNN neural network due to the added module that predicts segmentation masks for regions of interest (RoI). This module works together with the classification and regression of the Bounding box. One of the features of Mask R-CNN is pixel-to-pixel alignment, which is not implemented in Fast/Faster R-CNN.

The following tools, frameworks and libraries were used to create the project: CUDA, Jupyter Notebook, QGIS, PyTorch, Rasterio, fiona, and Matplotlib. The study was carried out using the Mask R-CNN model in the Python PyTorch machine learning framework, pre-trained on the *COCO* dataset which includes more than 330.000 images and 1.5 million objects. The pre-trained model is able to identify the boundaries of objects in images. However,

additional training is required for tree crown detection. This process is simpler and faster than learning from scratch, it reduces the risk of getting stuck in local minima and reduces the number of necessary adjustments. The neural network parameters were reconfigured for the classification of regions of interest, creation of bounding boxes, and mask segmentation.

*Creating a dataset for model training.*

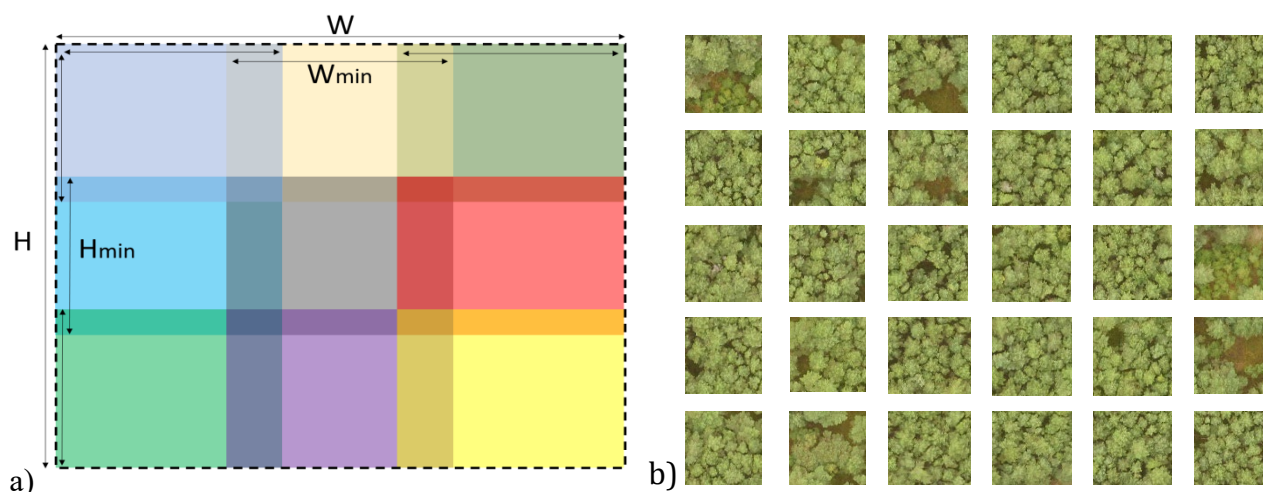
The first step is to create a specialized dataset. The following source data were used for this task: UAV orthophotos (in GeoTIFF .tiff format); manually selected vector boundaries of tree crowns (in .shp format); boundaries of the areas under study (in .shp format). The model was trained on the contours of seven thousand tree crowns (84%) obtained during visual annotation from all the studied areas. The size of the validation set was 1.337 crowns (16%) from key sites of pine forests, stratified taking into account the object and the age of the stand.

Preparation stages for the data set included checking and adjustment of the vector file geometry (irregular shapes of objects, self-intersections, etc.), data re-projection; conversion of images to 24-bit format; ensuring the correct alignment of

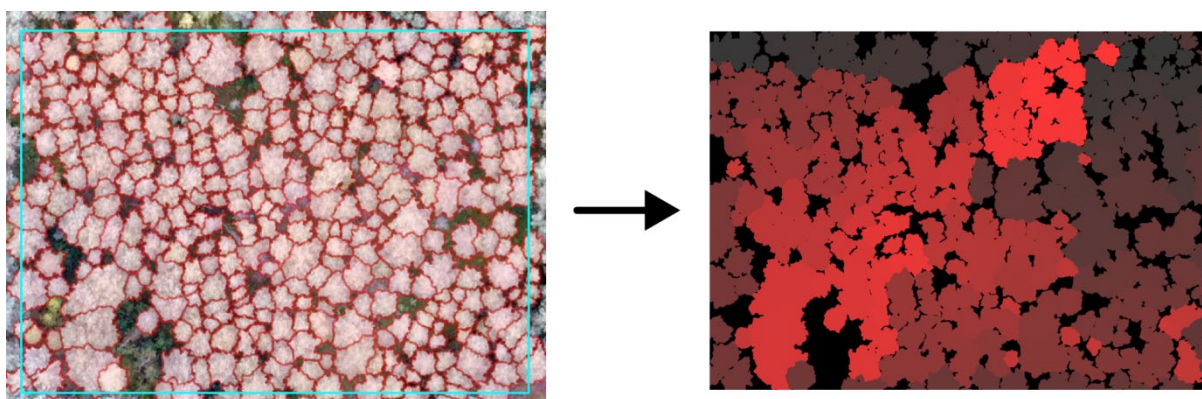
input data using uniform boundaries of the studied sites. The model accepts input images as a  $[W, H, 3]$  matrix (where  $W$  and  $H$  stand for width and height, respectively, and 3 is the number of RGB channels), therefore the original images were split into components of equal size based on the optimal grid. The width and height of each component were determined as the minimum among all images (Fig. 1).

A set of binary images of each segment (crown) where pixels have a value of [0] for the background and [1] for the crown was used for annotation (Fig. 2).

Thus, files with visual annotation were transformed into multidimensional arrays, with each tree crown saved as a separate image. The final dataset accepted by the model included raster orthophotos (.png), information about the boundaries of the sites with the transformation configuration (.json), annotation masks in the form of rasters where each crown corresponds to a certain numerical pixel value (.png), and orthophotos cropped along the extended boundary for the correct identification of marginal objects (.png). Additionally, vector annotation data were loaded, which are not involved in training and are used at the stage of evaluation of the model performance.



**Figure 1.** (a) A schematic diagram for splitting the source raster images into separate components, where  $W$  and  $H$  stand for width and height, respectively, and 3 is the number of RGB channels, and (b) individual components after splitting.



**Figure 2.** Converting an input dataset into a multidimensional array format.

*Training of the Mask R-CNN neural network model.* To train the model, a prepared dataset was used, which was divided into training and validation subsets. Validation data were not used in the learning process. Repeated  $k$ -fold cross-validation was used for training data ( $k=10$ ). The

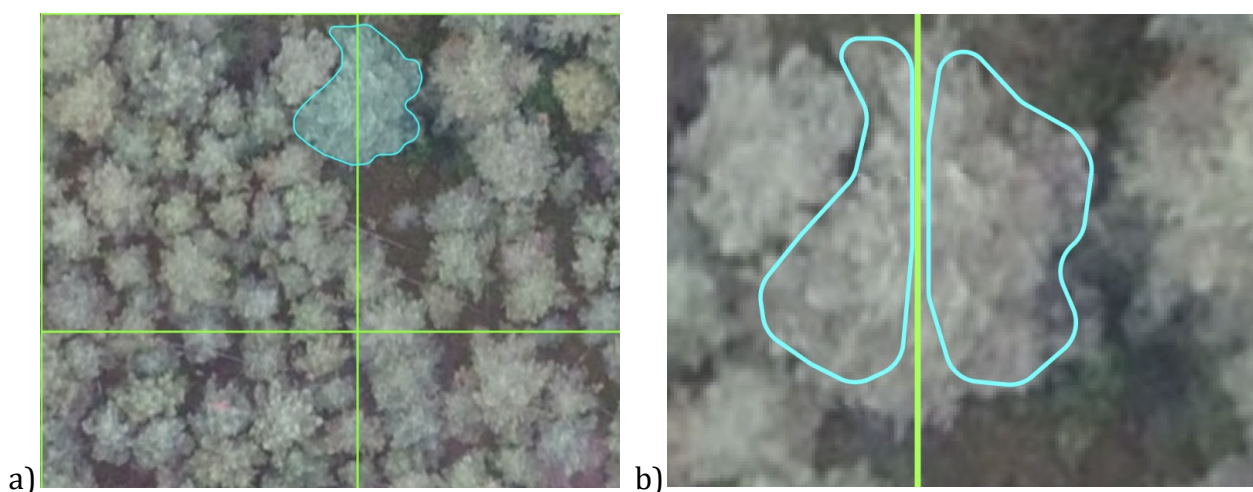
learning parameters of the neural network included the initial learning rate (LR), the LR schedule, the LR factor, the regularization coefficient, and the Stochastic Gradient Descent (SGD) parameter. The model was trained over 9 epochs, with the entire dataset passing through the neural network

in each epoch, and the model weights adjusted. Transformations of the input image to enlarge the dataset included rotations as well as contrast, saturation, and brightness adjustments with a total probability of change 0.1. To prevent overfitting, the model was regularly evaluated using a validation dataset. Neural network functioning resulted in the creation of a multidimensional stack of monochrome images of each object with an indication of the confidence level ranging from 0 to 1. This indicator reflects the probability that the output object belongs to a given class (crown).

*Processing of data obtained by the model.* To analyze the data of the trained neural network, orthophotos of various test areas in the .tiff format with the boundaries

of key sites and the results of visual interpretation were transmitted for subsequent calculation of model quality metrics. Considering that the model works with images in pixel coordinates, an additional metadata file was created with information about the coordinate system, coordinates of the corners of the segmented area and the geographical reference.

Due to the limitation of the model to process no more than 100 segments (crowns) per image, it was necessary to divide the source raster into parts (patches) so that each of them had less than 100 crowns. However, when dividing images strictly according to the grid, distortions in segments at the boundaries of the sections are possible (Fig. 3).



**Figure 3.** An example of a) splitting an image into patches and b) a possible error in segmentation of marginal areas.

The problem of marginal crowns which occurs when splitting an image in accordance with a grid was solved by creating a 50% overlap between patches and defining a zone of ignored boundaries to rule out edge effects. The offset information was saved to further restore the coordinates of the components in the original full-size image. The use of a pyramid of patches, where the size of components doubles at each level, provided adaptation to different scales of objects. The resulting segments were converted into images, and then into a vector format using the marching squares algorithm (confidence threshold = 0.5) implemented in the *Skimage* library. Crowns falling in the zone of ignored boundaries were removed from the results, thus reducing the number of duplicated and marginal crowns. The remaining segments were combined into a single dataset, where the score was indicated for each identified crown.

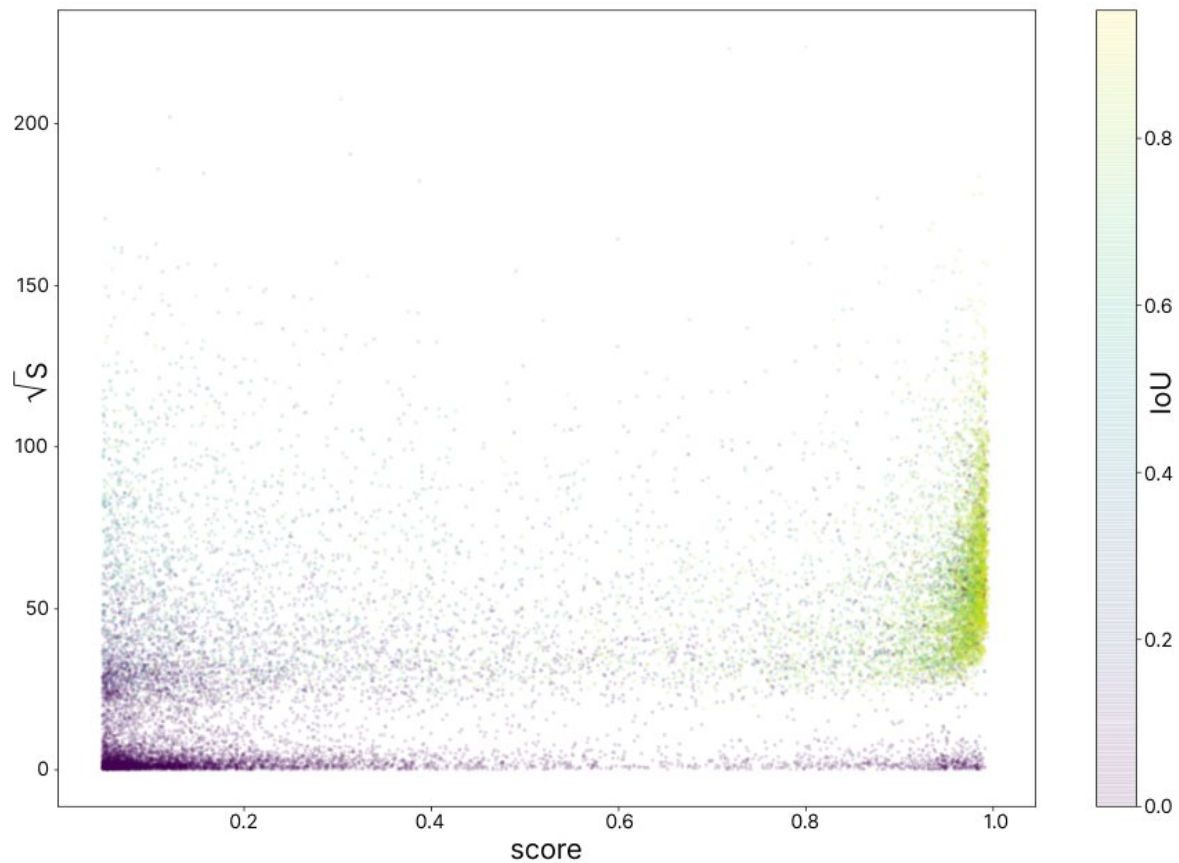
#### *Filtering of the segmentation data.*

Filtering in the context of neural network data processing is an important step aimed at improving the quality of results. During the splitting of images into separate components, a large overlap is assumed, which prevents the model from missing the marginal crowns of trees; however, when

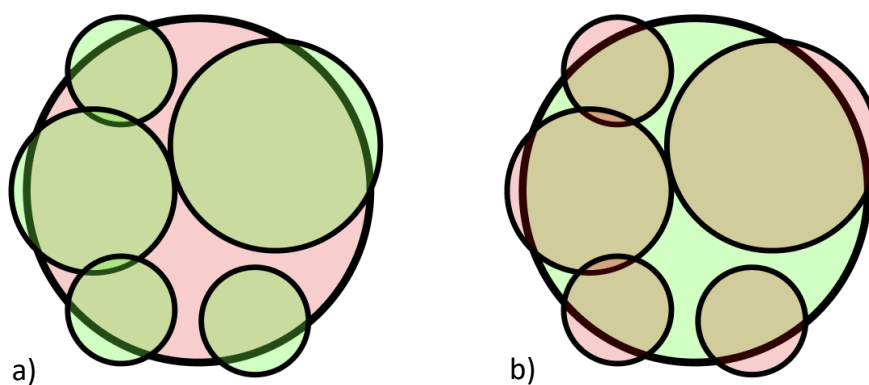
assembling the segmentation results into one dataset, duplicates of the same crown are created. In addition to possible duplicates, there may be objects with an irregular shape, unreliable area, or with a low confidence level. All this requires careful filtering which includes the analysis of various parameters in order to determine the optimal criteria for removing unwanted segments.

Optimal parameters for the filtering algorithm were calculated based on data with the calculated Intersection over Union (IoU) metric, which measures the degree of intersection between the crowns predicted by the neural network and the crowns identified visually. These data consisted of a set of points, where each point corresponded to a detected crown with several parameters (area, score, IoU). During the scatter plot analysis (Fig. 4), the threshold for filtering segments with a small area is clearly seen. This filtering step significantly reduced the number of excessively segmented crowns (28%) without significant loss of precision. The scatter plot also shows that the majority of segments with the minimal IoU values also have a low score, so it is advisable to use the score as a filtering parameter in the future.





**Figure 4.** A scatter plot showing the relationship between the area of a segment ( $\sqrt{S}$ , pixels), its score, and the degree of intersection of the segment predicted by the neural network with the reference segment of the annotation (IoU).



**Figure 5.** Filtering options for overlapping crowns with different confidence levels (green segments are saved in the final list, whereas red ones are removed).

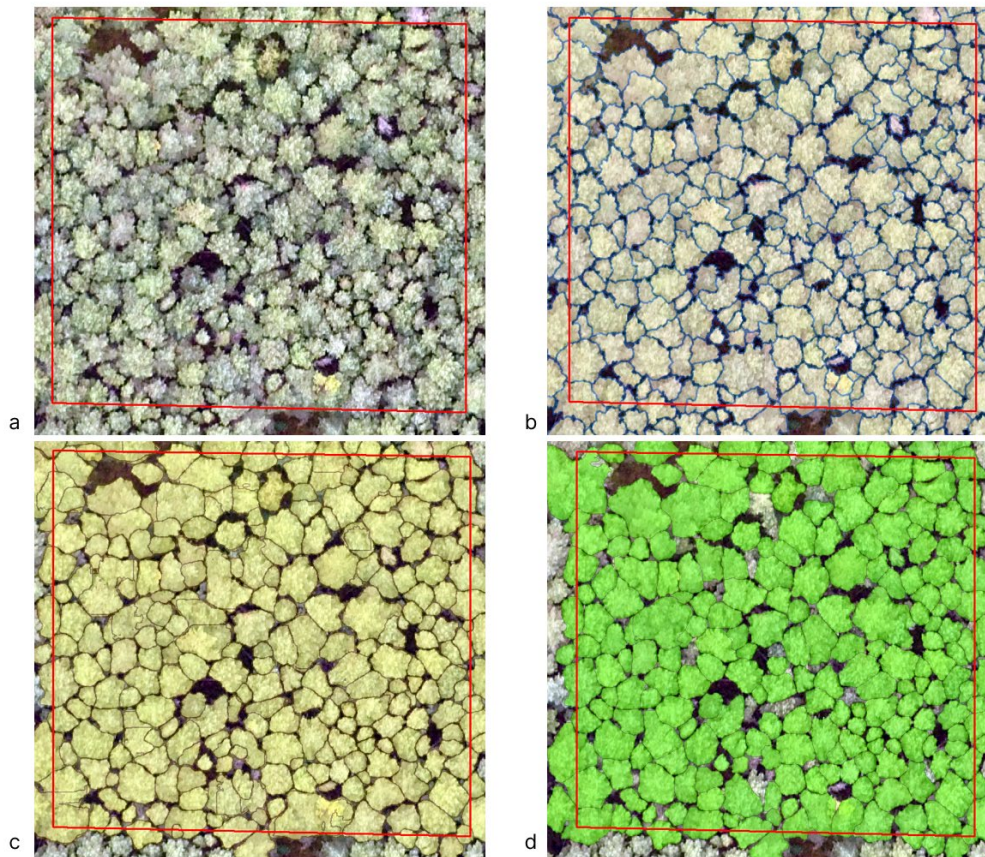
To remove duplicate crowns, filtering was used based on the degree of intersection and reliability of the data. If there is a significant overlap between two crowns, the one with a higher score of the neural network is selected. If there is one large and multiple small overlapping segments, only the large one remains if its score is higher. If its score is lower, such a segment is excluded. Therefore, if there is high confidence in smaller segments, a large crown is discarded (Fig. 5a), and vice versa (Fig. 5b). Removing duplicates also contributed to a significantly higher precision.

The developed sequential filtering included criteria for area and score, and then for segment duplicates. This made it possible to preserve high-quality segments, minimizing loss in recall.

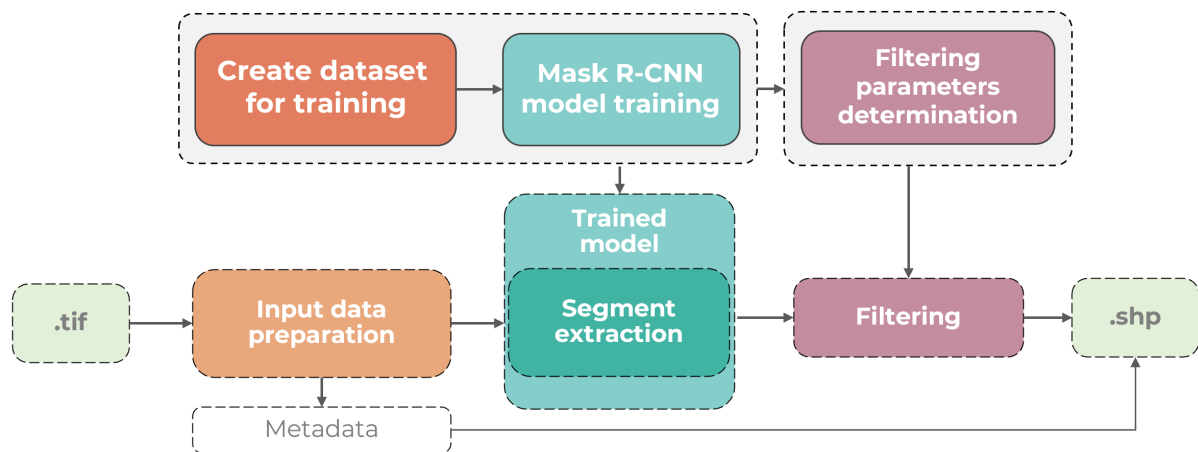
*Creation of the final vector layer.* To transform the coordinates of the segments into geographical ones and create the final

vector file in the .shp format, a metadata file created during the data processing stage was used. Figure 6 shows the neural network's crown segmentation results, demonstrating the precision of coordinate transformation.

The diagram (Fig. 7) shows the final data processing process for the segmentation of pine tree crowns using the Mask R-CNN algorithm. The upper part of the diagram reflects the research stages, including the creation of a training dataset, actual training of the Mask R-CNN model and selection of the filtering parameters. These steps are performed once to set up the model. The lower part of the diagram demonstrates the processing stage. It begins with uploading raster files, and then the input data is prepared, followed by segment selection and filtering of the results. As a result, a vector file is created that can be used for further analysis.



**Figure 6.** An example of a sample site, where (a) is an orthophoto without annotation, (b) is visual interpretation, (c) is the result of automatic segmentation without filtering, and (d) is the result of automatic segmentation after filtering.



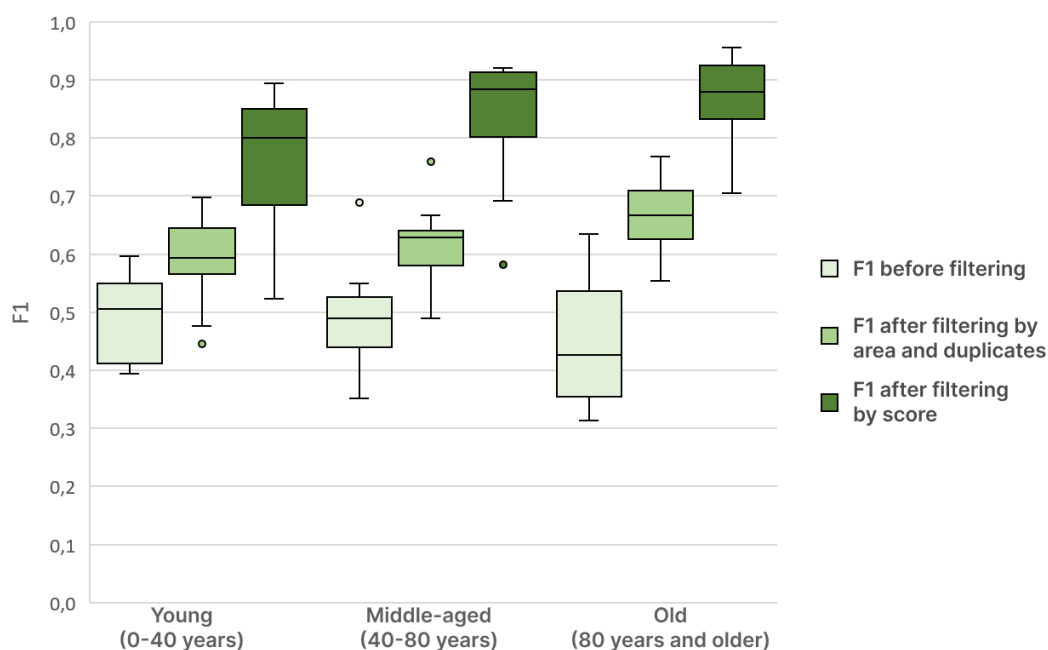
**Figure 7.** Data processing steps for automatic crown detection.

*Model quality metrics.* To evaluate the precision of the neural network in recognizing tree crowns, the results of visual interpretation of orthophotos were used. The crown was considered to be detected correctly if the IoU exceeded 0.5, which is the standard value found in studies on segmentation (Aubry-Kientz et al., 2019; Hao et al., 2021; Ball et al., 2023). For the model quality evaluation, standard error matrices were calculated: TP (true positive) — correct determination of a crown; FP (false positive) — incorrect determination of an object as a crown; FN (false negative) — incorrect exclusion of a crown; TN (true negative) — equal to 0 in segmentation tasks. Based on these data, key metrics were calculated as follows: precision — the proportion of correctly identified crowns among all recognized ones; recall — the proportion of correctly identified crowns from all actually existing ones; F1-score — the harmonic mean between precision and recall, which ensures a balance of these parameters.

## RESULTS AND DISCUSSION

At all key sites, the initial results of neural network segmentation had high recall (0.91 for all sites) values and low precision (0.31) and F1-score (0.46) values; segment redundancy can also be seen in the ratio of the number of segmented crowns to visual detection data. Filtering significantly improved the final average precision (0.87) and F1-score (0.83), while the final recall slightly decreased (recall = 0.81).

The chart (Fig. 8) reflecting the change in the F1-score at different filtering stages for different age groups of pine forest shows that after all filtering stages, the median values of the F1-score increase for all age groups, which is evidence of improved quality of segmentation for the entire sample. However, the improvement after filtering is most pronounced for old pine forests (over 80 years old). This is indicative of the effectiveness of the applied filtering approach to improve the quality of segmentation results, reducing redundancy and increasing data reliability.

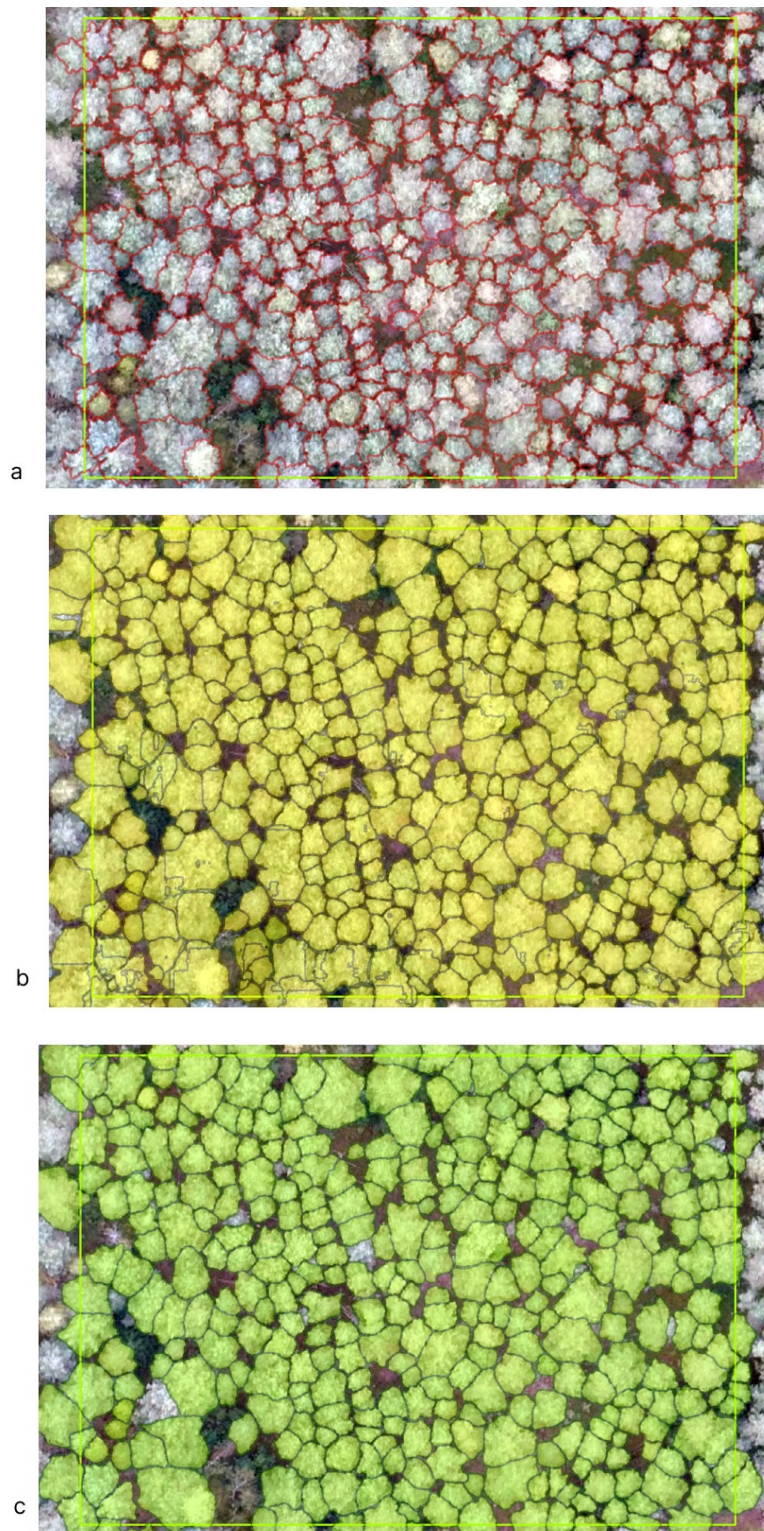


**Figure 8.** Changes in F1-score at different filtering stages for pine forests of different age groups.

The analysis of the training and validation samples used to control overfitting when setting up the model showed the greatest differences in the group of young pine forests ( $F1_{\text{training}} = 0.81$ ,  $F1_{\text{validation}} = 0.70$ ) with a median value of 0.8 for all key sites. Middle-aged and old pine forest have high F1-scores for both the training (0.84 and 0.88, respectively) and validation sets (0.83 and 0.82, respectively) of the sites, with a median F1-score of 0.88 for both groups.

The spread of model quality values in pine forests was 0.53–0.96 with an average value of 0.83 and a median value of 0.85. In

young forests, the spread of results ( $F1 = 0.53$ – $0.89$ ) shows lower adaptability of the model to some stands of this age, whereas the results are high on average ( $F1_{\text{average}} = 0.77$ ,  $F1_{\text{median}} = 0.8$ ). This may be due to the low quality of the survey (for low stands, it would make sense to carry out UAV surveys at an altitude below 120–180 m when using mass-market UAVs), difficulties with detecting individual trees in dense stands, and a smaller training sample for sample sites with young pine forests. The results were more stable ( $F1 = 0.7$ – $0.96$ ) for old forests with an average value of  $F1 = 0.86$  ( $F1_{\text{median}} = 0.88$ ).



**Figure 9.** An example of a key site where (a) is visual interpretation, (b) is the result of segmentation without filtering, and (c) is the result of segmentation after filtering.

The study showed that the adapted Mask R-CNN model provides high precision of results in different age groups of pine forest with different values of closeness, since the indicators of segmentation quality in key sites remain high for all datasets. An example of segmentation results is shown in Figure 9.

In studies on segmentation of individual trees in stands, closeness of the stands under study is an important factor. N. E. Ocer et al. (2020) carried out a study on detection of individual trees using the Mask R-CNN and Feature Pyramid Nets (FPNs) and obtained F1-scores within 0.82–0.91 for three test images. Sparse stands were analyzed in the study by N. V. Ivanova et al. (Ivanova et al., 2021), where watershed and region growing methods were applied and F1-scores of 0.7–0.9 were obtained. More closed stands are considered in the paper by X. Chen et al. (2023), with F1-scores between 0.71 and 0.79. The study by M. Beloiu et al. (2023) is focused on closed and diverse species stands with F1-scores ranging from 0.44 to 0.92. Studies show that the effectiveness of crown segmentation depends on the closeness of stands, and with its increase, the precision of segmentation becomes more variable.

## CONCLUSIONS

The method of automatic image segmentation using the Mask R-CNN neural network is an effective tool for research of pine forest that can reproduce the results of visual interpretation with high precision. The splitting of RGB orthophotos made it possible to take into account individual tree crowns in a closed canopy to the fullest extent possible. The initial results had high recall values; however, a result filtering unit was developed to increase precision. Filtering made it possible to eliminate redundant segments and improve the precision of the results, while maintaining a high degree of crown recognition. For all age groups of pine forests, there was an increase in F1-score after filtering. The final model demonstrates consistently high segmentation quality (F1-score = 0.83) of pine forest.

## FINANCING

The work was carried out with support from the Laboratory of Forest Climate-regulating Functions (project 122111500023-6) of the Center for Forest Ecology and Productivity of the RAS (CEPF RAS).

## REFERENCES

- Agisoft Metashape, available at: <http://www.agisoft.com> (2024, 01 June).
- Aubry-Kientz M., Dutrieux R., Ferraz A., Saatchi S., Hamraz H., Williams J., A comparative assessment of the performance of individual tree crowns delineation algorithms from ALS data in tropical forests, *Remote Sensing*, 2019, Vol. 11, No 9, pp. 1086 (1–21).
- Ball J. G., Hickman S. H., Jackson T. D., Koay X. J., Hirst J., Jay W., Coomes D. A., Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN, *Remote Sensing in Ecology and Conservation*, 2023, Vol. 9, No 5, pp. 641–655.
- 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 Sensing*, 2023, Vol. 15, p. 1463.
- Chen X., Shen X., Cao L., Tree Species Classification in Subtropical Natural Forests Using High-Resolution UAV RGB and SuperView-1 Multispectral Imageries Based on Deep Learning Network Approaches: A Case Study within the Baima Snow Mountain National Nature Reserve, China, *Remote Sensing*, 2023, Vol. 15, p. 2697.
- Diez Y., Kentsch S., Fukuda M., Caceres M. L. L., Moritake K., Cabezas M., Deep Learning in Forestry Using UAV-Acquired RGB Data: A Practical Review, *Remote Sens.*, 2021, Vol. 13, p. 2837.
- Espíndola R. P., Ebecken N. F. F., Advances in remote sensing for sustainable forest management: monitoring and protecting natural resources, *Revista Caribeña de Ciencias Sociales*, 2023, Vol. 12, No 4, pp. 1605–1617.
- Hao Z., Lin L., Post C.J., Mikhailova E.A., Li M., Chen Y. et al., 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, Vol. 178, pp. 112–123.
- He K., Gkioxari G., Dollár P., Girshick R., Mask R-CNN, *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.
- <https://cepl.rssi.ru/bio/forest/index.htm> (2024, 1 June).
- Ivanova N. V., Shashkov M. P., Shanin V. N., Study of pine forest stand structure in



- the priosko-terrasny state nature biosphere reserve (Russia) based on aerial photography by quadrocopter, *Nature Conservation Research*, 2021, Vol. 6, No 4, pp. 1–14.
- Medvedev A. A., Tel'nova N. O., Kudikov A. V., Alekseenko N. A., Analiz i kartografirovanie strukturnyh parametrov redkostojnyh severotajozhnyh lesov na osnove fotogrammetricheskikh oblakov toчек (Use of photogrammetric point clouds for the analysis and mapping of structural variables in sparse northern boreal forests), *Sovremennye problemy distancionnogo zondirovanija Zemli iz kosmosa*, 2020, Vol. 17, No 1, pp. 150–163.
- Nevalainen O., Honkavaara E., Tuominen S., Viljanen N., Hakala T., Yu X., Hyyppä J., Saari H., Pölönen I., Imai N. N., Tommaselli A. M. G., Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging, *Remote Sensing*, 2017, Vol. 9, No 3, p. 185.
- 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, Vol. 12, No. 7, p. 1070.
- Ocer N. E., Kaplan G., Erdem F., Matci D.K., Avdan U., Tree extraction from multi-scale UAV images using Mask R-CNN with FPN, *Remote Sensing*, 2020, Vol. 11, p. 847–856.
- Puliti S., Ene L. T., Gobakken T., Næsset E., Use of partial-coverage UAV data in sampling for large scale forest inventories, *Remote Sensing of Environment*, 2017, Vol. 194, pp. 115–126.
- Tuominen S., Näsi R., Honkavaara E., Balazs A., Hakala T., Viljanen N., Reinikainen J., Tree species recognition in species rich area using UAV-borne hyperspectral imagery and stereophotogrammetric point cloud, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2017, Vol. XLII-3/W3, pp. 185–194.
- Zhou J., Chen X., Li S., Dong R., Wang X., Zhang C., Zhang L., Multispecies individual tree crown extraction and classification based on BlendMask and high-resolution UAV images, *Journal of Applied Remote Sensing*, 2023, Vol. 17, No 1, p. 016503.

# АВТОМАТИЧЕСКАЯ СЕГМЕНТАЦИЯ КРОН СОСНОВЫХ ДРЕВОСТОЕВ С ИСПОЛЬЗОВАНИЕМ MASK R-CNN НА RGB-ОРТОФОТОПЛАНАХ БПЛА

**А. Д. Никитина**

*Центр по проблемам экологии и продуктивности лесов РАН  
Россия, 117997 Москва, ул. Профсоюзная, 84/32, стр. 14*

E-mail: nikitina.al.dm@gmail.com

Received: 18.05.2024

Revised: 05.06.2024

Accepted: 22.06.2024

В статье представлены результаты применения усовершенствованного метода автоматической сегментации RGB-ортофотопланов, полученных с помощью беспилотных летательных аппаратов (БПЛА) общедоступного сегмента, на основе нейронной сети архитектуры Mask R-CNN. Для работы с геопространственными данными разработаны блоки подготовки и постобработки растровых и векторных файлов. Модель обучена на 7000 кронах, выделенных в сосновых древостоях дренированных местообитаний подзоны хвойно-широколиственных лесов. Обучение выполнено с применением кросс-валидации. Для верификации дополнительно использованы данные 1337 кроны. На этапе последовательной фильтрации по площади, уровню достоверности и повторам сегментов качество итоговых результатов сегментации улучшилось для всех возрастных групп сосновых древостоев. Итоговая средняя точность (precision) равна 0.87, полнота (recall) – 0.81 и F1-мера (F-score) – 0.83. Полученные результаты свидетельствуют о высокой эффективности алгоритма фильтрации для уменьшения избыточности сегментов и увеличения надёжности данных. Метод автоматической сегментации Mask R-CNN представляет собой эффективный инструмент для исследований характеристик сосновых древостоев по RGB-ортофотопланам БПЛА-съёмки, способный с высокой точностью воспроизводить результаты визуального дешифрирования. Метод особенно эффективен при масштабировании исследований на большие территории, где ручное дешифрирование становится трудоемким.

**Ключевые слова.** *Mask R-CNN, автоматическая сегментация, выделение деревьев, сосновые древостои, RGB-ортофотопланы, БПЛА, экологический мониторинг, дистанционное зондирование*