COMPUTER RESEARCH AND MODELING
2025 VOL. 17 NO. 4 P. 641–663
DOI: 10.20537/2076-7633-2025-17-4-641-663

Ки&М

UDC: 004.415.2:004.932.1:582.47

Advanced neural network models for UAV-based image analysis in remote pathology monitoring of coniferous forests

C. R. Machuca[a], **N. G. Markov**[b]

National Research Tomsk Polytechnic University,
30 Lenina ave., Tomsk, 634050, Russia

E-mail: [a] kristianrodrigo1@tpu.ru, [b] markovng@tpu.ru

Received 19.05.2025, after completion — 12.07.2025.
Accepted for publication 21.07.2025.

The key problems of remote forest pathology monitoring for coniferous forests affected by insect pests have been analyzed. It has been demonstrated that addressing these tasks requires the use of multiclass classification results for coniferous trees in high- and ultra-high-resolution images, which are promptly obtained through monitoring via satellites or unmanned aerial vehicles (UAVs). An analytical review of modern models and methods for multiclass classification of coniferous forest images was conducted, leading to the development of three fully convolutional neural network models: Mo-U-Net, At-Mo-U-Net, and Res-Mo-U-Net, all based on the classical U-Net architecture. Additionally, the Segformer transformer model was modified to suit the task. For RGB images of fir trees *Abies sibirica* affected by the four-eyed bark beetle *Polygraphus proximus*, captured using a UAV-mounted camera, two datasets were created: the first dataset contains image fragments and their corresponding reference segmentation masks sized $256 \times 256 \times 3$ pixels, while the second dataset contains fragments sized $480 \times 480 \times 3$ pixels. Comprehensive studies were conducted on each trained neural network model to evaluate both classification accuracy for assessing the degree of damage (health status) of Abies sibirica trees and computation speed using test datasets from each set. The results revealed that for fragments sized $256 \times 256 \times 3$ pixels, the At-Mo-U-Net model with an attention mechanism is preferred alongside the Modified Segformer model. For fragments sized $480 \times 480 \times 3$ pixels, the Res-Mo-U-Net hybrid model with residual blocks demonstrated superior performance. Based on classification accuracy and computation speed results for each developed model, it was concluded that, for production-scale multiclass classification of affected fir trees, the Res-Mo-U-Net model is the most suitable choice. This model strikes a balance between high classification accuracy and fast computation speed, meeting conflicting requirements effectively.

Keywords: coniferous forests pathological monitoring, unmanned aerial vehicle, small spruce bark beetle *Polygraphus proximus*, multiclass classification of Siberian fir trees *Abies sibirica* images, fully convolutional neural networks, transformers

Citation: *Computer Research and Modeling*, 2025, vol. 17, no. 4, pp. 641–663.

КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ И МОДЕЛИРОВАНИЕ 2025 Т. 17 № 4 С. 641–663
DOI: 10.20537/2076-7633-2025-17-4-641-663

Ки&М

АНАЛИЗ И МОДЕЛИРОВАНИЕ СЛОЖНЫХ ЖИВЫХ СИСТЕМ

УДК: 004.415.2:004.932.1:582.47

Модели нейронных сетей для анализа изображений с БПЛА при дистанционном лесопатологическом мониторинге хвойных лесов

К. Р. Мачука[a], Н. Г. Марков[b]

Национальный исследовательский Томский политехнический университет,
Россия, 634050, г. Томск, пр. Ленина, д. 30

E-mail: [a] kristianrodrigo1@tpu.ru, [b] markovng@tpu.ru

Получено 19.05.2025, после доработки — 12.07.2025.
Принято к публикации 21.07.2025.

Рассмотрены основные задачи дистанционного лесопатологического мониторинга пораженных насекомыми-вредителями хвойных лесов. Показано, что при их решении необходимо использовать результаты мультиклассификации хвойных деревьев на изображениях высокого и сверхвысокого разрешения, оперативно получаемых при мониторинге путем съемки лесов с космических аппаратов или с беспилотных летательных аппаратов (БПЛА). Проведен аналитический обзор современных моделей и методов мультиклассификации изображений хвойных лесов и с учетом его результатов разработаны три модели полносверточных нейронных сетей Mo-U-Net, At-Mo-U-Net и Res-Mo-U-Net, основанные на классической модели U-Net, а также модифицирована модель трансформера Segformer. По RGB-изображениям поврежденных уссурийским полиграфом *Polygraphus proximus* деревьев пихты сибирской *Abies sibirica*, полученных с помощью фотокамеры на БПЛА, созданы два набора датасетов: первый набор включает фрагменты изображений и их эталонных масок сегментации размером $256 \times 256 \times 3$ пикселей, а второй — фрагменты размером $480 \times 480 \times 3$ пикселей. Проведены комплексные исследования каждой из обученных моделей нейросетей по точности классификации степени поражения (состояния здоровья) деревьев *A. Sibirica* на изображениях и по скорости вычисления моделей с использованием тестовых датасетов из каждого набора. Выявлено, что в случае фрагментов размером $256\times256\times3$ пикселей предпочтение наряду с моделью Modified Segformer следует отдать модели с механизмом внимания At-Mo-U-Net, а в случае фрагментов размером $480 \times 480 \times 3$ пикселей — гибридной модели с остаточными блоками Res-Mo-U-Net. Из результатов исследований точности классификации и скорости вычислений каждой из разработанных моделей сделан вывод о том, что при решении задачи мультиклассификации пораженных деревьев пихты в производственных масштабах предпочтение следует отдать модели Res-Mo-U-Net. Именно она является компромиссным вариантом, удовлетворяющим противоречащим друг другу требованиям высокой точности классификации деревьев на изображениях и высокой скорости вычислений модели.

Ключевые слова: патологический мониторинг хвойных лесов, беспилотный летательный аппарат, стволовой вредитель уссурийский полиграф *Polygraphus proximus*, мультиклассификация изображений деревьев пихты сибирской *Abies sibirica*, полносверточная нейронная сеть, трансформер

1. Introduction

Forests are crucial components of the Earth's biosphere, providing habitats for numerous animal and plant species and regulating global climate conditions. The health and viability of these ecosystems significantly affect the well-being of our planet's population. However, in recent decades, forests around the world have faced various negative challenges, including unprecedented climate change, the impact of insect pests, and harmful human activities. Consequently, monitoring and maintaining forest health has become increasingly important. Particularly critical is the monitoring of coniferous forests, which are susceptible to invasions by stem pests. Outbreaks of these pests can cause irreparable damage to coniferous forests in many countries [Chang et al., 2012; Lierop et al., 2015]. Forest management aims to align the ecological integrity of forests with human needs. This approach relies heavily on accurate and up-to-date information about forest conditions obtained through monitoring. Such information is essential for making informed decisions regarding conservation policies and harvesting plans, ensuring the long-term viability of forest ecosystems while addressing societal needs.

Forest monitoring technologies have undergone a significant evolution, shifting from labor intensive ground-based surveys to advanced remote sensing (RS) methods. This transformation was driven by the need for faster and more cost-effective approaches to assess forest conditions and manage resources across vast areas. Today, RS data are utilized to address various challenges within the forestry industry. Research indicates that operational forest monitoring can effectively be achieved by high-precision tree photography using specialized photo and video cameras mounted on spacecraft, aircraft (including helicopters), or unmanned aerial vehicles (UAVs). The resulting images can then be interpreted to gather valuable insights about forest health [Lierop et al., 2015; Кривец и др., 2015; Musolin et al., 2022].

However, forestry specialists often encounter a shortage of modern tools — such as models, methods, information systems, and technologies — that facilitate automatic and rapid analysis of these images for recognizing individual trees and assessing their condition. Moreover, remote sensing technologies improve the efficiency and accuracy of forest-related information. They allow for real-time data collection and analysis, enabling better decision-making in forest management. Despite the progress made, there remains a need for continued development of tools that can automate image analysis and enhance the recognition of tree species and health conditions.

When addressing the challenge of multi-classification of trees in images of coniferous forests affected by stem pests, specialists have found that traditional statistical classification methods and conventional machine learning techniques often fall short in terms of accuracy. However, recent studies have demonstrated that deep learning models, particularly convolutional neural networks (CNNs), can achieve significantly higher classification accuracy for this task. These advancements highlight the potential of deep learning to enhance the effectiveness of tree classification in affected forest areas.

This article presents an analytical review of modern models and methods for analyzing images obtained during the monitoring of coniferous forests, particularly in the context of forest pathology. It outlines potential directions for future development in this area. Additionally, the article discusses the results of our advanced CNN models and modified transformer Segformer in analyzing images of Siberian fir Abies sibirica (hereafter referred to as *A. sibirica*) tree crowns that have been affected by the four-eyed fir bark beetle *Polygraphus proximus* (hereafter referred to as *P. proximus*). These CNN models surpass the capabilities of traditional classification methods by enabling pixel-wise classification, which is essential for addressing various challenges in forest pathology monitoring. They facilitate the accurate identification and classification of individual crowns of affected fir trees within complex forest environments. By applying these models to high and ultra-high spatial resolution images captured by UAVs, we aim to establish a more rigorous methodology for assessing the pathological state of coniferous forests, specifically focusing on *A. sibirica* trees impacted by the *P. proximus* pest.

2. Forest pathology monitoring of coniferous forests

The growing threat of foreign insects invading forests poses serious risks to the biological security of many regions worldwide [Chang et al., 2012; Lierop et al., 2015]. Russia is no exception; for instance, the four-eyed bark beetle *P. proximus*, a well-known and harmful pest of fir forests, has been causing significant damage. Since 2007, outbreaks of varying severity have been reported in Siberia and several central regions of Russia, including the Moscow region, and more recently in the fir forests of Udmurtia and the Baikal region [Bystrov, Antonov, 2019; Dedyukhin, Titova, 2021]. Another aggressive pest, the small spruce bark beetle *Ips amitinus*, was discovered in 2019 in Siberian pine forests across multiple regions [Kerchev et al., 2019; Kerchev et al., 2022]. The introduction of this beetle has led to widespread tree mortality in Siberian pine forests near villages, raising concerns about the degradation of such valuable plantations. These examples highlight a troubling trend: experts warn that, without appropriate measures, the mass reproduction of forest pests could result in irreversible economic and environmental damage [Chang et al., 2012; Lierop et al., 2015; Dedyukhin, Titova, 2021; Kerchev et al., 2022]. Globally, outbreaks of forest pests lead to billions of dollars in economic losses each year due to the destruction of commercial timber [Chang et al., 2012; Lierop et al., 2015]. This underscores the urgent need for effective monitoring systems to detect and address breeding grounds for these pests early on.

The primary tasks addressed in the monitoring of coniferous forests for forest pathology are influenced by both the physiological processes occurring in trees of various species and the specific characteristics of the ecosystems containing these monitored forest areas. One significant aspect of insect pest infestations in coniferous forests is the sharp fluctuations in pest populations, which arise from the time it takes for a balance to be established between the invasive species and the native community. Following a latent phase during which the pest acclimatizes to its new environment and maintains low population levels, there is often an outbreak, followed by a rapid decline in numbers, which may subsequently be followed by another surge [Кривец и др., 2015; Dedyukhin, Titova, 2021; Kerchev et al., 2022]. Given the high variability in pest populations, two critical tasks emerge: the rapid identification of early-stage pest outbreaks and the monitoring of the health of infested trees. Additionally, a third important task in monitoring coniferous forests involves identifying dead or dying trees. Such trees may result from disease or climate change effects. For instance, dieback in spruce forests can occur due to sudden temperature changes and insufficient moisture during certain times of the year. Monitoring these forests to detect dead trees is essential for assessing biomass and carbon reserves within these areas. Ultimately, this information can help evaluate the contribution of carbon emissions from these areas to regional atmospheric carbon budgets and their negative environmental impacts (the contribution of dead trees to the region's "carbon problem").

The challenges associated with detecting pest outbreaks and obtaining reliable assessments of tree health and dead tree presence are compounded by the extensive areas that require monitoring and the diversity of tree species and their respective pathological conditions. This highlights the relevance of developing modern methodologies and tools for effective monitoring of pest-affected forests. Recent research has yielded promising initial results in addressing this issue, which will be discussed in further detail.

Monitoring forests using RS systems has become increasingly common worldwide, with the data obtained from such monitoring being utilized for various forestry management activities, including forest inventory and management operations. Several researchers have demonstrated that the challenge of operational forest monitoring can be effectively addressed through high resolution imaging of trees from spacecraft, airplanes (including helicopters), or UAVs, followed by expert interpretation of the resulting images [Lierop et al., 2015; Кривец и др., 2015; Musolin et al., 2022; Kerchev et al., 2022].

For instance, studies have shown that high-resolution (0.1–1.0 m/pixel) and ultra-high resolution (0.02–0.1 m/pixel) images obtained through RS can reveal even relatively minor damage to the crowns of pine, spruce, and fir trees caused by insect pests [Chenari et al., 2017; Кривец и др., 2015; Safonova et al., 2019; Керчев и др., 2021; Zhou et al., 2022a]. Furthermore, remote observations conducted via spacecraft and UAVs are significantly more cost-effective than traditional monitoring methods using airplanes and helicopters, making UAVs the preferred choice for many applications today.

The advantage of UAVs over spacecraft in forest monitoring lies in their ability to capture images quickly while being less affected by atmospheric conditions such as transparency and cloud cover. Moreover, modern digital cameras mounted on UAVs can produce ultra-high resolution images that allow specialists to analyze tree features at the branch level and, in some cases, at the leaf level. These images enable the identification of not only spectral characteristics of trees but also spatial attributes (such as textures and crown geometry), which are crucial for recognizing tree species and assessing the extent of pest damage.

These findings support the establishment of a foundational methodology for operational forest monitoring [Chenari et al., 2017; Safonova et al., 2019; Керчев и др., 2021; Zhou et al., 2022a]. Central to this methodology is a set of requirements and recommendations for using various classes of spacecraft and UAVs equipped with specialized equipment to obtain RGB and multispectral images of forests at high and ultra-high spatial resolutions. Additionally, since the early 2000s, compact LiDAR scanners have been increasingly installed on aircraft and more recently on UAVs for forest monitoring [Lierop et al., 2015; Gini et al., 2018; Musolin et al., 2022]. LiDAR technologies are now frequently replacing traditional forest inventory methods due to their ability to accurately determine tree height, crown dimensions, vertical and horizontal spatial structure of the crown, and tree trunk diameter.

The timely acquisition of forest inventory metrics such as tree height and trunk diameter enables effective management decisions regarding various forestry tasks (e. g., creating harvest plans). Initial research has also begun to integrate multispectral imagery with LiDAR data to address challenges related to detecting pest outbreaks and assessing the health of affected trees, indicating promising potential for this combined approach [Chenari et al., 2017; Lee et al., 2019; Onishi, Ise, 2018].

The analysis (interpretation) of images obtained during forest pathology monitoring is conducted to address each of the three previously described tasks, utilizing either a semi-manual approach (where an expert interprets the images using auxiliary software) or an automated method through specialized software or hardware-based classification algorithms. It is noteworthy that more advanced automatic image classification tackles the challenge of multi-classification, where each tree in the image must be assigned to one of several health classes based on various characteristics. This multi-classification task is particularly critical when assessing the health of trees affected by pests, as it requires the identification of multiple classes (conditions) of these trees within the images. For instance, in cases where Siberian fir (*A. sibirica*) trees are infested by the four-eyed bark beetle, it is necessary to recognize five classes: four representing different levels of damage to the fir trees and one representing the background (including other tree species and various surface objects).

By identifying multiple classes of damaged trees within the images, the analysis can yield more detailed insights into the extent and severity of infestations, thereby aiding in the formulation of management strategies aimed at mitigating damage caused by forest pests. It is evident that the results from multi-classification are also employed in addressing the other two tasks associated with forest pathology monitoring in coniferous forests.

The following sections discuss the main models and methods that facilitate the multiclass classification of trees in images from coniferous forests.

3. Models and methods for tree classification in images of coniferous forests

In recent years, the automated classification of trees in remote sensing imagery has received significant attention. Traditional classifiers, such as parametric and nonparametric statistical methods, as well as machine learning techniques, have been widely employed for tree classification on aerial and satellite images [Gini et al., 2018; Lee et al., 2019]. However, deep learning algorithms have proven to be accurate and effective methods for solving tree classification tasks in UAV imagery.

To the best of our knowledge, Onishi and Ise in [Onishi, Ise, 2018] pioneered the application of deep learning methods and UAV images in forestry, successfully addressing tree species classification. Safonova et al. in [Safonova et al., 2019] proposed a method for detecting and classifying *A. sibirica* trees infected by bark beetles in UAV images using CNNs. The authors employed a two-stage methodology: initially identifying candidate regions containing tree crowns in the image, followed by classifying the identified crowns using their CNN model. The results demonstrated the superior accuracy of their model in classifying damaged trees compared to well-known models such as Xception, VGG-16, VGG-19, ResNet-50, Inception–V3, Inception ResNet–V2, DenseNet-121, DenseNet-169, and DenseNet-201. However, the intricacy associated with implementing this two-stage approach, along with the limitation that the classifier can only handle images featuring a single tree at a time, may limit its practical applicability.

The You Only Look Once (YOLO) CNN architecture has gained relevance in tree identification tasks due to its good ability to simultaneously localize and classify multiple objects within an image. This capability has enabled researchers to effectively apply YOLO-based models for diverse tree analysis applications. Jintasuttisak et al. in [Jintasuttisak et al., 2022] employed the YOLOv5 model to detect damage to palm trees in UAV RGB-images. Zhou et al. in [Zhou et al., 2022b] further expanded the application of YOLO by analyzing multispectral UAV-images of damaged pine trees. Similarly, Yu et al. in [Yu et al., 2021] explored the detection of pine wilt disease using UAV-based multispectral images. The pine trees were categorized into four classes: early-infected trees, middle-infected trees, late-infected trees, and broad-leaved trees. Image analysis was performed using Faster R-CNN and YOLOv4 models, along with two traditional machine learning models: Support Vector Machine (SVM) and Random Forest (RF). In terms of classification accuracy, Faster R-CNN, SVM, and RF models outperformed YOLOv4, demonstrating superior performance in recognizing infected pine trees.

The study presented by Xie et al. in [Xie et al., 2024] addresses the early detection of the pine tree wilt disease caused by the pine wood nematode, utilizing images captured by UAVs. Various models from the YOLO family were employed, including YOLOv3, YOLOv4, YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. Despite achieving the best classification accuracy with the YOLOv5m model, the results were still deemed insufficiently acceptable by industry specialists. Consequently, the authors proposed a method that combines the original image with its representation in the frequency domain and modified the YOLOv5 models accordingly. Their additional research yielded improved classification accuracy for early-stage wilting trees; however, these results still did not meet the expectations of forestry professionals. This work highlights the ongoing challenges in accurately detecting early signs of tree wilt disease using UAV imagery and underscores the need for continual refinement of deep learning models to enhance their performance in practical applications within forestry.

The findings from studies [Yu et al., 2021; Jintasuttisak et al., 2022; Zhou et al., 2022b; Xie et al., 2024] suggest that YOLO-based architectures exhibit relatively lower classification accuracy compared to other CNN architectures in coniferous tree classification tasks. This performance gap can be attributed to YOLO's single-pass approach, which prioritizes speed over accuracy. As a result, YOLO models may not be suitable for applications that require high classification accuracy.

Wu and Jiang in [Wu, Jiang, 2023] highlighted the effectiveness of the Mask R-CNN model in detecting and extracting regions with pine wilt disease using UAV-based RGB images. This enhanced version of Faster R-CNN employs an intricate approach that employs a region proposal network in the first stage to generate object proposals. These proposals are then refined and classified in the second stage. This study, along with the findings in [Yu et al., 2021], illustrates that both Mask R-CNN and Faster R-CNN models, demonstrate commendable accuracy in classifying trees according to their health status. However, their computational efficiency remains a significant challenge, particularly for real-time applications.

Among CNNs, U-Net [Ronneberger et al., 2015] and U-Net-like architectures have demonstrated remarkable performance and computational efficiency in semantic image segmentation. In [Kocon et al., 2022] the authors carried out an evaluation of various CNN models for forest type classification using Sentinel-2 data. The U-Net model outperformed the SegNet, PSPNet, and FCN-8 models in terms of categorical accuracy. Similarly, Korznikov et al. in [Korznikov et al., 2021] demonstrated the superiority of the U-Net architecture in the binary classification of trees as either coniferous or deciduous in RGB satellite images. The U-Net model outperformed machine learning algorithms such as RF, K-Nearest Neighbor (K-NN) classifier, Naive Bayes classifier, and quadratic discrimination, delivering better results. The U-Net model exhibited promising outcomes in the multiclass classification of *P. proximus*-infected *A. sibirica* trees as suggested by [Керчев и др., 2021]. The mean intersection over union (mIoU) metric achieved a satisfactory value of 0.66 (values of this metric exceeding 0.5 correspond to high pixel-wise classification accuracy). Similarly, Markov et al. in [Марков и др., 2022] proposed employing the U-Net model for semantic segmentation of *I. amitinus*-infected P. sibirica trees in UAV images. The model demonstrated high segmentation quality achieving an mIoU of 0.61. However, despite achieving high classification accuracy for most classes of pest-affected coniferous trees, there is a noticeable drop in accuracy for one (in the case of *A. sibirica* trees) or two (in the case of *P. sibirica* trees) intermediate classes of trees. Recognizing trees in these intermediate states, which lie between healthy trees and old dead wood, is crucial for effective forest conservation, especially during the early stages of pest damage. Timely identification allows for prompt implementation of phytosanitary measures, contributing significantly to forest preservation. The low accuracy of identifying trees in intermediate states fails to meet the standards expected by forest industry specialists. Comparing classification results for affected *A. sibirica* trees, as reported by [Керчев и др., 2021] using the classical U-Net model on the dataset used in [Марков, Маслов, 2021], reveals significantly better performance than when employing the Gradient Boosting algorithm (with an mIoU metric value of 0.66 compared to 0.49 for the Gradient Boosting algorithm).

Analysis of modern models and methods for classifying coniferous trees in images from satellites and UAVs leads to several key conclusions. First, there is a growing scientific field focused on assessing the conditions of trees damaged by pests in such imagery, primarily through the application of various CNN models. However, the number of studies on coniferous tree classification remains limited, and comparing their classification accuracy becomes challenging due to variations in approaches, metrics and datasets. Second, it is evident, as supported by [Onishi, Ise, 2018; Safonova et al., 2019; Марков, Маслов, 2021; Yu et al., 2021; Korznikov et al., 2021], that deep learning models and methods outperform traditional classification approaches, including classical machine learning algorithms, particularly in addressing complex multiclass classification tasks. Third, within the specified scientific domain, there is potential for developing new U-Net like models to enable accurate multiclass classification of affected coniferous trees, even those in intermediate states of damage.

The goal of this research is to find the best CNN model (from a suite of models under development) that balances accurate classification of tree health with rapid processing speed. Specifically, the model must achieve an Intersection over Union (IoU) score above 0.5 for each health status class when classifying *A. sibirica* trees affected by bark beetles. Furthermore, the model needs to analyze tree images in quasi-real-time, which is essential for large-scale forest pathology monitoring of coniferous forests. For practical deployment in the forestry sector, using standard personal computers, quasi-real-time analysis means processing high- or ultra-high-resolution images covering 1 hectare (10 000 m^2) of affected coniferous trees within 12–15 seconds.

Within the framework of this scientific field, we develop and study new CNN models based on the classical U-Net architecture [Ronneberger et al., 2015], and a model based on the transformers architecture [Xie et al., 2021]. Below we present the results of the study of such models in solving relevant problems of multiclass classification of *A. sibirica* coniferous trees affected by pests.

4. Materials and methods

4.1. Study area and data collection

The study area for this research included locations within the Tomsk Region of Russia (Fig. 1). The study covered stands of *A. sibirica* trees, primarily located near the village of Parbig in the Bakchar district.

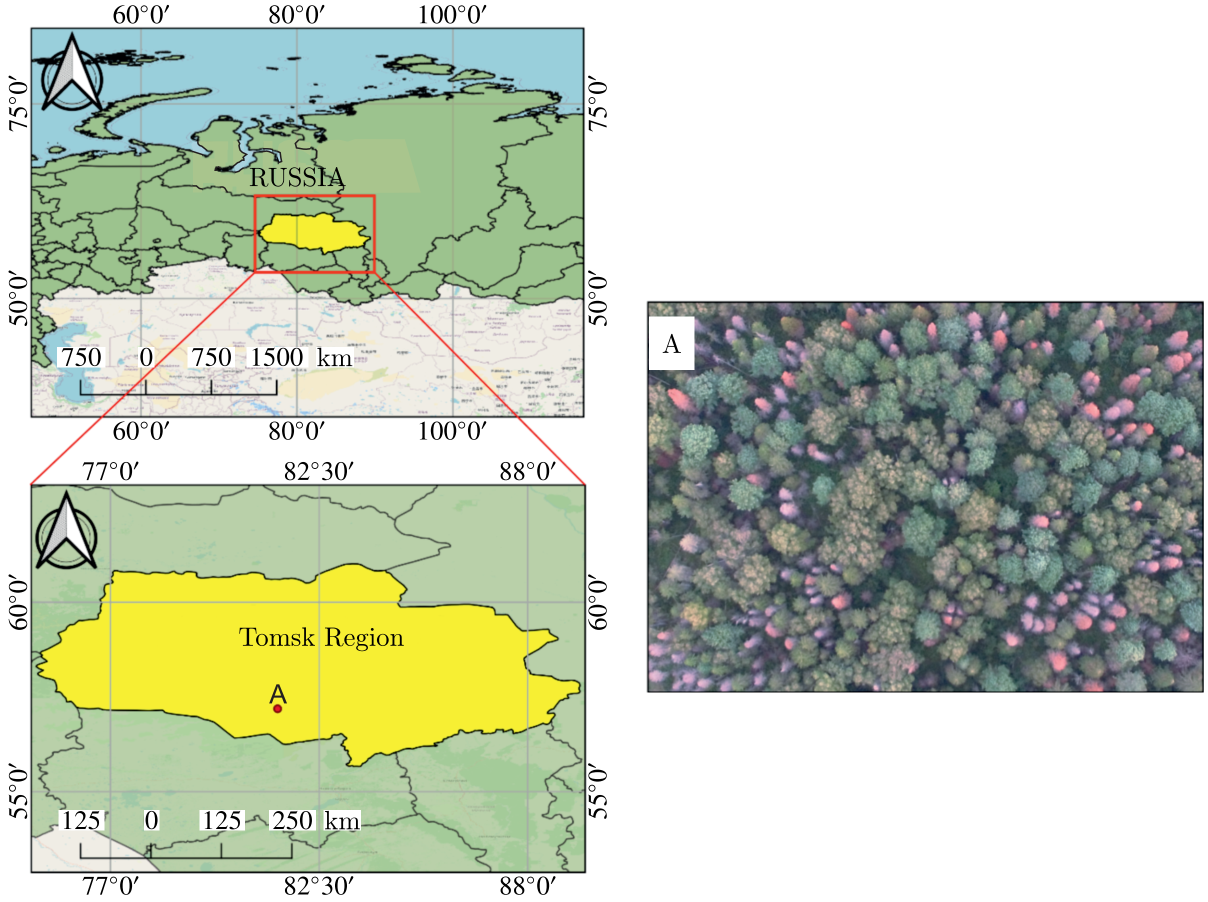


Figure 1. Location of the stud area, where A is a UAV-image fragment of *A. sibirica* stands

To assess the condition of *A. sibirica* trees affected by *P. proximus*, we collected five aerial panoramas, formed from images with a spatial resolution of about 0.1 m. These images were captured using a DJI Phantom 3 Standard UAV with a camera mounted on it, shooting in the visible light waves of the electromagnetic spectrum (RGB) from altitudes ranging between 365 and 388 m during the period from August 7 to August 28, 2017. The aerial survey was complemented by ground surveys, confirming the infestation of insect pests. The resulting panoramas had the following dimensions (in pixels):

- Panorama A: 1046 × 1912 × 3,
- Panorama B: 1536 × 1048 × 3,

- Panorama C: 1536 × 768 × 3,
- Panorama D: 768 × 1792 × 3,
- Panorama E: 1046 × 1912 × 3.

Panorama E served as testing data for our models.

The Institute for Monitoring Climatic and Ecological Systems of the Siberian Branch of the Russian Academy of Sciences devised an assessment scale based on the *A. sibirica* tree health condition [Musolin et al., 2022]. Four health categories of *A. sibirica* trees based on the level of *P. proximus* invasion into the trunk and its impact on the crown are recognized as seen in Fig. 2:

- Class 1 – Healthy (Fig. 2, *a*). Healthy trees, green needles.
- Class 2 – Dying (Fig. 2, *b*). The crown may look similar to a healthy tree, but more than half of the branches have drying needles. The needles in the upper part of the crown are still green, and the needles lower on the branches are reddish.
- Class 3 – Fresh dead wood (Fig. 2, *c*). The needles of the crown are dead, red, and partially fallen. The density of the crown is only a fraction of the crown of a healthy tree, withered branches throughout the crown. The tree died less than a year ago.
- Class 4 – Old dead wood (Fig. 2, *d*). the crown is gray and dead. The needles have completely crumbled. The tree died several years ago.

We included a fifth class “Background”, which includes trees of other species and objects present on the Earth’s surface.

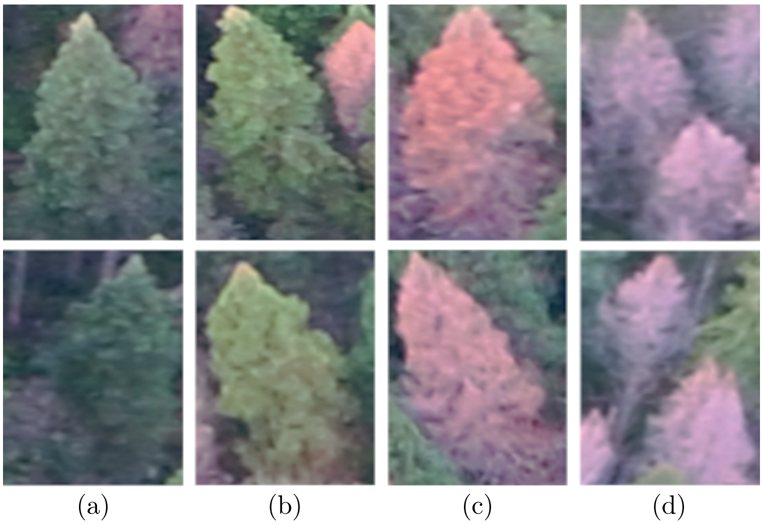
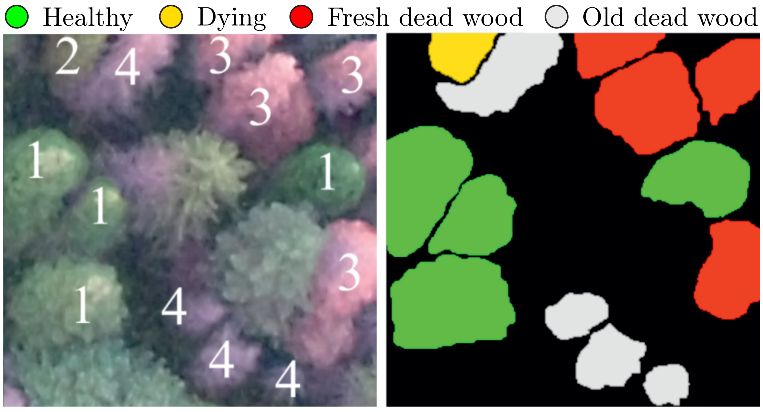


Figure 2. Examples of *A. sibirica* trees: (a) Healthy; (b) Dying; (c) Fresh dead wood; (d) Old dead wood

4.2. Datasets

The open-source image manipulation program, GIMP, was used to annotate the trees in the images based on the previously defined classes. A domain expert assigned numerical labels to the trees in the UAV images, indicating the health status of each *A. sibirica* tree. Then, another expert used the assigned numerical labels (Fig. 3, *a*) as a reference to accurately delineate the trees using colors representing the respective classes (Fig. 3, *b*). This process resulted in the generation of segmentation masks, which accurately depict *A. sibirica* trees belonging to four distinct classes and “Background”.

Four panoramas were selected for training and validation, and one for testing. To build the training, validation and test sets, we split the images and reference segmentation masks into fragments of size 256 × 256 × 3 pixels with an overlap of 128 pixels. To prevent data leakage, we applied Boolean masks to the original images. This approach prevented duplication of features between training and validation fragments. In order to augment the training and validation samples the images were rotated at different angles. In total, the *A. sibirica* dataset consisted of 2004 training fragments, 672 validation fragments, and 96 test fragments.



(a) (b)

Figure 3. Panorama fragment of *A. sibirica* trees labeled by classes (a) and its corresponding segmentation mask (b)

Similarly, a second dataset was created based on these expert-interpreted panoramas and the respective segmentation masks. The fragments and masks in this dataset were of size 480×480×3 pixels with an overlap of 240 pixels. In total, the sets consisted of 502 training fragments, 180 validation fragments, and 24 test fragments.

In order to gain a deeper understanding of the original monitoring data, the number of trees present in the training, validation, and test sets before data augmentation was counted. A significant class imbalance is observed, as shown in Table 1. For example, the “Dying” class is represented by only 80 trees, while the “Healthy” class contains 574 trees. This imbalance may negatively affect the performance of neural network models, resulting in low recognition accuracy of trees in minority classes such as “Dying” and “Fresh dead wood”. Indeed, when training network models, they may not be able to adequately extract the distinctive features of underrepresented trees of these classes, resulting in low recall and accuracy of tree classification when subsequently used in test samples. To mitigate the negative impact of tree class imbalance, it was necessary to increase the size of training fragments in each sample. For this purpose, various data augmentation methods were used: online augmentation was used during model training, including changes in image scale, brightness, contrast, vertical axis reflections, and elastic transformations.

Table 1. Number of *A. sibirica* trees by class

Dataset	Healthy	Dying	Fresh dead wood	Old dead wood	Total
Training	319	44	147	290	800
Validation	107	14	63	110	294
Testing	148	22	64	91	325
Total	574	80	274	491	1419

4.3. Proposed neural network models

To solve the problem of improving accuracy in the multiclass classification of pest-affected *A. sibirica* trees, three modifications of the classic U-Net model were proposed. In addition, a model based on the transformer's architecture [Xie et al., 2021] was used. Below we delve into their architecture and characteristics.

4.3.1. Mo-U-Net model

The developed modified U-Net (Mo-U-Net) model is based on one of the most well-known architectures — the classic U-Net. This architecture was originally designed for biomedical image segmentation tasks, where it demonstrated high quality results. A distinctive feature of the U-Net model is the presence of skip connections, which connect sets of feature maps from the encoder to sets of feature maps from the decoder in order to increase the detail of the resulting segmentation map [Ronneberger et al., 2015]. In our work [Markov, Machuca, 2024] a detailed explanation and description of the architecture of the proposed Mo-U-Net can be found.

The customized architecture incorporates several modifications to the original classic U-Net model, including:

1) The input image is represented by a $256 \times 256 \times 3$ tensor (or $480 \times 480 \times 3$), corresponding to an RGB fragment;

2) Convolutions do not reduce the feature maps size;

3) Cropped feature maps are not used for skip connections;

4) Batch normalization is applied after each nonlinear function.

The rectified linear unit (ReLU) is replaced by the ELU activation function [Clevert et al., 2015].

The application of the ELU activation function helps the model to learn more complex features, allowing negative values to improve the convergence during training. The output tensor is calculated by applying C convolutions with filters size 1×1, allowing one to directly classify pixels as one of the C classes (four classes of *A. sibirica* health status and "Background").

4.3.2. Res-Mo-U-Net model

The proposed hybrid fully convolutional network model, named residual-modified U-Net (Res-Mo-U-Net), is based on the Mo-U-Net architecture and incorporates several modifications. Specifically, the Dropout procedure has been replaced with a SpatialDropout procedure, and residual blocks have been added to enhance the model's performance [He et al., 2016]. These residual blocks are integrated into the Mo-U-Net architecture as part of the convolutional blocks in both the encoder and decoder. In each convolutional block of the encoder, standard convolutional layers have been supplemented with residual blocks, allowing for improved feature learning and better gradient flow during training. This modification allows for a deeper architecture without the risk of gradient vanishing. Similarly, in the decoder, residual blocks were used to process feature maps before upsampling, which helps preserve important details of tree crowns. Each residual block consists of two convolutional layers with a kernel size of 3×3. The output of these two convolutional layers is added to the original input of the block (or its projection if the dimensions do not match). This design enables the model to retain information about low-level features of crowns as they pass through the neural network, which is critically important for solving semantic segmentation tasks. The architecture of the Res-Mo-U-Net model is illustrated in Fig. 4 for an input fragment of an RGB image sized $H \times W \times 3$ pixels, where H is the height of the fragment and W is the width. C represents the number of classes for fir trees and "Background". It is important to note that the projection operations (Conv1x1) in the main branches of the residual blocks in this model do not include nonlinearities and biases. Therefore, it can be expected that such residual blocks may lead to a more stable training process for the model and simplify gradient backpropagation. Ultimately, this should contribute to better detection of subtle features of tree crowns when addressing the complex task of multiclass classification of infested *A. Sibirica* trees.

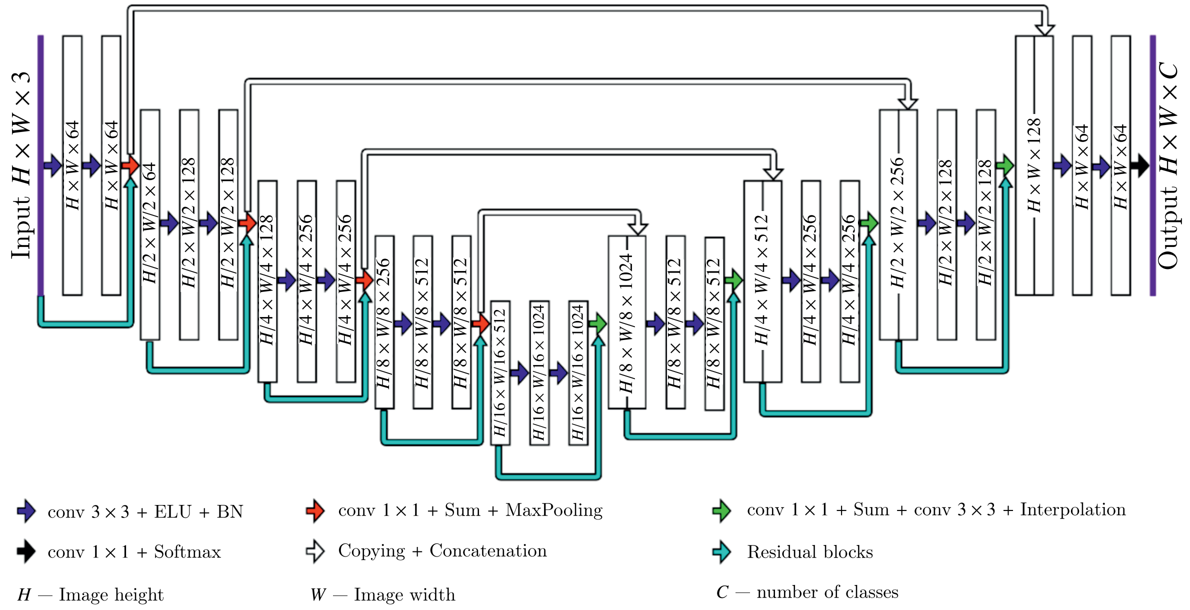


Figure 4. Res-Mo-U-Net architecture

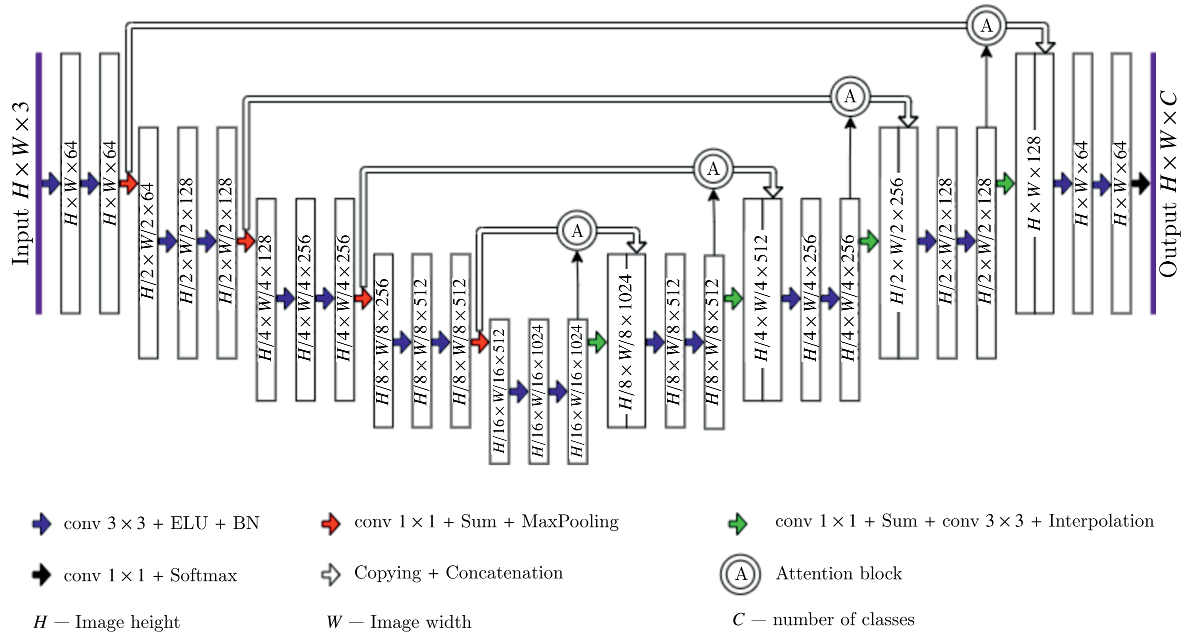


Figure 5. At-Mo-U-Net architecture

4.3.3. At-Mo-U-Net model

This proposed model is also based on the Mo-U-Net architecture, incorporating an attention mechanism [Vaswani et al., 2017]. The model is named At-Mo-U-Net. The architecture includes an encoder and a decoder (see Fig. 5), and features attention blocks. The encoder consists of a series of convolutional blocks followed by max-pooling layers. Each block increases the number of filters while reducing spatial dimensions, allowing the model to learn hierarchical features from the input image fragment. In turn, each convolutional block includes two consecutive convolutional layers with ELU activation and batch normalization, as well as Dropout to enhance the model's generalization ability. The decoder mirrors the structure of the encoder and includes upsampling and concatenation layers. The upsampling layers restore the spatial resolution of feature maps, achieved through nearest neighbor interpolation followed by a convolutional layer. Skip connections concatenate feature maps from the downsampling path with corresponding maps from the upsampling path passing through the attention block.

In the At-Mo-U-Net model, the attention mechanism enhances the neural network's ability to focus on the most significant features by dynamically determining the importance of various spatial positions on the feature maps. The attention mechanism is implemented using three layers: weight calculation, multiplication, and feature merging. The input to the weight calculation layer consists of feature maps obtained from the encoder and previous layers. This layer computes weight coefficients for each spatial position on the feature map. A sigmoid activation function is used for this purpose, which assigns values in the range [0, 1], reflecting the relative importance of each position. The output of this layer is a weight matrix, where each value corresponds to the level of attention for the respective position on the feature map. The multiplication layer receives both the original feature maps and the computed weight matrix as input, performing element-wise multiplication of the feature maps by the weight matrix. This operation amplifies more important features (with high weights) and suppresses less significant ones (with low weights). The output of this layer consists of refined feature maps where key features are emphasized. The merging layer receives the refined feature maps after applying the attention mechanism, along with additional feature maps from the decoder. It then merges (through concatenation) the feature maps that have passed through the attention mechanism with other feature maps in the decoder. This combination helps retain important details from previous stages and contributes them to the final segmentation result. Ultimately, a final feature map is produced, which is used to construct the segmentation mask. While attention mechanisms have been widely studied and implemented in various deep learning models, the specific combination of these techniques within the Mo-U-Net framework for tree crown segmentation represents a new approach. Previous models have utilized similar architectures, but the integration of attention in this specific context enhances its effectiveness for environmental and forestry applications

4.3.4. Modified Segformer model

In order to compare the results of studies using the proposed fully convolutional neural network models, it is worth to conducting similar studies and obtain results using modern transformer architectures, in this case the Segformer model. This architecture combines the advantages of transformer-based models in the encoder with lightweight multi-layer perceptrons (MLP) in the decoder. It is known to have been successfully applied to several practical tasks of semantic image segmentation [Xie et al., 2021]. The hierarchical architecture of the Transformer in the encoder of the Segformer model, known as Mix Transformer (MiT), allows the encoder to generate features at multiple scales, which is critically important for accurate segmentation of objects of various sizes. Unlike traditional Transformer architectures, the Segformer model does not require positional encoding. This simplifies its architecture and prevents performance degradation that may occur due to mismatches between training and test sample resolutions. Finally, layer aggregation is used, which facilitates the aggregation of information from different encoder layers and ultimately allows the model to effectively utilize both local and global attention. Second, structurally, the encoder consists of several blocks, each including the following elements:

- Overlap Patch Embedding Layer: It divides the image into overlapping patches, improving interaction between them. The patch size is larger than the stride, allowing information exchange between patches.
- Attention Layers: These process input data using attention mechanisms to highlight significant features. Each block includes Efficient Self-Attention and a Mix-Feedforward Network.
- Overlap Patch Merging Procedure: This implements the operation of merging patches back into feature maps. This process also includes layer normalization to improve model stability during training.

Third, the lightweight MLP decoder in the Segformer model is a key component of the architecture responsible for transforming multi-scale features obtained from the hierarchical transformer encoder into the final segmentation mask. This decoder is designed with an emphasis on simplicity and efficiency, allowing it to achieve high performance without using complex modules. Several changes were introduced to the model with the goal to improve the segmentation quality, such as:

- Instead of using standard MLPs, we incorporated spatial gating units (SGUs) to improve spatial feature interactions [Liu et al., 2021].
- Instead of relying on Efficient Self-Attention, we introduced an additional cross-level attention mechanism to facilitate information exchange between different scales [Han et al., 2021].
- Instead of simple upsampling, we used feature pyramids with attention layers to improve segmentation of fine details [Li et al., 2018].

4.4. Training and evaluation metrics

The training phase revolved around the semantic segmentation of *A. sibirica* trees in UAV images. Each of the models underwent 100 training cycles using the Focal Loss function, each fine-tuned with a distinct set of hyperparameters through Bayesian optimization. To avoid overfitting, an early stopping mechanism was integrated into the training process. Specifically, if the model failed to demonstrate improvement after 11 epochs on the validation set, the training was stopped, ensuring the model's adaptability to unseen data in practical scenarios. Adam, a widely adopted optimization algorithm, was employed to fine-tune the model weights during training. The hyperparameters included batch size, learning rate, and a suite of image transformations. These transformations included rotation angle, saturation, contrast, brightness, and sharpness, contributing to the model's ability to robustly handle diverse scenarios and variations present in the UAV images. This comprehensive training process was repeated for each training set.

Post-training, the models underwent an evaluation on the validation sets, aiming to identify the optimal models for semantic segmentation of *A. sibirica* trees. To assess the performance (semantic segmentation quality) of the proposed models for classifying damaged trees in UAV images, we adopt the Intersection over Union (IoU) metric. It is considered one of the most used and accepted evaluation metrics for image segmentation [Rahman, Wang, 2016; Bertels et al., 2019]. The IoU of each class *c* can be calculated as follows:

$$IoUc = \frac{TP}{TPc + FPc + FNc}. \tag{1}$$

In (1) TPc, FPc, and FNc denote the number of True Positives, False Positives, and False Negatives for class *c*, respectively. The mIoU metric was also used. It is calculated as the average of IoU values for all classes. IoUc and mIoU values exceeding 0.5 denote high segmentation quality, suggesting that CNN models and Modified Segformer models capable of achieving such accuracy have practical applications in the forestry industry.

The proposed models were implemented in Python 3 using the PyTorch framework.

5. Results and discussion

The training and validation of the proposed CNN models and the Modified Segformer model were performed in accordance with Section 4.4 twice: the first time using the training and validation sets from the first dataset (with fragments of size 256 × 256 × 3), and the second time using the corresponding sets from the second dataset (with fragments of size 480 × 480 × 3). Then, the models trained in this way were examined using test sets created based on a test panorama of *A. sibirica* trees affected by the four-eyed bark beetle.

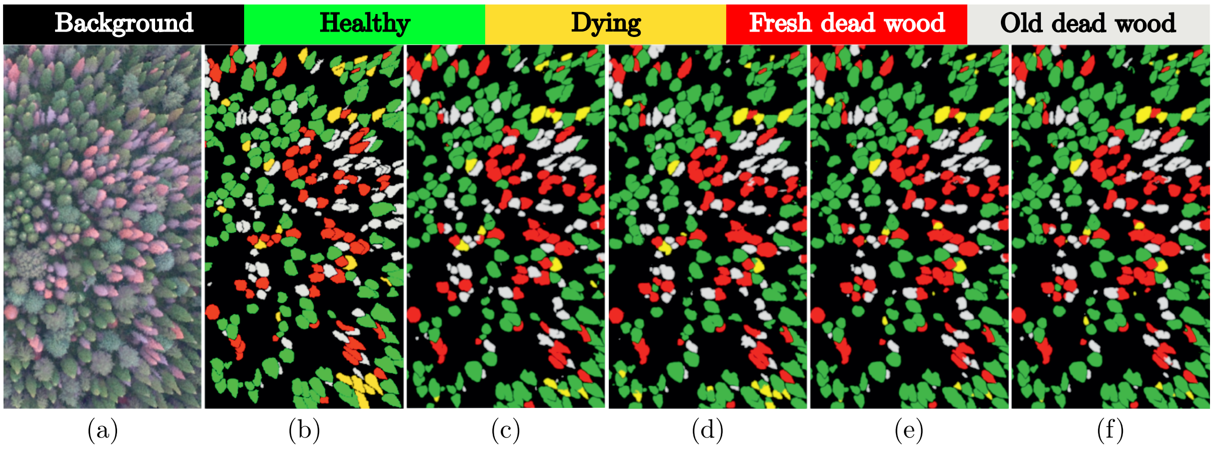


Figure 6. Semantic segmentation results for the *A. sibirica* test area using fragments of size 256 × 256 × 3 pixels: (a) Test area; (b) Ground truth; (c) Modified Segformer; (d) Mo-U-Net; (e) At-Mo-U-Net; (f) Res-Mo-U-Net

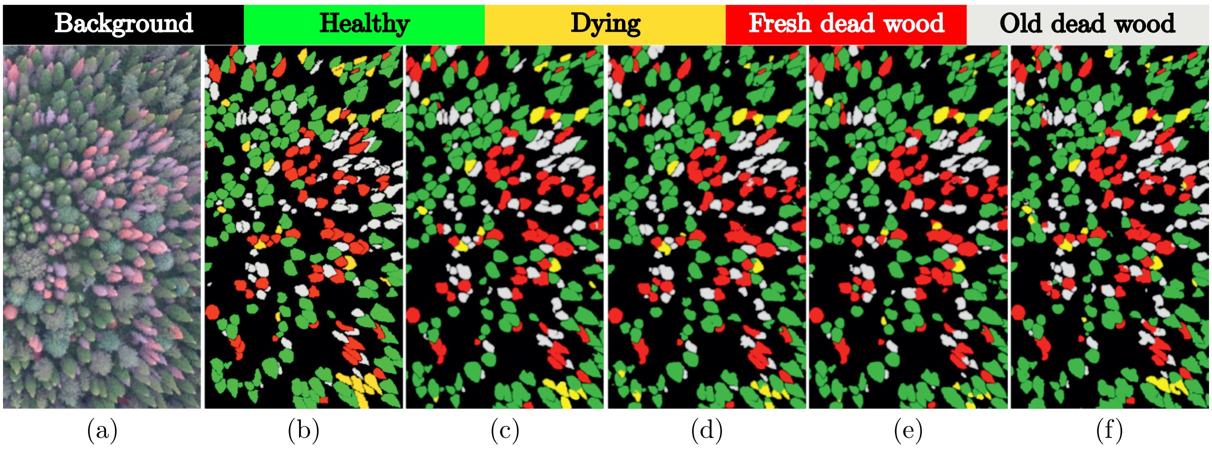


Figure 7. Semantic segmentation results for the *A. sibirica* test area using fragments of size 480 × 480 × 3 pixels: (a) Test area; (b) Ground truth; (c) Modified Segformer; (d) Mo-U-Net; (e) At-Mo-U-Net; (f) Res-Mo-U-Net

Figure 6 shows the results of the models in solving the problem of multiclass classification of infested fir trees on the test set (with fragments of 256 × 256 × 3 pixels). Upon visual analysis of the image of the test area (Fig. 6, *a*), its reference segmentation mask (Fig. 6, *b*) obtained during interpretation by experts, and the resulting output segmentation maps when using each of the models (Fig. 6, *c–f*), it is noticeable that the models are able to reproduce the borders of fir tree crowns and correctly classify a significant proportion of tree crowns.

Similar results of studying the trained models were obtained in the case of solving the same multiclass classification problem using the test set with fragments of 480 × 480 × 3 pixels (see Fig. 7).

Table 2 shows the classification accuracy of the proposed models based on the IoUc metric of each class of *A. sibirica* trees and the mIoU metric on the test set with fragments of 256 × 256 × 3 pixels. Table 3 shows the classification accuracy of the proposed models based on the IoUc metric of each class of *A. sibirica* trees and the mIoU metric on the test set with fragments of 480 × 480 × 3 pixels.

Table 2. Semantic segmentation quality of the models trained and tested using *A. sibirica* tree image fragments of size 256 × 256 × 3 pixels

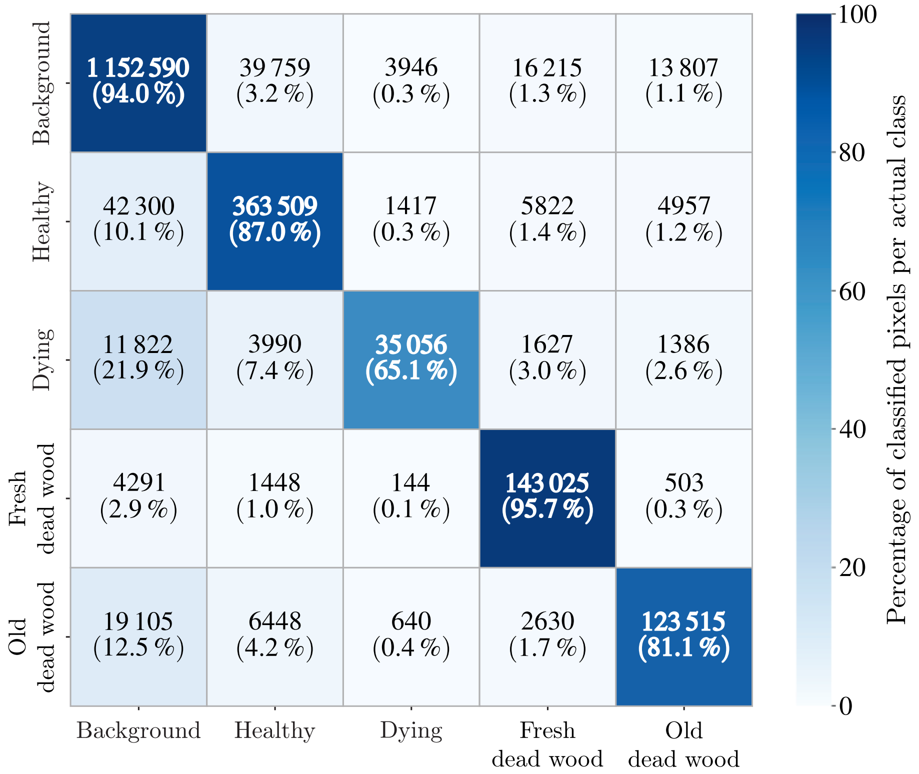
Model	IoUc					mIoU
	Background	Healthy	Dying	Fresh dead wood	Old dead wood	
Modified Segformer	0.87	0.76	0.56	0.80	0.70	0.74
U-Net	0.87	0.73	0.45	0.79	0.68	0.70
Mo-U-Net	0.86	0.74	0.51	0.75	0.65	0.71
At-Mo-U-Net	0.86	0.74	0.51	0.79	0.68	0.72
Res-Mo-U-Net	0.85	0.72	0.51	0.75	0.66	0.70

Table 3. Semantic segmentation quality of the models trained and tested using *A. sibirica* trees image fragments of size 480 × 480 × 3 pixels

Model	IoUc					mIoU
	Background	Healthy	Dying	Fresh dead wood	Old dead wood	
Modified Segformer	0.88	0.77	0.58	0.81	0.71	0.75
U-Net	0.87	0.73	0.46	0.79	0.69	0.71
Mo-U-Net	0.86	0.75	0.52	0.78	0.70	0.72
At-Mo-U-Net	0.86	0.75	0.52	0.80	0.70	0.73
Res-Mo-U-Net	0.86	0.76	0.57	0.80	0.70	0.74

It should be noted that Tables 2 and 3 also include the results of our studies of the classical U-Net model described in [Ronneberger et al., 2015] obtained on the same datasets. The analysis of the results presented in Table 2 indicates that the Modified Segformer model achieves the highest classification accuracy for fir trees, as measured by the IoUc and mIoU metrics, significantly surpassing the threshold value of 0.5 across all tree classes. This performance can be attributed to the hierarchical Transformer architecture in its encoder, known as MiT. This architecture enables the encoder to generate features at multiple scales, which is crucial for accurately segmenting tree crowns of varying sizes. In terms of mIoU, the second place is held by the At-Mo-U-Net model, followed closely by the Mo-U-Net model in third place. For the intermediate class of "Dying" trees, the Mo-U-Net, At-Mo-U-Net, and Res-Mo-U-Net models yield similar results, each slightly exceeding the 0.5 threshold according to the IoUc metric. In contrast, the U-Net model achieves an IoUc value of 0.45, which is significantly below the threshold and indicates a considerable lag in classification accuracy compared to these three models. We believe that the use of the ELU activation function in these models — compared to the ReLU activation function used in the U-Net model — enables them to extract more complex features from the crowns of infested trees. It is worth noting that the IoUc result for the U-Net model slightly exceeds the classification accuracy for intermediate "Dying" fir trees reported in [Керчев и др., 2021] for the same U-Net model. For other tree classes and background areas, all fully convolutional models achieve IoUc metric values above the threshold, with only minor differences among them.

The analysis of the results presented in Table 3 reveals that all IoUc and mIoU metric values have increased for all models and most tree classes compared to the results in Table 2, with only a few instances where they remained unchanged. This improvement can be attributed to the fragments from the second dataset, which are sized 480 × 480 × 3 pixels and contain a greater number of trees with complete (not cropped at the edges) outlines than the fragments from the first dataset, which are sized 256 × 256 × 3 pixels. This increase in well-defined tree contours contributes to enhanced classification accuracy. The Modified Segformer model continues to lead among all models; however, the Res-Mo-U-Net model has made significant improvement, particularly for the “Dying” tree class, coming close to matching the Modified Segformer accuracy. The At-Mo-U-Net model ranks third in terms of the mIoU metric, having previously held second place in the first test set.



	Background	Healthy	Dying	Fresh dead wood	Old dead wood
Background	**1 152 590 (94.0 %)**	39 759 (3.2 %)	3946 (0.3 %)	16 215 (1.3 %)	13 807 (1.1 %)
Healthy	42 300 (10.1 %)	**363 509 (87.0 %)**	1417 (0.3 %)	5822 (1.4 %)	4957 (1.2 %)
Dying	11 822 (21.9 %)	3990 (7.4 %)	**35 056 (65.1 %)**	1627 (3.0 %)	1386 (2.6 %)
Fresh dead wood	4291 (2.9 %)	1448 (1.0 %)	144 (0.1 %)	**143 025 (95.7 %)**	503 (0.3 %)
Old dead wood	19 105 (12.5 %)	6448 (4.2 %)	640 (0.4 %)	2630 (1.7 %)	**123 515 (81.1 %)**

Figure 8. Confusion matrix illustrating the pixel-wise classification performance of the Modified Segformer model on image fragments of size 480 × 480 × 3 pixels

The confusion matrix presented as an example in Fig. 8 for the Modified Segformer model allows for a deeper analysis of the effectiveness of this model, which demonstrated the best results in solving the task of classifying Siberian fir tree crown conditions. Similar confusion matrices were also obtained for the other neural network models which contain both normalized values (percentages) and pixel counts. This matrix shows that objects of the “Background” class are classified with high accuracy. The model also shows good accuracy results for trees of the “Healthy” class, while the main source of error is associated with the false assignment of pixels to the “Background” class. This likely indicates the presence of common textural and/or spectral features between the crowns of healthy *A. sibirica* fir trees and the crowns of other coniferous tree species.

Trees of the “Dying” class present the greatest challenge for classification in the images. Indeed, despite the correct recognition of a significant proportion of pixels (65.1 %), substantial errors are observed, mainly related to the false assignment of pixels from this class to “Background” (21.9 %) and to the “Healthy” class (7.4 %). Such significant error proportions indicate that the crowns of trees in the “Dying” class possess features similar to both the crowns of healthy trees and objects of the “Background” class. To a lesser extent, errors (3.0 %) are associated with assigning pixels of this class to the “Fresh dead wood” class.

In turn, trees of the “Fresh dead wood” class are generally classified successfully by the Modified Segformer model. A small portion of misclassified pixels is assigned to the “Background” class. Finally, for trees of the “Old dead wood” class, the model demonstrates relatively good results. However, errors are observed involving the assignment of pixels to the “Background” and “Healthy” tree classes.

The analysis of the research results presented in Tables 2 and 3 leads to the following conclusions. First, for the task of multiclass classification of affected fir trees with a specified accuracy threshold of 0.5, all three proposed fully convolutional models can be practically used alongside the Modified Segformer model, which demonstrates the best results. Second, depending on the size of the dataset fragments, preference should be given to the At-Mo-U-Net model with the attention mechanism or the hybrid model with residual blocks, Res-Mo-U-Net. Third, the classic U-Net model, which consistently demonstrates low classification accuracy for intermediate “Dying” class trees irrespective of fragment size, is not recommended for practical application. Fourth, to improve the classification accuracy of affected fir trees, it is advisable to prepare larger-sized fragments when forming datasets whenever possible.

It should be noted that the observed differences in the IoUc and mIoU metrics for evaluating the quality of semantic segmentation (pixel-wise image classification accuracy), obtained using the proposed neural network models, are not statistically significant. This is confirmed by the fact that these values were obtained through segmentation using models applied to only a single, contiguous test panorama. This panorama was split into fragments of fixed size (256 × 256 × 3 or 480 × 480 × 3 pixels). These fragments do not constitute classically independent elements, as they originate from overlapping regions of the test panorama.

To evaluate the computational speed of the proposed models, another experiment was conducted. The same two test sets used in the first experiment were employed. This experiment was performed on a computer with a graphics processing unit (GPU) with the following specifications: NVIDIA GPU RTX 3080 10 GB, Intel Core i7-14790F CPU 2.10 GHz, RAM 32 GB.

Table 4 details the computation time for each of the five models when processing the first test set, which comprises 96 image fragments of size 256 × 256 × 3 pixels. Table 5 presents the computation times for the same models, but using the second test set, consisting of 24 image fragments sized 480 × 480 × 3 pixels. The reported computation time for each model encompasses fragment loading, analysis via the model, and post-processing for visualization. Given that forest monitoring is typically conducted on a per-hectare basis (10 000 m^2), forestry specialists are primarily concerned with the time a model requires to analyze a one-hectare forest image. Consequently, the fourth column in both Tables 4 and 5 presents the experimentally determined processing time per hectare for each model. For reference, the third column in each table provides the total processing time for all fragments within each test set, representing the time required to analyze the entire study area.

Table 4. Computational time costs for CNN models and Modified Segformer model using the first test set

Model	One Fragment (ms)	All Fragments (ms)	Image Area of 1 ha (ms)
Modified Segformer	33.8	3242.2	515.5
U-Net	16.9	1620.9	257.7
Mo-U-Net	17.3	1660.5	263.8
At-Mo-U-Net	28.5	2740.0	434.6
Res-Mo-U-Net	19.0	1823.5	289.8

The processing time required to analyze a one-hectare image can be estimated using the following formula:

$$T_{ha} = T_f \frac{W_{ha} x H_{ha}}{W_f x H_f}. \tag{2}$$

In (2) T_f is the experimentally determined average time to analyze one fragment using the selected model (time values are given in the second column of Tables 4 or 5); W_{ha} and H_{ha} represent

Table 5. Computational time costs for CNN models and Modified Segformer model using the second test set

Model	One Fragment (ms)	All Fragments (ms)	Image Area of 1 ha (ms)
Modified Segformer	71.7	1720.4	311.2
U-Net	38.0	912.6	164.9
Mo-U-Net	38.3	918.4	166.2
At-Mo-U-Net	58.3	1398.9	253.0
Res-Mo-U-Net	41.6	999.3	180.5

the width and height, respectively, of the image of a one-hectare forest plot (in our case, the image size is 1000 × 1000 pixels, as the image has a spatial resolution of 0.1 m); and W_f and H_f are the width and height of the fragment (in our cases, these are 256 pixels or 480 pixels).

All results presented in Tables 4 and 5 were obtained using a GPU. Factors such as its load, degree of parallelism, and the hardware-dependent optimizations used affect the fragment analysis time observed in the experiments, leading to minor discrepancies in this time value for different fragments within the test set. Therefore, the second column of each of these tables provides the average T_f fragment analysis time value for the corresponding sample and model. As an example, the standard deviation (SD) of the calculation time for the first test set (96 fragments) using the Modified Segformer model is ±4.3 ms. For the second test set (24 fragments) evaluated with the same model, the SD is ±3.9 ms.

The T_{ha} values calculated using the provided formula (2) are also averaged. However, the time values presented in the third column of Tables 4 and 5, representing the time spent by the model analyzing all fragments of the corresponding test set, were obtained experimentally.

Analysis of the research results presented in Tables 4 and 5 reveals that the computation time for the Modified Segformer model is significantly higher than that of other models, a conclusion that holds true regardless of the analyzed fragment size. The At-Mo-U-Net model exhibits the second-highest computation time. These findings suggest that the attention mechanism implemented in the Modified Segformer and At-Mo-U-Net introduces algorithmic complexity, leading to increased computational costs. As a result, deploying these models on a production scale in quasi-real-time would require substantial computational resources. While the Res-Mo-U-Net model ranks third in computation time, it closely trails the Mo-U-Net and U-Net models in terms of computational speed.

A comparison of the time required to analyze a one-hectare image of a fir forest using any of the models considered reveals that computational costs are significantly lower for fragments sized 480×480×3 pixels compared to those sized 256×256×3 pixels. This finding is particularly important for the neural network analysis of forest pathology monitoring results for coniferous trees at a production scale, where rapid analysis of large volumes of images is essential.

Based on our classification accuracy studies and the computational time analysis of each model, we recommend the Res-Mo-U-Net model for multiclass classification of fir trees affected by the four-eyed bark beetle. Res-Mo-U-Net strikes a balance between high classification accuracy and computational efficiency. For production-scale coniferous forest monitoring, we advise analyzing original images with Res-Mo-U-Net, dividing them into 480×480×3 pixel fragments. This significantly reduces analysis time compared to 256 × 256 × 3 pixel fragments, while enhancing tree classification accuracy. Even on traditional personal computers, common in the forestry sector and typically an order of magnitude less performant than our experimental setup, Res-Mo-U-Net can analyze a one-hectare image of fir forest in 3–4 seconds, enabling quasi-real-time monitoring. This suggests strong potential for widespread adoption of a software implementation of Res-Mo-U-Net in forestry enterprises. Furthermore, the model achieves a 0.74 mIoU for classifying affected coniferous trees and a 0.57 IoUc for the “Dying” class, demonstrating exceptional performance.

6. Conclusions

The large-scale destructive impacts of invasive pest insects on coniferous forests now pose a serious threat to the biological security of several regions worldwide. This underscores the urgent need for forest pathology monitoring of coniferous forests to address three key objectives: 1) early detection of insect pest outbreaks, 2) monitoring the health status (degree of damage) of infested trees within these outbreaks, and 3) identifying dead trees and estimating their phytomass and carbon content, which is eventually released into the atmosphere as emissions. We have demonstrated that achieving these objectives requires multiclass classification of coniferous tree images captured in high and ultra-high resolution using satellites, aircraft (including helicopters), or UAVs.

An analytical review of existing models and methods for multiclass classification of coniferous forest images was conducted, leading to the identification of promising directions for future development. Based on this review, we developed three CNN models — Mo-U-Net, At-Mo-U-Net, and Res-Mo-U-Net — derived from the classical U-Net architecture, as well as a modified transformer-based model called Modified Segformer. Using RGB images of *A. sibirica* trees damaged by the *P. proximus*, captured by UAV cameras, we created two datasets: one with image fragments and segmentation masks sized $256 \times 256 \times 3$ pixels and another with fragments sized $480 \times 480 \times 3$ pixels. All models were trained and validated using these datasets. Comprehensive evaluations were conducted to assess the models' classification accuracy for tree health status and their computational efficiency using test sets from both datasets.

Multiclass classification tasks for damaged fir trees with an accuracy threshold of 0.5 can be effectively addressed using all three proposed CNN models alongside the Modified Segformer, which achieved the best overall performance. For datasets with $256 \times 256 \times 3$ pixel fragments, the At-Mo-U-Net model is recommended alongside the Modified Segformer due to its attention mechanism. For datasets with $480 \times 480 \times 3$ pixel fragments, the Res-Mo-U-Net hybrid model with residual blocks is preferred. Among the proposed models, Mo-U-Net demonstrated the fastest computation times, followed closely by Res-Mo-U-Net. Modified Segformer and At-Mo-U-Net were significantly slower in comparison. Processing time for analyzing a one-hectare fir forest image is substantially lower when using $480 \times 480 \times 3$ pixel fragments compared to $256 \times 256 \times 3$ pixel fragments. Considering both classification accuracy and computational speed, Res-Mo-U-Net is identified as the most suitable model for large-scale multiclass classification tasks involving fir trees affected by the Ussuri polygraph. It offers an optimal balance between high classification accuracy and fast computation times.

There are no limitations preventing our developed models from being adapted for other coniferous tree species affected by different pest insects. This could involve modifying class definitions for tree health conditions or retraining models on new datasets. As such, our models and research findings are highly transferable across various coniferous species and insect pests.

Acknowledgements

The authors extend their sincere appreciation to Ivan A. Kerchev, Cand. Sc., Head of the Laboratory at the Institute of Monitoring of Climatic and Ecological Systems at the Siberian Branch of the Russian Academy of Sciences, for generously providing images of forest areas.

References

Керчев И. А., Маслов К. А., Марков Н. Г., Токарева О. С. Семантическая сегментация поврежденных деревьев пихты на снимках с беспилотных летательных аппаратов // Современные проблемы дистанционного зондирования Земли из космоса. — 2021. — Т. 18, № 1. — С. 116–126.
Kerchev I. A., Maslov K. A., Markov N. G., Tokareva O. S. Semanticheskaya segmentatsiya povrezhdyonnykh derev'ev pikhty na snimkakh s bespilotnykh letatel'nykh apparatov [Semantic segmentation of damaged fir trees in unmanned

aerial vehicle images] // Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa. — 2021. — Vol. 18, No. 1. — P. 116–126 (in Russian).

Кривец С. А., Керчев И. А., Бисирова Э. М., Пашенова Н. В., Демидко Д. А., Петько В. М., Баранчиков Ю. Н. Уссурийский полиграф в лесах Сибири (распространение, биология, экология, выявление и обследование поврежденных насаждений). — М.: УМИУМ, 2015.

Krivets S. A., Kerchev I. A., Bisirova E. M., Pashenova N. V., Demidko D. A., Petko V. M., Baranchikov Yu. N. Ussuriyskiy poligraf v lesakh Sibiri (rasprostranenie, biologiya, ekologiya, vyyavlenie i obsledovanie povrezhdennykh nasazhdeniy) [Four-eyed fir bark beetle in Siberian forests (distribution, biology, ecology, detection and survey of damaged stands)]. — Moscow: UMIUM, 2015 (in Russian).

Марков Н. Г., Маслов К. А. Применение методов машинного и глубокого обучения в задачах семантической сегментации изображений лесного покрова // Молодежь и современные информационные технологии: Сборник трудов XVIII Международной научно-практической конференции студентов, аспирантов и молодых ученых. — Томск, 2021. — С. 55–57.

Markov N. G., Maslov K. A. Primenenie metodov mashinnogo i glubokogo obucheniya v zadachakh semanticheskoy segmentatsii izobrazheniy lesnogo pokrova [Application of machine and deep learning methods in problems of semantic segmentation of forest cover images] // Sbornik trudov XVIII Mezhdunarodnoy nauchno-prakticheskoy konferentsii studentov, aspirantov i molodykh uchenykh [Proc. 18th Int. Conf. “Adv. Inf. Technol. Robot.”]. — Tomsk, 2021. — P. 55–57 (in Russian).

Марков Н. Г., Маслов К. А., Керчев И. А., Токарева О. С. Модели U-Net для семантической сегментации поврежденных деревьев сосны сибирской кедровой на снимках с БПЛА // Современные проблемы дистанционного зондирования Земли из космоса. — 2022. — Т. 19, № 1. — С. 65–77.

Markov N. G., Maslov K. A., Kerchev I. A., Tokareva O. S. Modeli U-Net dlya semanticheskoy segmentatsii povrezhdyonnykh derev’ev sosny sibirskoy kedrovoy na snimkakh s BPLA [U-Net models for semantic segmentation of damaged Pinus sibirica trees in UAV imagery] // Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa. — 2022. — Vol. 19, No. 1. — P. 65–77 (in Russian).

Bertels J., Eelbode T., Berman M., Vandermeulen D., Maes F., Bisschops R., Blaschko M. Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice // Medical Image Computing and Computer Assisted Intervention. — 2019. — Vol. 11765. — P. 92–100.

Bystrov S., Antonov I. First record of the four-eyed fir bark beetle Polygraphus proximus Blandford, 1894 (Coleoptera, Curculionidae: Scolytinae) from Irkutsk province, Russia // Entomological Review. — 2019. — Vol. 99. — P. 54–55.

Chang W., Lantz V., Hennigar C., MacLean D. Economic impacts of forest pests: a case study of spruce budworm outbreaks and control in New Brunswick, Canada // Canadian Journal of Forest Research. — 2012. — Vol. 42, No. 3. — P. 490–505.

Chenari A., Erfanifard Y., Dehghani M., Pourghasemi H. Woodland mapping at single-tree levels using object-oriented classification of unmanned aerial vehicle (UAV) images // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. — 2017. — Vol. XLII-4/W4. — P. 43–49.

Clevert D. A., Unterthiner T., Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus) // arXiv preprint. — 2015. — arXiv:1511.07289

Dedyukhin S., Titova V. Finding of the bark beetle polygraphus proximus Blandford, 1894 (Coleoptera, Curculionidae: Scolytinae) in Udmurtia // Russian journal of biological invasions. — 2021. — Vol. 12. — P. 258–263.

Gini R., Sona G., Ronchetti G., Passoni D., Pinto L. Improving tree species classification using UAS multispectral images and texture measures // International Journal of Geo-Information. — 2018. — Vol. 7, No. 8. — P. 315.

Han X., He Z., Chen J., Xiao G. Cross-level cross-scale cross-attention network for point cloud representation // arXiv preprint. — 2021. — arXiv:2104.13053

He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — Las Vegas, 2016. — P. 770–778.

Jintasuttisak T., Edirisinghe E., Elbattay A. Deep neural network-based date palm tree detection in drone imagery // Computers and Electronics in Agriculture. — 2022. — Vol. 192. — P. 106560.

Kerchev I., Krivets S., Bisirova E., Smirnov N. Distribution of the small spruce bark beetle Ips amitinus (Eichhoff, 1872) in Western Siberia // Russian journal of biological invasions. — 2022. — Vol. 13. — P. 58–63.

Kerchev I., Mandelshtam M., Krivets S., Ilinsky Y. Small spruce bark beetle Ips amitinus (Eichhoff, 1872) (Coleoptera, Curculionidae: Scolytinae): a new alien species in West Siberia // Entomological review. — 2019. — Vol. 99. — P. 639–644.

Kocon K., Kramer M., Wurz H. Comparison of CNN-based segmentation models for forest type classification // AGILE: GIScience Series. — 2022. — Vol. 3. — P. 42.

Korznikov K., Kislov D., Altman J., Dolezal J., Vozmishcheva A., Krestov P. Using U-Net-like deep convolutional neural networks for precise tree recognition in very high resolution RGB (red, green, blue) satellite images // Forests. — 2021. — Vol. 12, No. 1. — P. 66.

Lee S., Park S., Baekm G., Kim H., Lee C. W. Detection of damaged pine tree by the pine wilt disease using UAV image // Journal of remote sensing. — 2019. — Vol. 35. — P. 359–373.

Li H., Xiong P., An J., Wang L. Pyramid attention network for semantic segmentation // arXiv preprint. — 2018. — arXiv:1805.10180

Lierop P., Lindquist E., Sathyapala S., Franceschini G. Global Forest area disturbance from fire, insect pests, diseases and severe weather events // Forest Ecology and Management. — 2015. — Vol. 352. — P. 78–88.

Liu H., Dai Z., So D., Le Q. Pay attention to mlps // Neural Information Processing Systems. — 2021.

Markov N., Machuca C. Deep learning models and methods for solving the problems of remote monitoring of forest resources // Bulletin of the Tomsk Polytechnic University Geo Assets Engineering. — 2024. — Vol. 335, No. 6. — P. 55–74.

Musolin D., Kirichenko N., Karpun N., Aksenenko E., Golub V., Kerchev I., Mandelshtam M., Vasaitis R., Volkovitsh M., Zhuravleva E., Selikhovkin A. Invasive insect pests of forests and urban trees in Russia: Origin, pathways, damage, and management // Forests. — 2022. — Vol. 13, No. 4. — P. 521.

Onishi M., Ise T. Automatic classification of trees using a UAV onboard camera and deep learning // arXiv preprint. — 2018. — arXiv:1804.10390

Rahman M. A., Wang Y. Optimizing intersection-over-union in deep neural networks for image segmentation // Advances in Visual Computing. — 2016. — Vol. 10072. — P. 234–244.

Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation // Medical Image Computing and Computer-Assisted Intervention — MICCAI. — 2015. — Vol. 9351. — P. 234–241.

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 Sensing. — 2019. — Vol. 11, No. 6. — P. 643.

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. Attention is all you need // Neural Information Processing Systems. — 2017.

Wu Z., Jiang X. Extraction of pine wilt disease regions using UAV RGB imagery and improved Mask R-CNN models fused with ConvNext // Forests. — 2023. — Vol. 14, No. 8. — P. 1672.

Xie E., Wang W., Yu Z., Anandkumar A., Alvarez J., Luo P. Segformer: simple and efficient design for semantic segmentation with transformers // NIPS'21. — 2021. — P. 1–14.

Xie W., Wang H., Liu W., Zang H. Early-stage pine wilt disease detection via multi-feature fusion in UAV imagery // Forests. — 2024. — Vol. 15, No. 1.

Yu R., Luo Y., Zhou Q., Zhang X., Wu D., Ren L. Early detection of pine wilt disease using deep learning algorithms and UAV-based multispectral imagery // Forest Ecology and Management. — 2021. — Vol. 497. — P. 119493.

Zhou H., Yuan X., Zhou H., Shen H., Ma L., Sun L., Fang G., Sun H. Surveillance of pine wilt disease by high resolution satellite // Journal of Forestry Research. — 2022a. — Vol. 33. — P. 1401–1408.

Zhou Y., Liu W., Bi H., Chen R., Zong S., Luo Y. A detection method for individual infected pine trees with pine wilt disease based on deep learning // Forests. — 2022b. — Vol. 13, No. 11. — P. 1880.