E3S Web of Conferences **390**, 07038 (2023) https://doi.org/10.1051/e3sconf/202339007038
AGRITECH-VIII 2023

Stem segmentation for sustainable forest management task

Fedor Svoikin[1*], *Kirill* Zhuk[1], *Vladimir* Svoikin[2], *Sergey* Ugryumov[1], and *Ivan* Bacherikov[1]

[1]St. Petersburg State Forest Technical University, 5, Institutskiy Lane, St. Petersburg, 194021, Russian Federation
[2]Syktyvkar Forest Institute (branch) of the Federal State Budgetary Educational Institution of Higher Education «Saint-Petersburg State Forest Technical University named after S.M. Kirov», 39, Lenin Street, Syktyvkar, 167982, Russian Federation

Abstract. Harvesting of forest raw materials, namely felling and bucking of log on forest areas, is the first and main stage in the logging chain. One of the problems in this industry is the shortage of highly qualified specialists-operators of forest machines, including feller-delimbing-bucking machines (harvesters). Operators who have just come to the industry or have insufficient experience (have worked for less than a year) cannot correctly configure harvesters, as a result of which the processes in the logging chain are disrupted. Thus, it becomes necessary to apply additional models of working with operators and forest machines to reduce the resulting costs for the company. Understanding how much wood raw material will be obtained from the logging site allows predicting not only the amount of equipment required, but also planning actions for the next stages of logging.

1 Introduction

Determining the length of a tree log before bucking, automatic classification of the tree species, calculating the volume of raw materials for a plot are relevant tasks now and research using machine learning methods helps to solve them. In the modern approach, for the most part, laser rangefinders are used, and the obtained data is used to build a point cloud in three-dimensional space, to which, subsequently, machine learning methods are applied, namely convolutional neural networks. Using VoxNet model allows creating a detailed representation of a free-standing tree with a clear drawing of the crown of the tree. However, the disadvantage of this approach is the low-quality display of the log in the butt cut, which contains the first, most valuable assortment for bucking. The aim of this study is to create a convolutional neural network for segmentation of stem parts that are in the butt.

Currently, at the legislative and executive level, special attention is paid to the timber industry complex of the Russian Federation, which, despite its raw material orientation, currently does not occupy one of the leading places in the formation of economic security of the Russian Federation as a whole and the revenue part of the budget of the Russian Federation in particular. An important stage in the formation of a sustainable timber industry cluster is the provision of raw materials and rhythmic transparent chain of supply of timber

* Corresponding author: fsvoikin@mail.ru

in round timber from the felling site to the end consumer. In 2020, according to various estimates, the volume of timber harvested in Russia was 217 000 000 – 242 000 000 m^3. However, the use of the estimated logging area did not exceed 30%, which is significantly lower than the same indicator in North American countries and the European Union, indicating a large growth potential for this indicator in Russia. According to modern ideas about the technological process and the level of technical development, the harvesting of such volume of timber implies the use of modern multi-operator forest machines: harvesters in the phase of felling-cutting-delimbing, subsorting (forming a micro-pack of similar in size and quality characteristics assortments directly on the apiary). It should be noted that in Russia the Scandinavian method of assortment harvesting of wood raw materials with the use of harvesters and wheeled assortment-picker (wheeled forwarders) mainly of medium class prevails.

When carrying out the process of assortment timber harvesting in round timber with the yield of assortments at the logging site using modern multi-operator forest machines, it is assumed to use trained and motivated highly qualified personnel who have specialized knowledge to operate such forest machines: harvester and forwarder operators. The experience and qualifications of the operator have a significant impact on the economic efficiency of such operations. As a rule, the more experience a particular operator has in operating and maintaining machinery in various natural-production conditions (as well as knowledge of effective logging, forestry knowledge, competence in hydraulics, electrics, control systems and base machine management), the better he will perform the task of developing the enterprise's timber harvesting fund. However, now, the logging industry is experiencing an acute shortage of qualified operators in the Russian Federation, so logging companies train their own staff (Ilim Group, Group of Companies “Ustyanskiy Lesopromyshlenniy kompleks”, etc.). This approach requires not only material investment, but also the need to spend a sufficiently large amount of time on the training of qualified personnel. Thus, the search and training of qualified personnel to carry out efficient logging of round timber by Scandinavian cut-to-length technology is a particularly urgent task.

Another solution to this problem is operator training on the basis of suppliers of solutions for Scandinavian logging technology (Ponsse Oyj, LLC ISTK, Group of Companies “Tractorodetal”, etc.) However, at the moment, training from the manufacturer is not actively implemented in the industry for a number of reasons: high cost of such training for the consumer and for companies providing such services.

One solution could be the implementation of training on the basis of state or non-state educational organizations on a hybrid basis with partial delegation of retranslation of knowledge and experience of forestry solutions provider companies by teachers of educational organizations, but due to lack of motivation of effective teachers (bad experience), the development of such effective systems, taking into account the earlier mistakes, is only in progress [1-4].

One of the relevant modern approaches for solving this problem, which do not require highly qualified (and therefore highly paid) personnel for retranslation of knowledge and experience on the part of the customer of such work and the contractor is the use of modern mathematical methods of machine learning to help operators of multi-operator forest machines. This approach is proposed to be used not only for logging round timber, but also at every stage of the logging process [5-7].

For example, Reksoft, a software application company, as a part of its activities, has developed and implemented a project to automatically determine the stacking density and species of wood when a timber truck passes through a frame that is installed at a wood processing plant. In this way, the logging company was able to speed up the calculation of the volume of imported wood raw materials, eliminating the human factor. Convergent neural

https://doi.org/10.1051/e3sconf/202339007038

networks were used to determine the type of wood raw material, making it possible to process incoming information efficiently [8, 9].

One of the global tasks in the logging complex of the Russian Federation to implement the decrees of the President of the Russian Federation on decriminalization of the industry and increasing the transparency of timber supply chains is the accurate calculation of the volume of wood raw material in the logging area. Quick and high-quality assessment of the volume of timber in the specified area allows effective planning of further actions in the chain of custody of the logging process.

Taking into account the fact that in Russia it is difficult or impossible to find highly qualified operators to operate logging machines due to a number of factors, it is necessary to hire less experienced workers, which affects the technical condition of the expensive logging equipment (1 complex: 1 harvester + 1 forwarder at the moment in Russia costs about 1 000 000 Euro and more), the cost of wood harvesting works and the profit of the company. However, in order to partially compensate for the lack of qualification of operators, it is necessary to go from two sides simultaneously: training of insufficiently qualified and unmotivated new operators and improvement of technical equipment, hardware and software.

It is worth noting the emerging trend in the logging industry to install video surveillance cameras in the cabins of modern forest machines in the Russian Federation to record the working time of forest machine operators.

This paper proposes to use machine learning methods to automatically segment tree trunks using a camera located in the cab of harvester operator. The purpose of this work was to test the approach of segmentation of individual tree trunks, to analyze the results for the quality of segmentation, which, in the future, will be used to implement automatic classification of tree species [10].

2 Theoretical basis

Modern approaches to logging processes allow decision makers planning and controlling all stages of round timber harvesting in a timely manner. Understanding how much wood resources are in a plot makes it possible to plan further actions and calculate the amount of logging equipment needed to harvest the specified volume of wood.

A lot of the research that has been done in this field involves estimating the amount of wood resources in a specified area using the technique of measuring the distance to objects by emitting light and measuring the time in which the beam travels the distance in two directions [11]. This approach is carried out using lidars [12]. The devices are embedded in unmanned aerial vehicles (UAV) and scan the terrain from the air. The result is a three-dimensional point cloud and an accurate map of the plot with found objects can be created [13, 14]. However, the resulting map will contain all objects together (land and trees), and in order to calculate the value of wood resources it is necessary to separate tree trunks from the landscape [15]. By applying machine learning techniques, namely convolutional neural networks, to the results obtained, the obtained trees can be segmented and separated from each other [16]. VoxNet convolutional neural network model [17] was used to solve a similar problem.

Figure 1 shows an example of how VoxNet segmentation neural network works.

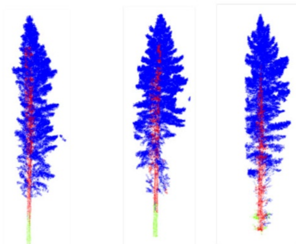

Fig. 1. Segmentation of individual tree trunks using VoxNet convolutional neural network model.

However, the disadvantage of this approach is that the part of the trunk that is in the butt is poorly segmented, and the calculated trunk volume will be approximate, while the recognition of the upper part of the tree is presented in a sufficiently good quality of the segmentation. The first grade in a cut tree is the most valuable, so it is important to use a method that will segment part of the tree in the butt cut with sufficient quality.

In order to accurately calculate the trunk volume of a tree, we need to know the diameter of its trunk at breast level (DBH). In the Russian Federation, this parameter is calculated at 1.3 meters. In a paper that was published in May 2020, a study was conducted to calculate DBH parameter using data obtained by UAV that is equipped with lidar. The values of the diameters of the original trees ranged from 70 to 96 cm. As a result, it was calculated that DBH on average differed by 15.85% from the real values. The error was calculated using the Root Mean Square Error (RMSE) formula, and its maximum value was 24.5 cm [18].

Research to find solutions that make the logging process more stable is relevant at this time. One such task is pre-segmentation of tree trunks, the results of which can be used for species classification and harvest counts [19]. A relevant task is the automatic classification of a tree trunk species as it is harvested. At the moment, the selection of the species is done directly by the operator of the forest machine. There are programmable buttons on the joysticks that control the machine, and a different species is recorded on each of them. Before starting the crosscutting process, the operator presses a button to select the species of the tree to be harvested. This process, although it takes very little time to carry out, has the disadvantage that the operator may make an unintentional mistake in selecting a species, and this will already affect the further processes of logging. Also, when the operator transfers to another machine, he adjusts the buttons to himself, and each time he changes to another operator, a similar process takes place. This process takes a different amount of time, which depends on the experience of the operator. It is worth noting that some operators are not experienced in programming buttons, which can lead to stopping the process of logging, or violation of Russian law, which can lead up to criminal liability (according to the existing regulatory and legal framework and law enforcement experience in Russia is not allowed a difference of more than two units by species and over 10% by volume of harvested wood for each species) [20, 21].

Thus, the initial task is to conduct a survey that is aimed at obtaining segmented tree trunks. The results of the segmentation must be suitable for a study on the classification of the segmented trunks.

3 Methodology

3.1 Data set

To perform the tree trunk segmentation work, a dataset of 100 images was collected. Each of them has a size of 400x400 pixels, and they are of RGB color model. The images were collected using Internet through Google search engine using the query "tree stems forest". The dataset contains images in daylight only. However, images with difficult working conditions for the operators were included (glare of glasses, protective grates on the windows, overlapping of the view by the boom of harvester). To train the segmentation neural network 20 images were also taken (included in the total set), the view of which was taken from the outside of harvester.

Fig. 2. Depiction of difficult working conditions from the dataset.

3.2 Image markup

In order to train a segmentation neural network to detect tree trunks in an image, we need to create a mask for each image in the dataset. The mask is a black and white image of 400x400 pixels. It has one channel (Grayscale) and the pixel value is 0 or 1. The value of 1 is put in the place where the trunk of the tree is depicted in the source picture. The Figure 3 shows a picture and its corresponding markup.

Fig. 3. Image with the tree trunks and the corresponding marked picture.

It is worth noting that the pixels of the markup for the image, which is shown in the Figure 3, have values equal to 0 or 255. This is done to make it visually clear where the marked areas are, because the value of 1 is almost impossible for the human eye to distinguish from 0. Also we mention that a value of 0 is pure black, while a value of 1 is pure white. So, before we start training the convolutional neural network, we need to replace all pixels in the marked-up pictures, which have values of 255, with 1. All the marked pictures were created in Photoshop CS6.

3.3 U-Net architecture

A neural network model with U-Net architecture [22] was chosen for the task of tree trunk segmentation. This model was created to solve the task of segmentation of biomedical images. The architecture is shown in the Figure 4.

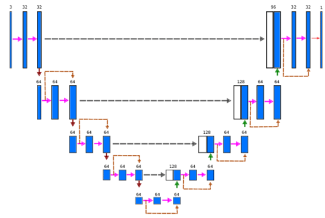

Fig. 4. Architecture of U-Net segmentation neural network.

U-Net model consists of two parts: convolution (left) and sweep one (right). The convolution part consists of 4 blocks, each of which applies two consecutive convolution operations with a kernel of 3x3, which are applied to each of the 32 filters. ReLU activation function is applied to the resulting matrices, and then a pulling operation with a maximum function and 2x2 kernel size in steps of 2 is applied. On each block, the number of filters is increased by a factor of 2. Sweep operation also consists of 4 blocks, in each of which is the operation of pooling, but in the opposite direction, and then the convolution with kernel 2x2 is applied to the obtained result, which allows changing the number of filters by 2 times. After it is done, the concatenation process is performed with layers from the corresponding block in the convolution part. In case the matrix dimensions do not match, a clipping operation is applied. At the end, two convolution operations with a 3x3 kernel are applied, to the result of which ReLU activation function is applied.

It is worth noting that in this article the architecture of the neural network from the first block uses 32 filters that are subsequently doubled in each subsequent block. In total the segmentation neural network contains 18 convolutional layers.

3.4 Sørensen–dice coefficient

In order to train the segmentation neural network it is necessary to designate the function by which the error will be calculated for each training example. In this case we will use Sørensen–Dice coefficient, because the initial image will be compared with the result, which the neural network predicts. The value of this index will be calculated by formula (1).

$$dice = \frac{2\left|X \cap Y\right|}{\left|X\right| + \left|Y\right|}, \tag{1}$$

To get the value of Sørensen–Dice coefficient, the marked-up picture and the matrix resulting from the segmentation neural network will overlap each other to get their common area of overlap. In the best case, Sørensen–Dice coefficient value will be equal to 1, when the marked area will coincide with the prediction of the neural network.

3.5 Data preparation

Immediately before training the segmentation neural network it is necessary to prepare the initial images and their marked images for work. For this purpose it is necessary to convert the initial data into data arrays in the numpy format and then make up picture-mask pairs.

 https://doi.org/10.1051/e3sconf/202339007038

The resulting data set of pairs is divided in the ratio 3:1, where 3 parts of the data goes to training the neural network, and 1 part remains for testing. Thus, 75% of the entire sample will be used for training. The second option of splitting the initial data set would be a ratio of 4:1, where 80% of the entire sample would be used for training.

Training of the neural network will be done using stochastic gradient descent, which means updating the parameters as each picture is passed. The number of filters starts at a value of 32, and there will be 512 filters on the last block, which connects the convolutional and sweeping parts. It is worth noting that the number of epochs will be a configurable parameter, and it is intended to take from 10 to 100 epochs. In this case, an epoch is the passing of all pictures once through the segmentation neural network.

3.6 Tools of analysis

Training and testing of the segmentation neural network were performed using Python programming language and additional libraries, which are downloaded and installed separately. The following third-party libraries will be needed for this study: numpy, tensorflow, keras and matplotlib. The numpy library was used for working with the dataset and manipulating it, while the system of two libraries was used for training and testing the neural network: tensorflow and keras. It is worth noting that keras library is installed only after downloading tensorflow.

4 Results

On the test sample, the average value of Sørensen–Dice coefficient was 0.58 for 80 epochs of segmentation neural network training. The comparison of the average value of Sørensen–Dice coefficient from the number of epochs is shown in the Table 1.

Table 1. Magnitude of the average value of Sørensen–Dice coefficient depending on the number of epochs of training.

Number of eras	Dice
10	0.43
20	0.37
30	0.47
50	0.54
70	0.53
80	**0.58**
100	0.52

It is worth recalling that the training was conducted using stochastic gradient descent and involved updating the parameters of the segmentation neural network after each passage of one picture from the training dataset.

The test portion was allocated 25% of the entire sample. For each value of the number of epochs shown in the Table 1, the neural network was trained 100 times, and each time the sample was shuffled randomly before a new training process. The results of each of the 100 trainings for each number of epochs were used to find the average value of Sørensen–Dice coefficient. Thus, it was found that with 80 epochs of training of the segmentation neural network, the best average value of Sørensen–Dice coefficient was 0.58. The worst mean coefficient value was 0.37 for 20 epochs of training, although the coefficient is 0.43 for 10 epochs of training.

The maximum values of Sørensen–Dice coefficient for the test sample for a single picture was about 0.81. This fact was recorded when training a segmentation neural network for 50 epochs. It is also worth noting that with 80 epochs of training, for individual images on the test sample the coefficient values of 0.75 and 0.79 were obtained.

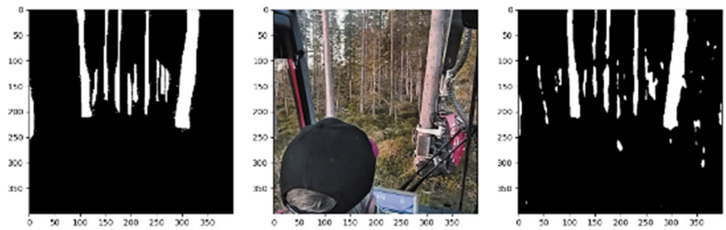

Fig. 5. Result of the segmentation neural network with the value of Sørensen–Dice coefficient equal to 0.81.

A comparison of Sørensen–Dice coefficient values when dividing the raw data into training and test samples in the ratio 3:1 and 4:1 is shown in the Table 2.

Table 2. Comparison of the average Sørensen–Dice coefficient scores for different partitions of the original data.

Number of eras	Dice (0.75)	Dice (0.8)
10	0.43	0.34
20	0.37	0.36
30	0.47	0.41
50	0.54	0.43
70	0.53	**0.46**
80	**0.58**	0.45
100	0.52	0.41

When the original sample was partitioned by 3:1 ratio, the average index value for each number of epochs was higher than when it was partitioned by 4:1. In the case where the sample was partitioned by 4:1 ratio, the best average value of Sørensen–Dice coefficient was achieved at 70 epochs, and the value was 0.46. At 20 epochs in the case of 3:1 split, the result was the worst, and when 4:1 ratio was used, the worst result was obtained at 10 epochs. However, when training at 20 epochs, the smallest difference in the average index values at different partitions was achieved and was 0.01, and the largest difference was obtained when training at 80 epochs and was 0.13.

5 Discussion

The value of Sørensen–Dice coefficient value for perfect segmentation will be as high as one. A coefficient value between 0.9 and 1 would be considered an excellent result, a good result between 0.7 and 0.9, and a satisfactory one between 0.5 and 0.7. Satisfactory results were obtained for the number of epochs equal to 50, 70 and 80. Such indicators are most likely related to the markup of the masks for the original images. All 100 images, which are involved in the training of the segmentation neural network, were collected from the Internet, and the markup of masks was done manually using the program Photoshop CS6. In some cases, artifacts in the form of separate areas of 1x1 pixels could occur during the markup, which affected the quality of neural network training. However, the pictures from the initial data set were not completely marked up, that is, a mask was not created for all of the tree trunks that were depicted on them, but only for those that stood out the most. And it turned out that the segmentation neural network resulted in images-masks, on which were found parts of the logs, which were not marked in the dataset, and thus, when calculating Sørensen–

Dice coefficient, the values could be different, because there would be more areas, which overlapped. Therefore, in this case, the resulting value of 0.58 for 80 training epochs can be considered good.

Dividing the original data set into training and test samples in the ratio 4:1 showed a worse result than 3:1 split. One would assume that by increasing the training sample size, the result of the segmentation neural network should be better, but because we increase the training sample, the test portion decreases. The original 100 images contain 20 images, which are taken from the outside of the harvester. Perhaps we should replace these 20 images with a view from the cabin of the forest machine, in which the operator sits.

6 Conclusions

This study was conducted in order to analyze the approaches that can be used for tree log segmentation. Understanding the value of the timber volume in a harvested area before the process begins allows for advance planning of further actions in the harvesting chain, which leads to sustainable management. If a quick estimation of available resources is needed, it is advisable to use UAV that are equipped with lidars [23]. Information from these devices is processed by modern machine learning methods and models such as VoxNet [24]. This approach identifies the tree crown and most of the log well, however, the part of the tree in the log is segmented residually poorly, and the average error, when calculating DBH parameter, was 15.85%.

Finding the log of a tree using a segmentation neural network on a picture that is obtained from the forest machine cab will be useful for automatic classification of species in the bucking process. This approach will save time in setting up the forest machine for a particular operator, and since some of the operators may not have enough knowledge to program the names of species on the joystick buttons, the use of machine learning methods to solve this problem will be optimal.

The obtained results of the segmentation neural network in this work can be used to calculate the volume of the specified tree before cutting, based on the calculated DBH parameter. This will allow comparing the volume of harvested wood raw material with the data, which is calculated by the onboard control system of harvester with the help of sensors measuring the log length and diameter, and if there are significant deviations, it will be necessary to pass the process of calibration with a measuring fork on harvester.

This research was supported by Grant of the President of the Russian Federation for state support of young Russian scientists – candidates of sciences № MK-2025.2022.4.

References

1. D. M. Kapustina, Persp. Nauki i Obr. **50(2)**, 118-29 (2021)
2. D. M. Kapustina, S. A. Makushkin, E. V. Danilova, T. I. Alexandrova, Y. M. Lagusev, V. D. Sekerin, Issues of Organizing and Monitoring the Performance of Labor Duties Web. **18(SI)**, 1160-1169 (2021)
3. E. G. Galizina, A. B. Feoktistova, S. A. Makushkin, I. E. Korotaeva, E. Y. Kartseva, N. Udaltsova, W. Center, Opportunities and Prospects Web. **18(SI)**, 420-435 (2021)
4. P. A. Ukhov, D. A. Borshchenko, D. D. Kabanov, M. E. Bergen, A. V. Ryapukhin, J. of Phys.: Conf. Ser. **1889(2)**, 022002 (2021)
5. M. V. Arkhipov, E. A. Vzhesnevsky, M. A. Eremushkin, E. A. Samorukov, L. B. Kocherevskaya, IOP Conf. Ser.: Mat. Sci. and Eng. **489(1)**, 012051 (2019)

6. M. V. Arkhipov, L. B. Kocherevskaya, V. V. Matrosova, Int. Conf. on Ind. Eng., Appl. and Man. **2021**, 630-635 (2021)
7. A. Burova, A. Ryapukhin, AIP Conf. Proc. **2402(1)**, 040002 (2021)
8. V. A. Tereshonkov, D. S. Shavelkin, I. V. Pocebneva, Int. Russ. Aut. Conf. **2021**, 274-289 (2021)
9. D. S. Shavelkin, V. A. Kaigorodova, J. of Phys.: Conf. Ser. **2094(4)**, 042054 (2021)
10. K. Olofsson, J. Holmgren, For. **7(9)**, 207 (2016)
11. P. Pueschel, G. Newnham, G. Rock, T. Udelhoven, W. Werner, J. Hill, ISPRS J. of Phot. and Rem. Sens. **77**, 44-56 (2013)
12. Q. Chen, D. Baldocchi, P. Gong, M. Kelly, Phot. Eng. & Rem. Sens. **72(8)**, 923-932 (2006)
13. U. Weiss, P. Biber, Rob. and Aut. Syst. **59(5)**, 265-273 (2011)
14. S. V. Novikov, D. S. Veas Iniesta, Am. Inv. **8(19)**, 298-307 (2019)
15. X. Liang, P. Litkey, J. Hyyppa, H. Kaartinen, M. Vastaranta, M. Holopainen, IEEE Trans. on Geosc. and Rem. Sens. **50(2)**, 661-670 (2011)
16. S. Ren, K. He, R. Girshick, J. Sun, IEEE Trans. on Patt. Anal. and Mach. Int. **39(6)**, 1137-1149 (2017)
17. L. Windrim, M. Bryson, IEEE/RSJ Int. Conf. on Int. Rob. and Syst. (IROS) **2019**, 3898-3904 (2019)
18. L. Windrim, M. Bryson, Rem. Sens. **12(9)**, 1469 (2020)
19. A. Bienert, S. Scheller, E. Keane, F. Mohan, C. Nugent, ISPRS worksh. on las. scann. **36**, 50-55 (2007)
20. I. S. Pinkovetskaia, D. F. Arbeláez-Campillo, M. J. Rojas-Bahamón, D. S. Veas Iniesta, Am. Inv. **9(29)**, 368-373 (2020)
21. J. Heinzel, M. O. Huber, Rem. Sens. **9(1)**, 9 (2016)
22. O. Ronneberger, P. Fischer, T. Brox, Int. Conf. on Med. Image Comp. and Comp-Ass. Int. **2015**, 234-241 (2015)
23. K. Olofsson, J. Holmgren, H. Olsson, Rem. Sens. **6(5)**, 4323-4344 (2014)
24. F. Milletari, N. Navab, S. A. Ahmadi, Int. Conf. on 3D vis. (3DV) **2016**, 565-571 (2016)