Computers and Electronics in Agriculture 166 (2019) 105012

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Classification of tree species and stock volume estimation in ground forest images using Deep Learning

Jiazheng Liu, Xuefeng Wang*, Tian Wang

Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China

ARTICLE INFO

Keywords:
Ground forest image
Deep Learning
UNET
Tree species classification
Nonlinear mixed effect model
Growing stock volume

ABSTRACT

Tree species classification and estimation of stock volume are two very important tasks in forest management. Currently, Ground Surveys' Development (GSD) is the basic and most common approach employed by foresters. However, GSD is time-consuming and inefficient as it requires great human effort. In this research, digital cameras have been used, to obtain images of the ground forest. The classification and accumulation of tree species is performed, by considering extracted relevant image information. The purpose of this effort is not only to improve research efficiency, but to reduce the consumption of human and material resources as well. This research uses the UNET network which is pre-trained by the VGG16 model. The aim is to semantically segment the image containing the ground forest and the species and then to accurately identify the number of trees contained in the image. The proportion of the number of pixels in the trunk of each segment is estimated by considering the total number of pixels in the image. The nonlinear mixed effect model is used to estimate the growing stock volume. The differences in the growing stock volume caused by different forest types, are resolved by using the growing stock volume estimation equations, related to different tree species. The experimental results show that the tree species' classification accuracy in testing is 96.03% and the average IoU (Intersection over Union) is 86%. The R^2 and RMSE of the growing stock volume prediction model are equal to 80.70% and 30.539 (m^3/ha) respectively. Therefore, it is concluded that the method proposed in this research can be used as an effective tool for tree species' image segmentation and classification, and that the growing stock volume is predicted accurately by the extracted tree pixel information. The combination of the two approaches provides a new method for forestry ground investigation work.

1. Introduction

Tree species classification and estimation of growing stock volume are not only very important forest management processes, but they can also be considered as main tasks of forest science in general. Tree species classification and growing stock volume estimation are important factors in the construction of forest resource surveys. They can also be used as important indicators for evaluating forest carbon sequestration capacity. Timely and accurate identification of tree species and acquisition of growing stock volume, can reflect the overall scale and level of forest resources in a country or in a region. They also provide important estimators for the abundance of forest resources and for the assessment of a regional ecological environment.

Obtaining information about forest growth factors and achieving accurate identification and location of trees is one of the important contents for studying growing stock volume and biomass. It is a fact that forest growth factors play a key role in assessing and evaluating the quantity and quality of the diverse services provided by forest ecosystems. Therefore, forestry personnel is required to carry out non-destructive monitoring of forest growing stock volume estimation through certain technical means (Olschofsky et al., 2016; Burkhart and Tomé, 2012).

In addition to the development of forest resources' ground surveys based on manpower, conducting in-depth research using information hidden in forest images is a hot approach. Remote sensing can be used to make such an exploration in a macroscopic scale. Wang et al. (2019) have studied tree species classification by focusing on urban environment. They actually combined airborne radar and optical imaging technology, to provide new ideas for improving urban tree species classification. Sačkov et al. (2017) have used airborne LIDAR data combined with multispectral aerial imagery, to evaluate the main tree species and the characteristics of the stand in the study area. This method has identified 73% of the upper layer trees and 28% of the lower layer ones. Alonzo et al. (2014) have tested 29 different tree

* Corresponding author.
E-mail address: xuefeng@ifrit.ac.cn (X. Wang).

https://doi.org/10.1016/j.compag.2019.105012
Received 17 May 2019; Received in revised form 12 September 2019; Accepted 13 September 2019
Available online 18 September 2019
0168-1699/

species in Santa Barbara California USA, by adding LIDAR structural indicators to AVIRIS image analysis. They have managed to increase the classification accuracy by 4.2%. Dechesne et al. (2017a,b) have used LIDAR and multi-spectral image data to optimize the semantic segmentation of pure forest images. The accuracy of the segmentation model was between 96% and 99%. Immitzer et al. (2012) have used the high spatial resolution 8-band WorldView-2 satellite data, to classify and identify 10 tree species in Austrian temperate forests, with a total recognition rate of 82%.

Of course, in addition to the classification of tree species, there are more studies on the estimation of growing stock volume and their respective biomass, through the consideration of stands' images. Bilous et al. (2017) have used Random Forest (RF) and k-Nearest Neighbors (k-NN) methods, in order to predict six tree species in the Polissya region of Ukraine. They have carried out growing stock volume and biomass mapping. The experimental results have shown that the K-NN method can estimate biomass with an accuracy of 3 m^3/ha. Hawryło et al. (2017) have used Sentinel-2 satellite imagery and on-board image point cloud, to predict the growing stock volume of Scots pine forests, and they have achieved R^2 equal to 0.82. Santi et al. (2017) have employed an Artificial Neural Network (ANN) inversion algorithm to estimate forest biomass in the Mediterranean region, using multi-frequency SAR images, and the experimental results were very good and reliable.

Recent developments in Artificial Intelligence and data mining have led to the employment of corresponding core algorithms like Deep Learning (DL). These technologies have been applied in many diverse areas, with remarkable results. Thus, new ideas for forest research have been introduced. Morales et al. (2018) have used a drone to shoot the image of the *Mauritia flexuosa* canopy. They have applied DL to automatically segment the image. The accuracy of the test set was equal to 98.14%. Freudenberg et al. (2019) have applied a UNET neural network to detect large-scale palm tree images, and they have achieved palm trees' detection accuracy between 89% and 92%. Li et al. (2016) have used a CNN to detect and count Oil palm trees in remote sensing images, and they have obtained an overall accuracy rate equal to 95%. Zakharova (2017) have developed an InceptionV3-CNN network model to detect coconut trees by means of migration learning, with a Recall rate of 93%. Mubin et al. (2019) have developed a CNN to detect juvenile oil palm and mature oil palm, with an accuracy of 95.11% and 92.96%, respectively.

The aforementioned remote sensing image is convenient for obtaining the distribution of the forest stand, the distribution of biomass, the productivity and the variation law at a large scale. However, the cost is high, and the versatility is poor. The actual problems related to the ground forest sample, such as segmentation and identification of forest trees, and estimation of the volume of accumulation in the sample, cannot be solved. In contrast, the image of the ground forest collected by a digital camera or a smart phone is not only simple to obtain, but it can also enable optimization of the tree species division and stand accumulation estimation in the sample.

This research paper proposes the use of Deep Learning to accurately segment and classify the trees of a ground forest image under a complex background. The aim is to extract the segmented and classified pixel, which is further applied to the prediction of the Growing Stock Volume (GSV). More specifically, the contributions of this paper are presented below in an algorithmic approach:

(1) First, the VGG16 pre-trained UNET was used to semantically segment the ground forest images in a complex background. The different tree species and their respective quantities in the single ground forest image were accurately segmented, detected and calculated. The Ratio of the Number of Pixels (RNP) related to each tree type after the segmentation, to the total pixel value of the entire image, was an important feature.
(2) The RNP of each tree species that was calculated in (1) was divided to the pixels of the whole image, and it was used as an independent variable. The GSV was the dependent variable and the nonlinear mixed effect model was used to determine the weight. The Pixel Percentage – Growing Stock Volume Estimation (PP-GSVE) model was used. The equations for the estimation of the GSV for different tree species, were used to resolve its variation under diverse forest types. New ideas and feasible approaches for the development of growing stock volume surveys were obtained.

2. Materials and methods

2.1. Study area and experimental data

The data in this paper have been collected in the northern mountainous area of Daxing'anling (120°48′01″ ~ 121°50′1.52″E, 51°20′59.18″ ~ 51°57′41.75″N), China. The main tree species in this area are *Larix gmelinii* (Rupr.) Kuzen., *Betula platyphylla* Suk, *Pinus sylvestris* L. var. *mongolica* Litv., *Populus davidiana*. This paper focuses on the classification of these four species and also on the estimation of the GSV.

Totally, 64 circular plots with a radius of 11.3 m were randomly set. The stands' variables and site factors of various plots were measured, including the canopy density, elevation, aspect and slope. The living wood with DBH (Diameter at Breast Height) greater than 3 cm was measured for its height and for its DBH. The image capture device was a Canon EOS 700D camera with a resolution of 3456 × 5184 pixels which had to stand in the center of the plot and to follow the order of "North → Northeast → East → Southeast → South → Southwest → West → Northwest" (Fig. 1). The lens was a vertical trunk, one for each longitudinal image. Totally 8 images were considered for each plot, in order to estimate the growing stock volume data set. The camera was placed at a height of 1.3 m. The user should keep to the center of the light and to the center of the image center, parallel to the ground in the direction of the photography.

A special effort was made to meet the sample size demand during the training of the Deep Learning model, and to avoid the occurrence of model overfitting. For this reason, 3000 ground forest images were collected in the study area according to the above shooting requirements and they were used as the training dataset. The verification dataset comprised of 512 images obtained in the 64 plots, following the same protocol. The background image of the ground forest had a great impact on the final experimental results. This was due to its high complexity, to the overlapping between the branches and the crown and due to the fact that the background was quite complex. Therefore, the trunk part of each forest image was used as the main research object to segment and classify.

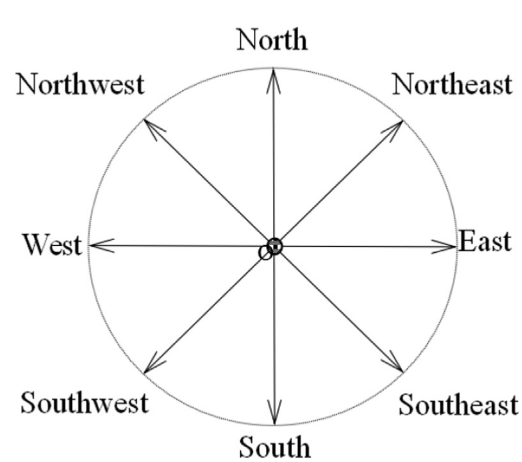


Fig. 1. The photographic process.

2.2. Methods

2.2.1. UNET architecture

From a methodological point of view, tree species' identification and detection can be seen as a semantic segmentation problem (Long et al., 2015; Dechesne et al., 2017a,b). Therefore, the model used for the semantic segmentation of tree species images was the UNET (Ronneberger et al., 2015) network. UNET is also a variant of the CNN. Specifically, it is based on the Full Convolutional Network (FCN), which was first applied to medical image segmentation studies with fewer samples. Due to the fact that its structure is similar to the shape of the letter U, it was named UNET.

The entire UNET model comprises of two main parts: the Contracting Path (CP) and the Expanding Path (EP). The CP is mainly used to capture the context information in the image. The extracted high pixel features are combined with the new feature map during the oversampling process, to maximize the front. Some important feature information are considered during the downsampling process. In order to obtain a network structure that can operate more efficiently, no fully connected layers are used. This can greatly reduce the number of training parameters and the model can benefit from the special U-shaped structure. All of the information are kept in the picture.

On the contracting path, each network layer contains a 3 × 3 Convolutional Layer (COL), and each COL is followed by a ReLU activation function. In order to achieve a better converge, the model proposed herein, has joined the Batch Normalization layer. The 2 × 2 maximum pooling approach, was employed and the step size was set equal to 2, in an effort to downsample the original image.

The expanding path performed localization on the part that needed to be segmented in the picture. In the expanding path, each layer of the network, contained a 3 × 3 COL. In the Upconvolution, each step comprised of a 2 × 2 convolution layer, whereas the step size was set to 2, and the activation function was also ReLU and joined the Batch Normalization layer. The architecture of the UNET proposed in this paper is shown in the following Fig. 2.

The data set used for image segmentation in this research consisted of 3000 images. Tus, if the model was to be trained from the beginning, the computer hardware requirements would have been very high and the training time longer. Therefore, in order to improve training efficiency and to save hardware resources, it has been decided to use Migration Learning (Yosinski et al., 2014) to migrate the knowledge gained from the VGG16 (Simonyan and Zisserman, 2014) model on the ImageNet (Russakovsky et al., 2015) image dataset, for the tree

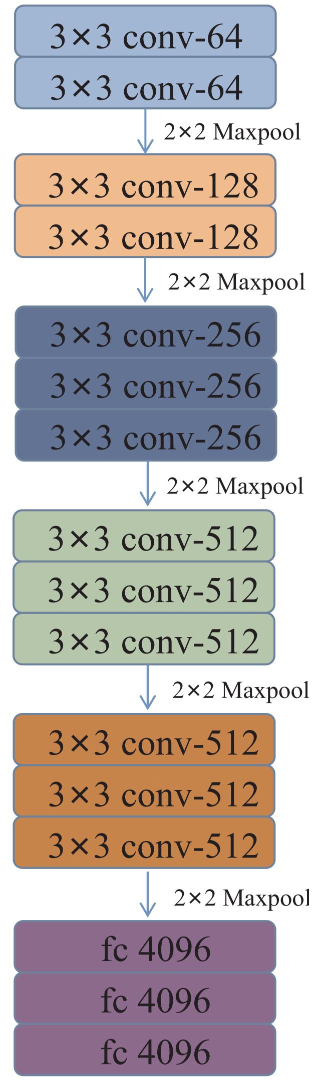


Fig. 3. The architecture of the VGG16 model.

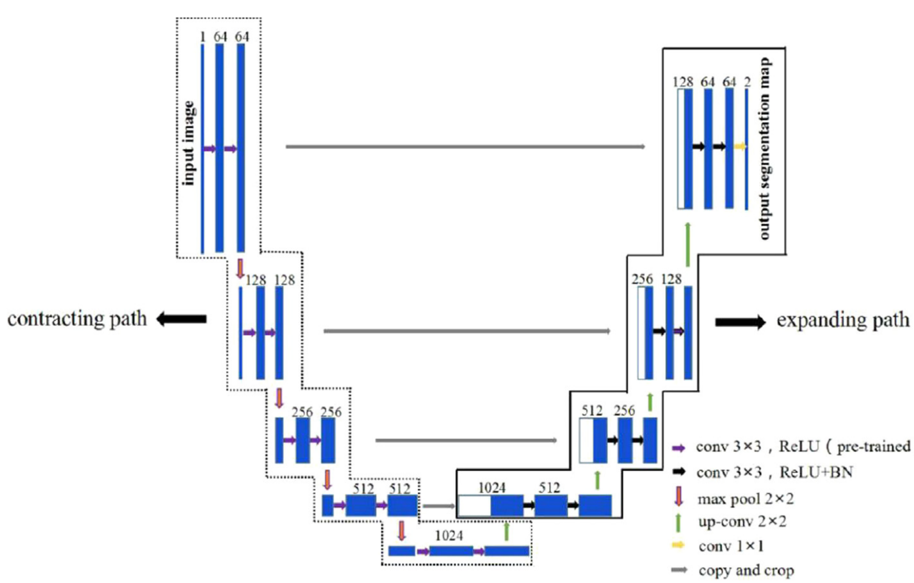


Fig. 2. UNET network architecture.

Table 1
Candidate volume models.

Model	Model form
M1	$V = a + bx$ (1)
M2(Curtis, R. O, 1967)	$V = a + bx + cx^2$ (2)
M3(Kittredge, J, 1944)	$V = ax^b$ (3)
M4(Schumacher, F. X, 1939)	$V = ae^{\frac{b}{x}}$ (4)
M5(Richards, F. J, 1959)	$V = a(1 - e^{(-bx)^c})$ (5)

segmentation image task.

The VGG16 model has been used as the encoder in the UNET network of this paper, and has been applied in the contracting path. Its specific structure is shown in Fig. 3. It comprised of 13 COLs, 5 downsampling layers and 3 fully connected layers in the network. All convolution kernels had a size of 3 × 3 and a convolution step size of one. A ReLU activation function (Dahl et al., 2013) has been used after each convolutional layer. Max-pooling has been employed for downsampling. The network comprised of 3 fully connected layers, and the number of corresponding neuron nodes was equal to 4096, 4096, 4096. The last fully connected layer neuron node corresponded to the number of labels in the ImageNet. There were 16 layers in the network containing 1,380,000,000 parameters.

2.2.2. Volume estimation – model selection

In this research, five commonly used candidate base models have been selected (Table 1). All candidate models have been fitted with experimental data and their fitting precisions have been compared. One candidate model with the highest degree of fitting was selected as the base model. Construct a weighted pixel percentage-predicted curve of the accumulated amount.

V represents the stands' volume, and x represents the pixel weight percentage of each tree in each of the 8 images. The specific calculation formulae are:

$$P_{ij}=\sum_{s=1}^{n_s} p_{ijs} \quad (6)$$

$$P_{i}=\sum_{j=1}^{n_i} p_{ij} \quad (7)$$

$$x_{is} = \frac{\sum_{j=1}^{n_i} p_{ijs}\hat{A}\cdot p_{ij}}{p_i} \quad (8)$$

P_{ij}: the percentage of all foreground pixels extracted from the jth image of the i-th sample.
P_{ijs}: the percentage of pixels of the sth tree species extracted from the jth image of the i-th sample.
P_i: the sum of the pixel percentages of all tree species extracted in the i-th sample.
n_s: tree type. In order to facilitate statistics, this paper has replaced the tree species with digital codes, s = 1 for *Larix gmelinii* (Rupr.) Kuzen., s = 2 for *Betula platyphylla* Suk, s = 3 for *Pinus sylvestris* L. var. *mongolica* Litv., and s = 4 for *Populus davidiana*.
n_j: number of images per plot, n_j = 1, 2, …8.
x_{is}: The i-th tree species weighted pixel percentage of the i-th sample.

2.2.3. Nonlinear mixed effects model

The NonLinear Mixed Effects Model (NLMEM) has been constructed based on the nonlinear relationship of the regression function, depending on the fixed effect parameters and on the random effect ones. Such models can be regarded as neglecting the group in the data analysis. In this way, a simple nonlinear model has been obtained. However, an important aspect is the lack of a nonlinear model describing the average trend of the population (Pinheiro and Bates, 2006).

The general form of the nonlinear mixed-effects model is as follows:

$$Y_i = f(\beta, u_i, x_i) + \varepsilon_i \quad (9)$$

In this research, Y_i and x_i are error vectors representing the growing stock volume of the i-th sample and the pixel weighting percentage respectively. Moreover, β and ui represent the fixed effect parameter vectors and the random parameter vectors respectively (Budhathoki et al., 2008) $f(\Delta)$. The base model has been identified in this research paper.

In order to solve the influence of the mixed forest and the inter-species effect on the growing stock volume estimation, the four species have been used as random effects. This was done to construct the corresponding growing stock volume estimation model of different tree species, and to resolve the problem of forest types' diversity, caused by the growing stock volume of forest stands.

2.2.4. Statistical analysis

(1) In the task of image segmentation of tree species, we need to train and test the UNET. Thus, the following indices were used to evaluate the model's training and testing process (Nevalainen et al., 2017):

$$\text{Accuracy} = \frac{t_p}{t_p + f_p + f_n} \quad (10)$$

$$\text{Precision} = \frac{t_p}{t_p + f_p} \quad (11)$$

$$\text{Recall} = \frac{t_p}{t_p + f_n} \quad (12)$$

$$\text{IoU} = \frac{\text{AreaofOverlap}}{\text{AreaofUnion}} \quad (13)$$

t_p: true positives, t_n: true negatives, f_p: false positives, f_n: false negatives. IoU refers to the overlap rate between the prediction result and the original data mark. The larger the IoU value, the better the prediction result.

(2) In the task of predicting the volume of stand, this study mainly used R^2 and RMSE metrics to evaluate the fitting effect and the prediction accuracy of the model:

$$R^2 = 1 - \frac{(V_i - \widehat{V_i})^2}{\sum_{i=1}^{n}(V_i - \bar{V_i})^2} \quad (14)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n}\frac{(\widehat{V_i} - V_i)^2}{n-p}} \quad (15)$$

where n is the plot number, V_i which is the measured value of the growing stock volume of the i-th plot, and $\widehat{V_i}$ is the estimated value of the growing stock volume of the i-th plot, V_i which is the average of the accumulated amount.

Moreover, the following indices were used for the evaluation and comparison of the fitting accuracy of the nonlinear hybrid model: The Akaike Information Criterion (AIC, denoted as m (AIC)) and the Bayesian Information Criterion (BIC, denoted as m (BIC)).

$$\text{m(AIC)} = \text{ICln}l + 2p \quad (16)$$

$$\text{m(BIC)} = \text{ICln}l + \ln n\hat{A}\cdot p \quad (17)$$

where p is the number of parameters in the model, n is the total number of samples, and l is the maximum likelihood function value of the model.

3. Results and discussion

The development environment of the proposed algorithm, comprised of the following: Windows10, Intel(R) Core(TM) i7-6700 CPU @

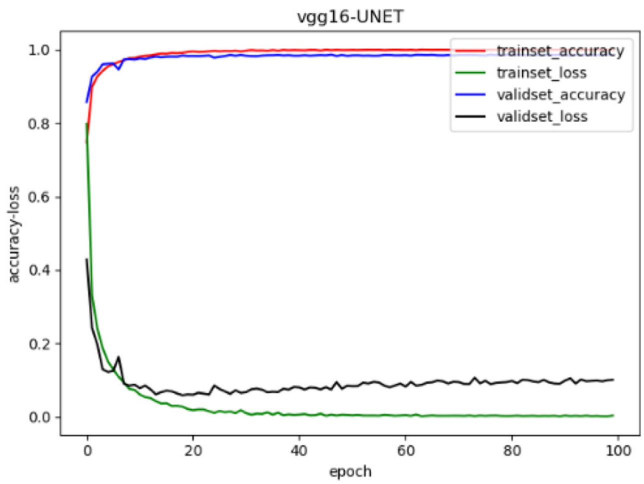


Fig. 4. The VGG16-UNET model.

3.4 GHz, and 8 GB memory. The CNN training algorithm was based on the Tensorflow-Keras framework and it was implemented in Python 3.6. To speed up the network's training, it was paired with the NVIDIA GeForce GTX 1080 Ti GPU. All of the growing stock volume model calculations in this paper were done in R (Team, 2013), where the nonlinear mixed effect model was calculated using the NLME package (Team et al., 2015).

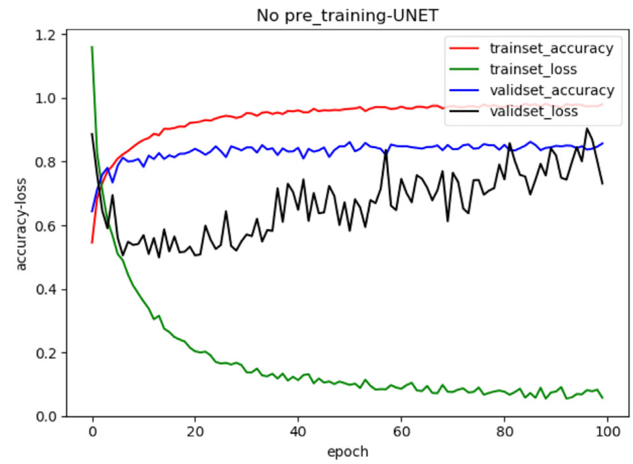
3.1. Model training

When training the CNN, the image input size was equal to 224 × 224, the learning rate was set to 0.001, the number of epochs was equal to 100, and the Batch_size had the value 8. The network model was optimized by employing the Adam optimizer (Kingma and Ba, 2014). The momentum term u was 0.9, and v was 0.999. A ReLU activation function was used after each convolutional layer. Additionally, in order to speed up the converge of the model in training, and to have a more stable distribution of the hidden output features, the Batch Normalization (BN) layer was added. The cross-entropy function used the binary crossentropy.

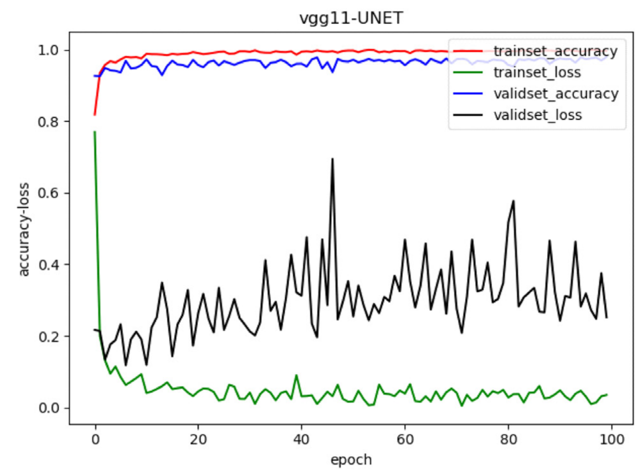
Fig. 4 shows the change of the model's accuracy and the loss rate in the training and in the verification processes. The whole network model changed according to the number of iterations. The purpose was to observe the model's training process in time and to avoid the occurrence of over-fitting and memorizing. It has been concluded from the above Fig. 3, that as the number of iterations increases slowly, the accuracy rate increases slowly as well. When the number of epochs is close to 20, the accuracy tends to stabilize. Similarly, as the number of iterations increases, the loss value slowly decreases, but there are still varying degrees of fluctuations in the verification set. When the number of epochs reaches the value 80 or so, the loss rate tends to be stable. At this point the model tends to be stable.

In order to prove the validity and the feasibility of the proposed method, we have chosen to compare it with other three established methods. Since this paper uses the VGG16 model to pre-train the UNET network by means of Migration Learning, the second method selects the non-pre-training approach and it performs the De Novo training on the same UNET. The other two methods are pre-training the UNET using the vgg11 and VGG19 models trained in Imagenet. The results of the comparative tests are shown in the following Fig. 5. It can be seen from (a), (b) and (c) that the common characteristics of the three methods are the following: the accuracy of the training set and the verification set are both ideal and the model fitting effect is good. However, the loss rate of the verification set is highly volatile and does not converge in 100 iterations, failing to meet the requirements of this experiment.

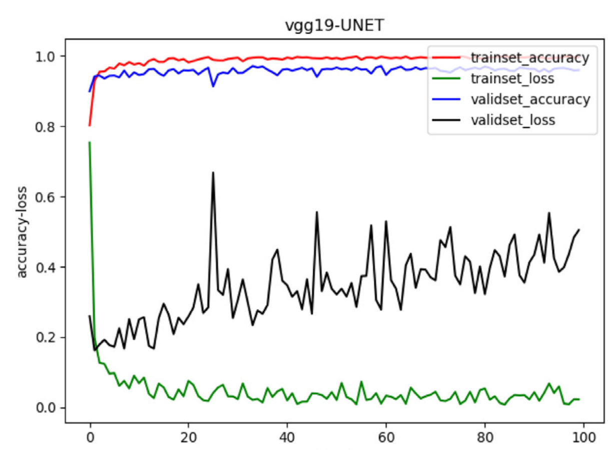
In order to visually compare the performance of the four network



(a) The No pre_training-UNET model



(b) The VGG11-UNET model



(c) The VGG19-UNET model

Fig. 5. Model training process of other methods.

forms, the following metrics were calculated by counting the results of the verification set: Accuracy, Precision, and Recall. The specific test results are shown in Table 2.

3.2. Segmentation results

In this study, the segmentation results of four tree species were displayed in different colors. Among them, red represents *Betula platyphylla* Suk, green represents *Populus davidiana*, blue represents *Larix*

Table 2
Metric Comparison of different methods.

Method	Metric		
	Accuracy	Precision	Recall
No pre-train_UNET	82.50%	84.45%	85.68%
vgg11_UNET	85.20%	92.10%	90.25%
vgg16_UNET	96.03%	97.25%	95.68%
vgg19_UNET	94.39%	96.80%	93.14%

gmelinii (Rupr.) Kuzen., purple represents *Pinus sylvestris* L. var. *mongolica* Litv., and black represents irrelevant background. Since 64 plots were randomly selected, they include both pure tree stands and mixed tree stands. Therefore, Fig. 6 shows the results of the segmentation of pure and mixed stands. According to the method proposed herein, the ground forest image has been imported into the trained UNET model for feature extraction and segmentation. It can be seen from the segmentation result graphs in (a)–(b) that the type and number of forest trees contained in the image can be accurately segmented and classified compared to ground truth.

The trees involved in an image can not only accurately identify the species, but they can also accurately segment the trunk. In addition, the pixel value of the trunk portion of each tree species included in the image, has been calculated as a percentage of the total pixels of the entire image. The data of the forest growing stock volume estimation model were constructed as the next part, and the IoU value was calculated as the test prediction result. The experimental results have shown that the accuracy of the tree species test set is 0.960, and the average IoU is 0.860. The specific extraction results and the evaluation indices are shown in Table 3.

In addition, the CNN network model is regarded as a “black box” model in the field of deep learning. Although we cannot completely “open” the “black box”, there are still ways to explore its internal structure (Krizhevsky et al., 2012). Therefore, this paper expresses the training process of each network layer of the model in a visual form. The visualization results are shown in Fig. 7. The figure shows the visualization results for the first convolutional layer, the first pooled layer, the first BN layer, the first ReLU activation function, and the sigmoid.

3.3. Estimation of stock volume

3.3.1. Determine the base model

The forest image pixel information extracted by the above image segmentation and classification means, has been used as an independent variable. The growing stock volume amount of 50 plots was used as a dependent variable to establish a storage amount estimation model. In view of the different fitting effects of various model forms, several common model fitting data have been selected, and the fitting results of each candidate model are shown in Table 4. It can be seen from the table that the M1-M4 model has a good fitting effect except for the M5 model calculation. Their parameter estimation tests have been performed, and R^2 was above 0.65. Although the fitting effect of M1 is Preferable, the estimated value of the constant term significantly deviated from the value 0. When the weighted pixel percentage x was equal to 0, the stands' volume was $-16.222\,m^3/h\,m^2$, which obviously is not in line with the actual situation. The biological significance of M2 is not obvious, however it should be mentioned that the R^2 for the cases of M4 and M2 were lower than the R^2 value of M3. In general, the M3 model had the best effect, and its decision coefficient was the largest. This model had the best goodness of fit. Therefore, it was used as the basic model to construct the stand growing stock volume estimation model.

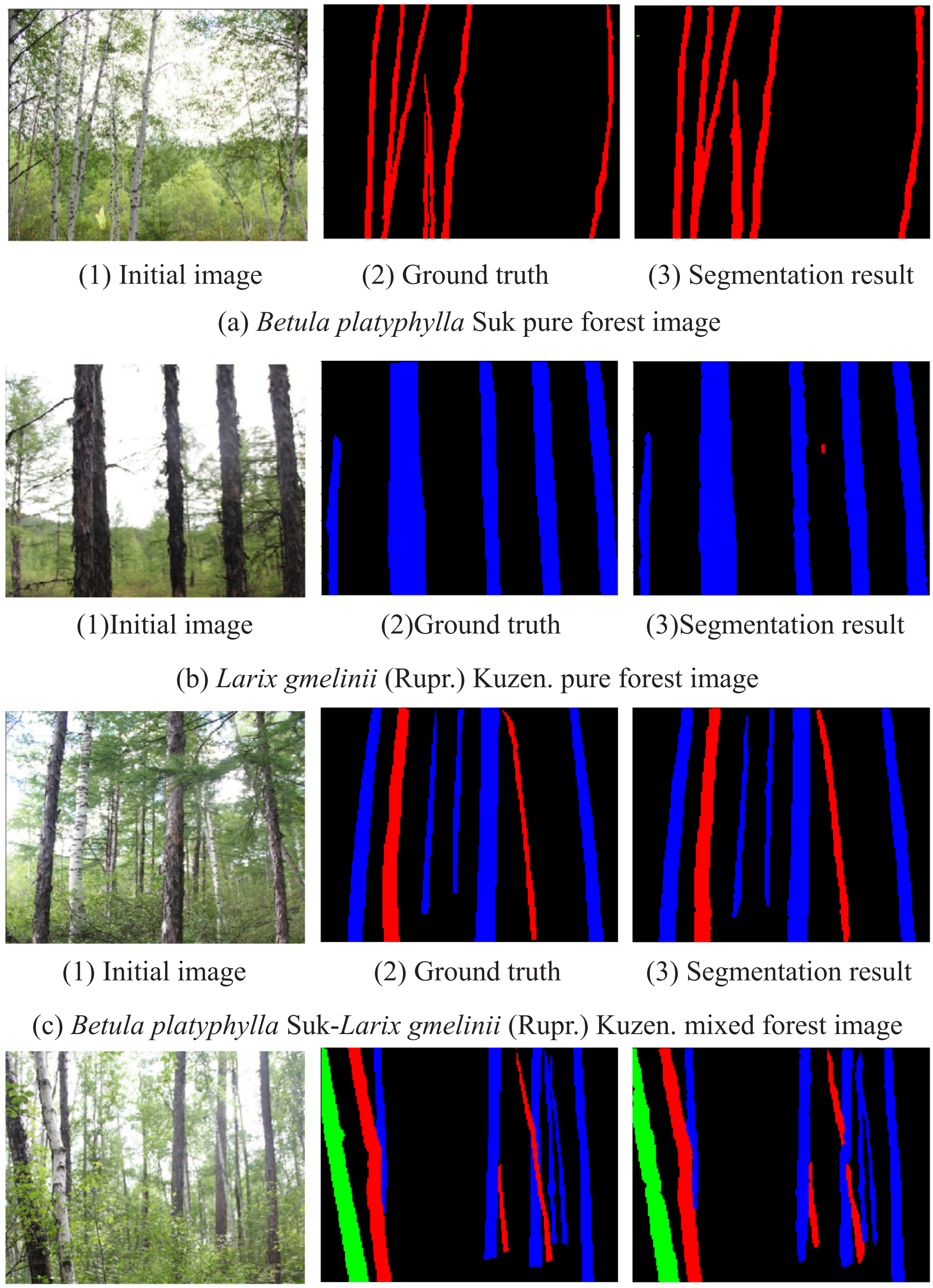
3.3.2. Nonlinear mixed effect model

This study employed the nonlinear mixed effect model, in order to use the tree species as a random effect to M3. This was done for the following reasons: (a) in order to solve the influence of different tree species on the stands' volume in a mixed forest, (b) to facilitate the prediction of the tree species composition in the stands and (c) to refine the model's application conditions. The growing stock volume estimation model is a basic obe, corresponding to different tree species. It was developed to solve the variations of the GSV under diverse forest types. In this research, random effects have been added to parameters a, b, and to a and b, respectively. The values of AIC and BIC were compared, and the model with the best fitting effect has been selected as the final model among the three schemes. The experimental results have shown that when the random effect term is added to the parameter a, the corresponding AIC value of the model is the smallest. Therefore, the study finally chooses the model based on the M3, and it adds the nonlinear mixed effect of the random effect, to the final forest. The subcumulative volume estimation, is the fitting result of the optimal nonlinear mixed-effects model (see Table 5).

Fig. 8 shows the fitted curve of pixel weighted percentage-forest growing stock volume when fixed effects and random effects are added. The Base model represents the fitted curve at the time of the fixed effect. The tree species effect is not considered when the growing stock volume amount of the entire stand level is studied. Model 1–4 represent the fitting curve of the growing stock volume model of each of the four tree species. When estimating the growing stock volume of different tree species in the forest, the model of the random effect of the tree species can be employed. The specific model is shown in Table 6. It can be found that when the pixel values are the same, the growing stock volume of *Pinus sylvestris* is the largest, above average, while the GSV of *Larix gmelinii*, *white Betula platyphylla* and *Populus davidiana* is below average. The volume of the four species was ranked as follows: *Pinus sylvestris* > *Populus davidiana* > *Larix gmelinii* > *Betula platyphylla*. The lowest GSV corresponds to the *Betula platyphylla*, and the difference with *Pinus sylvestris* var. Chinensis can reach the value of 467 $m^3/h\,m^2$. This is the main tree species in the Daxinganling forest area and the mostly clustered, but eventually it grows into little old trees. The mixed effect model with random effects refines the application conditions of the model. It can predict the growing stock volume of different tree species well, and it can provide a good means for judging the stand structure and estimating the growth of stand.

4. Discussion and conclusions

This research paper proposes a novel approach for tree species identification and accumulation estimation. The aim is the improvement of the ground survey development efficiency in the small forest scene and the reduction of the workload in the field. This effort is based on the UNET network that considers terrestrial digital images. Ground forest images were collected by using digital cameras. Each sample was collected in 8 different directions. The UGG network was pre-trained by using VGG16 and it was is employed to segment the trunk part of each tree in the ground forest image. The accuracy rate was as high as 96.03%, using the UNET network structure. It can combine the information of the bottom layer and the upper layer in the forest ground image and it can provide the context semantic information of the segmentation target in the whole image. In this way, it can contribute to the determination of the tree category. The different tree species in the image were statistically divided and the ratio of the number of trunk pixels of each tree species to the entire image was calculated and used as a factor for estimating the accumulation amount. The developed nonlinear mixed effect model was used to construct the estimation model of the stand volume. The obtained values of the R^2 and RMSE indices were equal to 80.70% and 30.539 (m^3/Ha) respectively. The tree species in the plot were accurately divided, and the stand volume was estimated.



(1) Initial image (2) Ground truth (3) Segmentation result

(a) *Betula platyphylla* Suk pure forest image

(1)Initial image (2)Ground truth (3)Segmentation result

(b) *Larix gmelinii* (Rupr.) Kuzen. pure forest image

(1) Initial image (2) Ground truth (3) Segmentation result

(c) *Betula platyphylla* Suk-*Larix gmelinii* (Rupr.) Kuzen. mixed forest image

(1) Initial image (2) Ground truth (3) Segmentation result

(d) *Populus davidiana-Betula platyphylla* Suk-*Larix gmelinii* (Rupr.) Kuzen. mixed forest image

Fig. 6. Segmentation results for different forest types.

In addition, the results of the study indicate that it is feasible to transfer the pre-training model from one data set to another, through the consideration of a reasonable amount of training data. The comparison of specific models, has proven that the model using Migration Learning (MIL) was better than the ones without MIL. This is due to the fact that by using less model parameters, model over-fitting is less likely to happen. This proves that migration learning plays an important role in improving the accuracy of the model.

In view of the fact that there is no research on the identification and detection of tree species and stand volume using ground forest images, the proposed method has certain novelty, feasibility and application value. Of course, our research still has some shortcomings and needs to

Table 3
(a)–(b) Four images segmentation results.

Image	Tree code	Pixel ratio		IoU
		Ground truth	Segmentation result	
(a)	2	10.64%	12.87%	80.21%
(b)	1	27.36%	28.27%	93.42%
(c)	1	17.58%	17.97%	90.52%
	2	6.04%	6.78%	
(d)	1	13.21%	13.74%	87.34%
	4	5.38%	5.44%	
	2	6.42%%	6.86%%	

be further improved. For example, in the face of a more complex forest image, there is still a misclassification phenomenon, and the segmentation accuracy needs to be further improved. Regarding the model of the accumulation estimation that was proposed herein, only the pixel weight percentage of the tree species was added as a single variable. In future research, we will try to add more variables such as canopy closure, to further improve the method, and to build a more accurate estimation model of stand accumulation, under diverse ecosystems and stand conditions. More experiments and tests will be carried out to further verify the feasibility of the method.

Table 4
Parameter estimates and evaluation indices for the candidate models.

Model	Parameters			Evaluation indices
	a	b	c	R^2
M1	−16.222	704.739		0.650
M2	2.621	218.154	1689.370	0.678
M3	1405.137	1.547		0.680
M4	598.282	0.313		0.678
M5	–	–	–	–

Note: – indicates the models without reaching convergence.

Funding

This study was supported by the National Key Research and Development Program of China (Grant No. 2017YFC0504106).

Declaration of Competing Interest

The authors declare no conflict of interest.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2019.105012.

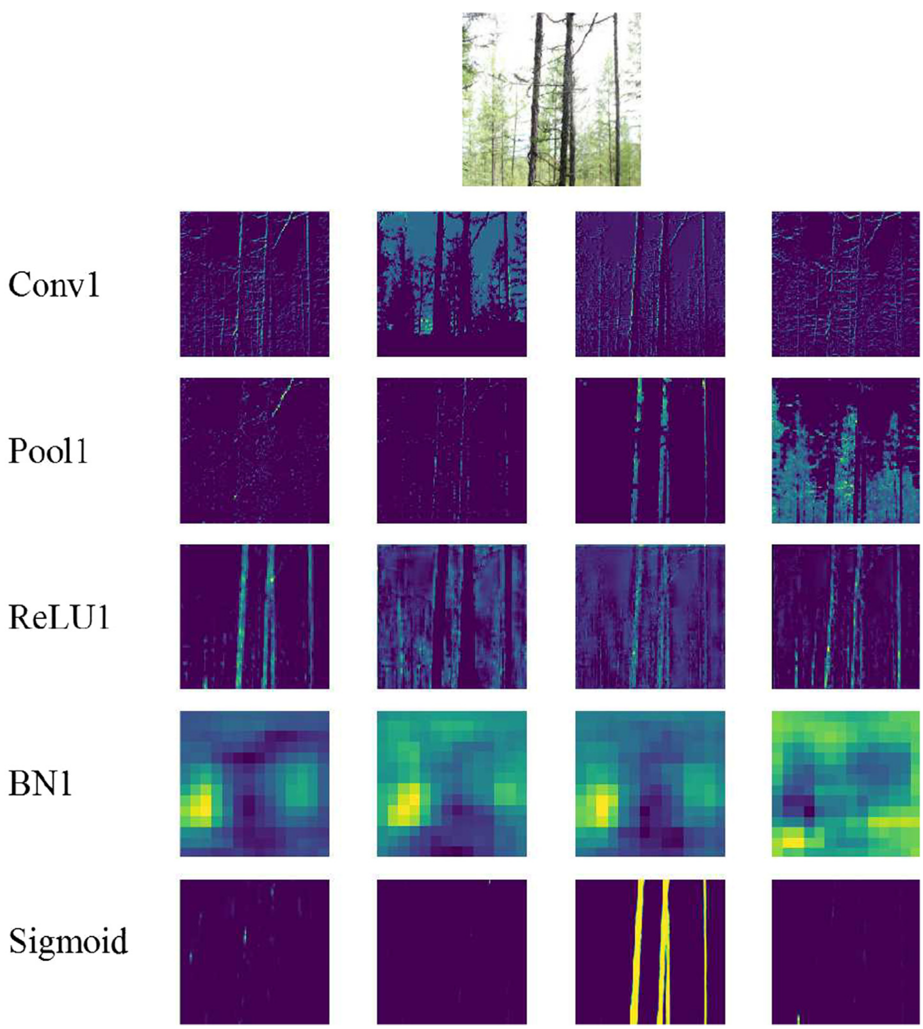


Fig. 7. Model visualization results.

Table 5
NLMEM parameter estimates and evaluation indices for the candidate models.

	Parameters		Significant	R^2	RMSE (m^3/ha)	AIC	BIC
Fixed effects	a	1062.31	***				
	b	1.348	***				
Random effects	1. | a	−66.538	***	0.807	30.539	759.700	778.760
	b	1.348					
	2. | a	−379.991	***				
	b	1.348					
	3. | a	457.201	***				
	b	1.348					
	4. | a	−10.672	***				
	b	1.348					

Note: *** indicates that the progressive *t*-test is significant at a significance level of 0.05.

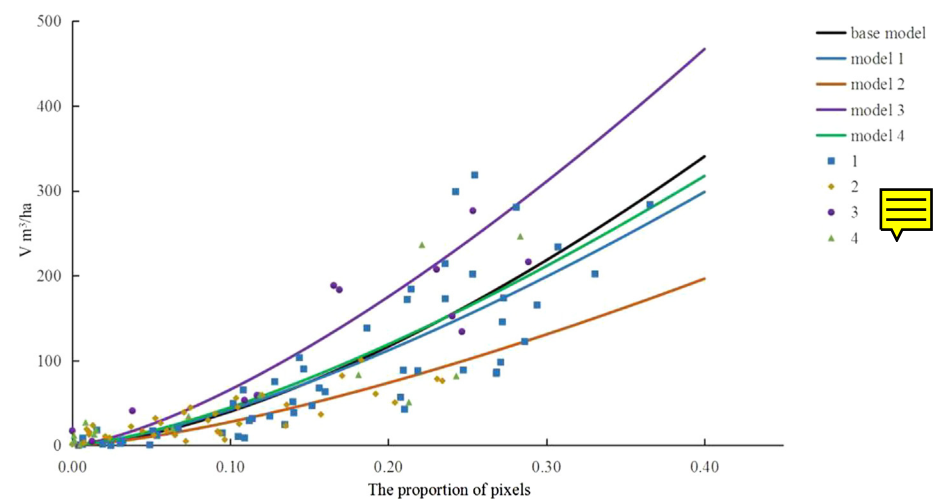


Fig. 8. Pixel Weighted Percentage – GSV Fit Curve.

Table 6
Estimation model of fixed effect and random effect growing stock volume.

	Model form
Fixed effects	$y = 1062.309x^{1.348}$
Random effects	model 1$y = 995.771x^{1.348}$
	model 2$y = 682.318x^{1.348}$
	model 3$y = 1519.510x^{1.348}$
	model 4$y = 1051.637x^{1.348}$

References

Alonzo, M., Bookhagen, B., Roberts, D.A., 2014. Urban tree species mapping using hyperspectral and LIDAR data fusion. Rem. Sens. Environ. 148, 70–83.

Bilous, A., Myroniuk, V., Holiaka, D., Bilous, S., See, L., Schepaschenko, D., 2017. Mapping growing stock volume and forest live biomass: a case study of the Polissya region of Ukraine. Environ. Res. Lett. 12 (10), 105001.

Budhathoki, C.B., Lynch, T.B., Guldin, J.M., 2008. Nonlinear mixed modeling of basal area growth for shortleaf pine. Forest Ecol. Manag. 255, 3440–3446.

Burkhart, H.E., Tomé, M. (Eds.), 2012. Modeling Forest Trees and Stands. Springer Science & Business Media, Berlin, Germany.

Curtis, R.O., 1967. Height-diameter and height-diameter-age equations for second-growth Douglas-fir. Forest Sci. 13 (4), 365–375.

Dahl, G.E., Sainath, T.N., Hinton, G.E., 2013. Improving deep neural networks for LVCSR using rectified linear units and dropout. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 8609–8613.

Dechesne, C., Mallet, C., Le Bris, A., Gouet-Brunet, V., 2017a. Semantic segmentation f forest stands of pure species as a global optimization problem. ISPRS Ann. Photogram., Rem. Sens. Spatial Inform. Sci. IV-1/W1, 141–148.

Dechesne, C., Mallet, C., Le Bris, A., Gouet-Brunet, V., 2017b. Semantic segmentation of forest stands of pure species combining airborne LIDAR data and very high resolution multispectral imagery. ISPRS J. Photogramm. Remote Sens. 126, 129–145.

Freudenberg, M., Nölke, N., Agostini, A., Urban, K., Wörgötter, F., Kleinn, C., 2019. Large scale palm tree detection in high resolution satellite images using U-Net. Remote Sens. 11 (3), 312.

Hawryło, P., Tompalski, P., Wężyk, P., 2017. Area-based estimation of growing stock volume in Scots pine stands using ALS and airborne image-based point clouds. Forestry Int. J. Forest Res. 90 (5), 686–696.

Immitzer, M., Atzberger, C., Koukal, T., 2012. Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. Rem. Sens. 4 (9), 2661–2693.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kittredge, J., 1944. Estimation of amount of foliage of trees and shrubs. J. Forest 42, 905–912.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inform. Process. Syst. 25, 1097–1105.

Li, W., Fu, H., Yu, L., Cracknell, A., 2016. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. Rem. Sens. 9 (1), 22.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.

Morales, G., Kemper, G., Sevillano, G., Arteaga, D., Ortega, I., Telles, J., 2018. Automatic segmentation of mauritia flexuosa in Unmanned Aerial Vehicle (UAV) imagery using deep learning. Forests 9 (12), 736.

Mubin, N.A., Nadarajoo, E., Shafri, H.Z., Hamedianfar, A., 2019. Young and mature oil palm tree detection and counting using convolutional neural network deep learning method. Int. J. Remote Sens. https://doi.org/10.1080/01431161.2019.1569282.

Nevalainen, O., Honkavaara, E., Tuominen, S., Viljanen, N., Hakala, T., Yu, X., Hyyppä, J., Saari, H., Pölönen, I., Imai, N.N., Tommaselli, A.M.G., 2017. Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging. Remote Sens. 9, 185.

Olschofsky, K., Mues, V., Köhl, M., 2016. Operational assessment of aboveground tree volume and biomass by terrestrial laser scanning. Comput. Electron. Agric. 127, 699–707.

Pinheiro, J., Bates, D., 2006. Mixed-effects Models in S and S-PLUS. Springer Science & Business Media.

Richards, F.J., 1959. A flexible growth function for empirical use. J. Exp. Bot. 10, 290–301.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Li, F.F., 2015. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 115 (3), 211–252.
Sačkov, I., Sedliak, M., Kulla, L., Bucha, T., 2017. Inventory of close-to-nature forests based on the combination of airborne LiDAR data and aerial multispectral images using a single-tree approach. Forests 8 (12), 467.
Santi, E., Paloscia, S., Pettinato, S., Fontanelli, G., Mura, M., Zolli, C., Maselli, F., Chiesi, M., Bottai, L., Chirici, G., 2017. The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas. Remote Sens. Environ. 200, 63–73.
Schumacher, F.X., 1939. A new growth curve and its application to timber yield studies. J. For. 37, 819–820.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
Team, R.C., 2013. R development Core team. RA Language and Environment for Statistical Computing, 55, 275–286.
Team, R.C., Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., 2015. nlme: Linear and nonlinear mixed effects models. R Package Version 3-1.
Wang, K., Wang, T., Liu, X., 2019. A review: Individual tree species classification using integrated airborne LiDAR and optical imagery with a focus on the urban environment. Forests 10 (1), 1.
Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? Adv. Neural Inform. Process. Syst. 27, 3320–3328.
Zakharova, M., 2017. Automated Coconut Tree Detection in Aerial Imagery Using Deep Learning. Diss. Master's Thesis. The Katholieke Universiteit Leuven, Löwen, Belgium.