Skolkovo Institute of Science and Technology

MASTER'S THESIS

Deep learning-based individual tree crown delineation from remote sensing RGB imagery

Master's Educational Program:
Space and engineering systems

Student: ______________________

Batyrzhan Alikhanov

Research Advisor: ______________________

Anton Ivanov
Associate Professor

Co-Advisor: ______________________

Alexey Trekin
Research scientist

Moscow 2020

All rights reserved.©

The author hereby grants to Skoltech permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole and in part in any medium now known or hereafter created.

Skoltech

Skolkovo Institute of Science and Technology

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Распознавание крон деревьев с RGB-изображений дистанционного зондирования на основе глубоких нейронных сетей

Магистерская образовательная программа:
Космические и инженерные системы

Студент: ____________________
Алиханов Батыржан Газизович

Научный руководитель: ____________________
Иванов Антон Борисович
Профессор

Со-руководитель: ____________________
Трекин Алексей Николаевич
Научный сотрудник

Москва 2020

Все права защищены.©
Автор настоящим дает Сколковскому институту науки и технологий разрешение на воспроизводство и свободное распространение бумажных и электронных копий настоящей диссертации в целом или частично на любом ныне существующем или созданном в будущем носителе.

Deep learning-based individual tree crown delineation from remote sensing RGB imagery

Batyrzhan Alikhanov

Submitted to the Skolkovo Institute of Science and Technology
on May 29, 2020

Abstract

Emerging remote sensing technologies and state-of-the-art deep learning methods demonstrated a big potential in forest inventory, transforming the scale, cost and speed of forestry assessments. This research presents a novel approach for Individual Tree Crown (ITC) detection and delineation from RGB satellite images using the aerial imagery data for model training. Individual tree crown detection and delineation pipeline was built during this project, and the dataset, containing 0.31m and 0.5m RGB images and their segmentation maps was created from high-resolution aerial imagery. The hypothesis, stating that CNN model trained on rescaled aerial imagery is able to delineate the tree crowns directly from the satellite RGB imagery was confirmed. F1 score of 72% and tree detection rate of 89% was obtained from the satellite test image, which demonstrates promising perspective for the proposed approach to substitute the manual field assessments in forest inventory related tasks.

Research Advisor:
Name: Anton Ivanov
Degree: PhD
Title: Associate Professor

Co-Advisor:
Name: Alexey Trekin
Degree: PhD
Title: Research scientist

Acknowledgments

I would like to express my sincere gratitude to the Skolkovo Institute of Science and Technology (Skoltech) for granting me an opportunity to study and conduct my Master Thesis research at top-tier university.

I sincerely thank my research advisor Professor Anton Ivanov for supporting me during my research. I am also grateful to the research scientists from AeronetLab: Alexey Trekin and Vladimir Ignatiev for their patience and support, prompt and useful advice during my research.

I want to thank my closest friends and Family for their love, companionship, encouragement and trust.

Contents

Glossary

ALS Airborne laser scanning. 13

CHM Canopy Heigh Model. 13, 18, 23

CNN Convolutional Neural Network. 10, 11, 15, 20, 24

DEM Digital Elevation Model. 11, 13, 23

DSM Digital Surface Model. 8, 13, 18, 19, 23

IoU Intersection over Union. 15, 21–23

ITC Individual Tree Crown. 3, 5, 10–16, 20, 23, 24

LIDAR Light Detection and Ranging. 12, 13, 15, 16, 23

LM Local Minima. 14

LMX Local Maxima. 14

NIR Near-infrared. 10

ResNet Residual Convolutional Neural Network architecture. 8, 20

ROI Region of Interest. 11, 21

UAV Unmanned Aerial Vehicle. 12, 16, 17, 23, 24

Unet Unet Convolutional Neural Network architecture. 8, 20

List of Tables

List of Figures

1. INTRODUCTION

Forests are important part of the world's natural ecosystem, maintaining sustainability in terms of economy and ecology in different scales: both locally and globally. Exploitation of the forest resources, forest fires and further other reasons lead to deforestation, and this is a common problem at different parts of the world. Russia is no exception here, up to 50% of its territory is covered by forest, which makes up more than 20% of the World's forestry resources (Figure 1.1). Moreover, Russia is amongst the top 5 wood exporters, the industry exports are booming in the last decade, and ongoing investments will keep raising output [21, 22, 28]. Considering the aforementioned facts, it is obvious that efficient management of the forest sector should be established for sustainable and innovative development of forest use, conservation, and reproduction.

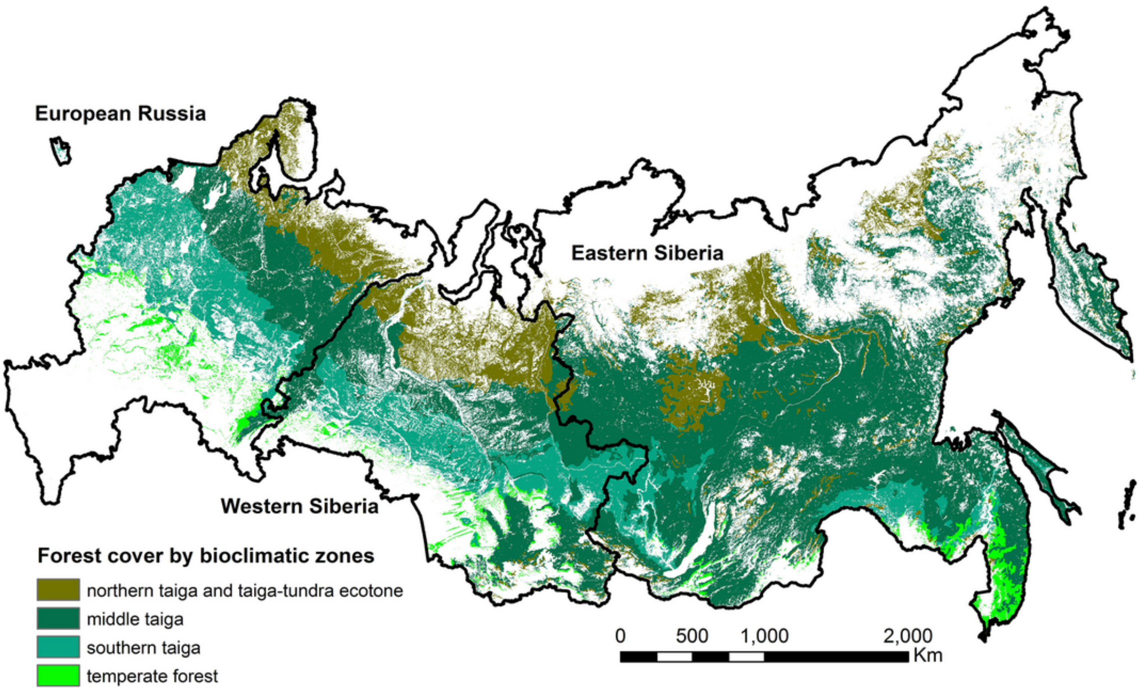

Figure 1.1: Distribution of forest cover across Russia [15]

Systematic collection of forest information is one of the main tools in effective forest management. Collection of forest information, such as tree counting, tree height and location, species classification, and damage area mapping allows governments to perform statistical analysis to understand the state and dynamics for future planning.

While in-situ field assessment is a good option to obtain comprehensive information about small forest patches, it is still very inefficient method in terms of cost, time and assessment fre-

Table 1.1: High resolution satellites, their spectral resolution and bands

Satellite	Launch year	Panchromatic resolution (m)	Multispectral resolution (m)	Bands
GeoEye1	2008	0.46	1.84	RGB-NIR
WorldView2	2009	0.46	1.85	RGB-NIR, RedEdge, Coastal, Yellow, NIR2
Pleiades1A	2011	0.5	2.0	RGB-NIR
Pleiades1B	2012	0.5	2.0	RGB-NIR
Kompsat-3	2012	0.7	2.8	RGB-NIR
SkySat1	2013	0.9	2.0	RGB-NIR
SkySat2	2014	0.9	2.0	RGB-NIR
WorldView3	2014	0.31	1.24	RGB-NIR, RedEdge, Coastal, Yellow, NIR2
Kompsat3A	2015	0.55	2.2	RGB-NIR
WorldView4	2016	0.34	1.36	Not available

quency. On the other hand, emerging remote sensing technologies demonstrated a big potential in different fields, such as precision agriculture, land cover changes detection, urban environment monitoring, etc. It can also be beneficial in forest inventory tasks, transforming the scale, price and speed of forestry assessments. High-resolution remote sensing imagery is becoming one of the most extensively used data types in forestry analysis [23]. Nowadays, there are a lot of satellites, which can obtain sub-meter resolution imagery with limited number of spectral bands (Table 1.1) [4]. Combined with artificial intelligence remote sensing data can be applied to more challenging problems, since neural networks allows to extract non-trivial features and patterns from the satellite or aerial imagery.

The goal of this research is to develop a pipeline for Individual Tree Crown (ITC) detection and delineation from RGB satellite imagery using the processed aerial data for training. The hypothesis is to check that rescaled high-resolution aerial imagery and satellite imagery will have similar domains for the neural network, thereby we will confirm that state-of-the-art Convolutional Neural Network (CNN) trained on rescaled aerial RGB imagery is able to delineate the tree crowns directly from the real satellite RGB imagery. We will try to confirm the applicability of the proposed methodology, demonstrate its practical usage for forest inventory to substitute manual tree delineation. For this purpose, the existing methods for automatic tree delineation are to be analyzed, the new dataset will be created and used in the experiments by the following framework. First of all, UAV-carried aerial imagery from Ust-Ilimsk region (Irkutsk oblast, Russia) was processed to obtain the high resolution orthophotos and Digital Elevation Models (DEM) files. Applying different morphological operations and algorithms on generated data ground truth masks of ITC were created for the Regions of Interest (ROI). After that, the aerial orthophotos and ground truth masks are to be rescaled to two different satellite resolutions (namely 0.5m and 0.31m, which correspond to WV-2 and WV-3 satellite resolutions respectively) to imitate the satellite imagery. Two independent experiments were performed on these different spatial resolution datasets. The data was used for training the CNN to delineate individual trees, and then validated on a separate part of the dataset and several manually delineated satellite images. Overall, remote sensing and deep learning-based framework for ITC was built and validated. Results, obtained from the experiments, confirm the feasibility of the proposed approach. Also, the limitations and challenges were described, and possible developments pathways are discussed in the further sections of this research.

The first part of this work describes essential notions and previously performed research in forest inventory. The second part of the thesis describes obtained data and the proposed methodology. The description and architecture of the deep neural network used in this research is presented in the chapter. The final part presents the results, discussion of findings, limitations, and possible applications.

2. BACKGROUND LITERATURE

Forest inventory-related research papers have increased significantly in the last decade because of the rapid development of computational capacity and availability of high-resolution remote sensing imagery. The detection and delineation of ITCs from remote sensing imagery is one of the initial steps needed to perform forest inventory on a bigger scale. For instance, delineated tree crowns can be used in object-based image analysis and classification, giving necessary information about the number of trees, size of the crowns, geolocation, forest type and tree species. Accurate tree delineation increases the classification rates for specie recognition, since it improves spectral signature characterization of individual trees by reducing the number of pixels, representing the tree crown [26]. Moreover, additional features, such as textural information and reflectance distributions, could be obtained from the data.

Figure 2.1: Example of ITC detection

2.1 UAV-carried systems

Currently, it is possible to obtain and analyze the data on a sub-stand level from Unmanned aerial vehicles (UAVs) with multispectral and Light detection and ranging sensors (LIDAR). There has been a decent amount of research related to forest inventory tasks with the mentioned devices, since UAVs with multispectral and LIDAR cameras become a better alternative to Airborne laser scanning (ALS) in terms of density of point clouds, cost, and flexibility [8]. Characteristics of individual tree structures, such as crown size, height, can be obtained from the segmented trees.

2.2 ITC delineation methods

Several approaches of ITC delineation were proposed in the articles. The first type of data used for tree segmentation is 2D raster images, which is called Digital elevation model (DEM). DEM data can be either Digital surface model (DSM) or Canopy height model (CHM). CHM represents the distance between the top surfaces of trees and the terrain for every pixel in the raster image, so basically it shows the real height of the trees, while DSM represents only top part of the surface (Fig. 2.2). The second data type uses 3D LIDAR point clouds to extract the tree crown information. Regardless of the data type, tree delineation approaches can be classified as parametric and non-parametric [7].

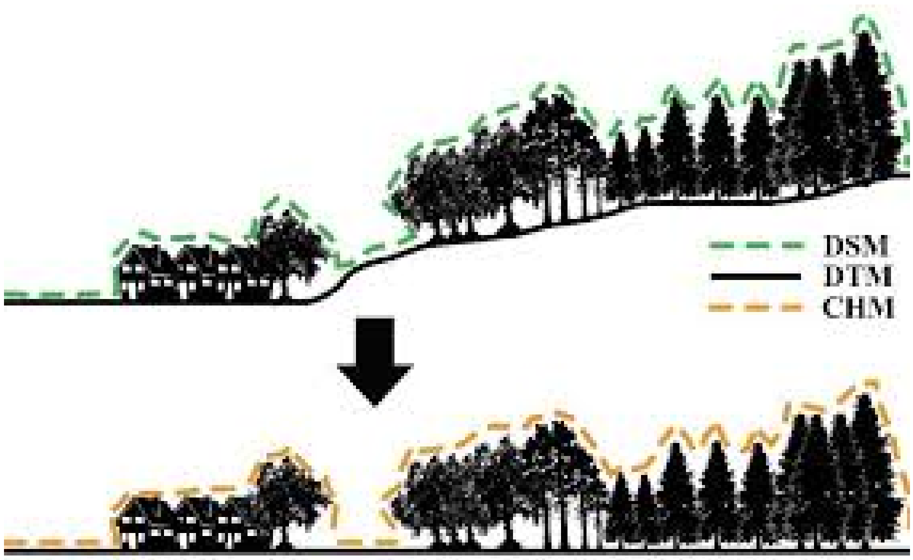

Figure 2.2: Representation of DEM

Parametric methods perform multi-stage filtering from the 3D points, where the filtering is based on our prior assumptions about the tree crown geometry. These methods include rule-based distance and height thresholding, voxel-, graph-, or kernel-based approaches [30]. The parameters of these methods are selected manually based on the field assessment of the forest structure and tree crown shapes. For instance, the Mean shift method, which was originally used for clustering the raster data, was adapted for direct 3D point cloud data segmentation by seeking the mode in feature space [17]. The mode represents the maxima density function, it can be found by shifting the weighted mean determined by predefined kernel. Since the kernel can be extended into 3D space, the mean also can be calculated from the 3D point clouds. Due to this particular ability to go through the 3D space, the above-mentioned methods demonstrated good results for various forest types, such as boreal coniferous, mixed-species, or multi-layered forests [30]. Hamraz et al. suggest that these methods give better results in comparison to the 2D Digital elevation model [7]. However, there are limitations due to the fact that these methods need predefined parameters (kernel shape and size, weighting parameters).

Non-parametric methods find the tree apexes, which are the Local maxima (LMX) in the raster image, or Local minima (LM), which correspond to boundaries between trees. By applying LMX methods top of tree crown can be found within neighborhood area, followed by region-growing or clustering methods. The drawback is that it might be challenging to identify proper kernel size, due to varying tree crown sizes, or spatial resolutions [9], thus it can easily lead to loss of some tree crowns in image. Regression models based on tree heights of the region can be used to adaptively change the kernel size [7], but good results are obtained mostly from the coniferous forest, where tree crowns are homogeneous. The newer approaches use the non-parametric segmentations with a variety of kernel sizes, which creates a number of segmentation maps at different scales. After that, the results are interpreted based on the proposed scoring system [7].

Local minima (LM) based approaches usually use watershed, region growing, hill climbing, and valley following algorithms [18, 5].

The template-matching algorithm compares the spectral characteristics of sample tree with the pixels by sliding through the image, which is not applicable to deciduous forest due to the variability of tree crown size and shape.

Watershed segmentation is the most widely used algorithm, which has problems with under- or over- segmentation mainly due to the changing vegetation indices inside the tree crown area. This can be tackled by marker-controlled watershed algorithm, where the tree apexes are marked in advance prior to the actual delineation procedure. Since manual annotation of tree apexes is inefficient and time-consuming, the problem is solved by applying different mathematical morphological operations [7]. Combination of Local maxima (LMX) and watershed algorithm (LM) performed well, and gave better results in this particular research.

Non-parametric methods described above are mostly applied to coniferous or sparse forests, where certain assumptions can be made about tree and forest type [33]. These factors impose limitations to the variability of trees, and shall be carefully applied to deciduous forest. Increasing number of species, dense forest structure, changing tree sizes and shape, and overall complex vegetation conditions of deciduous forests challenge the application of ITC delineation algorithm. The literature suggests that the aforementioned methods' performance radically varies (tree detection rate from 50% to 90%) depending on forest conditions [7]. In addition, these methods are weakly generalizable on larger scales, since the parameters are tuned for distinct forest type, which confirms the need for a superior approach, which can be applied to different forest types while ensuring solid ITC delineation results.

2.3 Deep Learning in Forest Inventory

Deep learning is state-of-the-art method for object detection and segmentation in RGB images, however it has only recently been applied to remote sensing data [27]. The main benefit of this approach is in Convolutional neural network's CNN ability to automatically calculate and obtain the needed features and parameters, with nominal prior information about the task.

In the case of image segmentation task, initial data is only provided as original images and ground truth masks of the objects for training. This approach is commonly used for land cover, scene classification and object extraction [26, 25]. In the field of remote sensing, deep learning-based applications are at the beginning, there are few, but perspective applications, such as tree delineation, as well as tree type classification [13], however the spatial resolutions of the training data in the most of the papers was considerably higher than the satellite imagery resolution.

The authors in this article [27] used the LIDAR-based methods to obtain the training set for ITC delineation neural network model. Despite drawbacks of LIDAR-based unsupervised delineation method, described in the previous section, the deep neural network was successfully trained on the noisy dataset. Using Intersection Over Union (IoU) threshold set to 0.5, results obtained from the trained model were: 0.69 recall, 0.61 precision, and 0.82 tree detection rate for manually annotated test data. Tree detection is considered to be correct if manually labeled tree apex point lies inside the predicted bounding box. This is logically more suitable approach in comparison to other researches, where, for example, tree detection is considered to be correct if an edge of the bounding box lies within search radius. Because of these differences in measurement approaches, it was hard to locate the best performance, but overall 60%-80% tree detection rate between the ground truth and predicted trees is typical [29, 30, 19, 25, 11, 2, 8, 13, 27, 6].

However, in spite of all of the advantages of DL methods, the need for large and diverse training dataset is still the biggest problem. One approach to tackle this problem proposed by Weinstein et al. was to use "semi-supervised learning" approach, where firstly unsupervised methods are applied to gather initial training dataset [27]. It was recently applied to remote sensing for hyperspectral image classification. Combination of unlabeled and labeled data represents semi-supervised approach, which can leverage the results of deep neural networks trained on a limited amount of data. The model obtains the initial generalized features on a wide range of training samples, and then can be retrained on a smaller, but high quality dataset. In detail, the semi-supervised pipeline for ITC detection developed by Weinstein et al. proposes the following steps: first, LIDAR unsupervised algorithm produces low quality tree crowns segmentation maps, which was used for initial training. Following this, the model is then retrained on smaller manually annotated data to improve the model parameters. As a result, the combination of differently obtained data: by unsupervised methods and manual delineation, allows the model to obtain good quality segmentation maps from new unseen RGB imagery [27]. The problem arising from the lack of datasets is that the performance of the model will severely depend on the region, and results will change among different environments [27]. However, the availability of UAV-carried multispectral or LIDAR sensors will lead to cost-efficient acquisition of datasets from different regions. Overall, it can be concluded that growing use of UAVs in remote sensing grants promising opportunities for combining very high resolution local data with large scale imagery obtained from satellites.

2.4 ITC Classification

The results of individual tree crown detection and delineation (ITCd) can be applied in forest classification, a particular case of land cover classification [18]. In general, any land cover classification task can be categorized either as pixel-based or as object-based, but the latter usually shows better results [20]. Moreover, in forestry applications object-based tree species classification is preferable, since it gives more detailed information about a forest structure. There are two approaches for object-based forest classification: classification at the individual tree level, classifying every tree separately, or area-based classification, when dominant tree species are predicted for different areas of a forest. Forestry legislation seeks to obtain information for each tree rather than general statistics for whole parcels of a forest, thus the case of individual tree level forest species classification is more valuable.

3. DATA AND METHODOLOGY

3.1 Study area

Geographical location of the experimental area is shown in Figure 3.1. The forests, which are located in Ust-Ilimsky district, Itkutsk oblast (southeastern Siberia, 31N and 64E approximately), are classified as boreal deciduous, the terrain is pretty plain, with minor gently sloping areas. This choice of the study area was due to the fact that there was already an ongoing work on logging and forest inventory. Besides, trees in this area are good representatives of the majority of all forest stands in the region and characterize their diversity quite well, so the results of the model will have good generalization for the region.

10 RGB-camera mounted drone flights were performed on the experimental region. Total area covered by the UAVs is 1020 hectares (Figure 3.1), 486 hectares out of the total amount was used in dataset generation.

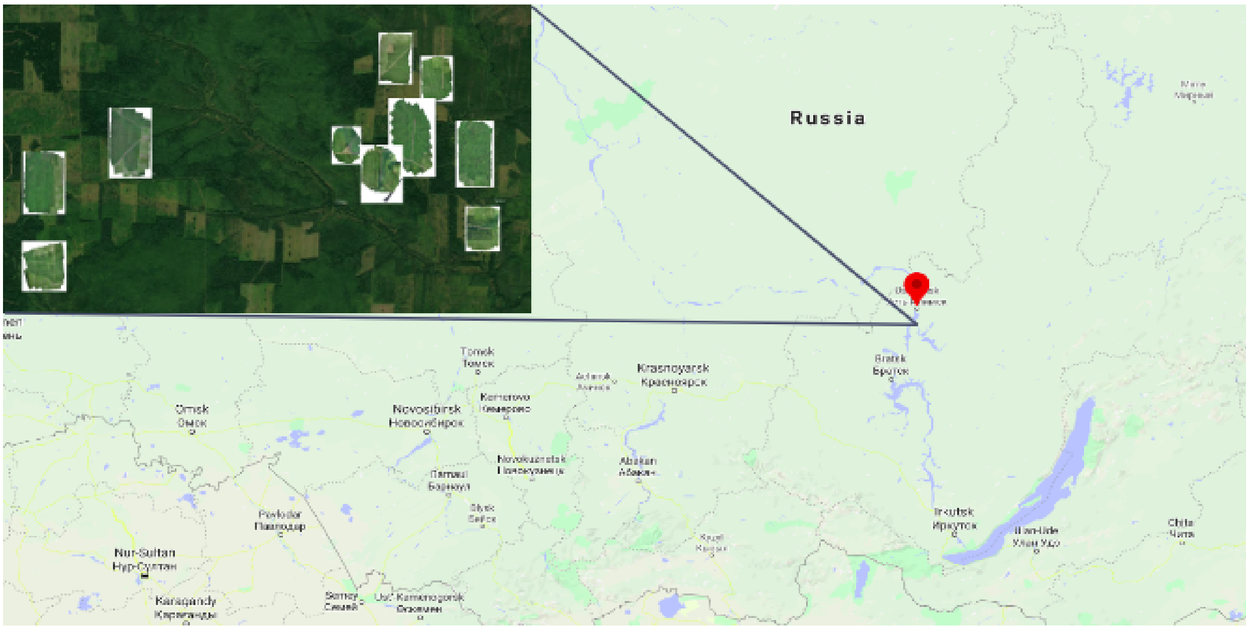

Figure 3.1: Geographical location of the experimental area

3.2 Dataset generation

The orthophotos and DSM files of the 10 experimental regions were processed and stitched using the Agisoft Metashape software from the raw RGB images, obtained from the drone. Agisoft Metashape is a special software product, which performs photogrammetric processing of digital images and generates spatial data. Aerial RGB orthophoto resolution was around 0.05m, DSM images resolution varies from 0.1m to 0.2m. After that, 10 rectangular polygons were cut from the original orthophotos and DSM files, due to the reason that the quality of photogrammetry at the edges of the images were lower, and contained artifacts. This happens due to the fact that the amount of overlapping images at the edges is significantly lower in comparison to the middle part, and the software cannot obtain proper orthographic view. However, the off-nadir imagery with different perspective views are also highly useful for diversity of the data, since most of the satellite imagery is collected from the off-nadir perspective, thus we tried to make the rectangular polygons as big as possible.

Next step was to obtain the ground truth masks of delineated trees from the DSM files. From the DSM raster image (first image of Figure 3.2a) we can clearly observe the round shapes, which correspond to the tree crowns on RGB orthophotos. The main idea was to create 3 segmentation masks, which correspond to Individual trees, Background and Boundary (between the trees, or between the trees and background). Individual trees mask is binary raster image, where distinct contours are the tree crowns, background is the mask with non-forested region in the images. Firstly, non-forested area needs to be thresholded from the DSM. This would be trivial if we had CHM rasters, or the terrain had no curvature and slope, however in our case we need to perform terrain flattening operation. This can be done by applying Flat-field correction technique, which can be performed by subtracting the Gaussian blurred image from the original one. Thereby we create pseudo flat field reference image, by correcting the illumination intensity levels. After that, the background is thresholded from the forested regions. By applying the Local maxima method highest points of tree crowns are determined from the thresholded forestry regions (Figure 3.2b), then using these points as markers we apply the Watershed algorithm to delineate the tree crowns from each other (Figure 3.2c). This is how we obtain the needed masks from DSM raster images. Different mathematical morphological operations, such as adaptive histogram equalization, erosion, dilation, opening and closing, etc., were used in the process. The main challenge here is that all of the operations were performed by visual observation and interpretation of the obtained masks, since, as it was addressed in Section 2.2, the proper delineation can be made only by accurate selection of parameters for all of the mentioned algorithms. Most of the proposed algorithms and methods were carried out using the Skimage and OpenCV libraries.

The final step of dataset generation pipeline was to rescale the RGB aerial orthophotos

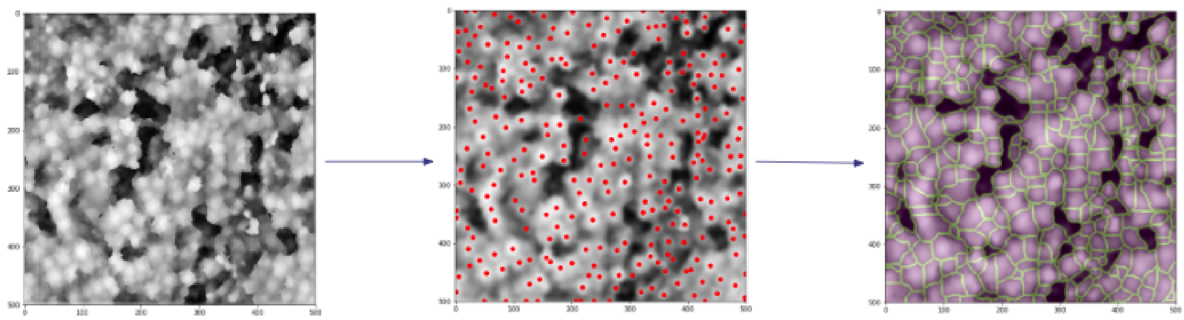

Figure 3.2: The process of dataset generation. a) DSM raster image, b) Local maxima of DSM, c) Watershed tree delineation

Table 3.1: Number of trees (contours) in ground truth masks delineated from the aerial imagery

Experimental area	**Number of trees in 0.31m rescaled ground truth**	**Number of trees in 0.5m rescaled ground truth**
Region1	18481	18478
Region2	20876	20485
Region3	5167	5171
Region4	15534	15487
Region5	14259	14220
Region6	19090	19089
Region7	14829	14820
Region9	5871	5875
Region10	7335	7327

and ground truth masks to match the satellite imagery spatial resolutions (0.5m and 0.31m). The number of trees for every experimental area is presented in Table 3.1, it can be seen that the number of trees slightly differs for different resolutions. Important issue to note here is the thickness of the boundaries: too thick boundaries lead to decrease or loss of the tree crown area. On the other hand, some part of thin boundaries will disappear in the rescaled ground truth masks. Moreover, too thin boundaries may not be enough for deep neural network to perform the proper tree delineation. The optimal boundary thickness for 0.5m resolution was 2-3 pixels, for 0.31m resolution 3-4 pixels. Overall, 10 polygons of aerial imagery were processed by proposed methodology, and rescaled to corresponding spatial resolutions.

3.3 Deep neural network: model training

Deep convolutional neural network is state-of-the-art method for object detection and segmentation in RGB imagery. For our particular case U-Nets with ResNet Encoders and cross connections was chosen, since this is currently one of the most popular architectures for semantic and instance segmentation tasks, showing decent results in different competitions. Convolutional networks can be substantially deeper, more accurate, and more efficient to train if they contain shorter connections between layers close to the input and those close to the output. Thus, the backbone of this architecture is build on Residual Networks ResNet with 34 layers, which is made from a series of residual blocks with skip connections (Figure 3.3 (left)), so basically ResNet is used for the encoder/down sampling part of our architecture. Model with 34 layers was chosen to be optimal for our dataset in terms of training time and memory consumption.

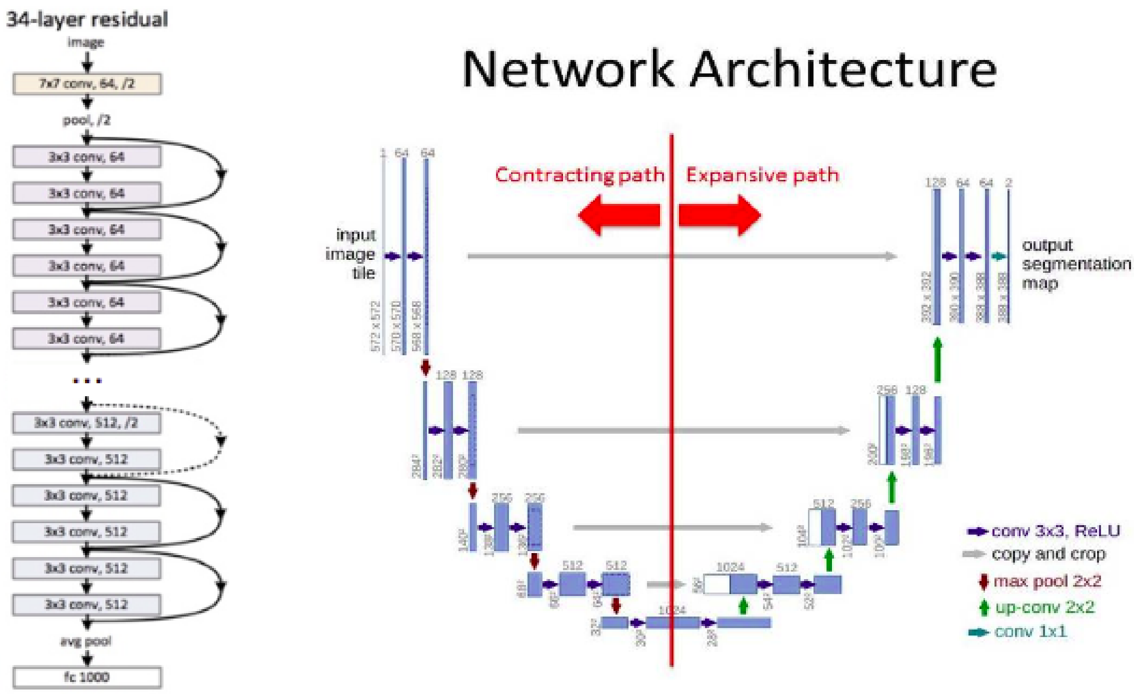

Figure 3.3: ResNet (left) and Unet (right) architectures

The decoding part of our CNN is U-Net architecture, which initially was developed for biomedical image segmentation. U-Nets have been found to be very effective for tasks where the output is of similar size as the input and the output needs that amount of spatial resolution. This makes them very good for creating segmentation masks and for image processing/generation. The Unet is used for the upsampling, where the reverse of the downsampling path is carried out to retain the original image resolution.

The initial model was created from Segmentation Models library (Python library with Neural Networks for Image Segmentation based on Keras and TensorFlow). Initially, the model was pre-trained for 2-class segmentation task to classify forest/non-forest regions with 2m spatial resolution imagery. Then the model structure was changed to perform 3-class segmentation task (tree, boundary, background) for ITC delineation, retaining the weights of pre-trained model. This procedure is know as transfer learning, where the knowledge (weights) for some task is obtained, and then applied to related problem. By doing so we can significantly improve the model efficiency and training time.

The initial dataset imagery, containing 10 experimental regions of interest ROI was divided into 30 images of smaller size, shuffled, and then split into Train/Validation/Test datasets with 70/20/10 percentages. Since entire images cannot fit into GPU memory, 256*256 pixel size batches were used to provide enough spatial context for tree delineation, solving the memory constraints. The model with batch size of 16 was trained on nVidia GeForce GTX 1080Ti for 100 epochs. Different types of image augmentations were used from Albumentations library. The loss function used for training is Categorical Cross-Entropy, which is basically combination of Softmax activation function and Cross-Entropy loss. The metrics used for training the model are Intersection over Union IoU and F1 score. These are the most common used metrics for semantic segmentation. IoU, also known as the Jaccard Index, is calculated by the following formula:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

For multi-class segmentation, the mean IoU of the image is calculated by taking the IoU of each class and averaging them.

F1 score, also known as the Dice Coefficient, is calculated by the following formula:

$$\text{F1 score} = \frac{\text{Area of Overlap}}{\text{Total number of pixels combined}}$$

Model evaluation is done by standard segmentation metrics, described above, plus 2 custom metrics, which are 1) Number of trees (contours) comparison between the ground truth and predicted images, 2) Tree detection rate, where the tree centers points from ground truth are obtained and checked if they fall inside the predicted tree contours.

4. RESULTS AND DISCUSSION

This section presents the results collected in our experiments. First Table 4.1 reports the performance of model, trained on 0.31m imagery, second table 4.2 stands for model, trained on 0.5m imagery, both models were tested with corresponding resolution images from test regions. Third one 4.3 presents the the performance of the latter model on satellite imagery. To assess the performance of proposed tree delineation approach quantitatively, IoU, F1 score, number of trees and detection rate was calculated through comparison with the ground truth.

Table 4.1 presents the results obtained from the model, which was trained and tested on 0.31m resolution imagery. Although, deep CNNs learnt and performed better on higher resolution images, from the comparison of number of contours in ground truth and prediction it can be seen that the model tends to over-segment tree crowns. It may happen due to the changing vegetation indices inside the tree crown area. Visual observation shows that these predictions represent several main branches of individual large tree, rather than multiple small trees.

Table 4.1: Model evaluation metrics for 0.31m resolution test images

Metric names	**Score**	
IoU	0.6	
F1 score	0.7	
	Ground truth	**Predicted mask**
Number of trees	7416	9036
Tree detection rate	0.89	

The results of model, trained on 0.5m resolution RGB aerial imagery, is reported in Table 4.2. Overall performance of this model is slightly lower than the results, obtained from higher resolution.

Table 4.2: Model evaluation metrics for 0.5m resolution test images

Metric names	**Score**	
IoU	0.58	
F1 score	0.68	
	Ground truth	**Predicted mask**
Number of trees	7430	7816
Tree detection rate	0.88	

Google Earth engine was used to download RGB images with different forest types across the Russia to test the model on real satellite imagery. Several samples were manually annotated by visual observation and interpretation of RGB images. The spatial resolution of downloaded data was 0.6m per pixel. Since all of the satellite test images were previously unseen by the model from completely different regions, applying the model, trained on 0.5m aerial dataset, controversial results were obtained. The model performed poorly on sparse forests, complex terrain, or high off-nadir images. However, on some samples the model did pretty good job. For instance, Table 4.3 presents the performance of the model on deciduous forests from Perm region. An illustrative examples of the above-mentioned model predictions are presented in Appendix A.

Table 4.3: Model evaluation metrics for satellite RGB image from Perm

Metric names	**Score**	
IoU	0.42	
F1 score	0.72	
	Ground truth	**Predicted mask**
Number of trees	4014	3913
Tree detection rate	0.89	

As it was addressed before, it is clear that accurate tree detection will be region specific, and that the model performance will change among environments, so there is need for more diverse datasets. Another issue here is the use of DSMs for ITC delineation, which were obtained using photogrammetry on RGB orthophotos. It requires a lot of workload and parameters tuning to obtain accurate delineations. LIDAR data would significantly facilitate the process of tree delineation. if available, including LIDAR CHM into neural networks will likely allow the model to learn the altitudinal features of trees, in addition to the 2D color features in the RGB data, providing opportunities for multi-sensor modeling. These enhancements might be a key in segmenting the dense deciduous forests, where tree crowns overlap each other in 2D RGB imagery.

Overall, using the UAVs for capturing local DEM at high resolution and combining it with the tree delineation approaches will advance cost effective development of regional tree detection models. At the same time, there is increasing number of commercial sub-meter RGB data at large spatial extents, which could be used to detect and delineate ITC at unprecedented extents.

5. CONCLUSION

The scope of this thesis was related to the usage of remote sensing and deep learning in forest inventory. The importance of forest taxation as well as the drawbacks of in-sute field assessments were discussed at the beginning of this work. A detailed literature review was done in order to estimate the current state of technology for trees detection and delineation. Automatic trees detection delineation pipeline, which was built during this project, demonstrates a promising perspectives for forest inventory. The hypothesis, stating that rescaled aerial imagery and satellite imagery will have similar domains for the neural network was confirmed, results show that state-of-the-art Convolutional Neural Network (CNN) trained on rescaled aerial RGB imagery is able to delineate the tree crowns directly from the real satellite RGB imagery. ITC delineation F1 score of 72% and tree detection rate of 89% confirm that it is possible to substitute the manual field assessments by the proposed technology in forest inventory. 486 hectare of experimental area (out of 1020 hectare) was used to generate dataset. Local maxima and Watershed algorithm altogether with different mathematical morphological operations were successfuly applied to obtain satellite spatial resolution (0.31m and 0.5m) imagery from the UAV-carried high-resolution aerial images. Fully-convolutional neural network UResnet34 model with pretrained values was trained on the training data, and hypothesis was confirmed by the test dataset and real satellite imagery. As it was discussed in Background literature section, there is a problem arising from the lack of datasets for various forest types and the performance of the model tested on different satellite images show that the model severely depends on the region, and it will change among different environments. However, the results of this research show that growing use of UAVs and availability of high-resolution imagery combined with generalization ability of deep neural networks will eventually make it possible to obtain good quality forest inventory information from large scale satellite imagery.

Bibliography

[1] Michele Dalponte, Hans Ole Ørka, Liviu Theodor Ene, Terje Gobakken, and Erik Næsset. Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. *Remote Sensing of Environment*, 140:306–317, 2014.

[2] Anderson Aparecido dos Santos, José Marcato Junior, Márcio Santos Araújo, David Robledo Di Martini, Everton Castelão Tetila, Henrique Lopes Siqueira, Camila Aoki, Anette Eltner, Edson Takashi Matsubara, Hemerson Pistori, Raul Queiroz Feitosa, Veraldo Liesenberg, and Wesley Nunes Gonçalves. Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVS. *Sensors (Switzerland)*, 19(16), aug 2019.

[3] W Eckstein. Segmentation and texture analysis. *Proceedings of the International Archives of Photogrammetry and Remote Sensing*, (December 2000):165–175, 1996.

[4] Marilia Ferreira Gomes and Philippe Maillard. Detection of Tree Crowns in Very High Spatial Resolution Images. In *Environmental Applications of Remote Sensing*. InTech, jun 2016.

[5] Marília Ferreira Gomes, Philippe Maillard, and Huawu Deng. Individual tree crown detection in sub-meter satellite imagery using Marked Point Processes and a geometrical-optical model. *Remote Sensing of Environment*, 211:184–195, jun 2018.

[6] Emilio Guirado, Siham Tabik, Domingo Alcaraz-Segura, Javier Cabello, and Francisco Herrera. Deep-learning Versus OBIA for Scattered Shrub Detection with Google Earth Imagery: Ziziphus lotus as Case Study. *Remote Sensing*, 9(12):1220, nov 2017.

[7] Hamid Hamraz, Marco A. Contreras, and Jun Zhang. A robust approach for tree segmentation in deciduous forests using small-footprint airborne LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, 52(859):532–541, 2016.

[8] Hamid Hamraz, Nathan B. Jacobs, Marco A. Contreras, and Chase H. Clark. Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:219–230, dec 2019.

[9] Linhai Jing, Baoxin Hu, and Jili Li. INDIVIDUAL TREE CROWN DELINEATION USING MULTI-SCALE SEGMENTATION OF AERIAL IMAGERY. Technical report.

[10] Yinghai Ke and Lindi J. Quackenbush. A comparison of three methods for automatic tree crown detection and delineation from high spatial resolution imagery. *International Journal of Remote Sensing*, 32(13):3625–3647, 2011.

[11] S. Khan and P. K. Gupta. COMPARITIVE STUDY OF TREE COUNTING ALGORITHMS IN DENSE AND SPARSE VEGETATIVE REGIONS. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5:801–808, nov 2018.

[12] Morten Larsen. Single tree species classification with a hypothetical multi-spectral satellite. *Remote Sensing of Environment*, 110(4):523–532, oct 2007.

[13] Weijia Li, Haohuan Fu, Le Yu, and Arthur Cracknell. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sensing*, 9(1), 2017.

[14] Tao Liu, Jungho Im, and Lindi J. Quackenbush. A novel transferable individual tree crown delineation model based on Fishing Net Dragging and boundary classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 110:34–47, dec 2015.

[15] Tatiana V. Loboda and Dong Chen. Spatial distribution of young forests and carbon fluxes within recent disturbances in Russia. *Global Change Biology*, 23(1):138–153, jan 2017.

[16] Julia Maschler, Clement Atzberger, and Markus Immitzer. Individual tree crown segmentation and classification of 13 tree species using Airborne hyperspectral data. *Remote Sensing*, 10(8), 2018.

[17] Julien Michel, David Youssefi, and Manuel Grizonnet. Stable mean-shift algorithm and its application to the segmentation of arbitrarily large remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(2):952–964, 2015.

[18] Vasilii Mosin, Roberto Aguilar, Alexander Platonov, Albert Vasiliev, Alexander Kedrov, and Anton Ivanov. Remote sensing and machine learning for tree detection and classification in forestry applications. page 14, 2019.

[19] Omkar Narasipura. Semantic Segmentation and Individual Crown. 2017.

[20] D. A. Pouliot, D. J. King, F. W. Bell, and D. G. Pitt. Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sensing of Environment*, 82(2-3):322–334, 2002.

[21] Russ Taylor. Russian forestry industry on the move - International Forest Industries, 2018.

[22] Russ Taylor. 2020 global lumber markets outlook - Wood Business, 2020.

[23] Dong Tianyang, Zhang Jian, Gao Sibin, Shen Ying, and Fan Jing. Single-tree detection in high-resolution remote-sensing images based on a cascade neural network. *ISPRS International Journal of Geo-Information*, 7(9), 2018.

[24] Jari Vauhkonen, Liviu Ene, Sandeep Gupta, Johannes Heinzel, Johan Holmgren, Juho Pitkänen, Svein Solberg, Yunsheng Wang, Holger Weinacker, K. Marius Hauglin, Vegard Lien, Petteri Packalén, Terje Gobakken, Barbara Koch, Erik Næsset, Timo Tokola, and Matti Maltamo. Comparative testing of single-tree detection algorithms under different types of forest. *Forestry*, 85(1):27–40, jan 2012.

[25] Fabien H. Wagner, Alber Sanchez, Yuliya Tarabalka, Rodolfo G. Lotte, Matheus P. Ferreira, Marcos P.M. Aidar, Emanuel Gloor, Oliver L. Phillips, and Luiz E.O.C. Aragão. Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Remote Sensing in Ecology and Conservation*, 2019.

[26] Fabien Hubert Wagner, Matheus Pinheiro Ferreira, Alber Sanchez, Mayumi C.M. Hirye, Maciel Zortea, Emanuel Gloor, Oliver L. Phillips, Carlos Roberto de Souza Filho, Yosio Edemir Shimabukuro, and Luiz E.O.C. Aragão. Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145(October):362–377, 2018.

[27] Ben G. Weinstein, Sergio Marconi, Stephanie Bohlman, Alina Zare, and Ethan White. Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks. *Remote Sensing*, 11(11), jun 2019.

[28] WWF. The Russian-Chinese timber trade: Export, Supply chains, Consumption, and Illegal logging, 2007.

[29] C. Xiao, R. Qin, X. Huang, and J. Li. A STUDY of USING FULLY CONVOLUTIONAL NETWORK for TREETOP DETECTION on REMOTE SENSING DATA. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume 4, pages 163–169. Copernicus GmbH, sep 2018.

[30] Wen Xiao, Aleksandra Zaforemska, Magdalena Smigaj, Yunsheng Wang, and Rachel Gaulton. Mean shift segmentation assessment for individual forest tree delineation from airborne lidar data. *Remote Sensing*, 11(11), jun 2019.

[31] Jian Yang, Yuhong He, and John Caspersen. A multi-band watershed segmentation method for individual tree crown delineation from high resolution multispectral aerial image. In *Inter-*

national Geoscience and Remote Sensing Symposium (IGARSS), pages 1588–1591. Institute of Electrical and Electronics Engineers Inc., nov 2014.

[32] Dameng Yin and Le Wang. Individual mangrove tree measurement using UAV-based LiDAR data: Possibilities and challenges. *Remote Sensing of Environment*, 223:34–49, mar 2019.

[33] A. Zaforemska, W. Xiao, and R. Gaulton. Individual tree detection from uav lidar data in a mixed species woodland. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, volume 42, pages 657–663. International Society for Photogrammetry and Remote Sensing, jun 2019.

A. Figures

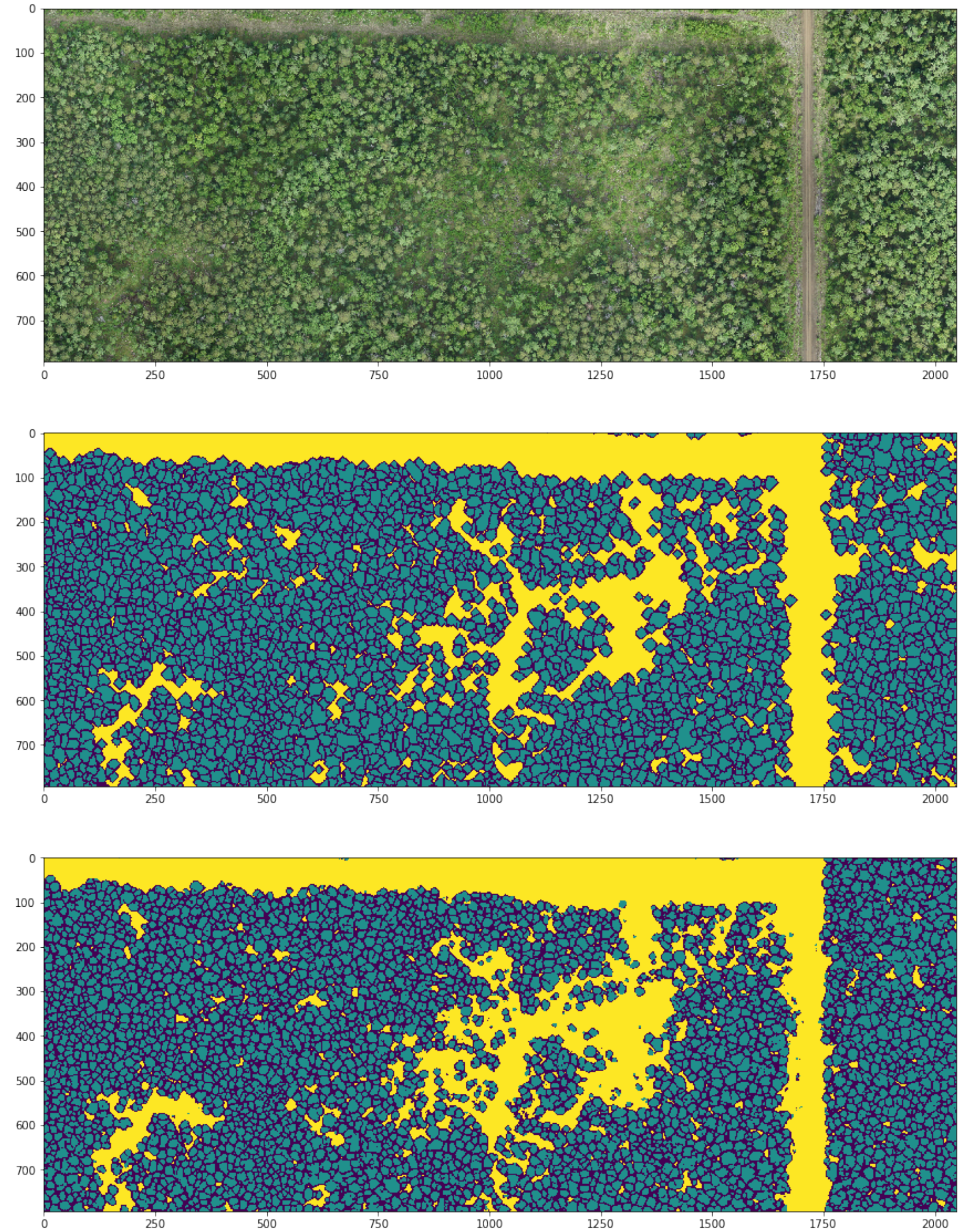

Figure A.1: a) Test region aerial RGB orthophoto (0.31m resolution), b) Ground truth mask c) Prediction of model, trained on 0.31m resolution imagery

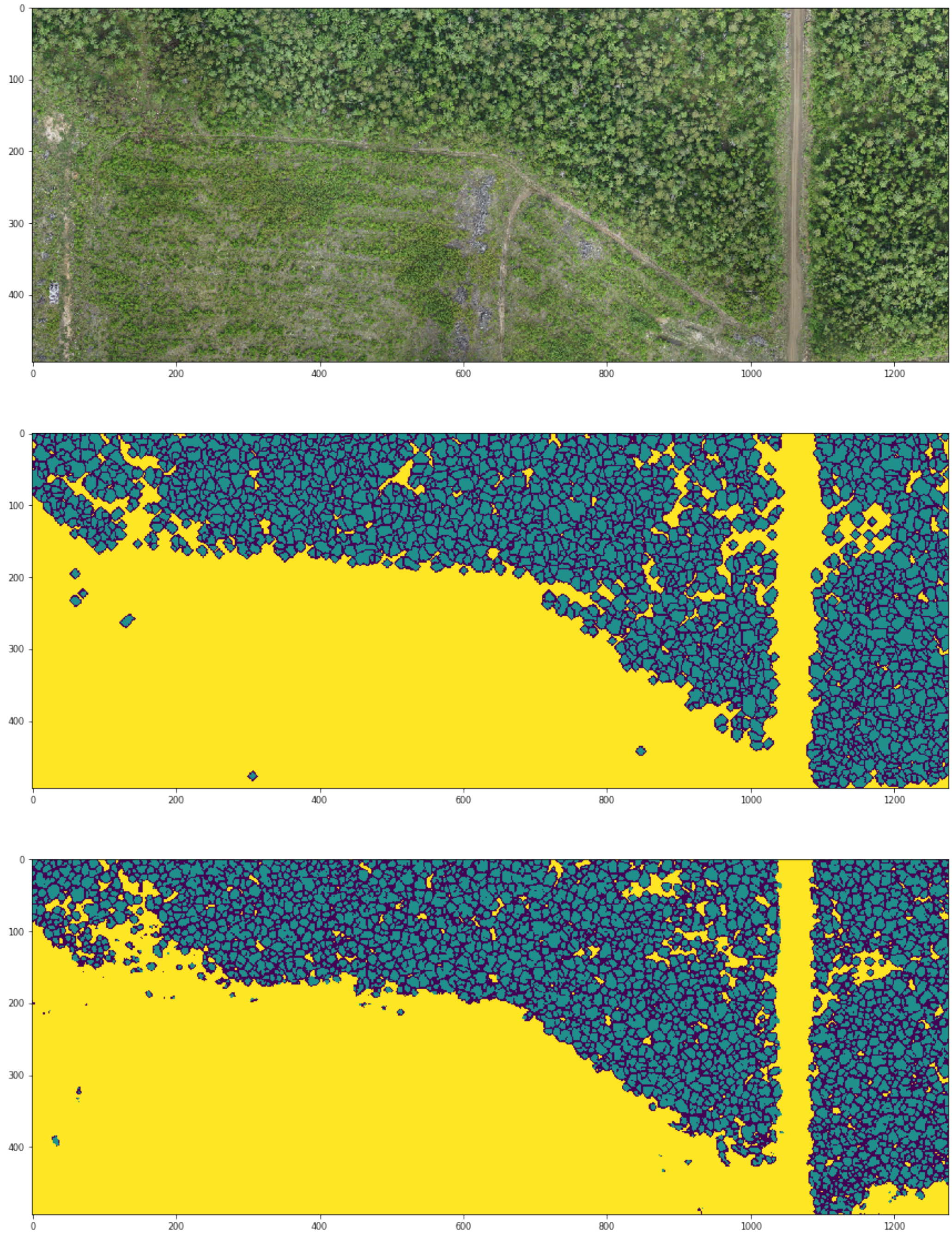

Figure A.2: a) Test region aerial RGB orthophoto (0.5m resolution), b) Ground truth mask c) Prediction of model, trained on 0.5m resolution imagery

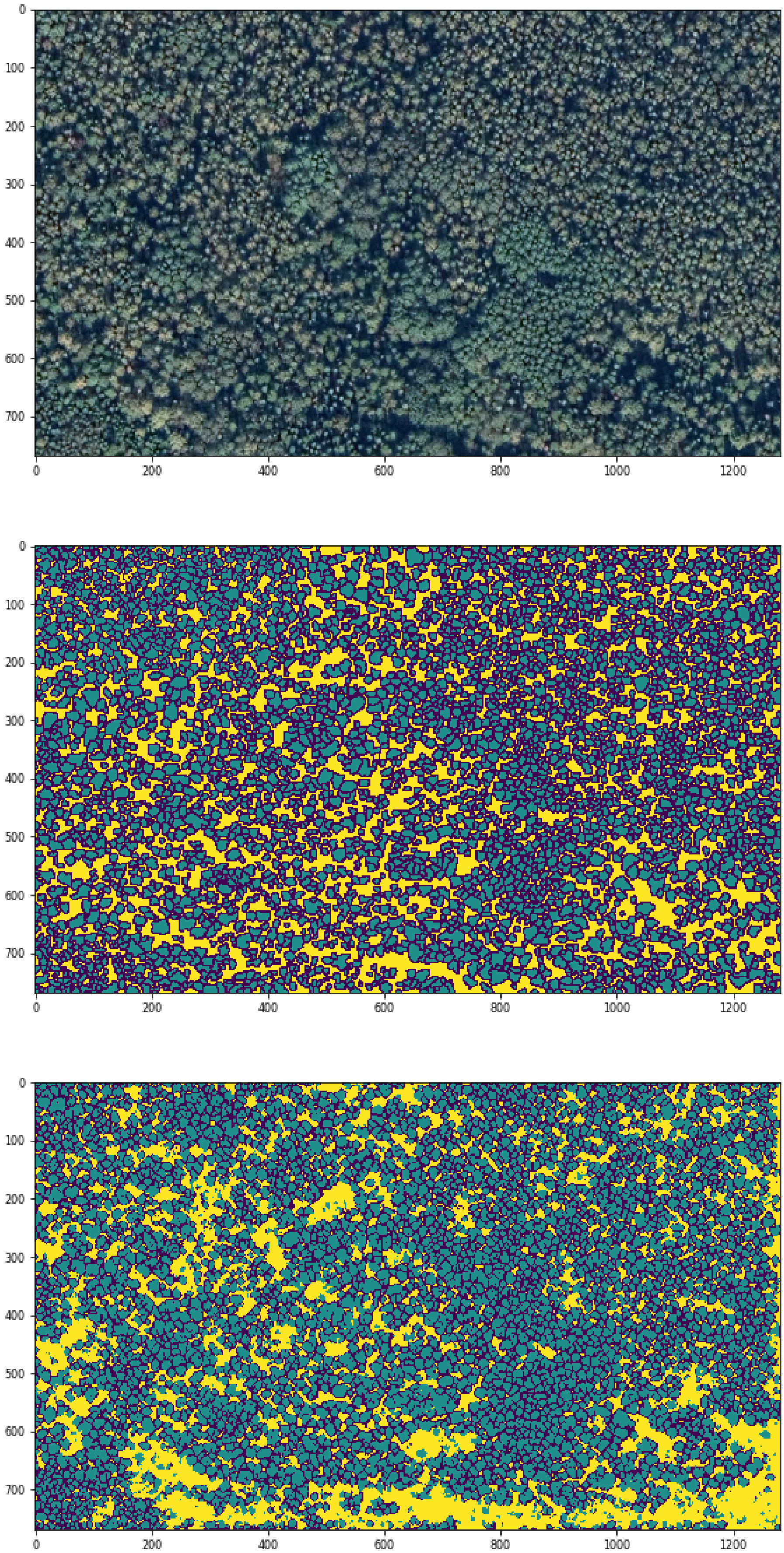

Figure A.3: a) Test satellite RGB orthophoto (0.6m resolution), b) Ground truth mask c) Prediction of model, trained on 0.5m resolution imagery